`

Computer Science &Information Technology 154

Machine Learning Techniques and Data Science

`

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 2nd International Conference on Machine Learning Techniques and Data Science (MLDS 2021), November 20~21, 2021, Zurich, Switzerland
- 10th International Conference of Networks and Communications (NECO 2021)
- 2nd International Conference on Software Engineering and Managing Information Technology (SEMIT 2021)
- 2nd International Conference on IoT, Blockchain & Cloud Computing (IBCOM 2021)
- 10th International Conference on Signal, Image Processing and Pattern Recognition (SPPR 2021)
- 10th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2021)
- 12th International Conference on Communications Security & Information Assurance (CSIA 2021)
- 11th International Conference on Computer Science, Engineering and Applications (ICCSEA 2021)
- 13th International Conference on Wireless, Mobile Network & Applications (WiMoA 2021)

Published By

AIRCC Publishing Corporation

`

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Dhinaharan Nagamalai (Eds),
Wireilla Net Solutions, Australia
E-mail: dhinthia@yahoo.com

ISSN: 2231 - 5403
ISBN: 978-1-925953-52-7
DOI: 10.5121/csit.2021.111801- 10.5121/csit.2021.111823

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

`

Preface

The International Conference on 2nd International Conference on Machine Learning Techniques and Data Science (MLDS 2021), November 20~21, 2021, Zurich, Switzerland, 10th International Conference of Networks and Communications (NECO 2021), 2nd International Conference on Software Engineering and Managing Information Technology (SEMIT 2021), 2nd International Conference on IoT, Blockchain & Cloud Computing (IBCOM 2021), 10th International Conference on Signal, Image Processing and Pattern Recognition (SPPR 2021), 10th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2021), 12th International Conference on Communications Security & Information Assurance (CSIA 2021), 11th International Conference on Computer Science, Engineering and Applications (ICCSEA 2021) and 13th International Conference on Wireless, Mobile Network & Applications (WiMoA 2021) was collocated with 2nd International Conference on Machine Learning Techniques and Data Science (MLDS 2021). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The MLDS 2021, NECO 2021, SEMIT 2021, IBCOM 2021, SPPR 2021, SCAI 2021, CSIA 2021, ICCSEA 2021 and WiMoA 2021 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically.

In closing, MLDS 2021, NECO 2021, SEMIT 2021, IBCOM 2021, SPPR 2021, SCAI 2021, CSIA 2021, ICCSEA 2021 and WiMoA 2021 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the MLDS 2021, NECO 2021, SEMIT 2021, IBCOM 2021, SPPR 2021, SCAI 2021, CSIA 2021, ICCSEA 2021 and WiMoA 2021.

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld,
Dhinaharan Nagamalai (Eds)

`

General Chair	Organization
David C. Wyld,	Southeastern Louisiana University, USA
Dhinaharan Nagamalai (Eds)	Wireilla Net Solutions, Australia

Program Committee Members

Abdalhossein Rezai,	ACECR, Iran
Abdel-Badeeh M. Salem,	Ain Shams University, Egypt
Abdelbaky Hamadene,	AASTMT, Egypt
Abdelhadi Assir,	Hassan 1st University, Morocco
Abdelhalim Kessal,	University of Bordj Bou Arreridj, Algeria
Abdellatif I. Moustafa,	Umm AL-Qura University, Saudi Arabia
Abdelsalam Maatuk,	University of Benghazi, Libya
Abderrahim Siam,	University of Khenchela, Algeria
Abderrahmane EZ-Zahout,	Mohammed V University, Morocco
Abdessamad Belangour,	University Hassan II Casablanca, Morocco
Abdullah,	Adigrat University, Africa
Abhishek Shukla,	R.D. Engineering College, India
AbtoyAnouar,	Abdelmalek Essaadi University, Morocco
Addisson Salazar,	Universitat Politècnica de València, Spain
Adeyanju Sosimi,	University of Lagos, Nigeria
AdrianOlaru,	University Politehnica of Bucharest, Romania
Ahmad A. Saifan,	Yarmouk University, Jordan
Ahmad Yarahmadi,	Tarbiat Modares University, Iran
Ahmed Mehaoua,	Universite Paris Descartes, France
Ajit Kumar Singh,	Patna Women's College, India
Ajit Singh,	Patna Women's College, India
Alessandro Massaro,	Dyrecta Lab, Italy
Alessio Ishizaka,	NEOMA Business School, France
Ali A. Amer,	Taiz University, Yemen
Ali Abdrhman Mohammed Ukasha,	Sebha University, Libya
Alireza Valipour Baboli,	University Technical and Vocational, Iran
Amari Houda,	Networking & Telecom Engineering, Tunisia
Amel Ourici,	University Badji Mokhtar Annaba, Algeria
Amizah Malip,	University of Malaya, Malaysia
Amrita Agarwal,	Sikkim Manipal Institute of Technology, India
Andrea Visconti,	University of Milan, Italy
Anghelescu Petre,	University of Pitesti, Romania
Anita Yadav,	Harcourt Butler Technical University, India
Aridj Mohamed Hassiba,	Benbouali University, Algeria
Arjav Bavarva,	RK University, India
Arnold Kwofie,	University for Development Studies, Ghana
Assia Djenouhat,	University of Algiers 3, Algeria
Asssem Abdel Hamied Moussa,	Chief Eng, Egypt
Attila Kertesz,	University of Szeged, Hungary
Azah Kamilah Muda,	UTeM, Malaysia
Azeddine Wahbi,	Hassan II University, Morocco
Baihua Li,	Loughborough University, UK
Bala Modi,	Gombe State University, Nigeria
Basanta Joshi,	Tribhuvan University, Nepal

`

Benyettou Mohammed,	University center of Relizane, Algeria
Beshair Alsiddiq,	Prince Sultan University, Saudi Arabia
Bin Zhao,	Northwestern Polytechnical University, China
Bouchra Marzak,	Hassan II University, Morocco
Brahami Menaouer,	Qatar University, Qatar
Brahim Lejdel,	University of El-Oued, Algeria
Bremiga Gopalan,	Anna University, India
Carlos O. Rolim,	Federal University of Rio Grande do Sul, Brazil
Cheng Siong Chin,	Newcastle University, Singapore
Cherkaoui Leghris,	Hassan II university of Casablanca, Morocco
Christian Mancas,	Ovidius University, Romania
Christos J. Bouras,	University of Patras, Greece
Chuan-Ming Liu,	National Taipei University of Technology, Taiwan
Claude Tadonki,	MINES ParisTech, France
Dac-Nhuong Le,	Haiphong University, Vietnam
Dadmehr Rahbari,	University of Qom, Iran
Dalila Guessoum,	Saad Dahleb University, Algeria
Dario Ferreira,	University of Beira Interior, Portugal
Dharmendra Sharma,	University of Canberra, Australia
Dhruv Arya,	University of Pennsylvania, Philadelphia
Dinesh Reddy V,	SRM University, India
Domenico Rotondi,	Fincons SpA, Italy
El Murabet Amina,	Abdelmalek Essaadi University, Morocco
Endre Pap,	Singidunum University, Serbia
Eng Islam Atef,	Alexandria University, Egypt
Eric Renault,	Institut Telecom - Telecom SudParis, France
Eva Shayo,	University of Dar es Salaam, Tanzania
Felix J. Garcia Clemente,	University of Murcia, Spain
Francesco Zirilli,	Sapienza Universita Roma, Italy
Gabor Kiss,	J. Selye University, Slovakia
Gabriel Badescu,	University of Craiova, Romania
GálZoltán,	University of Debrecen, Hungary
Giuliani Donatella,	University of Bologna, Italy
Gniewko Niedbała,	Poznan University of Life Sciences, Poland
Grigorios N. Beligiannis,	University of Patras, Greece
Grzegorz Sierpiński,	Silesian University of Technology, Poland
habil Gabor Kiss,	Obuda University, Hungary
Hala Abukhalaf,	Palestine Polytechnic University, Palestine
Hamid Ali Abed AL-Asadi,	Basra University, Iraq
Hamid Khemissa,	USTHB University Algiers, Algeria
Hatem Yazbek,	NSU, USA
Heba Elgazzar,	Morehead State University, USA
Hedayat Omidvar,	National Iranian Gas Company, Iran
Heldon Jose,	Professor of Integrated Faculties of Patos, Brazil
Hemavathi P,	Bangalore Institute of Technology, India
Henok Yared,	Mettu University, Ethiopia
Hiromi Ban,	Sanjo City University, Japan
Homero Toral Cruz,	University of Quintana Roo, México
Hong Shen,	Sun Yat-sen University, China
Hossein Jadidoleslami,	MUT University, Iran
Hui Song,	Frostburg State University, USA
Ihab Sbeity,	Lebanese University, Lebanon

`

Ihtiram Raza Khan,	Jamia Hamdard, India
Ilango Velchamy,	CMR Institute of Technology, India
Ilham Huseyinov,	Istanbul Aydin University, Turkey
Isidro Calvo,	University of the Basque Country, Spain
Ismael Etxeberria-Agiriano,	UPV/EHU, Spain
Israa Shaker Tawfic,	VIT University, India
Israel Goytom,	Chapa Financial Technologies, Ethiopia
Iyad Alazzam,	Yarmouk University, Jordan
Jasmin Cosic,	DB AG, Germany
Jesuk Ko,	Universidad Mayor de San Andres, Bolivia
Jia Ying Ou,	York University, Canada
Jonah Lissner,	Israel Institute of Technology, Israel
Juntao Fei,	Hohai University, P. R. China
Jyotsna Kumar Mandal,	University of Kalyani, India
Kai Pan,	University of North Carolina at Charlotte, USA
Kai-Long Hsiao,	Taiwan Shoufu University, Taiwan
Kamel Benachenhou,	Blida University, Algeria
Karim El Moutaouakil,	FPT/USMBA, Morroco
Karthikeyan V,	SVS College of Engineering, India
Ke-Lin Du,	Concordia University, Canada
Khaleefah Al-Janab,	Alhikma College University, Iraq
Khalid M.O Nahar,	Yarmouk University, Jordan
Kirtikumar Patel,	Chemic Engineers, USA
Klenilmar L. Dias,	Federal Institute of Amapa, Brazil
Koh You Beng,	University of Malaya, Malaysia
Liquan Chen,	Southeast University, China
Loc Nguyen,	Independent scholar, Vietnam
Luisa Maria Arvide Cambra,	University of Almeria, Spain
M V Ramana Murthy,	Osmania University, India
MA.Jabbar,	Vardhaman College of Engineering, India
Mabroukah Amarif,	Sebha University, Libya
Mahdi Niamanesh,	Imam Reza University, Iran
Mamoun Alazab,	Charles Darwin University, Australia
Mansi Subhedar,	University of Mumbai, India
Marcin Paprzycki,	Adam Mickiewicz University, Poland
Marco Favorito,	Sapienza University of Rome, Italy
María Hallo,	Escuela Politécnica Nacional, Ecuador
Maryam Rastgarpour,	Azad University, Iran
Marzak Bouchra,	Hassan II University, Morocco
Mehdi Soltani,	Qazvin Islamic Azad University, Iran
Mervat Bamiah,	Alnahj for IT Consultancy, Saudi Arabia
Michail Kalogiannakis,	University of Crete, Greece
Min Guk I. Chi,	S P Jain School of Global Management, Philippines
Mirsaeid Hosseini Shirvani,	Islamic Azad University, Iran
Mohamed Arezki Mellal,	M'Hamed Bougara University, Algeria
Mohamed Gasmi,	University of Larbi Tebessi, Algeria
Mohamed Ismail Roushdy,	Ain Shams University, Egypt
Mohamed Yacoab,	The New College, India
Mohammad A. Alodat,	Sur University College, Oman
Mohammad Abdallah,	Al-Zaytoonah University of Jordan, Jordan
Mohammad Jafarabad,	Iran University of Science & Technology, Iran
Mohammadreza Balouchestani,	Ryerson University, Canada

`

Shing-Tai Pan,	National University of Kaohsiung, Taiwan
Shoeib Faraj,	Technical and Vocational University Of Urmi, Iran
Shridhar B. Devamane,	Tecsec Technologies Bangalore, India
Siarry Patrick,	Universite Paris-Est Creteil, France
Sidi Mohammed Meriah,	University Of Tlemcen, Algeria
Sikandar Ali,	China University of Petroleum, China
Smain Femmam,	UHA University, France
Solley Joseph Thomas,	Goa University, India
Stefano Michieletto,	University of Padova, Italy
Subhendu Kumar Pani,	Krupajal Computer Academy, India
Suhad Faisal Behadili,	University of Baghdad, Iraq
sukhdeep kaur,	punjab technical university, India
Sun-yuan Hsieh,	National Cheng Kung University, Taiwan
Taha Mohammed Hasan,	University of Diyala, Iraq
Tanzila Saba,	Prince Sultan University, Saudi Arabia
Tatyana A. Komleva,	Odessa State Academy, Ukraine
V. Dinesh Reddy,	SRM University, India
V.Ilango,	CMR Institute of Technology, India
Valentina Emilia Balas,	University of Arad, Romania
Vinay S,	PES College of Engineering, India
Vishal Polara,	BVM engineering college, India
Waleed Bin Owais,	Qatar University, Qatar
Wei Cai,	Qualcommm technology, USA
Wei lu,	Early Warning Academy, China
Wenwu Wang,	University of Surrey, UK
Xiao-Zhi Gao,	University of Eastern Finland, Finland
Yahya Slimani,	Faculty of Sciences of Tunis, Tunisia
Yogeshver Khandagre,	RGPV, Bhopal, India
Youssef Taher,	Center of Guidance and Planning, Rabat -Morocco
Youye Xie,	Colorado School of Mines, United States
Yu-Chen Hu,	Providence University, Taiwan
Yung Gi Wu,	Chang Jung Christian University, Taiwan
Zainab S. Attarbashi,	Universiti Utara Malaysia, Malaysia
Zakaria Kurdi,	University of Lynchburg, USA
Ze Tang,	Jiangnan University, China
Zhou Rou Gang,	HangZhou DianZi University, China
Zoran Bojkovic,	University of Belgrade, Serbia

`

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Artificial Intelligence Community (AIC)

Soft Computing Community (SCC)

Digital Signal & Image Processing Community (DSIPC)

`

2nd International Conference on Machine Learning Techniques and Data Science (MLDS 2021)

10th International Conference of Networks and Communications (NECO 2021)

2nd International Conference on Software Engineering and Managing Information Technology (SEMIT 2021)

`

2nd International Conference on IoT, Blockchain & Cloud Computing (IBCOM 2021)

10th International Conference on Signal, Image Processing and Pattern Recognition (SPPR 2021)

10th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2021)

12th International Conference on Communications Security & Information Assurance (CSIA 2021)

`

11th International Conference on Computer Science, Engineering and Applications (ICCSEA 2021)

13th International Conference on Wireless, Mobile Network & Applications (WiMoA 2021)

CHARACTERISTICS OF SUPER-UTILIZERS OF CARE AT THE UNIVERSITY HOSPITALS OF GENEVA USING LATENT CLASS ANALYSIS

Gilles Cohen[1], Pascal Briot[2] and Pierre Chopard[2]

[1]Finance Directorate Geneva University Hospitals Geneva, Switzerland
[2]Quality and Patient Safety Division Geneva
University Hospitals Geneva, Switzerland

ABSTRACT

In hospitalized populations, there is significant heterogeneity in patient characteristics, disease severity, and treatment responses, which generally translates into very different related outcomes and costs. A better understanding of this heterogeneity can lead to better management, more effective and efficient treatments by personalizing care to better meet patients' profiles. Thus, identifying distinct clinical profiles among patients can lead to more homogenous subgroups of patients. Super-utilizers (SUs) are such a group, who contribute a substantial proportion of health care costs and utilize a disproportionately high share of health care resources. This study uses cost, utilization metrics and clinical information to segment the population of patients (N=32,759) admitted to the University Hospital of Geneva in 2019 and thus identifies the characteristics of its SUs group using Latent Class Analysis. This study shows how cluster analysis might be valuable to hospitals for identifying super-utilizers within their patient population and understanding their characteristics.

KEYWORDS

Latent Class Analysis, Clustering, Super-Utilizers, Inpatient Segmentation, Hospital Efficiency, Quality Improvement.

1. INTRODUCTION

The ongoing increase in healthcare expenditures [1] [2] and the introduction of new payment incentives which favor reductions in avoidable admissions and reoperations [3] [4][5] are forcing hospitals to develop new quality improvement strategies and improve their efficiencies. Since the greater share of hospital expenditure is often directed toward a limited number of patients commonly referred in the literature as super-utilizers (SUs)[6] [7] [8] [9], identifying these patients and designing better targeted interventions for them have the potential to increase appropriateness of care, improve outcomes and reduce costs. This study aims to stratify the population of patients admitted for more than 24 hours to the University Hospitals of Geneva and discharged between January 1, 2019 and December 31, 2019 applying cluster analysis on utilization data using demographics, admission and medical data.

The proposed approach uses Latent Class Analysis (LCA)to identify distinct patient clusters within our inpatient data. LCA is a model-based method that determines clusters of patients by common underlying unobserved characteristics. It is an iterative, maximum likelihood method that estimates how patterns in patient characteristics can be summarized into a finite number of

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 01-15, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111801

groups, or latent classes, by producing a probability distribution over the cluster allocation for each patient.LCA is convenient for analysis of categorical variables that are commonly found in clinical settings.

Clustering has been used to identify new disease subgroups in a diverse range of conditions, such as asthma, chronic lung disease (COPD), chronic heart failure (CHF) and neurological disorders. Nevertheless, the application of clustering to health care delivery is still emerging.

2. METHODS

2.1. Data and Variables

The Hôpitaux Universitaires de Genève (HUG) in Geneva is the largest academic medical center in Switzerland with approximately 2,000 acute care beds and 47,000 admissions per year. Located on 8 different sites, the hospital offers acute, intensive and long-term inpatient care, including pediatric and psychiatric care as well as rehabilitation and ambulatory care. All data for the study were collected from the HUG Enterprise Data Warehouse (EDW). The EDW contains information from several information systems including the patient administrative file (DPA - Dossier Patient Administratif), the clinical data repository which includes data from the HUG electronic medical record system (DPI - Dossier Patient Integré), the accounting costing system, and other operation tracking systems. Case costing at the HUG is determined using the standardized cost accounting model known by the acronym REKOLE developed by the Swiss Hospital Association (H+) [10]. It is based on real and normative costs which provide detailed information on the direct and indirect costs associated to each patient hospital stay. All costs are quoted in Swiss francs (CHF). From the EDW we used patient hospital utilization data. Detailed admission data were gathered from hospital discharge summaries comprising admission and discharge dates, admission and discharge disposition, length of stay (LOS), level of care provided (standard care or intensive care), category of services provided including surgical interventions, medications, tests, imaging and both primary and secondary diagnoses. The Elixhauser comorbidity index was calculated for every admission using the International Classification of Diseases, Version 10 [11] using a coding algorithm. DiagnosesICD-10 codes were matched with chapter headings. These data are gathered and coded systematically for each admission by coding service. Patients with missing data were excluded from the analysis.

Since we focused on high-cost patients according to the costs charged, we examined the distribution of health care costs in our data set representing all patients with a non-psychiatric inpatient admission discharged between January 1, 2019, and December 31, 2019. This rapid analysis confirms that the distribution of health care costs is highly concentrated on a small number of patients. In Figure 1, the population on the horizontal axis is segmented into deciles, starting from the decile of patients with the lowest costs consumption on the left to the decile of patients with the highest costs consumption on the right. The vertical axis shows the cumulative costs consumption for all patients. Thus, the 58.5% indicated above the 90% on the horizontal axis signifies that 90% of individuals (the least costly) accounted for only 58.5% of the total costs of the population. While the other 10% of the population generated 41.5% of the total costs. Therefore, the high-costs group was defined as the top 10 percentile of patients incurring the largest total (direct and indirect) admission costs.

The primary objective of this study was to characterize the high-costs users compared to the remaining 90% of patients according to patient characteristics, primary diagnoses, as well as their admission (emergency department) and discharge dispositions (e.g., home, acute care transfer, and long-term care transfer)and selected comorbidity score.

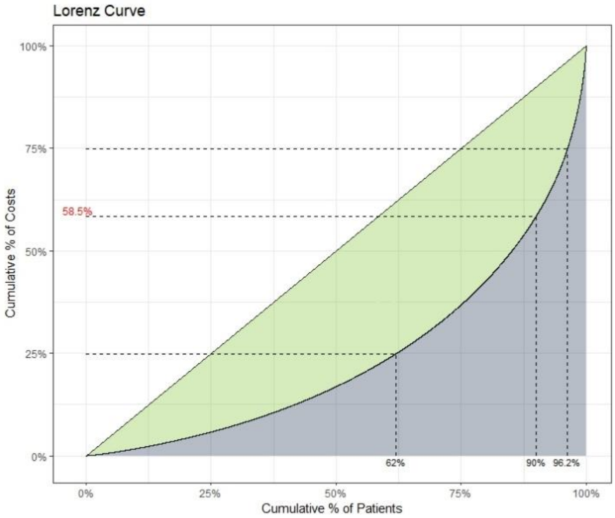


Figure 1. Lorenz curve for the total costs distribution. The diagonal line is the line of equality (for a perfectly equal distribution of costs per patient). Greater distance from the equality line indicates greater disparity in the distribution of total HUG costs.

2.2. Methodology

2.2.1. Latent Class Analysis

Let **X**be the $N \times M$ data matrix, where each row $\mathbf{X}_n$is the realization of an M-dimensional vector of random dichotomous or polytomous variables $\mathbf{X}=(X_{n1},\ldots,X_{nM})$. Model based clustering assumes that each $\mathbf{X}_n$ comes from a finite mixture of G probability distribution in the exponential family, such as Bernoulli or Multinomial, each representing a different cluster, class or group. The general form of finite joint distribution of observed variables is as follow:

$$p(\mathbf{X}_n) = \sum_{g=1}^{G} \tau_g p(\mathbf{X}_n|\theta_g) \qquad (1)$$

where the τ_g are the mixing probabilities and θ_g is the parameter set corresponding to component g. The component densities completely describe the cluster structure of the data and each observation belongs to the respective cluster in accordance with a set of unobserved cluster membership indicators $\mathbf{z}_n=(z_{n1},z_{n2},\ldots,z_{nG})$ such that $z_{ng}=1$if $\mathbf{X}_n$ arises from the gth subpopulation.

When grouping multivariate categorical data, a prevalent model-based approach is the latent class analysis (LCA) model. In this setting, within each class, each variable $\mathbf{X}_m$ is modelled using a multinomial distribution, thus

$$p(X_m|\theta_g) = \prod_{c=1}^{C_m} \theta_{gmc}^{\mathbb{1}\{X_m=c\}}$$

where c=1,…,C_m are the possible categories values for variable m, θ_{gmc} is the probability of the variable taking value c given class g, and $\mathbb{1}\{X_m = c\}$ is the indicator function that is 1 if the variable takes value c, and 0 if not. In LCA, the variables are considered to be statistically independent given the class value of an observation. This is a primary assumption referred to as the local independence assumption [12].Transgressions of this assumption often cause the incompatibility of latent class models. The variables are then modelled for each variable within each group with a multinomial density giving the following factorization of the joint component density:

$$p(\mathbf{X}_n|\theta_g) = \prod_{m=1}^{M} \prod_{c=1}^{C_m} \theta_{gmc}^{\mathbb{1}\{X_{nm}=c\}}$$

accordingly the overall density in (1) turn into

$$p(\mathbf{X}_n) = \sum_{g=1}^{G} \tau_g \prod_{m=1}^{M} \prod_{c=1}^{C_m} \theta_{gmc}^{\mathbb{1}\{X_{nm}=c\}}$$

For a specified value G, the set of LCA model parameters is typically estimated by maximum likelihood by means of the Expectation-Maximization (EM) [13]. The algorithm is initialized with a random set of starting values. Therefore, it is usually recommended to run the procedure a bunch of times and then to pick the best solution [14]. More information about the model and parameter estimation can be found in [15][16] [17] and [14]. Concerning parameters interpretation, in the LCA model the parameter θ_{gmc} represents the probability of occurrence of attribute c for variable $\mathbf{X}_{\text{m}}$ in class g. Hence for the variables in the HUG dataset, θ_{gmc} will stand for the probability of having a certain criterion for each patient who belongs to the class g.

2.2.2. Model Selection

Various LCA models are being specified by the assignment of different values to G. For the purpose of selecting the optimal clustering model, various measures have been considered [18] and their performance were compared [19][20].Selecting the number of classes usually requires estimating models with incremental numbers of latent classes, and picking the number of classes based on the model that best fit the observed data. However, statistical criteria must always be assessed in combination with interpretability[21]. A class solution with better statistics is not of any use if it does not make any sense theoretically. Most current ways to decide the number of classes can be broken down into three categories: information-theoretic methods, likelihood ratio statistical test methods, and entropy-based criterion. Information criteria (ICs) are fitted indices that are frequently considered in a broad variety of statistical models and are used to make comparisons between a set of models. ICs consider model complexity into account and are also used to assess statistical fit. These indices comprise the Akaike Information Criterion (AIC) [22], the Consistent Akaike Information Criterion (cAIC)[23], the Bayesian Information Criterion (BIC) [24] and the adjusted Bayesian Information Criterion (aBIC)[25], where lower values denote a better fitting model. The AIC can be defined as:

$$AIC = -2LL + 2p,$$

where p is the number of free model parameters and LL the log-likelihood. The cAIC is a variant of the AIC but also punishes the value of -2 times the log-likelihood of the model for the number of free model parameters and sample size (Bozdogan, 1987). The cAIC is described as:

$$cAIC = -2LL + p[log(n) + 1],$$

where p is the number of free parameters and n is the sample size. The BIC also incorporates an adaptation for sample size and is given as follows:

$$BIC = -2LL + 2p\log(n),$$

where p is the number of free parameters and n is the sample size. Finally, aBIC is a by-product of BIC that decreases the penalty related to the sample size. The aBIC is defined as:

$$aBIC = -2LL + 2p\log[(n+2)/24],$$

where again p is the number of free parameters and n is the sample size.

The second type of methods for assessing model fit in latent class models involves likelihood ratio (LR) statistic tests. These tests compare the relative fit of two models that disagree in a set of parameter restrictions. For example, it compares a nested (n-1)-class solution to an (n)-class solution. The final category of the fit tests used to evaluate latent class models is the measure of entropy. The entropy index is based on the uncertainty of classification [26] [27]. Basically, the uncertainty of classification is evaluated at the individual level using the posterior probability; thus, entropy is a measure of the aggregated classification uncertainty. The uncertainty of classification is raised when the posterior probabilities are very close across classes. The normalized version of entropy, which scales to the interval [0, 1], is commonly used as a model selection criteria indicating the level of separation between classes. A higher value of normalized entropy represents a better fit; values > 0.80 indicate that the latent classes are highly discriminating [28].

3. RESULTS

3.1. LCA Results

A sequence of models was fitted to the data with the number of classes ranging from 1 to 12. The number of classes was determined by the evaluation of model fit indices (Table 1). Smaller values indicate better latent class separation except for entropy where values near 1 indicate better latent class separation. Regarding the relative goodness-of-fit indices, the value of BIC, aBIC, and cAIC continued to decline for the estimated models from the single-class model to the twelve-class model, while they reached a flattening from the five-class model onwards. However, there was no substantial improvement in either BIC, aBIC or cAIC fit beyond models with nine to twelve classes indicated by the elbow-shaped curve in Figure 2.Moreover, upon examination, the eight-class model appeared to have a meaningful interpretation. Therefore, based on a trade-off between several fitting indices, parsimony, and interpretability of the model, the eight-class model was retained as the final model.

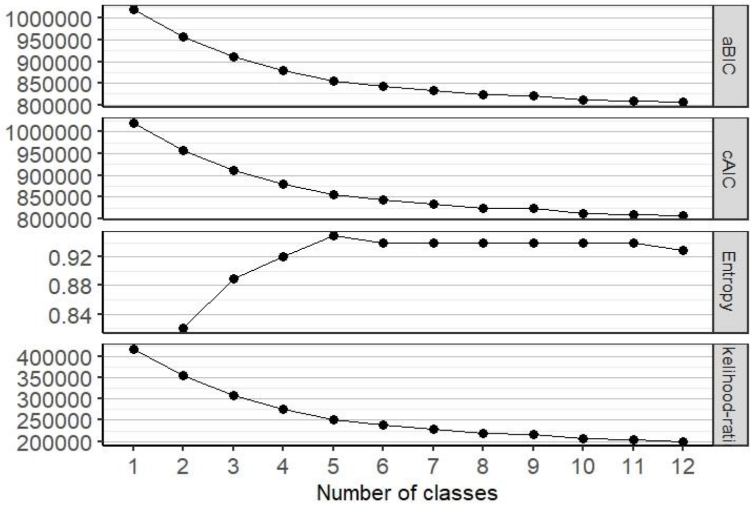


Figure 2. Plots showing goodness of fit with varying number of classes

Table 1. Fit Statistics for Latent Class Analyses

Number of latent class	**Number of parameters estimated**	**BIC**	**aBIC**	**cAIC**	**Entropy**	**likelihood-ratio**
1	32701	1020364.3	1020179.9	1020422.3	-	419028.1
2	32642	957342.3	956970.4	957459.3	0.82	355392.7
3	32583	910366.9	909807.6	910542.9	0.89	307804.0
4	32524	880006.5	879259.7	880241.5	0.92	276830.1
5	32465	855091.5	854157.2	855385.5	0.95	251301.7
6	32406	842763.4	841641.6	843116.4	0.94	238360.2
7	32347	834598.9	833289.6	835010.9	0.94	229582.2
8	32288	824737.8	823240.9	825208.8	0.94	219107.7
9	32229	818230.5	816546.2	818760.5	0.94	211987.0
10	32170	813011.7	811139.9	813600.7	0.94	206154.8
11	32111	810219.2	808159.9	810867.2	0.94	202748.9
12	32052	808030.0	805783.2	808737.0	0.93	199946.2

BIC: Bayesian information criterion; aBIC: adjusted Bayesian information criterion; cAIC: consistant Akaike information criterion

3.2. Results for the groups

3.2.1. Results for demographics and mode of admission and discharge from hospital

32,759 unique patients across 8 groups were identified by the clustering method. The number of patients per group ranges from 2,735 (8.4%) to 5,711 (17.4%) with an average of 4,095. Groups 6 and 8 have only single patients (N = 5,927; 18.1%) and group 3 has only women patients (N = 3,981; 12.2%) as described in table 2 below.

Table 2. Gender and status distribution of patients per groups

	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
Men	2622 (45.9%)	1694 (43.3%)	0 (0%)	2236 (55.2%)	2084 (60.0%)	1586 (49.7%)	2774 (48.7%)	1589 (58.1%)	14585 (44.5%)
Women	3089 (54.1%)	2215 (56.7%)	3981 (100%)	1818 (44.8%)	1392 (40.0%)	1606 (50.3%)	2927 (51.3%)	1146 (41.9%)	18174 (55.5%)
Single	2706 (47.4%)	2504 (64.1%)	1570 (39.4%)	2495 (61.5%)	1889 (54.3%)	3192 (100%)	3214 (56.4%)	2735 (100%)	20305 (62.0%)
Couple	3005 (52.6%)	1405 (35.9%)	2411 (60.6%)	1559 (38.5%)	1587 (45.7%)	0 (0%)	2487 (43.6%)	0 (0%)	12454 (38.0%)

The patients' age showed a bimodal distribution with a first mode in the 0 to 18 age range (N = 6781; 20.7%) and the second mode in the 75 and above age range (7107; 21.7%). Groups 6 and 8 include mostly young patients less than 19 years of age (99.8% and 81.7% respectively). Group 3 includes nearly only young adult patients from age 19 to 44 (99.2%) while group 7 has a majority of older adults from age 75 and above (57.5%) and no young patients (less than 18 years old) as described in table 3 below.

Table 3. Age bracket distribution of patients per groups

Age bracket	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
[0,18]	161 (2.8%)	373 (9.5%)	13 (0.3%)	566 (14.0%)	248 (7.1%)	3185 (99.8%)	0 (0%)	2235 (81.7%)	6781 (20.7%)
(18,34]	858 (15.0%)	540 (13.8%)	2687 (67.5%)	699 (17.2%)	109 (3.1%)	0 (0%)	182 (3.2%)	347 (12.7%)	5422 (16.6%)
(34,44]	1008 (17.7%)	333 (8.5%)	1260 (31.7%)	503 (12.4%)	145 (4.2%)	4 (0.1%)	285 (5.0%)	78 (2.9%)	3616 (11.0%)
(44,54]	1152 (20.2%)	371 (9.5%)	21 (0.5%)	663 (16.4%)	373 (10.7%)	3 (0.1%)	451 (7.9%)	67 (2.4%)	3101 (9.5%)
(54,64]	1107 (19.4%)	407 (10.4%)	0 (0%)	704 (17.4%)	600 (17.3%)	0 (0%)	648 (11.4%)	2 (0.1%)	3468 (10.6%)
(64,74]	766 (13.4%)	409 (10.5%)	0 (0%)	476 (11.7%)	750 (21.6%)	0 (0%)	857 (15.0%)	6 (0.2%)	3264 (10.0%)
(74,150]	659 (11.5%)	1476 (37.8%)	0 (0%)	443 (10.9%)	1251 (36.0%)	0 (0%)	3278 (57.5%)	0 (0%)	7107 (21.7%)

Admissions to the HUG were done majorly via the emergency department (ED) for all the groups (55.7%) with groups 2 and 7 at 92.5% and 93.3% respectively. Group 6 was the exceptionwith only 50 patients out of 3,142 (1.6%) admitted via the ED. On the average 78.3% of all patients (N = 25,654) were discharged to home with the exception of group 5 with only 49.5% discharged to home (N = 1719). Groups 5 and 7 had the most patients transferred to rehabilitation with 32.1% and 23.4% respectively; while groups 1, 3,6 and 8 had the least with 0.4%, 0.1%, 0.0% and 0.8% respectively. These results are tabulated in table 4 below.

Table 4. Mode of admission and discharge from hospital for patients per groups

	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
ED	1766 (30.9%)	3617 (92.5%)	2686 (67.5%)	1313 (32.4%)	2212 (63.6%)	50 (1.6%)	5319 (93.3%)	1299 (47.5%)	18262 (55.7%)
Not ED	3945 (69.1%)	292 (7.5%)	1295 (32.5%)	2741 (67.6%)	1264 (36.4%)	3142 (98.4%)	382 (6.7%)	1436 (52.5%)	14497 (44.3%)
Home	5547 (97.1%)	2185 (55.9%)	3947 (99.1%)	3510 (86.6%)	1719 (49.5%)	3160 (99.0%)	3037 (53.3%)	2549 (93.2%)	25654 (78.3%)
Rehab	21 (0.4%)	669 (17.1%)	2 (0.1%)	414 (10.2%)	1116 (32.1%)	1 (0.0%)	1336 (23.4%)	22 (0.8%)	3581 (10.9%)
Others	143 (2.5%)	1055 (27.0%)	32 (0.8%)	130 (3.2%)	641 (18.4%)	31 (1.0%)	1328 (23.3%)	164 (6.0%)	3524 (10.8%)

3.2.2. Results for diagnoses, procedures and Elixhauser index

Groups 1 and 4 show a range of precisely targeted procedures (such as obstetric technics and operations on musculoskeletal system) and primary diagnoses (such as diseases of the digestive system) while groups 2 and 6 show no procedures done in 2019. In addition, group 6 shows a majority (61.1%) of diagnoses related to factors influencing the health status and reasons to access the health system.

35.4% of group 1 patients received operations of the digestive system with 30.9% of patients diagnosed with a disease of the digestive system. Of all patients with operations of the digestive systems (N = 3003), group 1 includes 67.4% patients (N = 2024) and of all patients with a primary diagnosis of disease of the digestive system (N = 2774), group 1 includes 63.6% patients (N = 1763).

90.3% of group 4 patients received operations of the musculoskeletal system with 49.5% of patients diagnosed with a disease of the musculoskeletal system and 47.1% with traumatic lesions. Of all patients with operations of the musculoskeletal system (N = 4473) group 4 includes 81.8% patients (N = 3660) and of all patients with a primary diagnosis of disease of the musculoskeletal system or traumatic lesions (N = 6407), group 4 includes 61.1% patients (N = 3917).94.0% of the patients (N = 3,742) in group 3 (women only patients) received obstetric procedures.

These results are summarized in tables 5 and 6 below.

Table 5. Distribution of procedure categories for patients by groups

Procedure categories	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
Operations on the nervous system	196 (3.4%)	0 (0%)	10 (0.3%)	111 (2.7%)	177 (5.1%)	0 (0%)	116 (2.0%)	83 (3.0%)	693 (2.1%)
Operations on the urinary system	440 (7.7%)	0 (0%)	0 (0%)	1 (0.0%)	129 (3.7%)	0 (0%)	195 (3.4%)	37 (1.4%)	802 (2.4%)
Operations on male genital organs	314 (5.5%)	0 (0%)	0 (0%)	0 (0%)	15 (0.4%)	0 (0%)	0 (0%)	47 (1.7%)	376 (1.1%)
Operations on female genital organs	688 (12.0%)	0 (0%)	198 (5.0%)	0 (0%)	25 (0.7%)	0 (0%)	0 (0%)	2 (0.1%)	913 (2.8%)
Obstetric techniques	0 (0%)	0 (0%)	3742 (94.0%)	0 (0%)	1 (0.0%)	0 (0%)	0 (0%)	0 (0%)	3743 (11.4%)
Operations on musculoskeletal system	0 (0%)	0 (0%)	0 (0%)	3660 (90.3%)	504 (14.5%)	0 (0%)	294 (5.2%)	15 (0.5%)	4473 (13.7%)
Operations on integumentary system	424 (7.4%)	0 (0%)	1 (0.0%)	204 (5.0%)	73 (2.1%)	0 (0%)	107 (1.9%)	29 (1.1%)	838 (2.6%)
Diagnostic and therapeutic techniques	572 (10.0%)	0 (0%)	24 (0.6%)	22 (0.5%)	828 (23.8%)	0 (0%)	4158 (72.9%)	1841 (67.3%)	7445 (22.7%)
Operations of the nose, mouth and pharynx	317 (5.6%)	0 (0%)	0 (0%)	0 (0%)	23 (0.7%)	0 (0%)	8 (0.1%)	370 (13.5%)	718 (2.2%)
Operations of respiratory system	147 (2.6%)	0 (0%)	0 (0%)	10 (0.2%)	127 (3.7%)	0 (0%)	107 (1.9%)	49 (1.8%)	440 (1.3%)
Operations of cardiovascular system	195 (3.4%)	0 (0%)	3 (0.1%)	41 (1.0%)	672 (19.3%)	0 (0%)	228 (4.0%)	103 (3.8%)	1242 (3.8%)
Operations of digestive system	2024 (35.4%)	0 (0%)	2 (0.1%)	0 (0%)	736 (21.2%)	0 (0%)	142 (2.5%)	99 (3.6%)	3003 (9.2%)
Other classified procedures	377 (6.6%)	0 (0%)	0 (0%)	5 (0.1%)	103 (3.0%)	0 (0%)	0 (0%)	57 (2.1%)	542 (1.7%)
Procedures non classified elsewhere	17 (0.3%)	0 (0%)	1 (0.0%)	0 (0%)	62 (1.8%)	0 (0%)	346 (6.1%)	3 (0.1%)	429 (1.3%)
No procedure	0 (0%)	3909 (100%)	0 (0%)	0 (0%)	1 (0.0%)	3192 (100%)	0 (0%)	0 (0%)	7102 (21.7%)

Table 6. Distribution of diagnosis categories for patients by groups

	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
Certain infectious and parasitic diseases	34 (0.6%)	117 (3.0%)	0 (0%)	0 (0%)	202 (5.8%)	2 (0.1%)	255 (4.5%)	77 (2.8%)	687 (2.1%)
Tumors	1426 (25.0%)	42 (1.1%)	0 (0%)	29 (0.7%)	775 (22.3%)	2 (0.1%)	195 (3.4%)	0 (0%)	2469 (7.5%)
Diseases of the blood, hematopoietic organs, immunity system	22 (0.4%)	23 (0.6%)	0 (0%)	0 (0%)	29 (0.8%)	0 (0%)	91 (1.6%)	27 (1.0%)	192 (0.6%)
Endocrinien, metabolic and nutritionel diseases	274 (4.8%)	103 (2.6%)	0 (0%)	0 (0%)	98 (2.8%)	6 (0.2%)	154 (2.7%)	35 (1.3%)	670 (2.0%)
Diseases of the circulatory system	208 (3.6%)	410 (10.5%)	0 (0%)	0 (0%)	771 (22.2%)	0 (0%)	1820 (31.9%)	32 (1.2%)	3241 (9.9%)
Mental and behavior diseases	4 (0.1%)	409 (10.5%)	0 (0%)	0 (0%)	15 (0.4%)	0 (0%)	177 (3.1%)	41 (1.5%)	646 (2.0%)
Diseases of the nervous system	231 (4.0%)	122 (3.1%)	0 (0%)	9 (0.2%)	85 (2.4%)	0 (0%)	345 (6.1%)	119 (4.4%)	911 (2.8%)
Diseases of the eyes	95 (1.7%)	69 (1.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	20 (0.4%)	50 (1.8%)	234 (0.7%)
Diseases of the respiratory system	235 (4.1%)	495 (12.7%)	0 (0%)	0 (0%)	130 (3.7%)	2 (0.1%)	909 (15.9%)	808 (29.5%)	2579 (7.9%)
Diseases of the digestive system	1763 (30.9%)	286 (7.3%)	0 (0%)	20 (0.5%)	421 (12.1%)	4 (0.1%)	193 (3.4%)	87 (3.2%)	2774 (8.5%)
Diseases of the skin and subcutaneous tissue	117 (2.0%)	84 (2.1%)	0 (0%)	28 (0.7%)	36 (1.0%)	1 (0.0%)	68 (1.2%)	36 (1.3%)	370 (1.1%)
Diseases of the musculoskeletal system	0 (0%)	214 (5.5%)	0 (0%)	2006 (49.5%)	157 (4.5%)	1 (0.0%)	170 (3.0%)	38 (1.4%)	2586 (7.9%)
Diseases of the urinary track system	970 (17.0%)	237 (6.1%)	0 (0%)	1 (0.0%)	107 (3.1%)	6 (0.2%)	214 (3.8%)	66 (2.4%)	1601 (4.9%)
Traumatic lesions, poisoning and other external cause of illness	81 (1.4%)	609 (15.6%)	0 (0%)	1911 (47.1%)	461 (13.3%)	1 (0.0%)	619 (10.9%)	139 (5.1%)	3821 (11.7%)
Pregnancy and delivery	0 (0%)	101 (2.6%)	3981 (100%)	0 (0%)	4 (0.1%)	1 (0.0%)	1 (0.0%)	0 (0%)	4088 (12.5%)
Perinatal related illness	0 (0%)	1 (0.0%)	0 (0%)	0 (0%)	12 (0.3%)	1037 (32.5%)	0 (0%)	782 (28.6%)	1832 (5.6%)
Genetic malformations and chromosomic abnormalities	26 (0.5%)	1 (0.0%)	0 (0%)	49 (1.2%)	110 (3.2%)	154 (4.8%)	0 (0%)	212 (7.8%)	552 (1.7%)
Abnormal results from exams and labs non classified elsewhere	92 (1.6%)	526 (13.5%)	0 (0%)	1 (0.0%)	53 (1.5%)	26 (0.8%)	467 (8.2%)	118 (4.3%)	1283 (3.9%)
Factors influencing health status and reasons to access health system	133 (2.3%)	60 (1.5%)	0 (0%)	0 (0%)	10 (0.3%)	1949 (61.1%)	3 (0.1%)	68 (2.5%)	2223 (6.8%)

The Elixhauser comorbidity index was calculated for each patient based on their diagnosis codes. The distribution per group for chronic heart failure (CHF), cardiovascular disease (CARIT), chronic obstructive pulmonary disease (COP), and diabetes (DIABC) do not show any significance difference across the groups. The proportion of patients across the groups exhibiting each conditions are very homogeneous as described in table 7 below.

Table 7. Distribution Elixhauser comorbidity index for selected conditions for patients by groups

	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
CHF									
0	5306 (92.9%)	3632 (92.9%)	3744 (94.0%)	3794 (93.6%)	3260 (93.8%)	2982 (93.4%)	5340 (93.7%)	2574 (94.1%)	30632 (93.5%)
1	405 (7.1%)	277 (7.1%)	237 (6.0%)	260 (6.4%)	216 (6.2%)	210 (6.6%)	361 (6.3%)	161 (5.9%)	2127 (6.5%)
CARIT									
0	5062 (88.6%)	3432 (87.8%)	3545 (89.0%)	3583 (88.4%)	3080 (88.6%)	2827 (88.6%)	5059 (88.7%)	2421 (88.5%)	29009 (88.6%)
1	649 (11.4%)	477 (12.2%)	436 (11.0%)	471 (11.6%)	396 (11.4%)	365 (11.4%)	642 (11.3%)	314 (11.5%)	3750 (11.4%)
COPD									
0	5412 (94.8%)	3716 (95.1%)	3801 (95.5%)	3866 (95.4%)	3313 (95.3%)	3049 (95.5%)	5435 (95.3%)	2589 (94.7%)	31181 (95.2%)
1	299 (5.2%)	193 (4.9%)	180 (4.5%)	188 (4.6%)	163 (4.7%)	143 (4.5%)	266 (4.7%)	146 (5.3%)	1578 (4.8%)
DIABC									
0	5143 (90.1%)	3549 (90.8%)	3589 (90.2%)	3663 (90.4%)	3192 (91.8%)	2898 (90.8%)	5182 (90.9%)	2497 (91.3%)	29713 (90.7%)
1	568 (9.9%)	360 (9.2%)	392 (9.8%)	391 (9.6%)	284 (8.2%)	294 (9.2%)	519 (9.1%)	238 (8.7%)	3046 (9.3%)

3.2.3. Results for top 10 percentile of cost and clinical outcomes

Group 5 (N = 3,476) had 80.5% of its patients in the top 10 percentile for total costs compared to all the other groups combined with 3.0% of their patients (N = 883).

Group 5 patients had the most number of patients with more than 10 ambulatory visits (42.9%), more than 10 different diagnoses (69.9%), more than 3 procedures (90.5%), more than 10 lab tests (80.2%), more than 10 medications (96.3%), and more than 2 hospitalizations (23.9%).

Group 5 had also the most number of patients who were discharged to rehabilitation facilities after their hospital stay (32.1%).

More group 5 patients were 65 years and older (N = 2,001; 57.6%) than any other groups except group 7 (N = 4,135; 72.5%). While group 7 had more patients 65 years and older than group 5, it also had no patient less than 19 years of age while group 5 had 248 patients (7.1%).

Group 7 provides some other results which are noteworthy. After group 5, it has the most number of patients (N = 371; 6.5%) in the top 10 percentile of costs; with more than 10 diagnoses (N = 2,028; 35.6%); with more than 10 tests (N = 2,649; 46.5%); and with more than 10 medications (N = 4,628; 81.2%).

These results are tabulated in table 8 below.

Table 8. Distribution of costs percentile and clinical outcomes for patients per groups

	Group 1 (N = 5,711)	Group 2 (N = 3,909)	Group 3 (N = 3,981)	Group 4 (N = 4,054)	Group 5 (N = 3,476)	Group 6 (N = 3,192)	Group 7 (N = 5,701)	Group 8 (N = 2,735)	All Groups (N = 32,759)
Percentile distribution of costs									
Top 10th percentile	133 (2.3%)	32 (0.8%)	13 (0.3%)	157 (3.9%)	2797 (80.5%)	11 (0.3%)	371 (6.5%)	166 (6.1%)	3680 (11.2%)
Bottom 90th percentile	5578 (97.7%)	3877 (99.2%)	3968 (997%)	3897 (96.1%)	679 (19.5%)	3181 (99.7%)	5330 (93.5%)	2569 (93.9%)	29079 (88.8%)
Ambulatory visits									
0 - 4	2320 (40.6%)	2660 (68.0%)	1998 (50.2%)	1073 (26.5%)	1075 (30.9%)	3076 (96.4%)	3828 (67.1%)	2115 (77.3%)	18145 (55.4%)
5 - 10	1758 (30.8%)	647 (16.6%)	1294 (32.5%)	1907 (47.0%)	910 (26.2%)	95 (3.0%)	1085 (19.0%)	407 (14.9%)	8103 (24.7%)
> 10	1633 (28.6%)	602 (15.4%)	689 (17.3%)	1074 (26.5%)	1491 (42.9%)	21 (0.7%)	788 (13.8%)	213 (7.8%)	6511 (19.9%)
Hospital admissions									
1	4919 (86.1%)	3359 (85.9%)	3663 (92.0%)	3695 (91.1%)	1623 (46.7%)	3117 (97.7%)	4495 (78.8%)	2362 (86.4%)	27233 (83.1%)
2	619 (10.8%)	420 (10.7%)	273 (6.9%)	311 (7.7%)	1021 (29.4%)	71 (2.2%)	860 (15.1%)	333 (12.2%)	3908 (11.9%)
> 2	173 (3.0%)	130 (3.3%)	45 (1.1%)	48 (1.2%)	832 (23.9%)	4 (0.1%)	346 (6.1%)	40 (1.5%)	1618 (4.9%)
Number of diagnoses									
1	1433 (25.1%)	600 (15.3%)	10 (0.3%)	553 (13.6%)	3 (0.1%)	1680 (52.6%)	38 (0.7%)	547 (20.0%)	4864 (14.8%)
2 - 10	4210 (73.7%)	2773 (70.9%)	3804 (95.6%)	3489 (86.1%)	1043 (30.0%)	1512 (47.4%)	3635 (63.8%)	2124 (77.7%)	22590 (69.0%)
> 10	68 (1.2%)	536 (13.7%)	167 (4.2%)	12 (0.3%)	2430 (69.9%)	0 (0%)	2028 (35.6%)	64 (2.3%)	5305 (16.2%)
Number of treatments									
0	0 (0%)	3908 (100.0%)	0 (0%)	0 (0%)	0 (0%)	3192 (100%)	0 (0%)	0 (0%)	7100 (21.7%)
1 - 2	4059 (71.1%)	0 (0%)	2665 (66.9%)	3194 (78.8%)	329 (9.5%)	0 (0%)	4499 (78.9%)	2151 (78.6%)	16897 (51.6%)
> 3	1652 (28.9%)	1 (0.0%)	1316 (33.1%)	860 (21.2%)	3147 (90.5%)	0 (0%)	1202 (21.1%)	584 (21.4%)	8762 (26.7%)
Number of labs (Tests)									
1 - 10	5241 (91.8%)	3275 (83.8%)	3884 (97.6%)	3948 (97.4%)	689 (19.8%)	3181 (99.7%)	3052 (53.5%)	2639 (96.5%)	25909 (79.1%)
> 10	470 (8.2%)	634 (16.2%)	97 (2.4%)	106 (2.6%)	2787 (80.2%)	11 (0.3%)	2649 (46.5%)	96 (3.5%)	6850 (20.9%)
Number of medications									
1 - 10	2526 (44.2%)	2100 (53.7%)	2859 (71.8%)	2294 (56.6%)	130 (3.7%)	3192 (100%)	1073 (18.8%)	2315 (84.6%)	16489 (50.3%)
> 10	3185 (55.8%)	1809 (46.3%)	1122 (28.2%)	1760 (43.4%)	3346 (96.3%)	0 (0%)	4628 (81.2%)	420 (15.4%)	16270 (49.7%)

4. DISCUSSION

This study was conducted to determine how cluster analysis using the SU criterion used in the literature and by Grafe et al [29]might applied to the inpatient population of the Hôpitaux Universitaires de Genève. The results show that the LCA clustering model is able to generate 8 groups with distinctive characteristics. In particular, the algorithm was able to identify a group with mostly patients less than 19 years of age who use the hospital for health related factors but not serious illness as well as a group with only women who use the hospital for only women related procedures and diagnoses and two other groups whose patients are greater utilizers of digestive and musculoskeletal procedures with consistent related diagnoses. Across and among the groups the results for the variables studied appear highly coherent and as would be expected demonstrating that the clustering algorithm appears robust in stratifying the population of patients admitted to the Hôpitaux Universitaires de Genève in 2019.

Most important among the 8 groups was the group of patients whose costs are in the top 10th percentile (Group 5) and for whom the use of ambulatory and inpatient services is the greatest as well as the use of treatments, test (labs) and medications. Given the consistency of the results for these patients and the coherence we observed across the other groups (described above), we are confident that group 5 represent the super-utilizers of care for the HUG in 2019.

In this investigation, we demonstrated the use of cluster analysis to identify distinct subgroups of patients with specific combinations of co-occurring conditions in a large academic medical center.

The model revealed the expected segmentation by age brackets and gender such as with group 6 (patients less than 19 years of age) and group 3 (women only patients) along with the expected utilisation of care services such as pregnancy and delivery for group 3. The identification of these expected groups in our analysis provide assurance of the validity of our data mining method.

The cluster analysis provided also a data driven approach to identifying at least 3very distinctive clinically relevant groups of patients with patterns of care utilization that could be targeted with new, enhanced care management strategies. The super-utilizers (group 5, N = 3476, 10.6% of all patients) including mostly SUs (N = 2797, 8.5% of all patients). The patients who consistently (93.3% of patients) are admitted via the ED (group 7, N = 5701, 17.4% of all patients) including at least some SUs (N = 371, 1.0% of all patients). The musculoskeletal patients (group 4, N = 4054, 12.4% of all patients) whose care and costs are mostly related to problems associated with the musculoskeletal system including at least some SUs (N = 157, 0.5% of all patients).Together these 3 groups (N = 12947, 39.5% of all patients) which alone contain the majority of SUs (N = 3325, 10% of all patients), and considering only the SUs, account for all costs above the 90% percentile which means that targeted intervention to improve the care of these patients will have the most impact on total costs for the HUG.

While the model appears coherent and robust to further assess the stability of these clusters over time, analyses should be conducted on cohorts from different years. A larger population of patients (over multiple years) might also provide more power to detect significant difference in the Elixhauser comorbidity index across groups.

Like any investigation, the characteristics of our clusters are constrained to our data and setting. Reproducing these analyses in different settings and different patient populations may potentially yield different clusters. However, these differences would and should nevertheless inform on different management strategies specific to populations in those settings.

In this study we showed how cluster analysis can be used to identify homogeneous groups of complex patients from a large heterogeneous population. Such data science methods demonstrate that it is possible to use the conceptual findings of this investigation to raise awareness of the need for a more personalized approach of care management services for patients with high levels of healthcare utilization (super utilizers). However, further understanding of the care management needs of clusters of patients with similar comorbidities and care utilization is warranted before designing specific tailored interventions.

5. CONCLUSIONS

This study identified SU criterion that have commonly been used in the literature and applied these criterion to the inpatient population of a large academic medical center. The procedures and results reported illustrate how cluster analysis can be helpful in differentiating homogeneous groups of complex patients from a large heterogeneous population. These results should help in the application of more targeted interventions per subgroups to improve appropriateness of care, improve outcomes and reduce costs.

ACKNOWLEDGEMENTS

The medical directorate, the service for quality and patient safety, and the department of finance of the Hôpitaux Universitaires de Genève provided useful insights into the broad focus and goals of this study. We appreciate their inputs, especially from our colleagues in the medico-economic unit, which were most helpful and constructive for the analysis and writing of this paper. This study will not have been possible without the assistance and collaboration of all these people.

REFERENCES

[1] Spending on health: Latest trends. Health Spending Latest Trend Brief.2018. [En ligne]. Available: https://www.oecd.org/health/health-systems/Health-Spending-Latest-Trends-Brief.pdf.

[2] R. W. Raghupathi. Healthcare Expenditure and Economic Performance: Insights From the United States Data. Front Public Health, vol. 8, 2020.

[3] R. Axon, M. Williams. Hospital readmission as an accountability measure. JAMA, vol. 305, pp. 504-5, 2011.

[4] H. A. Purdey. Predicting and preventing avoidable hospital admissions: a review. The journal of the Royal College of Physicians of Edinburgh, vol. 43, pp. 340-4, 2013.

[5] C. Schwierz. Cost-Containment Policies in Hospital Expenditure in the European Union. 2016.

[6] M. Berk, A. Monheit. The concentration of health care expenditures, revisited. Health Affairs (Project Hope)., vol. 20, pp. 9-18, 2001.

[7] S. Rais, A. Nazerian, S. Ardal, Y. Chechulin, N. Bains, K. Malikov, A. Nazerian. High-cost users of Ontario's healthcare services.Healthc Policy., vol. 9, pp. 44-51, 2013.

[8] W. P. Wodchis, P. C. Austin, D. A. Henry. A 3-year study of high-cost users of health care.CMAJ : Canadian Medical Association journal., vol. 188, pp. 182-188, 2016.

[9] S. B. Cohen. The Concentration of Health Care Expenditures and Related Expenses for Costly Medical Conditions, 2012.Statistical Brief (Medical Expenditure Panel Survey (US)), Agency for Healthcare Research and Quality (US), 2014.

[10] P. Besson. Manuel Rekole® Comptabilité Analytique A L'hôpital. Les Hôpitaux de Suisse, 2013.

[11] H. Organization. ICD-10 Classification of Mental and Behavioral Disorders (The): Diagnostic Criteria for Research. World Health Organization, 1993.

[12] C. C. Clogg. Latent Class Models for Measuring. Latent Trait and Latent Class Models, pp. 173-205, 1988.

[13] A. Dempster, N. Laird, D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society, series B, vol. 39, pp. 1-38, 1977.

[14] D. J. Bartholomew, K. Martin, I. Moustaki. Latent Variable Models and Factor Analysis: A Unified Approach. Wiley Series in Probability and Statistics, 2011.

[15] P. F. Lazarsfeld, N. W. Henry. Latent Structure Analysis, Houghton, Mifflin, 1968.

[16] C. C. Clogg. Latent class models, chez Handbook of statistical modeling for the social and behavioral sciences, 1995.

[17] A. Agresti. Categorical data analysis, New York: Wiley, 2002.

[18] G. McLachlan, D. Peel. Finite mixture models, New York: Wiley, 2000.

[19] J. G. Dias. Latent Class Analysis and Model Selection, chez From Data and Information Analysis to Knowledge Engineering, 2006.

[20] C. Biernacki. Model selection theory and considerations in large scale scenarios, chez Research Summer School on Statistics for Data Science - S4D, Caen France, 2018.

[21] B. O. Muthén, L. K. Muthén. Integrating person-centered and variablecentered analyses: Growth mixture modeling with latent trajectory classes.Alcoholism: Clinical & Experimental Research, vol. 24, pp. 882-891, 2000.

[22] H. Akaike. A new look at the statistical model identification. IEEE Transactions on Automatic Control, vol. 19, pp. 716-723, 1974.

[23] H. Bozdogan. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika, vol. 52, pp. 345-370, 1987.

[24] G. Schwarz. Estimating the dimension of a model. Annals of Statistics, vol. 6, pp. 461-464, 1978.

[25] S. L. Sclove. Application of model-selection criteria to some problems in multivariate. Psychometrika, vol. 52, pp. 333-343, 1987.

[26] M.-C. Wang, Q. Deng, X. Bi, H. Ye, W. Yang. Performance of the entropy as an index of classification accuracy in latent profile analysis: A Monte Carlo simulation study. Acta Psychologica Sinica, vol. 49, pp. 1473-1482, 2017.

[27] G. Celeux, G. Soromenho. An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification, vol. 13, pp. 195-212, 1996.

[28] C. Larose, O. Harel, K. Kordas, D. Dey. Latent Class Analysis of Incomplete Data via an Entropy-Based Criterion. Stat Methodology, vol. 32, pp. 107-121, 2016.

[29] Grafe, Carl J et al. “How to Classify Super-Utilizers: A Methodological Review ofSuper-Utilizer Criteria Applied to the Utah Medicaid Population, 2016-2017.Population health management vol. 23,2 (2020): 165-173.

A COMPREHENSIVE STUDY ON VARIOUS STATISTICAL TECHNIQUES FOR PREDICTION OF MOVIE SUCCESS

*Manav Agarwal, Shreya Venugopal, Rishab Kashyap, R Bharathi

Department of CSE, PES University, Bangalore, India

ABSTRACT

The film industry is one of the most popular entertainment industries and one of the biggest markets for business. Among the contributing factors to this would be the success of a movie in terms of its popularity as well as its box office performance. Hence, we create a comprehensive comparison between the various machine learning models to predict the rate of success of a movie. The effectiveness of these models along with their statistical significance is studied to conclude which of these models is the best predictor. Some insights regarding factors that affect the success of the movies are also found. The models studied include some Regression models, Machine Learning models, a Time Series model and a Neural Network with the Neural Network being the best performing model with an accuracy of about 86%. Additionally, as part of the testing data for the movies released in 2020 are analysed.

KEYWORDS

Machine Learning models, Time Series, Movie Success, Neural Network, Statistical significance.

1. INTRODUCTION

One of the most important contributing factors to the entertainment industry are movies, which is one of the highest revenue-generating businesses from a commercial perspective [1]. A majority of the population love to watch a variety of movies, and their choices are determined based on the various factors that contribute to the type of movie such as the genre of the movie. Most people thus look into the ratings of a given movie before they proceed to watch it to identify it as a movie which is worth watching. These ratings come from a variety of sources, some of which include popular websites such as Rotten Tomatoes, IMDb and many more. Thus, the analysis involves the study of these user ratings and the other factors that affect the movie and this enables the prediction of whether a movie is truly a successful one or not.

2. LITERATURE SURVEY

We begin our survey by selecting an appropriate dataset [2], since it proves to be the foundation for our project. We also apply data mining [3-4] to extract movies released in 2020. The steps that follow involve creating a study on the various features that complement a movie's success, and performing a thorough study on them as done by Abidi et. al [1]. These attributes are fed into a variety of regression [5] Machine Learning techniques [6-7] and Neural Network [8] models. We thus create a comparative study between them as done by Dhir and Raj [9] who have created a comparison specifically among Machine Learning models. Looking into the variety of models provided in the various papers [10-11], we predict the success rate [4] using a subset of these

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 17-30, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111802

with respect to the attributes chosen earlier. Furthermore, for each model, proving their statistical significance using a variety of tests [12-13] proves to be the highlight of our paper. In comparison with the accuracies provided in the papers [14] and their respective models [9], our models end up with a similar accuracy of 86% as shown through our papers.

3. PROPOSED METHODOLOGY

Similar to the methods mentioned above, through our paper we perform similar analytics on the IMDb dataset and create a comparison between the various models used [9], [11], and try to find an appropriate model as shown in Figure 1.

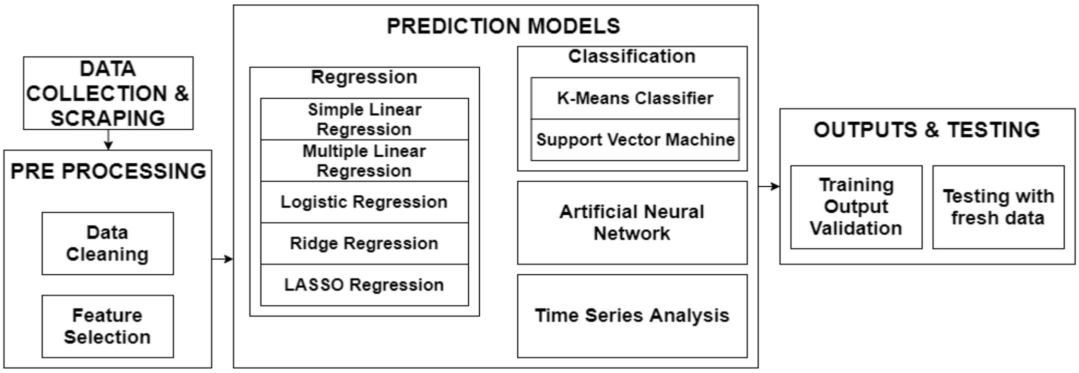


Figure 1. Proposed Methodology

4. DATASET DESCRIPTION

As mentioned above, we have considered the IMDb dataset [2] for our project. Some of the major attributes include the movie rating itself, followed by the genre, top voter ratings, total votes, duration of the movie and the release date. The minor attributes are cleaned in the pre-processing stage. We consider a total of 81274 movies for this process, split into training and testing according to the year of release in the steps that follow. To add on to the testing dataset, we have mined the official IMDb website for all the movies released in the year 2020, consisting of the same attributes to maintain uniformity.

5. PRE-PROCESSING

Traditional pre-processing methods such as removing null values, outliers and other basic steps have been applied. Apart from these, another pre-processing step used is the MultiLabelBinarizer [15] which is a method used to encode attributes that can belong to more than one category such as genre. Genres belonging to a movie are marked with the value 1, and the rest of the genres are marked with 0. This process is repeated for all movies until we have a list of numeric values depicting the genres that belong to the movie, and those that do not. We took all movies from 1990 to 2015 as the training dataset and all movies after 2015 as our validation dataset. The dataset contains an attribute called the “metascore” which ranges from 0 to 100 and is used as a measure to depict the success of a movie. A higher metascore implies that the movie was more successful. Hence this is used as the dependent variable for the complete study. To classify a movie as a success or a failure, the metascore value is divided into classes or bins, representing movies that are a hit or a flop or mediocre [16]. The predicted values can be classified under this partitioning.

6. REGRESSION METHODS

6.1. Simple Linear Regression

As we all know, Simple Linear Regression (SLR) is the most basic regression form. Since it involves only one independent variable against the metascore the most suitable attribute is selected. To do so, we use the Variance Inflation Factor (VIF) method to show the attributes with their multicollinearity with respect to each other. The highest value is considered to be the best prediction value, and in this case, we obtain the attribute to be the top1000_voters_ratings. We hence use this attribute in the SLR model and train the model with the existing values. The VIF ranges from values 0 in the case of "budget" to around 0.55 in the case of "top1000_voters_rating" as depicted in Figure 2.

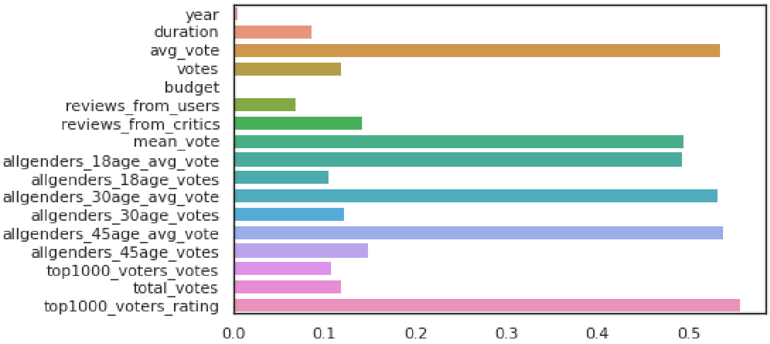


Figure 2. Diagram to show the VIF values for all attributes

Once we have our trained model, we use the testing dataset to predict the future values, and then make a comparison between the true and predicted values. The accuracy of the model is calculated using a confusion matrix. On completion of this, we proceed to perform tests on the model to prove its statistical significance.

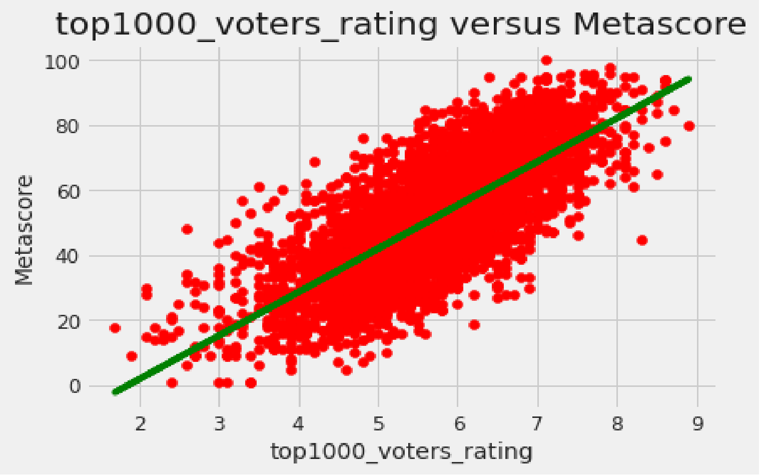


Figure 3. SLR model

We then test the same on our 2020 dataset, taking the "avg_vote" attribute i.e. the average rating of a movie instead of the "top100_voters_ratings" attribute in the previously trained model and get an accuracy of 0.5833. The model has already been tested for its significance and hence we do not test it again for the 2020 dataset.

6.2. Multiple Linear Regression

Multiple Linear regression (MLR) performs a process almost identical to SLR, but with multiple independent variables used to predict the same target variable. Similar to the procedure above, the target variable is predicted using the top1000_voters_rating attribute along with fifteen other attributes. Here, twelve of these attributes come under the MultiLabelBinarized values of the genre. The VIF test performed on them shows us the initial attributes that can be considered to plot the model. Once again, we perform the OLS tests and several hypothesis tests such as Jarque Bera test and Lagrange’s Multiplier test for attribute selection to show the significance of the values. The model that we get gives us an accuracy of 0.7116. An example of the MLR model in three dimensions using two independent variables is shown in Figure 4. We perform the same analysis using our 2020 dataset, taking the attributes avg_vote and duration for plotting our MLR model, and result with an accuracy of 0.608.

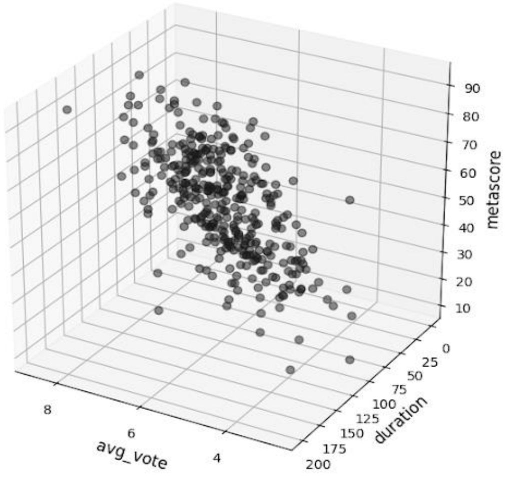


Figure 4. MLR model.

6.3. Logistic Regression

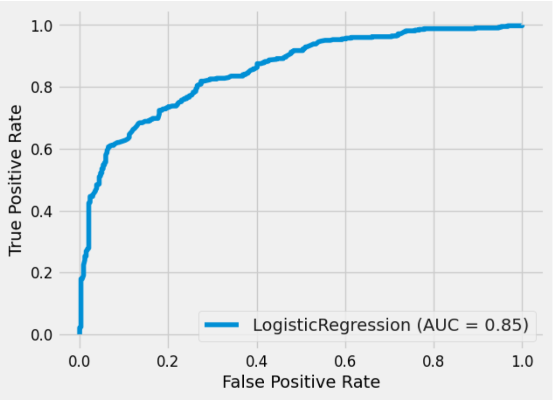


Figure 5. ROC-AUC curve for the logistic regression model

Logistic Regression is a non-linear regression and a statistical technique for finding the existence of a relationship between a qualitative and a quantitative dependent variable and several independent variables or explanatory variables [13]. The deviation we take from our originally defined steps here is, we divide our metascore into two domains, a successful or an unsuccessful movie based on the metascore value. This model results in an accuracy of 0.76. We plot an ROC-AUC curve as shown in Figure 5 using the confusion matrix that we have obtained. The same way we perform the Logistic analysis on our 2020 dataset to get an accuracy of 0.6833.

6.4. Regularization Techniques

Regularization is required to penalize certain features, and we have two regression methods for the same, namely the Ridge, and the LASSO regression models. Ridge regression is a technique used when the data suffers from high multicollinearity [12]. It uses the L2 regularization method to perform this process. The regularization parameter for Ridge Regression was 1150. The accuracy we get from this model comes up to 0.74. Plotting the same as in Fig. 6 using our 2020 dataset we thus get an accuracy of 0.61 where it can be inferred that the model is somewhat accurate however tests are conducted to validate the same.

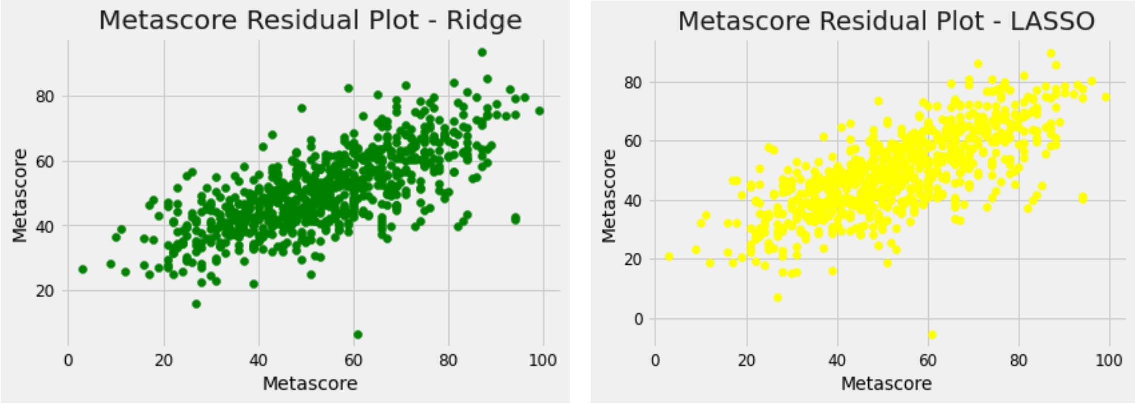


Figure 6. Ridge and LASSO regression model residual plot

As we did for Ridge, we follow the exact same procedure in LASSO regression, which along with L2 regularization, uses central tendency to penalize. The regularization parameter for LASSO was 0.145. For the training and testing process, calculating the statistical values as well as obtaining an accuracy of 0.72. Plotting the same using our 2020 dataset we thus get an accuracy of 0.59.

7. CLASSIFICATION METHODS

The classification models that we have used follow the unsupervised learning algorithm, rather they use self-learning techniques where there is no training data. The two classification models that we use here are the Support Vector Machine and the K-Means classifier.

7.1. K-Means Classifier

K-Means is a popular clustering technique and can also be used as a classifier where each cluster is considered as a class. Using the K-Means classifier in this case would involve making the value of K as 3 in this case for each of the metascore bins [12]. The accuracy of the K-Means model is 0.5. Statistical tests such as the silhouette test for a better model is performed. We plot as shown in Figure 7 the same for our 2020 dataset using our previously trained model and hence get an accuracy of 0.42.

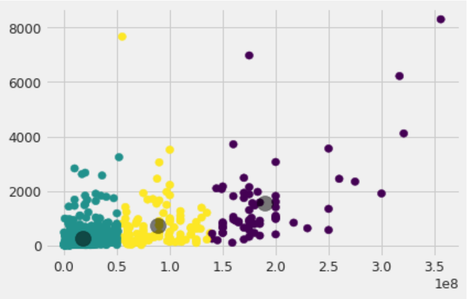


Figure 7. K-Means plot to show the final centroid positions

7.2. Support Vector Machine

Instead of the conventional binary Support Vector Machine (SVM) the one that classifies into 3 categories with two hyperplanes of separation is used. This is because the data is divided into 3 categories and it is such that they somewhat fall in a sequence. We can see that the model thus results in an accuracy of 0.71. The trained model is then used against the 2020 testing dataset to calculate the same and thus results in an accuracy of 0.62.

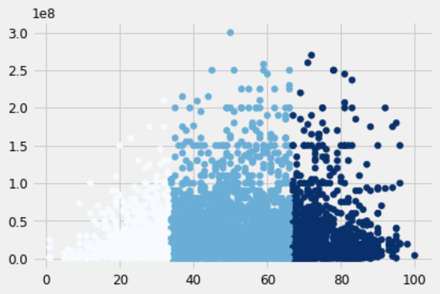


Figure 8. SVM models for our initial dataset

8. Time Series Analysis

We begin by defining the method of Forecasting, which is one of the most important and frequently used applications in analytics, which focuses on the prediction of future values based on the present and past values. We say that the variable is forecasted into future values to perform analysis on patterns such as trend, seasonality and so on. We make use of the SARIMAX model for our analysis. We first establish all the necessary parameters to plot our time series graphs. We then initialize the various attributes that will be used to determine the forecasted values for metascore. The time-wise analysis is based on the release date of the movie.

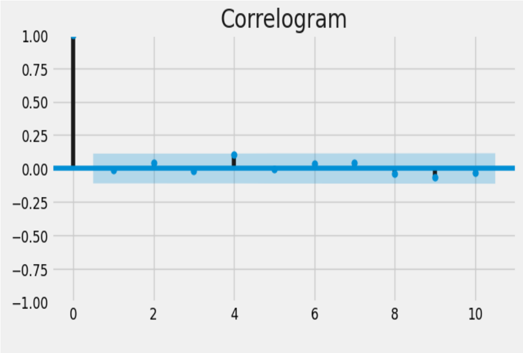


Figure 9. Correlogram Plot

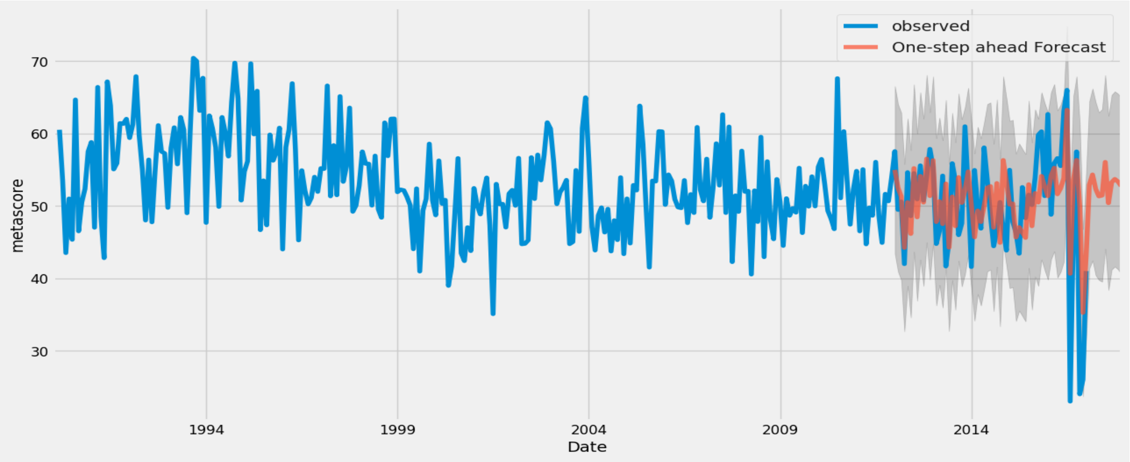


Figure 10. Residual plot with Forecasted values

On plotting the metascore value against the date_published attribute to get the time-series graph for the entire dataset, and then plotting the decomposition of the data into the trend, seasonality as well as the residual variations we obtain Figure 9 and Figure 10. To prove the stationarity of the data we use the Augmented Dickey-Fuller Test which revealed that the data is stationary. The presence of seasonality in the plot leads us to use the SARIMAX model. Initially we define the function to execute all possibilities for SARIMAX from which one model is selected based on its AIC value. We can see that the model gives us a roughly accurate result for the forecast. We plot the same for our 2020 dataset, using the SARIMAX model we used before to get similar outputs as in Figure 11.

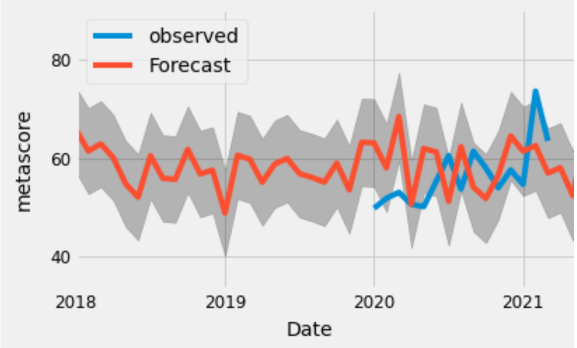


Figure 11. Plot of output using the model against the actual values.

9. Artificial Neural Network

In the Artificial Neural Network model, we have proposed, we give the inputs based on the statistical significance they hold from the tests we have performed in the previous cases. We obtain a final result of 86.16% accuracy from the model as well as minimal loss incurred through the process. The loss is also displayed in Figure 12. We have also tested this with the 2020 dataset obtained resulting in a 88.056 % accuracy.

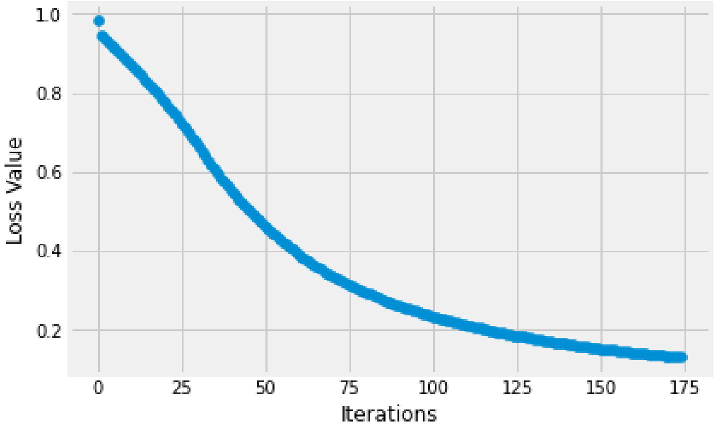


Figure 12. Loss Curve for the ANN

10. Results and Interpretation of Values

10.1. Regression

Wald's test for Logistic Regression proved the statistical significance of the model in Table 1.

Table 1. Wald's Test for Logistic Regression

X	Chi^2	p-value > Chi^2	df constraint
const	1062.614071	4.411953e-233	1
top1000_voters_rating	1104.475166	3.517505e-242	1
Action	53.033523	3.279040e-13	1
Crime	22.853796	1.748038e-06	1
Drama	24.430425	7.704234e-07	1
Fantasy	16.520951	4.811547e-05	1
Mystery	25.968786	3.469824e-07	1
Romance	3.914710	4.786528e-02	1
Sport	7.101317	7.702733e-03	1
Thriller	9.629585	1.914678e-03	1
War	4.919469	2.655568e-02	1

Hence the analysis shifts to the other regression models that are compared in Table 2. The Durbin Watson Test is a hypothesis test that checks for autocorrelation between error terms, the null hypothesis indicating that there is no autocorrelation. As a rule of thumb, a Durbin-Watson statistic close to 2 implies no autocorrelation. This is the case for all our regression models since it has Durbin-Watson statistic values ranging from 1.7163 as in the case of LASSO regression to 1.963 for Multiple Linear Regression. In terms of R2 values and its variants Multiple Linear Regression seems to be showing the highest value. This is supported by the F statistic or ANOVA

(Analysis of Variance) which is a hypothesis test whose null hypothesis states that all regression coefficients of the model should be zero. Hence it is a test that ensures the overall regression of the model. In Multiple Linear Regression the F-statistic is high unlike that for Ridge and Lasso Regression indicating that the results are not significant. However, the F-statistic value is significantly higher for SLR which seems to give it along with MLR more credit compared to the other Regressions in the ranking of models as its R2 value was only marginally lower. The Jarque-Bera test is a goodness-of-fit test of whether sample data has the skewness and kurtosis matching a normal distribution. Lagrange Multiplier test (LM test also known as Breusch Godfrey test) is a hypothesis test for autocorrelation in the errors in a regression model, the null hypothesis being the absence of autocorrelation. Hence, in terms of normal distribution of errors all regressions except SLR are statistically significant with respect to their Jarque-Bera and Lagrange Multiplier values. Therefore, it can be interpreted that out of these regressions only MLR and Logistic Regression can go for further analysis since they are statistically significant for all parameters.

Table 2. Statistical tests for Regression Analysis

X	Simple Linear	Multiple Linear	Ridge	LASSO
R2/ Pseudo R2	0.556	0.619	0.4855	0.46
Adjusted R-Square	0.556	0.618	0.48	0.4553
F-Statistics	6007	485.0	2.449e-37	2.477e-27
Durbin-Watson Test	1.952	1.963	1.79127	1.7163
Jarque-Bera (JB) Test	9.617	0.91	34.778	90.18286
Lagrange Multiplier Statistic	14.237	145.0	208.301	185.15
Accuracy	0.71	0.711	0.72	0.72

10.2. Classification

The parameters have been tuned to obtain maximum Silhouette Score which ranges from -1 to 1 and a higher Silhouette Score for a given model indicates better performance within that model. Based on accuracy it is seen that K-Means is not as accurate as SVM.

Table 3. Statistical tests for Classification Analysis

X	K-Means	SVM
Silhouette Test	0.7021	0.13387
Accuracy	0.49934	0.71

10.3. Time Series Analysis

From the autocorrelation and partial autocorrelation plot results it is evident that the AR and MA parameters of the Time Series to be considered should be 1 each. This is supported by the Durbin Watson Statistic being 1.4928194722663908 which indicates positive autocorrelation. The Augmented Dickey Fuller Test was showing stationarity based on the value given. The presence of seasonality was confirmed when the model with seasonality was giving a lower AIC value. The presence of exogenous variables was confirmed when they resulted in a reduction in the RMS value. The model was tuned to get the lowest possible Likelihood, AIC, BIC and HQIC. The Ljung Box Statistic leading to a p-value greater than the significance level also shows the validity of the model. The Jarque-Bera Statistic confirms the heteroscedasticity.

Table 4. Time Series Analysis Statistical Test Observations

X	Older movies analysis
Autocorrelation	Cuts of to 0 after 1 lag
Partial Autocorrelation	Cuts of to 0 after 1 lag
Augmented Dickey-Fuller Test Statistic	-10.462528698062133
RMSE	62.19
Log-Likelihood	-965.539
AIC	1949.079
BIC	1982.679
HQIC	1962.512
Ljung-Box (Q)	24.64
Skewness	-0.26
Kurtosis	4.13
Jarque-Bera (JB) Test	19.9
Heteroscedasticity	0.95
Durbin-Watson Test	1.4928194722663908

10.4. Artificial Neural Network

The neural network shows its superiority by giving a very high accuracy with the minimum possible loss. The tuned hyper-parameters are given below. The model does not show overfitting.

Table 5. Artificial Neural Network

Attributes	'duration', 'avg_vote', 'Action', 'Adventure', 'Animation', 'Biography', 'Comedy', 'Crime', 'Drama', 'Family', 'Fantasy', 'Horror', 'Mystery', 'Thriller'
Type	Multi-layer Perceptron Classifier
Architecture	İnput Layer: 14 Hidden Layer: 100 Output Layer: 3
Output Type	Ternary
Initial Loss	0.6915657421532461
Final Loss	0.1624720046108523
Activation Function	Logistic
Optimizer	Adam
Early Stopping	True
Validation Fraction	0.1
Number of training examples	4796
Loss Curve Type	Strictly Decreasing
Testing results with 2020 dataset	93.055555556 %

10.5. Comparison of Valid Models with Similar Accuracy

SLR and Ridge and Lasso Regression were not considered as they proved to be invalid by the normality test and F Statistic respectively. Now 3 models of comparable accuracy exist as shown in Table 5. Hence, a Jaccard Index is used to find the similarity between the attributes obtained from the dataset and that scraped from the IMDb website. The Jaccard Index is used as an indication of the degree to which attributes used in the model are available. From this it can be concluded that SVM is the better model in terms of availability among the 3 followed by Multiple Linear Regression and Logistic Regression respectively.

Table 6. Comparison of Models with Similar Accuracy

Model Name	Attributes	Attributes 2020	Jaccard Index
Multiple Linear regression	‘budget’,‘reviews_from_users’,‘review_from_critics’, ‘top1000_voters_ratings’, ‘Action’,‘Animation’,‘Crime’, ‘Drama’,‘Family’,‘Fantasy’, ‘Horror’, ‘Music’, ‘Musical’, ‘Mystery’,‘Sport’,‘Thriller’	‘duration’,‘Action’,‘Animation’,‘Biography’,‘Drama’, ‘Horror’	4/18=0.222
Logistic regression	‘top1000_voters_rating’, ‘Action’,‘Crime’,‘Drama’, ‘Fantasy’,‘Mystery’,‘Romance’, ‘Sport’, ‘Thriller,’ ‘War’	‘avg_vote’,‘Action’,‘Crime’, ‘Fantasy’,‘Mystery’	4/11=0.18
SVM	'top1000_voters_rating', 'Action', 'Crime', 'Drama', 'Fantasy', 'Mystery','Romance', 'Sport','Thriller', 'War'	'avg_vote', 'Action','Crime', 'Drama','Fantasy','Mystery',' Thriller'	6/11=0.545

10.6. Some Movie Success Prediction Examples

Table 7 shows all the model results for the most recent movies in the testing dataset used. Here H stands for Hit, F stands for Flop and N stands for Neutral.

Table 7. Movie Success Prediction Examples

Movie Name	True Success	SLR	MLR	KMeans	Logistic	Ridge	Lasso	SVM	ANN
Jeanne	F	N	N	N	N	N	N	N	N
I Trapped the Devil	N	N	N	N	F	N	N	N	H
Midsommar	H	N	N	N	F	N	N	N	H
Knives Out	H	H	H	N	N	H	H	H	N
Sextuplets	F	N	N	N	F	N	N	N	H
Unplanned	F	N	F	N	F	F	F	F	F
Cold Blood Legacy	F	N	F	N	F	N	F	N	F
Playing with Fire	F	H	F	N	F	F	F	F	H
Jexi	N	N	N	N	F	N	F	N	H
Tomasso	N	N	N	N	F	N	N	N	N

11. DISCUSSION

Our analysis on the prediction of a movie success using traditional regression methods was inspired by the fact that there were plenty of papers that used various machine learning models to predict the success of a movie. Given that these returned accurate results, we wanted to explore these techniques and the statistical significance of using such models. The tests were aimed at checking for the basic assumptions that follow the application of a model such as the normal distribution of errors for regression and so on. The reason for doing so was that we did not want any inconsistency or overfitting to affect our predictions. The end goal was to make the predictor useful in the real world. To corroborate this, the most recent 2020 data was scraped. This gave a sense of the attributes that will be available on immediate release of a movie.

12. CONCLUSION AND FUTURE WORK

The prediction values from each of the models thus varies depending on their structural build as well as the method they use for prediction, and analysing these changes through comparison as well as statistical tests to prove its significance. By doing so it is seen that the Artificial Neural Network is the best model for prediction followed by the Support Vector Machine, Multiple Linear Regression and Logistic Regression that give comparable performance in terms of accuracy Logistic Regression being slightly higher. The availability of the attributes of these 3 models have also been analyzed where the Support Vector Machine has better availability. Following these 3 models, comes the K-Means with a much lower accuracy. Simple Linear Regression and the Regularization techniques are deemed invalid due to statistically insignificant results. Further, a Time Series Analysis that uses a SARIMAX model forecasts the metascore value effectively. From the models it is evident that some attributes like the top1000_voters_rating and the genres of the movie play a major role in the prediction of movie success. Also, some genres have more predictability than others as is evident from the analysis shown previously. In the near future we aim to look into the various improvements that can be performed on the individual models to boost their performance, and hence broaden our perspective on the classification of a variety of other models with comparable performances to show the significance of each one of them.

REFERENCES AND BIBLIOGRAPHY

[1] Abidi, S.M.R., Xu, Y., Ni, J. *et al.* Popularity prediction of movies: from statistical modeling to machine learning techniques. *Multimed Tools Appl* **79,** 35583–35617 (2020). (Abidi et al. 2020)

[2] IMDb extensive dataset, Kaggle

[3] Ahmad J., Duraisamy P., A. Yousef and Buckles, B.: Movie success prediction using data mining, pp. 1-4. In: 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), doi: 10.1109/ICCCNT.2017.8204173. (Ahmad et al. 2017) Delhi, India (2017)

[4] Nithin, Vr, M. Pranav, Pb Sarath Babu and A. Lijiya. "Predicting Movie Success Based on IMDB Data." *International journal of business* 003 (2014): 34-36. (Bristi et al. 2019; Subramaniyaswamy et al. 2017; Nithin et al. 2014) (2014)

[5] Subramaniyaswamy V., Vaibhav M. V., Prasad R. V. and Logesh R.: Predicting movie box office success using multiple regression and SVM, pp. 182-186. In: 2017 International Conference on Intelligent Sustainable Systems (ICISS), doi: 10.1109/ISS1.2017.8389394. (Bristi et al. 2019; Subramaniyaswamy et al. 2017) Palladam (2017)

[6] Lee K, Park J, Kim I & Choi Y: Predicting movie success with machine learning techniques: ways to improve accuracy. In: Information Systems Frontiers, vol (in press), doi: 10.1007/s10796-016-9689-z (Lee et al. 2018)(2016)

[7] Darapaneni N. et al.: Movie Success Prediction Using ML, pp. 0869-0874. In: 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), doi: 10.1109/ UEMCON51285.2020.9298145. (Darapaneni et al. 2020) New York City, NY (2020)

[8] Quader N., Gani M. O., Chaki D. and Ali M. H.: A machine learning approach to predict movie box-office success. In: 2017 20th International Conference of Computer and Information Technology (ICCIT), pp. 1-7, doi: 10.1109/ICCITECHN.2017.8281839. (Quader et al. 2017) Dhaka (2017)

[9] Dhir R. and Raj A.: Movie Success Prediction using Machine Learning Algorithms and their Comparison, pp. 385-390. In: 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), doi: 10.1109/ICSCCC.2018.8703320 (Dhir and Raj 2018) Jalandhar, India (2018)

[10] Ericson, J., & Grodman, J.: A predictor for movie success. CS229, Stanford University.(2013)

[11] Quader N., Gani M. O. and Chaki D.: Performance evaluation of seven machine learning classification techniques for movie box office success prediction. In: 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), pp. 1-6, doi: 10.1109/EICT.2017.8275242. (Quader et al. 2017; Quader et al. 2017) Khulna (2017)

[12] Kumar, U.D., Wiley, Business Analytics: The Science of Data-Driven Decision Making(2017).

[13] Navidi, W., McGraw Hill: Statistics for Engineers and Scientists:Indian Edition(2011)

[14] Sharma, T., Dichwalkar, R., Milkhe, S., and Gawande, K.: Movie Buzz - Movie Success Prediction System Using Machine Learning Model. In: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 111-118, doi: 10.1109/ICISS49785.2020.9316087. Thoothukudi, India (2020)

[15] Multi-Label Binarizer Python Documentation: MultiLabelBinarizer

[16] Verma, G., and Verma, H.: Predicting Bollywood Movies Success Using Machine Learning Technique, pp. 102-105. In: 2019 Amity International Conference on Artificial Intelligence (AICAI), doi: 10.1109/AICAI.2019.8701239. (Verma and Verma 2019) Dubai, United Arab Emirates, (2019)

[17] Lash, Michael T., and Kang Zhao. "Early predictions of movie success: The who, what, and when of profitability." Journal of Management Information Systems 33, no. 3 (2016): 874-903. (Lash and Zhao 2016)

[18] Na, S., Xumin, L., and Yong, G.: Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm, pp. 63-67. In: 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, doi: 10.1109/IITSI.2010.74 (Na et al. 2010) Jinggangshan, (2010)

[19] Bristi, W. R., Zaman, Z., and Sultana, N.: Predicting IMDb Rating of Movies by Machine Learning Techniques, pp. 1-5. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), doi: 10.1109/ICCCNT45670.2019.8944604. (Bristi et al. 2019) Kanpur, India, (2019)

[20] S. F. Ershad and S. Hashemi, "To increase quality of feature reduction approaches based on processing input datasets," *2011 IEEE 3rd International Conference on Communication Software and Networks*, 2011, pp. 367-371, doi: 10.1109/ICCSN.2011.6014289. (Ershad and Hashemi 2011)

[21] Fekri-Ershad, S. (2019). Gender classification in human face images for smartphone applications based on local texture information and evaluated Kullback-Leibler divergence. Traitement du Signal, Vol. 36, No. 6, pp. 507-514. (Fekri-Ershad 2019)

AUTHORS

Shreya Venugopal
Student at PES university pursuing B-Tech in Computer Science, member of the ACM student chapter, specialization in Machine Intelligence and Data Science and an avid coder.

Rishab Kashyap
Student at PES university pursuing B-Tech in Computer Science, member of the ACM student chapter, Specialization in Machine Intelligence and Data Science along with minors degree in Electronics and Communication.

Manav Agarwal
Undergraduate Student at PES university pursuing B-Tech in Computer Science, Specialization in Machine Intelligence and Data Science with an innate interest in gadgets and electronics.

R Bharathi
Associate professor at PES University specialisation in Data Science, Machine learning and Data Analytics.

FINDING CLUSTERS OF SIMILAR-MINDED PEOPLE ON TWITTER REGARDING THE COVID-19 PANDEMIC

Philipp Kappus and Paul Groß

Department of Computer Engineering, Baden-Wuerttemberg Cooperative State University, Friedrichshafen, Germany

ABSTRACT

Two clustering methods to determine users with similar opinions on the Covid-19 pandemic and the related public debate in Germany will be presented in this paper. We believe, they can help-gaining an overview over similar-minded groups and could support the prevention of fake-news distribution. The first method uses a new approach to create a network based on retweet-relationships between users and the most retweeted users, the so-called influencers. The second method extracts hashtags from users posts to create a "user feature vector" which is then clustered, using a consensus matrix based on previous work, to identify groups using the same language. With both approaches it was possible to identify clusters that seem to fit groups of different public opinions in Germany. However, we also found that clusters from one approach cannot be associated with clusters from the other due to filtering steps in the two methods.

KEYWORDS

Data Analysis, Twitter, Covid-19, Retweet network, Hashtags.

1. INTRODUCTION

During the years of 2020 and 2021 the Covid-19 pandemic and the resulting legal regulations polarized the society and led to a huge discussion dominating the social networks, including Twitter (most used hashtag in Germany of 2020: "corona" [1]). In Germany, the so called "Querdenker" (eng.: lateralthinkers) gained nation-wide attention by organizing protests against Covid-19 related regulations, partly denying the existence of the virus and spreading various conspiracy theories on Twitter.

The aim of this work was to identify groups of similar opinions on this topic by analysing the tweets posted on Twitter in Germany regarding the Covid-19 pandemic. To achieve this, a novel method was developed, establishing a communication network based on relations between users, where one retweeted the other similar to [2]. A new way of filtering the data regarding only connection to influential users was used to distil clusters from the otherwise chaotic and over-connected network. This led to the possibility of grouping "normal users" to a so-called "superuser", if they retweet the same influencers. The DBSCAN algorithm is then used to detect communities inside the filtered network.

Furthermore, an implementation of the previously published hashtag-based clustering method of [3], where a combination of several rounds of the k-means algorithm and the DBSCAN algorithm is proposed, has been applied to the dataset. The latter approach applied to the tweets about the

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 31-40, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111803

pandemic and its visualization yields insight into the different hashtags used by different groups in the debate.

Results from both methods can be viewed on https://andfaxle.github.io/twitteranalysis/.

2. LITERATURE REVIEW

Twitter, with its 192 million daily active (monetizable) users [4] expressing their views and opinions in shorttexts, has been a popular playground for data research analysing people interest in topics or products, performing personality studies and even predict flu outbreaks [5]. Previous work in this area can be divided into two groups of clustering approaches: Network Clustering and Content Clustering. While Network Clustering establishes a graph between users with retweets, mentions or followers connecting them, Content Clustering uses NLP to analyse the actual text that have been posted in terms of keywords used, sentiment or various other parameters.

[6] clustered the content of tweets by hashtags with the k-means algorithm, agglomerative hierarchical clustering and a fuzzy neighbourhood model. Similar to that, [3] analysed 30,000 tweets from just before a world cup to cluster the content in order to extract topics from the tweets. To reduce noise, they proposed four different algorithms one of which runs several rounds of the k-means algorithm with varying k using the cosine distance on keywords extracted from the tweets. A so-called consensus matrix is then created stating in how many rounds two users end up in the same cluster. Users that have been clustered together in more than 50 per cent of the k-means rounds, can now be considered a community.

[7] used a combination of both content and network approaches. They developed a classifier grouping users into political-left and political-right by first establishing a network of reference users based on retweets (and mentions). Afterwards they assigned features for each group, consisting of keywords extracted from the users tweets in the cluster.

[2] examined the influence of Russian trolls in the context of #BlackLivesMatter by setting up a retweet network that clearly showed two distinct clusters (political-left, political-right) and continued to analyse the influence of trolls on each of these clusters.

3. DATA ACQUISITION

Since the terms of service are accepted, users agree that Twitter can make their content “available to other companies, organizations or individuals”[8] by providing access over the Twitter API. An AWS architecture was used to automatically retrieve and store tweets regarding the Covid-19 pandemic. An EC2 instance (t2) running a Python script registered on Twitters filtered-stream API to retrieve tweets ingerman language and containing the keywords “covid” and “corona”. Through a Kinesis Data Firehose, the tweets are stored as bundles in a S3 - Bucket. Over the course of March 2021, a total of 2,955,282 tweets posted by 260,954 different users were collected.

4. NETWORK CLUSTERING

Users can be regarded as nodes of a graph. A relationship between two users can be interpreted as edges and is established when one of them retweets the other. The approach to create a complex, has been used several times as noted in section 2 by [7] or [2]. Both regard all retweets as valuable connection but this makes the graph unfeasible large and complex. We propose to regard only connections to “influencers” in order to distil a subgraph that is nearly as meaningful but offers great and more nuanced insights into communities in the Twitter landscape.

The whole process was implemented using Python orchestrating basic file and folder operations, no external database or tools has been used.

4.1. Influencers

Analysing the data set, we found that 31.54 per cent of all retweets are originally posted by the same 100 users (0,04 per cent of all users), who, as of now, will be called influencers. We can therefore assume, that these influencers primarily shape the opinion-landscape and are core users of possible clusters. A graph $G = (N, E)$ can be build regarding only relationships between users U and the 100 influencers I with edges $(u, i, g) \in E$ and $g$representing how often the user u retweeted the influencer i. Testing with a subset of four days and only three influencers using matplotlib and an implementation of the NEATO-layout algorithm [9] takes 35.2 seconds on a standard linux computer and produces a graph depicted in figure 1.

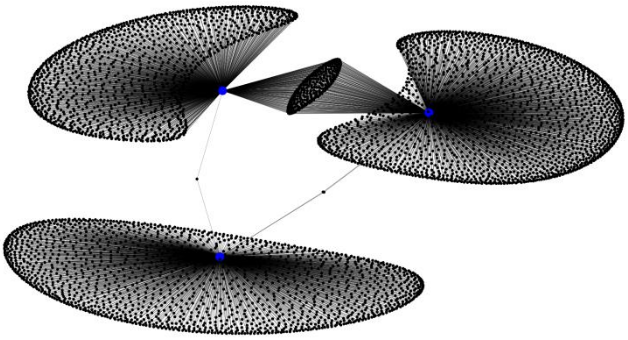


Figure 1. Graph using three influencers (blue)

4.2. Superusers

As seen in figure 1, there are many users only retweeting one of the influencers and some retweeting two. To further reduce complexity, we can aggregate all users that retweet the same influencers to one superuser s.

$$s_{i^n .. i^m} = \{u \in U \mid \forall(u, i^n .. \; i^m) \in E\} \; (1)$$

This limits the number of possible nodes in the graph to:

$$|K| = |S| + |H| = 2^{|H|} - 1 \; (2)$$

The weights g of the edges are summed up.

4.3. Thresholding

With 100 influencers there are still more than 10^{30} possible nodes. To further minimize complexity a threshold is used to cut edges with weights lower than a threshold T:

$$E_{filtered} = \{(s, i, g) \in E | g > T\} \; (3)$$

A good method of determining the best threshold value has not been established but using 0.65 per cent of the maximum weight in the graph seems to produce clear clusters.

4.4. Normalizing weights

While analysing the count of retweets on the 100 most retweeted users (influencers), we found out ? that they are distributed according to an inverse power law:

$$N_{retweets} = b * x^{-m}$$
$$m = 0.65$$
$$b = 39{,}215$$

Where x is the rank, m is the slope and b is the scaling factor or the number of retweets of the influencer at rank 0.Following the nature of an inverse power-law and as seen in figure 2, the higher ranks have significantly greater retweet counts, making it hard to find a threshold that on the one hand minimizes complexity regarding connections to the higher ranked influencer while keeping groups consisting of lower ranked influencers in the graph. To counteract this problem, weights of edges to influencers are multiplied by the common logarithm of the rank of this influencer before applying the threshold. This does not completely normalize the retweets as it would overvalue retweets of lower ranked influencer but decreases the dominance of higher ranked influencer to a certain amount.

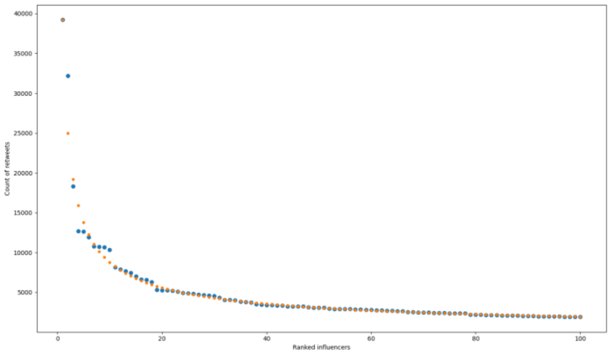


Figure 2. The distribution of retweets (blue) and the power-law

4.5. Clustering

These steps applied on the full dataset of one month, 100 influencers and a threshold of 61 took 2 hours and 58 minutes to calculate and yielded a graph depicted in figure 3.

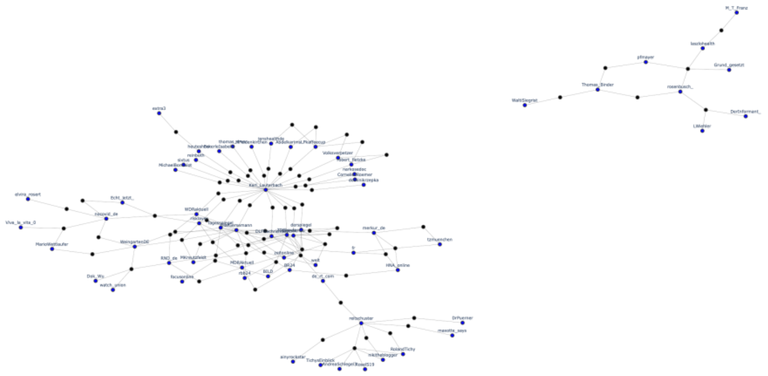


Figure 3. Full graph from one month and |I| = 100 and T = 61

A human can identify a main cluster in the middle, a smaller one connected to the main one on the bottom and a separated cluster on the upper right. To automatically identify these, aDBSCAN algorithm was used where a core influencer is defined if it is connected to $minPts = 2$ other influencers (over superusers). A minimal distance ε does not have to be defined since by applying a threshold, irrelevant connections are already filtered out. To prevent the algorithm to cluster all nodes that are in any way connected (the lower cluster has one connection to the main cluster) the algorithm is modified in such a way, that already visited nodes do not count as a new neighbour, reducing the number of core points. The resulting clustered graph can be seen in figure 4.

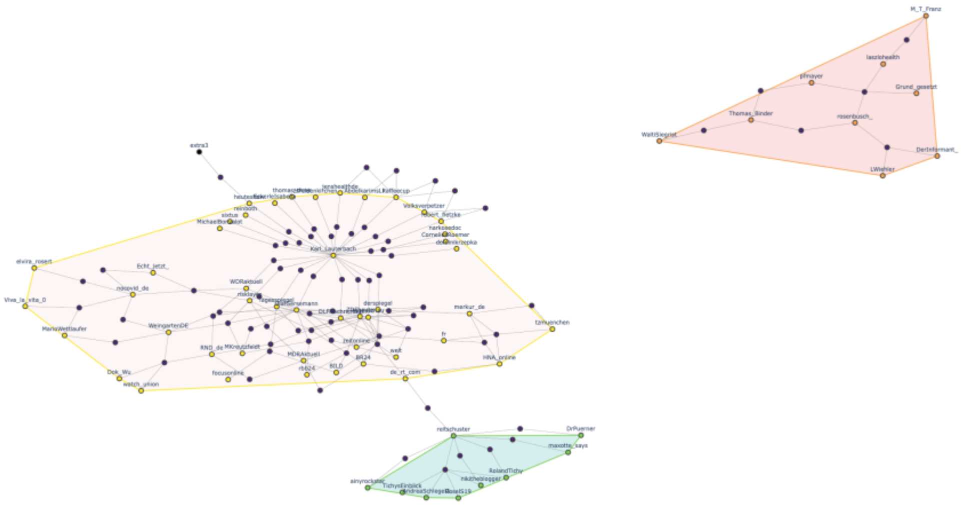


Figure 4. The same graph as in figure 3 but clustered using DBSCAN

4.6. Looking into the clusters

Let’s have a look on some of the influencers of each cluster. The main cluster seen in yellow consists of public and private news agencies like “ZDF”, “tagesschau”, “derspiegel” or “BILD” as well as the most retweeted user in the dataset: “Karl Lauterbach”. He is a present figure in the public debate on the pandemic based on his background in the social democratic party (SPD) and as a medical practitioner.

Some influencers from the green cluster are: “maxotte_says” (Max Otte) former leader of the “Werteunion” a faction within the Christian Democratic Union (CDU), known for advocating more conservative positions. “ainyrockstar” is a right-winged journalist [10]and “RolandTichy” former editor of “Impuls” and “Euro”.

An influencer from the upper right, orange cluster is “rosenbusch” (Henning Rosenbusch), an independent journalist, advocating the “Swedish-way”. Also, the user called “laszlohealth” (unknown) has a pinned tweet: “Corona has become a strange mixture of religious and political war by all means. The mask is the symbol of belonging. PCR mass testing the weapon. Objectivity, freedom of expression and normal interaction no longer exist.” ([11] translated with google translation).Another example is “Thomas Binder”, a swiss doctor who has been advocating an anti-regulation position and was arrested and admitted to psychiatry because of suspected threats against politicians [12] and whose account has been blocked.

5. Language Clustering

In this approach we implement the algorithm of [3] discussed in section 2, assuming that users sharing the same opinion on the topic will also use the same hashtags. Summing these hashtags up, a map of hashtags can be assigned to each user. Creating a user feature vector and comparing them using the cosine distance, the k-means algorithm and the DBSCAN algorithm can be used to detect similar language preferences among users.

5.1. Hashtag Extraction and Preparation

Initially, hashtags are extracted from the individual tweet object as the Twitter API already delivers them in a separate field. For data preparation and to reduce variance among all extracted hashtags they are traced backed to their root word, aka lemmatized. This was implemented using the HanTa library for Python [13]. Furthermore, words that have no semantic value for the sentence and are only included for grammatical reasons (e.g.: “like”, “the” etc.), are filtered out. These words are called “stop words” and are based on a detailed list of German stop words from [14].

5.2. User Feature Vector Creation

In order to find communities of users, a list associated which each user is created stating which hashtags he used and how often. Since the amount of all words across all users is very large, it must be reduced first. This involves a loss of information but is necessary to make the data processable. All hashtags which are present only once and whose overall count does not move within the 97 per cent and 99.98 per centquantile, are removed. Thus, all hashtags which are not frequently used or are used by everyone (and therefore provide no information) are sorted out. Theupper limit is necessary because the tweets were collected according to the hashtags concerning the Covid-19 pandemic (#covid, #corona, #covid-19) and are therefore contained in all tweets.

Making the list of hashtags comparable using the cosine distance, a binary feature vector is created for each user. The number of dimensions on this vector is equal to the count of distinct hashtags in the whole dataset. If a user has used a hashtags more than three times, a 1 is put in the dimension corresponding to this hashtag, otherwise a 0.The threshold of at least three hashtags is a hyperparameter.

5.3. User Clustering

A well-established clustering method is the k-means algorithm. A variable k is used to determine howmany clusters a dataset should be divided into. The Centers of each cluster are chosen randomly and every user is assigned to the cluster where the cosine distance to its center is lowest. In the next step new centers are calculated as the average characteristic values of all in the cluster contained users. Every user is then reassigned again. This is done until there are no more changes. As the number of clusters is predetermined by k and the centers are initialized randomly, the algorithm is not optimal, since the number of clusters cannot be determined beforehand. Furthermore, all users are assigned to clusters, whichmakes it impossible to exclude noise.

As [3] proposed these issues can be solved in creating a consensus matrix by running k-means multiple times with different values for k. A null matrix of size $n \times n$ is created, where n is the number of users and each row and column represent a particular user. For each run of k-means, the value within the matrix at (User A, User B) and (User B, User A) is increased by one if these users endup in the same cluster. The resulting matrixholds a value for each pair of users stating how often they ended up in the same cluster and therefore how similar the characteristics of the users are.

From the consensus matrix a graph can be imagined between all users where the value in the matrix defines the weight of the edge between these two users. To find communities of users that are densely connected (they used the same hashtags), the DBSCAN algorithm is used, where users are regarded as points. A core point is defined as a point that has $minPts$ of connected points with an edge-weight of ε or greater. Both are hyperparameters but setting $minPts$ to 2 per cent of the total users in the filtered dataset and ε to 80 per cent of the times k-means was run (the maximal possible value in the matrix) worked well. In the case of March 2021: $minPts = 20, \varepsilon = 15$. An advantage of the DBSCAN is that the aforementioned noise points are labelled as such. With the found clusters the graph was formed and visualized using Gephi and the ForceAtlas Layout Algorithm [15]. Figure 5 shows the graph using 15 iterations of k-means.

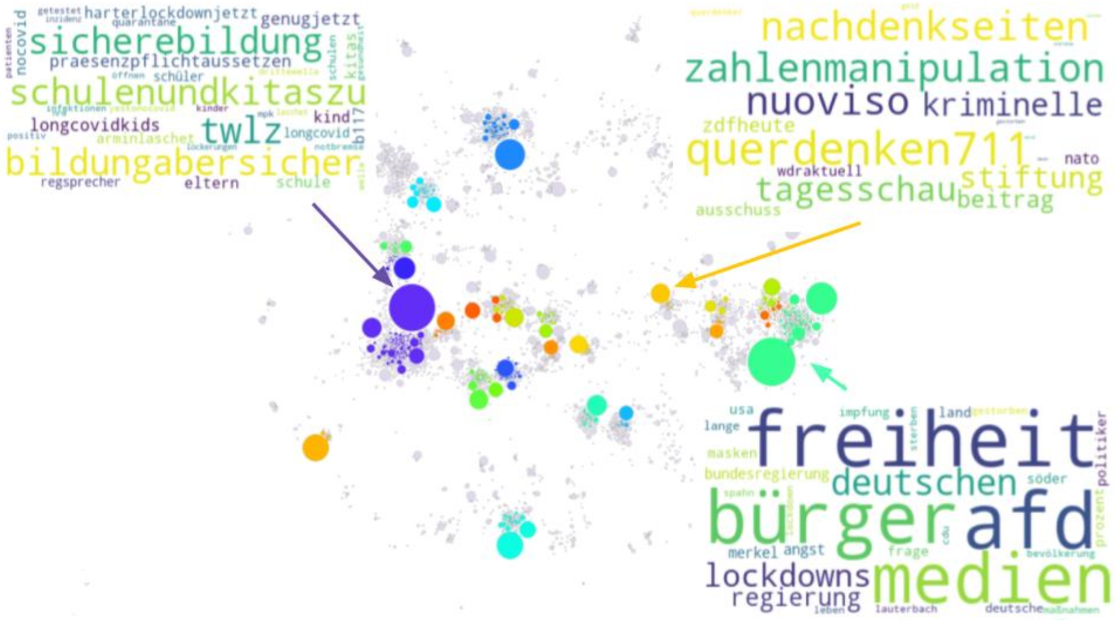


Figure 5. A graph created by evaluating the similarity of different user.

5.4. Interpreting Clusters

In order to find the hashtags that constitute each cluster, we compared the relative frequency of a hashtag used within a cluster to the relative frequency of that hashtag in the whole data set. The resulting word clouds of a sample of three clusters is depicted in figure 6.

The first one features the right-wing AfD together with the hashtags “freiheit”, “medien” and “bürger” (eng.: freedom, media and citizens).

In the second cluster, no unambiguous subject can be found.

An example hashtag from the third cluster is “nachdenkseiten” which refers to a germanjournalistic web page. In recent history it was labelled as a “Conspiracy ideological and / or right-wing open media” ([16] translated with google translation). Further hashtags are “kriminelle” and “zahlenmanipulation” (eng.: criminals and manipulation of numbers).

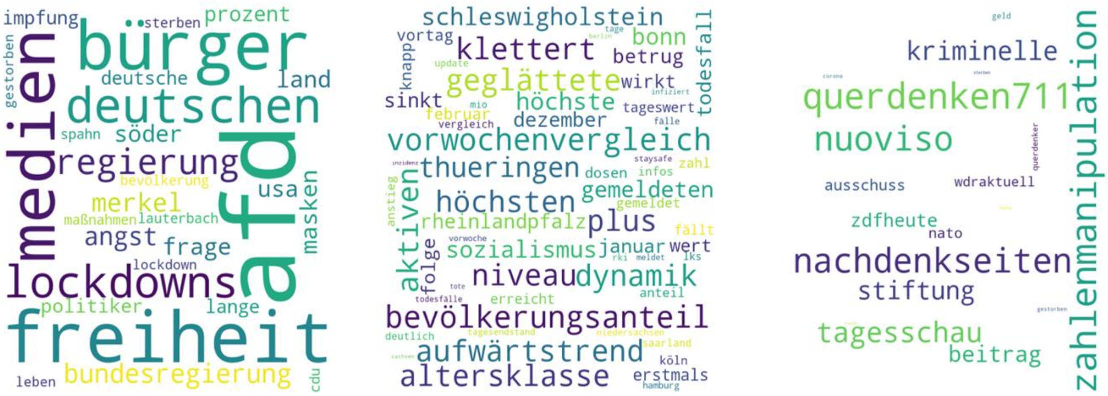


Figure 6. Word clouds of different user groups clustered by hashtag**.**

6. Comparison of Clustering Methods

To further consolidate the clusters found using both approaches, we compared the users belonging to each cluster in order to investigate whether it is possible to associate clusters from the retweet network approach to the language clustering approach. To illustrate the relationship a Sankey diagram (figure 7) was build depicting the network clusters on the left, language clusters on the right and users that are part of two clusters as agrey flow. Users from the network clusters, that do not belong to any language cluster flow to “undefined”. More than 50 per centof the users from the network clusters have not been clustered in the language approach (also vice versa). This can be explained by the steps in both approaches filtering users:

Filtering steps in the network approach

The raw dataset holds 260,954 users.

- Only considering users that retweet: 37 per cent (966,322) users are filtered out.
- Only considering users that retweet influencers: 57 per cent of the users that retweet (94,730) are filtered out.

- Applying a threshold to superuser connections: Since some superusers will be deleted when all its connections are below the threshold their associated users are filtered from the cluster. 96 per cent of the users that retweet influencers (67,119) are filtered out.

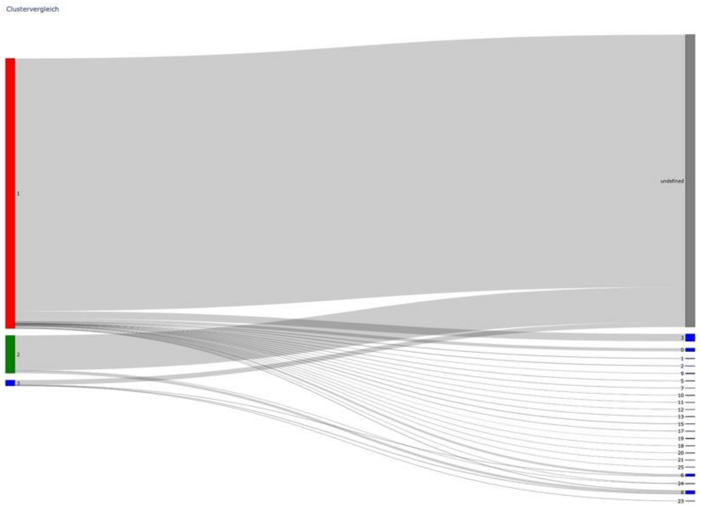


Figure 7. Comparison of users associations between clusters of the two approaches

Filtering steps in the language approach

- Only the hashtags whose counts are between the 97 per cent and 99.98 per cent quantile of the frequency distribution are used: 98 per cent (404,905) hashtags are filtered out.
- Only the users that have used those hashtags more than 6 times within the month are kept: 99.5 per cent (238,717) of all users are filtered out.

7. CONCLUSION AND FUTURE WORK

Twitter is a platform where millions of people publicly share their feelings and opinions every day. Analysing even large amounts of this information has become possible through the increase in computer power but also with the help of many dedicated research, coming up with better and better ways to structure this otherwise chaotic data. This paper presented two approaches to the problem of finding clusters in this unstructured data set. We showed that there are clusters re-tweeting only themselves and using the same language regarding the debate on the Covid-19 pandemic. In developing a new approach and expanding methods already in place to distil this data and find clusters of similar-minded people we hope to distribute important information about the structure of the Twitter ecosystem and hope that further research can be conducted on top of our work.

Finding clusters in the retweet network heavily depends on the number of influencers and the threshold chosen. In future works, a method of choosing these parameters to reduce complexity only as much as necessary while keeping as many users in the data set as possible will mark a step ahead. Expanding the language clustering method to keywords and fine-tuning the parameters for k-means and DBSCAN can yield clearer clusters. An important factor is the number of iterations of the k-means algorithm. More iterations with values fork ≥ 20 or more would give more detail to the results.

Furthermore, since both methods used simple folder and file operations, a more sophisticated architecture could be set up to decreases computing time.

BIBLIOGRAPHY

[1] H. Eberhardt, “Absatzwirtschaft,” [Online]. Available: https://www.absatzwirtschaft.de/twitter-2020-ein-jahresrueckblick-mit-trends-und-hashtags-176871/. [Accessed 8 7 2021].

[2] A. A. a. K. S. L. G. Stewart, ““Examining trolls and polarization with a retweet network,” 2018.

[3] Godfrey, “A Case Study in Text Mining: Interpreting Twitter Data From World Cup Tweets,” 2014.

[4] Twitter, Q1 2021 Letter toShareholders, 2021.

[5] A. G. R. L. S.-H. Y. a. B. L. H. Achrekar, “Predicting Flu Trends using Twitter data,” 2011.

[6] B. S. a. I. B. G. Ifrim, “Event detection in Twitter using aggressive filtering and hierarchical tweet clustering,” in CEUR Workshop Proceedings, 2014.

[7] B. G. J. R. A. F. a. F. M. M. D. Conover, “Predicting the Political Alignment of Twitter Users,” in IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 2011.

[8] Twitter, “Twitter Terms of Service,” [Online]. Available: https://twitter.com/en/tos. [Accessed 8 7 2021].

[9] S. North, “Drawing Graphs with Neato, 2004,” 2004.

[10] A. Graen, “Focus,” [Online]. Available: https://www.focus.de/panorama/welt/panorama-anabel-schunke-ist-eine-der-wichtigsten-figuren-der-neurechten-szene-wir-waren-mit-ihr-feiern_id_10281656.html. [Accessed 8 7 2021].

[11] laszlohealth, “Twitter,” [Online]. Available: https://twitter.com/laszlohealth/status/1319338449149874181. [Accessed 8 7 2021].

[12] Medinside, “Medinside,” [Online]. Available: https://www.medinside.ch/de/post/verhafteter-aargauer-arzt-in-der-psychiatrie. [Accessed 8 7 2021].

[13] C. Wartena, “ A Probabilistic Morphology Model for German Lemmatization,” in 15th Conference on Natural Language Processing, 2019.

[14] J. Oppenlaender, “Github,” [Online]. Available: https://github.com/solariz/german_stopwords/. [Accessed 8 7 2021].

[15] V. T. H. S. B. M. Jacomy M, “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software,” 2014.

[16] M. Schwarzer, “Redaktionsnetzwerk Deutschland,” [Online]. Available: https://www.rnd.de/panorama/esoteriker-auf-corona-demos-tanzende-hippies-neben-rechtsextremen-und-verschworungstheoretikern-was-will-sie-esoterikszene-QMRYIWRQLNCP5N5GGWMK7V53LM.html. [Accessed 8 7 2021].

AUTHORS

Philipp Kappus and Paul Groß are computer science students at the Baden-Wuerttemberg Cooperative State University in Friedrichshafen, Germany, spending half their study time at the campus and the other half at Airbus Defence and Space GmbH in Immenstaad. Due to the corona virus outbreak, both could only visit the courses online for more than one and a half years. That’s why they became interested in analysing twitter to get insight on how different people think about and handle the pandemic.

Paul Groß

AN ANALYSIS OF FACE RECOGNITION UNDER FACE MASK OCCLUSIONS

Susith Hemathilaka[1] and Achala Aponso[2]

[1]Department of Computer Science and Software Engineering,
University of Westminster, London, United Kingdom
[2]Department of Computer Science, Informatics Institute of Technology,
Colombo, Sri Lanka

ABSTRACT

The face mask is an essential sanitaryware in daily lives growing during the pandemic period and is a big threat to current face recognition systems. The masks destroy a lot of details in a large area of face and it makes it difficult to recognize them even for humans. The evaluation report shows the difficulty well when recognizing masked faces. Rapid development and breakthrough of deep learning in the recent past have witnessed most promising results from face recognition algorithms. But they fail to perform far from satisfactory levels in the unconstrained environment during the challenges such as varying lighting conditions, low resolution, facial expressions, pose variation and occlusions. Facial occlusions are considered one of the most intractable problems. Especially when the occlusion occupies a large region of the face because it destroys lots of official features.

KEYWORDS

CNN, Deep Learning, Face Recognition, Multi-Branch ConvNets.

1. INTRODUCTION

In uncontrolled environments face recognition systems drastically decrease their performance, due to various challenges such as facial expressions, pose variations, face occlusions, varying lighting conditions and low-resolution image inputs and different scales of images, partial captures, etc.[3]. Sometimes there are more than one challenges that make an incapable face recognition system detect and identify persons. During the pandemic, Occlusion problem is well highlighted when incapability of face recognition systems to recognize faces due to growing incentive of wearing a face mask to control spread of coronavirus. Even our mobile device's face recognition is incapable to identify the owners. This problem is more critically highlighted in airports, border control systems and security optimized premises.

Occlusions often happen in natural environments and they are most challenging and problematic in many fields of computer vision and object detection because any object can be occluded in the unconstrained environment and they destroy all details of the subject [4]. Because of that face recognition under occlusions remains a major challenge because of the unpredictable nature of occlusions. When it comes to facial recognition it's a very critical challenge because the occluded area varies in position, size and shape in face image [5]. Facial occlusions are unavoidable in unconstrained environments and there can be millions of different occlusion scenario which can't countable Practically, collecting a large dataset with all possible occlusion scenario to train a

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 41-48, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111804

deep neural network is not feasible[2]. There are several scenarios which occur Facial occlusions can be categorized into four aspects[6].

- Facial accessories: occluded by eyeglasses, facemasks, hat and hair.
- External occlusions: occluded by hands or other random objects.
- Partially captured faces: partially captured due to Limited field of view.
- Artificial occlusions: occluded with random white or black rectangles, random salt & pepper noise.

2. FACE DETECTION UNDER OCCLUSIONS

Occluded object detection is challengeable in unconstrained environments when a large area is occluded because of intraclass similarity and intraclass variations. Many approaches are taken to occluded object detection. Convolutional correlational filters are one of them to address the problem successfully. Binary segmentation is useful in the same task.

The existing face detection models less accurate in sometimes when faces occluded. To address this problem, they introduced special algorithms for occluded face detection. General face detection algorithms give an optimized performance in unconstrained environments. There are three main categories in face detection from approaches. Rigid templates, deformable part models and ConvNets. The popular viola-jones face detection model, Harr like feature and AdaBoost comes under rigid template category which can be drop performance in real-time applications. DPM based suited better in real-time application, but their computational complexity is too much. Promising DCNN Approaches can solve the various problems that occurred in an unconstrained environment [1]. DCNN provide a solid solution for various A-PIE problems up to date.

Handling Occlusions in faces are difficult because of occlusion variation and unknown locations. Efforts taken to detect occluded faces can be clustered into three main categories. Locating visible facial segments, Discarding the features taken from occluded sub-regions, Use the occlusion information. Attribute aware CNN like categorized face features according to the property of face part like big lips, long eye likewise to create facial response map. Most probably those approaches achieve good accuracies around 99%.FAN, LLE-CNN, AdaBoost cascade classifier are the approaches usually trained as segment-based face detectors from discarding occluded face sub-regions [2]. They achieve good accuracy, and their simple architecture makes them speed because of low computational complexity.

In the third approach, there are trying to minimize the occlusion damage by extracting features in the near area of occlusion. Novel grid-loss, AOFD, faster-CNN and LSTM hierarchical Attention Mechanism used to give better performance than other approaches.

3. OCCLUSION ROBUST FACE RECOGNITION

When Occlusion is in a small context and extracted features of the non-occluded area are more robust than occlusion is suitable for current face recognition systems. But the unpredictability of occlusion occurs problems in occlusion-robust approaches in practical environments. Occlusion robust approaches suggested new similarity measurement techniques and loss layers to deal with inter-class similarity. Those approaches try to keep the robustness at the same level when happening an occlusion but sometimes it’s challenging. They try to leverage discriminative feature learning ability of face recognition systems. With the obtained success of deep learning approaches in face recognition systems handcrafted or engineering features are not satisfactory because those approaches are not adoptable into state-of-the-art systems. Learning-based features are the most suitable to address the problem in current deep learning face recognition systems. Facial descriptors are key players in handcrafted engineering-based approaches. These approaches are not realistic in practical situations; their poor alignment makes difficulties in feature extraction in a meaningful way. Local Binary Patterns and (SIFT) descriptor highly used in these applications. SIFT descriptors invariancy of illumination, pose, scale and rotation are very beneficial in practical environments.

The distance metric used in patch-based matching is important in those approaches. Elastic Bunch Graph Matching represents extracted facial features corresponding to the face representation of in a graph. Each node represents correspondence location features of the face and edges represent a relationship between different facial features in a related area. Learning-based features can be clustered into four. As appearance-based learning approaches, fisher face and eigenface are discriminately learn the subspace using in order PCA and LDA. Their limitation is proper alignment needed based on eye location. when eye location is occluded, they are not effective. Statistical learning approaches are introduced to account occlusion probability in different areas of the face. These approaches consider the knowledge of occluded facial parts [7]. introduced a scoring-based approach to select appropriate areas to validate to face recognition. They improved robustness of facial recognition systems considering the largest matching area of the occluded face. Sparse representation classifier leverages discriminative power than statistical learning. It combines linear combination and sparse errors to leverage discriminative power [8] study effort to overcome shortcomings of sparse representation approaches which perfectly works in laboratory conditions but dramatically fails in unconstrained environments. They introduce a structured occlusion coding approach to solve the limitation of the existing approach. They divide and recognize the occlusion separately to provide good classification to the probe image and reconstruction is learned by using mask strategy and a dictionary learning technique. Their work can be categorized into a robust face recognition approach than a recovery-based method. Their work has a limitation of collecting all occlusion scenarios that can happen in a practical environment and are not effective. When considering Discriminative power deep learning-based approaches are pioneered with comparing other learning-based approaches. The necessity of a lot of training data is a limitation of these approaches. Considering Deep learning approaches their architecture and loss functions to play an important role to leverage the robustness of the algorithms. Feeding a large dataset of collected occlusion can help relieve the occlusion problem but it’s not practical and cost-effective. Some prior efforts have taken on the augmented face with occlusion used to train algorithms and they are not sufficient to improve the robustness into the satisfactory level.

Solving data deficiency problems using data augmentation is not a practical solution and augmenting all the possible occlusion scenarios are inevitable.[9] the study found the sensitivity of the occlusion area to the facial recognition system. They conclude occlusion of the middle of the face is more difficult to handle. From this conclusion, they try to increase the discriminative power of the outer face to get better robustness. Researchers introduce some dictionaries that occlusion can happen in which area of the face and their effect according to occlusion location and it helps to identify occlusion patterns.

4. OCCLUSION RECOVERY FACE RECOGNITION

Most of the approaches attempt to solve the occlusion problem in the low level of face recognition pipeline which is known as feature space. Occlusion recovery approaches explicitly try to solve it at a higher level also known as image space. Occlusion recoveries have been done before to add data to face recognition algorithms. Then it becomes a more complex task than other approaches. There are several efforts and various approaches to recover clean faces from an occluded face. Reconstruction of the occluded area is another approach more competitive and better than inpainting. Space representation classification is the long-studied approach in reconstruction approaches. It overcomes the flaws of linear reconstructions efforts which used principal component analysis, recursive error compensation and Markov random field networks. It's an extension of linear combination by adding sparse errors accounting into the approach. SRC uses an identity matrix as an occlusion dictionary able to tackle occlusion precisely. Those approaches are not capable of contributing to current deep neural network-based systems. Few studies found which use deep learning techniques to tackle face de-occlusion problems. They usually learn encoding data from clean data and transfer them to corrupted or noisy data. Learnt decoding parameters are clarified to obtain clean data [10]. introduced the LSTM network and autoencoders to address shortcomings. Two LSTM components are introduced. One encodes face patches and other autoencoders decode the reconstruction representation. Adversarial ConvNets are capable to enhance the discriminative features more. Image inpainting is an emerging area in computer vision and deep learning. Those approaches are adopted to face de-occlusion problems as well.

Most studies focus on realistic image inpainting than their accuracy. Because several in painted approaches can't contribute to face recognition efforts. approaches can be clustered into two groups which are blind and non-blind inpainting considering the awareness of the location of corrupted pixels of an image. Most deep learning approaches are usually blind inpainting because deep learning is not able to discriminately learn the corrupted pixels like handcrafted approaches. In non-blinding inpainting, it copied mostly used textures in the image and replaced them in the occluded areas. In filling the pixels, the confidence value of the pixel is considered. As an extension of studies hybrid approaches are introduced to leverage face reconstruction. Finally, all recovery-based models are not provided with any promising results and improvements in face recognition evaluations.

There are a lot of new loss layers introduced in previous research to improve the similarity measurements on the feature extraction phase. Grid loss, SoftMax, Arc Face are some of them used in the latest facial recognition systems. Loss layers can improve the discriminative power of feature extraction and it improves the face recognition accuracy.

5. OCCLUSION DISCARD FACE RECOGNITION

Occlusion aware approaches are highly aware of the occlusion. Some of the approaches are discarded the occluded part of the face which is known as partial face recognition approach. Other Approaches considering the occlusion and try to minimize the effect of occlusion to face verification. We can divide Occlusion aware face recognition into two sections such as partial face recognition and context-aware feature extraction.

Capturing Partial faces is often incident in the unconstrained environment due to occlusions and limited field of view or non-frontal faces. In security-critical situations such as surveillance identifying mugshots and criminal's investigation used it most. Some of the Occlusion based face recognition systems follow a similar approach to recognized faces can categorize into partially occluded face recognition. In this approach only qualifies the non-occluded area of the face to be used to face verification. Partial face matching is not satisfactory every time especially when occlusion free face appears. Partial face recognition approaches are not suitable in unified real-time systems due to its less efficiency and less-robustness to security breaches. [11] is a partial recognition approach using sparse representation classification combining to a Fully connected convolutional network to propose a novel approach called dynamic feature mapping (DFM). It can address partial faces regardless of size. They understood the necessity of partial face recognition in situations like criminal investigations and built the system for partial face recognition. Their work did not need fixed size and aligned facial images like others. Computational efficiency is high. But partially feature extraction is less secure when a holistic face is available. Furthermore, Partial face recognition needs a lot of training data and computational cost to train them. In our work, it discards non-discriminative features like jawline and chin shapes.

Robust face recognition has shown some influence in occluded face recognition. To date, it’s not evaluated the capability of recognizing large occlusions like face masks. These approaches have some limitations when occlusion is varied, it challenges the system’s robustness and usability. These approaches are more suited than partial face recognition in a unified face recognition system.[12] is a work introducing pairwise differential Siamese network (PDSN) with the intention for addressing the underlying problem of deep CNN’s incapability of recognition occluded faces due to large intra-class variations and higher inter-class similarity caused by occlusions. With the inspiration of the human visual system works they detect occluded areas and discard them from face verification. They attempt to address the shortcomings of current face recognition systems when varying the occluded facial area on the face. They detect occlusion space in the image features using mask learning strategy to avoid corrupted feature vectors and introduced feature discarding mask and removed corrupted features by identifying correlations in top convolutional layers of corrupted face images. This work has given state-of-the-art results in general Occlusions.

6. DATA AUGMENTATION TOOLS FOR OCCLUSION GENERATION

[13] introduce tools for generating masked faces to existing face data by adding different types of masks. Their motivation comes because most face recognition systems already have in-house databases of faces that are not operable when the probe face is wearing a face mask. They avoid the threat of invalidating already collected databases of face recognition systems and without taking new pictures and recreating existing databases. It is a computer vision-based script to mask the face. They use DILB for facial key point detection to apply the mask to perfectly fit the face. It provides 100 different mask variations, and it can be used to convert any face dataset to masked face dataset. It supports multiple images in the same image, and it can bulk masking an entire dataset. They introduced a small dataset (MRF2) due to the lack of masked faces in datasets to retrain existing models and experimentally retrain existing face net models and reported the accuracy of face net improved up to 35% when wearing a face mask. It contains 53 identities and 269 images. They addressed the problem of invalidating existing databases of existing systems and evaluate performance improvement when training with masked faces.

[14] introduce three datasets for masked face detection (MFDD), masked face recognition (RMFD) and simulated masked face dataset (SMFRD). They identified the urgency of making a masked face dataset and introduced a real-world masked face dataset containing around 24,771 masked faces. MFDD mainly scrawled on the internet by using their introductory tool called RMFRD which can crawl the frontal facial images. It’s capable of to train a masked face detection purposes and it used to detect when a person is wearing a mask or not as it is beneficial to control the epidemic situation. They introduce a tool to simulate and apply face masks to expand their dataset and add more diversity. The RMFD contains 5000 face images of 525 identities when wearing masks. We can consider their dataset as a good contribution to the addressing data insufficiency problem in masked face recognition.

[15] proposed a novel GAN based method to recreate the occluded area. They detect the masked face and then try to complete the image of the removed masked region. The first face they use binary segmentation to detect masked faces and then remove the area and synthesize. They use

two discriminators to synthesize the face area, one discriminator for learning face structure and another to focus on learning the missing regions. Technically it finds similar patterns of facial features from a database of images and pastes in the occluded part. Their proposed approach removes masked areas automatically using gradually learning two discriminators. To address data insufficiency synthetical masked face dataset using CELEB-A face dataset. It's not qualitative enough for secure authentication and can't be preferable in real-time application because removing masked regions is not efficient.

7. IMPACT OF FACE MASK ON FACE RECOGNITION

Recognizing masked faces is considered as the most difficult facial occlusion challenge because it occludes a large area usually covering around 60% of the frontal face which contains rich features including the nose and mouth. Face mask makes higher inter-class similarity and inter-class variations due to covering a large area of the face which tricks the facial verification process of face recognition systems [14].

The study [16], concludes their results on the effect of wearing a face mask which given strong points of performance degrade on face recognition systems, showing the necessity of development of mask capable face recognition solutions. They evaluate the performance of the two state-of-the-art academic algorithms ArcFace [17] and SphereFace [18] and one non-academic algorithm which is state-of-the-art in face recognition performance. Their experimental setup is used to evaluate performance in face recognition without masks, when wearing a mask, with additional illumination (room light) and without illumination. Data taken under the mentioned four scenarios are used to evaluate three face recognition algorithms. Their results explicitly illustrated the degradation in the verification performance when considering masked face probes. These results illustrated the limitations of current face recognition solutions in matching masked probe faces with unmasked gallery faces.

Ongoing Face Recognition Vendor Test (FRVT) [6] is a performance evaluation report based on 89 face verification algorithms when wearing face masks [6] They baselined the performance of all algorithms using the original unmasked images and then applied synthetic masks digitally by varying mask shape, colour and coverage on acquired datasets from authorized travel or immigration processes. False rejection performance is highly affected when wearing face masks according to evaluation matrices. If masks cover around 70% of the face area most algorithms give false rejection rates between 20% and 50%. Minimum failure rate given by the quite competitive algorithm is 5% when face mask occlusion is less than 70%. In cooperative access control the personal log-on user can be prompted to the second attempt but it's not effective since the failure happens in algorithm level. Furthermore, false accepting rates 1 in 100000 impostors which is very critical in security consequences for verification.

8. DISCUSSION

Future dataset and analysis issues are addressed in this section. In certain cases, a modern study problem necessitates the use of complex datasets. Datasets, on the other hand, represent fundamental issues that must be addressed in the real world. Future challenges in the sense of dataset and research are mentioned in discussion.

The datasets have three major problems: dataset size, diversity of occlusions, and default standards. The occluded facial recognition datasets are on a small scale. AR is one of the few that has actual occlusions, with photographs of only 126 people included. Sunglasses and scarf occlusions are often regarded when it comes to occlusion diversity. Occlusions in real life, on the other hand, are far more varied. Unconstrained occluded facial detection will become a tricky challenge in the future that must be solved.

9. CONCLUSIONS

In this paper presents a comprehensive analysis on occluded face recognition strategies under face mask occlusion within detail comparisons, and we systematically classify approaches into occlusion robust, occlusion discard and occlusion recovery. Recently published and ground-breaking papers were discussed, including novel deep learning approaches. In addition, we demonstrate how face mask occlusions affect face recognition. Reader's attention is brought into room of improvements and future enhancement. Finally, we address upcoming dataset and analysis problems (along with possible solutions) that will help to advance the sector of face recognition under face mask occlusions.

REFERENCES

[1] R. Ranjan *et al.*, "Deep Learning for Understanding Faces," no. January, 2018.

[2] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," pp. 1–23, 2020, [Online]. Available: http://arxiv.org/abs/2006.11366.

[3] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep Face Recognition: A Survey," *Proc. - 31st Conf. Graph. Patterns Images, SIBGRAPI 2018*, pp. 471–478, 2019, doi: 10.1109/SIBGRAPI.2018.00067.

[4] I. Kim and D. Kim, "Optimal design of extended slot allocation for multiuser relay networks," *2012 1st IEEE Int. Conf. Commun. China, ICCC 2012*, pp. 475–480, 2012, doi: 10.1109/ICCChina.2012.6356930.

[5] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Underst.*, vol. 189, no. July, p. 102805, 2019, doi: 10.1016/j.cviu.2019.102805.

[6] M. Ngan, P. Grother, and K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms," 2020, [Online]. Available: https://doi.org/10.6028/NIST.IR.8311.

[7] N. Mclaughlin, J. Ming, and D. Crookes, "Largest Matching Areas for Illumination and Occlusion Robust Face Recognition," pp. 1–13, 2016.

[8] J. Li, T. Qiu, C. Wen, K. Xie, and F. Q. Wen, "Robust face recognition using the deep C2D-CNN model based on decision-level fusion," *Sensors (Switzerland)*, vol. 18, no. 7, pp. 1–27, 2018, doi: 10.3390/s18072080.

[9] D. Sáez Trigueros, L. Meng, and M. Hartnett, "Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss," *Image Vis. Comput.*, vol. 79, pp. 99–108, 2018, doi: 10.1016/j.imavis.2018.09.011.

[10] J. Zhang, C. Zhao, B. Ni, M. Xu, and X. Yang, "Variational few-shot learning," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 1685–1694, 2019, doi: 10.1109/ICCV.2019.00177.

[11] F. Ding, P. Peng, Y. Huang, M. Geng, and Y. Tian, "Masked Face Recognition with Latent Part Detection," pp. 2281–2289, 2020, doi: 10.1145/3394171.3413731.

[12] L. Song, Di. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 773–782, 2019, doi: 10.1109/ICCV.2019.00086.

[13] A. Anwar and A. Raychowdhury, "Masked Face Recognition for Secure Authentication," pp. 1–8, 2020, [Online]. Available: http://arxiv.org/abs/2008.11104.

[14] Z. Wang *et al.*, "Masked Face Recognition Dataset and Application," pp. 1–3, 2020, [Online]. Available: http://arxiv.org/abs/2003.09093.

[15] N. Ud Din, K. Javed, S. Bae, and J. Yi, "A Novel GAN-Based Network for Unmasking of Masked Face," *IEEE Access*, vol. 8, pp. 44276–44287, 2020, doi: 10.1109/ACCESS.2020.2977386.

[16] N. Damer, J. H. Grebe, C. Chen, F. Boutros, F. Kirchbuchner, and A. Kuijper, "The Effect of Wearing a Mask on Face Recognition Performance: an Exploratory Study," 2020, [Online]. Available: http://arxiv.org/abs/2007.13521.

[17] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting Ultimate Accuracy: Face Recognition via Deep Embedding," pp. 1–5, 2015, [Online]. Available: http://arxiv.org/abs/1506.07310.

[18] G. Milhaud and H. Bloch-Michel, “SphereFace: Deep Hypersphere Embedding for Face Recognition,” *Therapeutique*, vol. 47, no. 5, pp. 517–518, 1971.

AUTHORS

Susith Hemathilaka
Final year undergraduate at informatics institute of technology, Sri Lanka

Achala Aponso
Senior lecturer at informatics institute of technology, Sri Lanka

CLASSIFICATION METHODS FOR MOTOR VIBRATION IN PREDICTIVE MAINTENANCE

Christoph Kammerer[1], Michael Gaust[2], Pascal Starke[2],
Roman Radtke[1] and Alexander Jesser[1]

[1]University of Applied Sciences Heilbronn,
Max-Planck-Str. 39, 74081 Heilbronn, Germany
[2]CeraCon GmbH, Talstraße 2, 97990 Weikersheim, Germany

ABSTRACT

Reducing costs is an important part in todays business. Therefore manufacturers try to reduce unnecessary work processes and storage costs. Machine maintenance is a big, complex, regular process. In addition, the spare parts required for this must be kept in stock until a machine fails. In order to avoid a production breakdown in the event of an unexpected failure, more and more manufacturers rely on predictive maintenance for their machines. This enables more precise planning of necessary maintenance and repair work, as well as a precise ordering of the spare parts required for this. A large amount of past as well as current information is required to create such a predictive forecast about machines. With the classification of motors based on vibration, this paper deals with the implementation of predictive maintenance for thermal systems. There is an overview of suitable sensors and data processing methods, as well as various classification algorithms. In the end, the best sensor-algorithm combinations are shown.

KEYWORDS

Predictive Maintenance, Industry 4.0, Internet of Things, Big Data, Industrial Internet, ARMA.

1. INTRODUCTION

The topic of predictive maintenance (PMA) is becoming more and more important for industrial plants and is the key topic in mechanical engineering from the Industry 4.0 aspect [1]. PMA is defined as condition-based maintenance which is carried out on the basis of a wear or service life forecast [2]. PMA uses methods that allow for individual maintenance intervals of an industrial plant to be determined and the maintenance process to be initiated automatically. As part of a R&D cooperation project between CeraCon GmbH and the Heilbronn University of Applied Sciences, a thermal system is to be set up under automation and a PMA strategy is to be implemented, which should then be adaptable to other industrial plants 1. Due to the complexity of industrial plants, an intelligent solution is required in order to be able to offer individual maintenance strategies depending on the state of the plant. For this reason, the project uses machine learning (ML) methods. The essential steps of an intelligent PMA strategy are the digital acquisition of (sensor) data, their evaluation, the analysis of the acquired data and the prediction of probable events. First, possible component defect combinations (CDC) of the industrial plant were analyzed using standard technical risk analysis methods (FMEA, risk graph, fault tree analysis) [3]. CDC is the assignment of a wear component of the industrial system to a potentially occurring defect. Depending on the number of possible defects, a component can therefore have several CDCs. Each CDC was assigned an potential detection measure, e.g. physical vibration measurement or electrical current measurement. Suitable sensors were selected for the analyzed

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 49-60, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111805

detection measures and analyzed with regard to the PMA strategy. CDC’s with the same detection methods were combined and measurement data recorded with the respective sensors. The core of this work is the evaluation of a combination of detection measures for data processing methods and ML algorithms. The optimal combination of these is a prerequisite for an efficient PMA strategy that can be used for the respective industrial plant.

2. STATE OF THE ART

A study by Bearingpoint [4] shows that PMA implementations capture 76% of the relevant data using suitable sensors, although only 59% of the process, measurement and machine data are evaluated in a targeted manner. There are three basic approaches to implementing a PMA strategy [5]. A basic approach is to use the already implemented sensors of the plant for process monitoring. This passive method is particularly suitable for systems that are already in operation. Another passive approach is to introduce dedicated sensors into the system. The additional sensors are introduced to monitor defined wear components and to detect potential defects. In the third approach, a test signal is actively fed into the system. The degree of wear of the components to be monitored can be deduced from the feedback. An example of this is Time Domain Reflectometry (TDR) [5].

3. DATA COLLECTION

3.1. Sensor Resolution

When buying industrial sensors, you often have to commit to a sensor resolution. This requires that you have a basic understanding of what accelerations occur on the component. For this purpose, the effects were previously considered in an experiment when an accelerometer with an insufficient resolution is used. In this case, the sensor generates vibrations that exceed the sensor resolution. A CDC of the fan motor is that the fan wheel has an imbalance. This fault situation was simulated by attaching an unbalance to the fan blade. The result of this simulation is shown in Figure 1(a). There are shown the measured acceleration values in x- and y- axis of an acceleration sensor with a maximum resolution of ±2G. The red values show the vibrations of the motor without an imbalance and the blue values show the vibrations which occurs with an imbalance. It can be clearly seen that the vibrations on the motor increased due to the imbalance. It can also be seen that vibrations that go beyond the set sensor resolution of ±2G were not recorded correctly. They are in line with the maximum acceleration of ±2G. The measured values that did not exceed the maximum resolution were not affected by this. The experiment shows that the CDC "imbalance" can cause very strong vibrations. The vibrations are so strong that they exceed a sensor resolution of ±2G. If a sensor is used that can only record values up to a resolution/acceleration of ±2G, these are recorded incorrectly. The values that exceed the maximum resolution are then incorrectly saved in the data record [6]. To prevent such problems, it is important to see how large the vibrations can be. The sensor resolution should have at least this value with a safety buffer. In Figure 1(b), instead of the resolution of ±2G, the double resolution of ±4G was chosen for the same motor level. In the picture you can see that no "lines" have formed and therefore the vibrations were not greater than the sensor resolution. The resolution of ±4G is therefore much more suitable than the resolution of ±2G. The experiment has shown that a correctly selected sensor resolution is a prerequisite for obtaining meaningful results. If the vibrations are greater than the resolution of the sensor, the incorrectly stored measured values cannot be classified correctly [6].

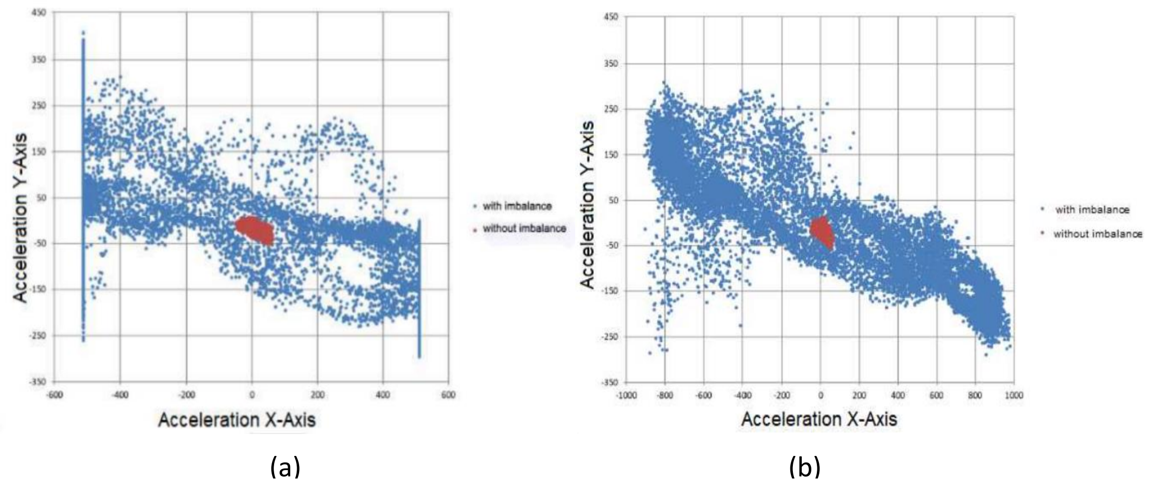


Figure 1. Comparison of an unbalanced fan with a resolution of ±2G (a) and a resolution of ±4G (b).

3.2. Sensors and Test Set-Up

The requirement for a condition-based PMA is a structured data collection of sensor values. The following sensors were used to obtain status data:

- Three-axis acceleration sensors (Accelerometer):
    - LIS 3DH, MMA 8451, ADXL 343, ADXL 345
- Three-axis acceleration sensors with three-axis yaw rate sensor (gyroskope)
    - MPU 60.50
- Three-axis magnetic field sensor (magnetometer)
    - MLX 90393
- Multi sensors with three-axis acceleration, three-axis yaw rate and three-axis magnetic field measurement
    - MPU 92.65, BNO 055, GY 250, GY 521

These recorded the acceleration, the rotation rate and the surrounding magnetic field of the fan motor R3G180-AJ11-XF from ebm-papst Mulfingen GmbH & Co. KG used in the thermal system. Figure 2 shows the measurement set-up with the selected three-axis acceleration sensors [6].

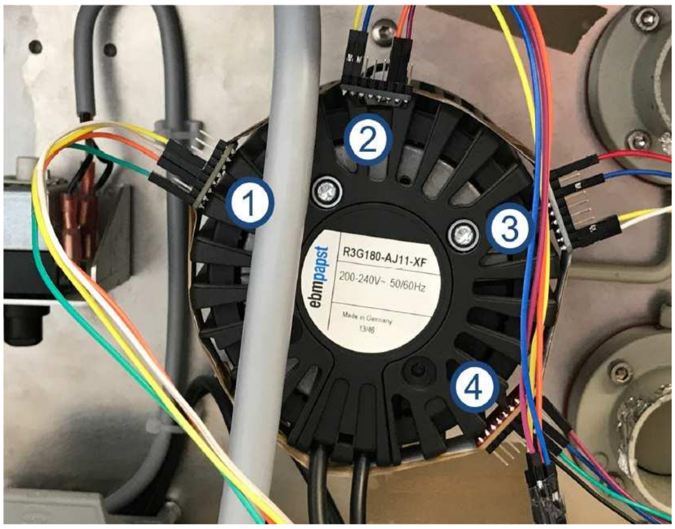


Figure 2. Measurement setup with the three-axis acceleration sensors LIS3DH, MMA 8451, ADXL 345 and ADXL 343.

The fan motor was operated at fixed speeds, which were divided into 7 classes. This classification was based on the specific values 0%, 50%, 60%, 70%, 80%, 90% and 100% of the maximum engine speed. During the operation of the fan motor the vibration of the crankcase was sensed and recorded by the sensors. More than 980,000 structured sensor data sets per measurement series and sensor type were recorded. A total of more than 2.6 million data sets have thus been recorded for all sensor types. A section of a full data set is shown in Table 1.

Table 1. A Section of the Data measured with the MPU 60.50.

AccelX	**AccelY**	**AccelZ**	**GyroX**	**GyroY**	**GyroZ**	**Target**
416	6088	60146	2709	62478	65187	Motor 000
426	6026	60148	2733	62469	65185	Motor 000
404	6110	60146	2727	62478	65192	Motor 000
470	6046	60140	2720	62486	65188	Motor 000

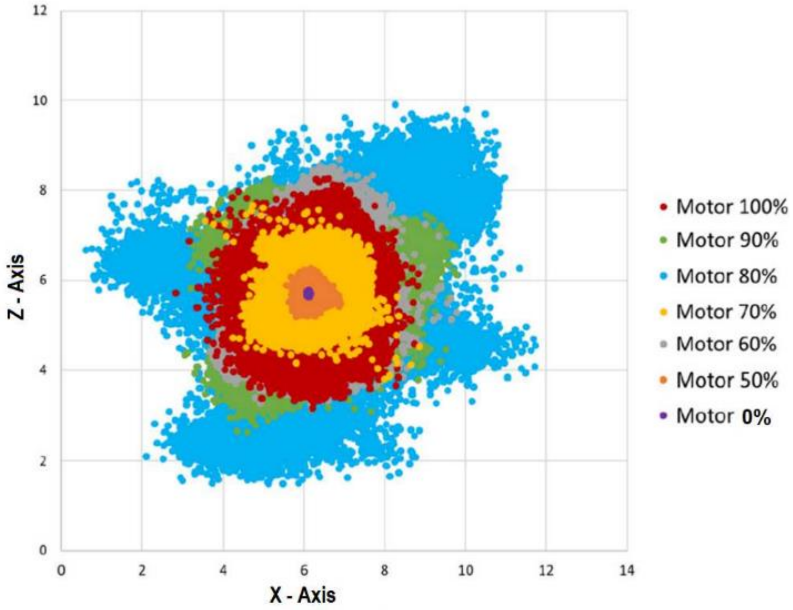


Figure 3. Measurement results from the acceleration sensor MMA 8451.

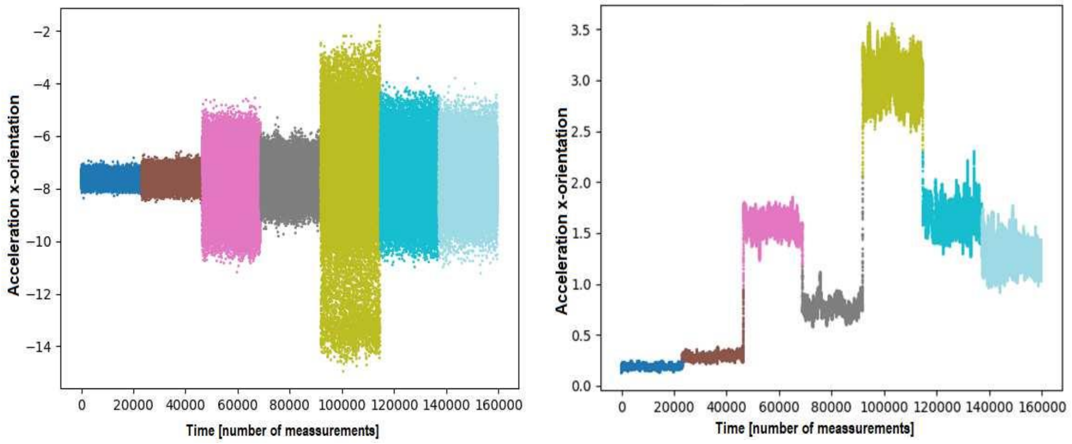
An example of a recorded data set is shown in Figure 3. It shows the measured acceleration from the housing vibration in the spatial x- and z- orientation. The individual classes are highlighted in color to make a distinction possible. Due to the highest spatial coverage, it can be seen that the measurement results for class 80% can be assigned to the resonance range of the fan motor, since the acceleration values in the x- and z- alignment are at their maximum values here.

4. DATA CONDITIONING

In order to be able to better differentiate the individual classes, it is in some cases advantageous if the data records are processed before classification. The methods used for data conditioning are presented here: One possibility to process the data sets consists of the differencing and absolute value formation of subsequent values according to (1)

$$X_{idif} = |X_i - X_{i+1}| \qquad (1)$$

X_i and X_{i+1} are the successive sensor values. Another processing method is the integration of the data according to (2). Here the area under two successive values X_i and X_{i+1} is calculated.



(b)

Figure 4. Comparison of raw data (a) and data prepared by differencing (b).

$$X_{iint} = \begin{cases} X_i + 0.5 * (X_{i+1} - X_i) if X_i < X_{i+1} \\ X_i - 0.5 * (X_i - X_{i+1}) if X_i > X_{i+1} \\ X_i if X_i = X_{i+1} \end{cases} \qquad (2)$$

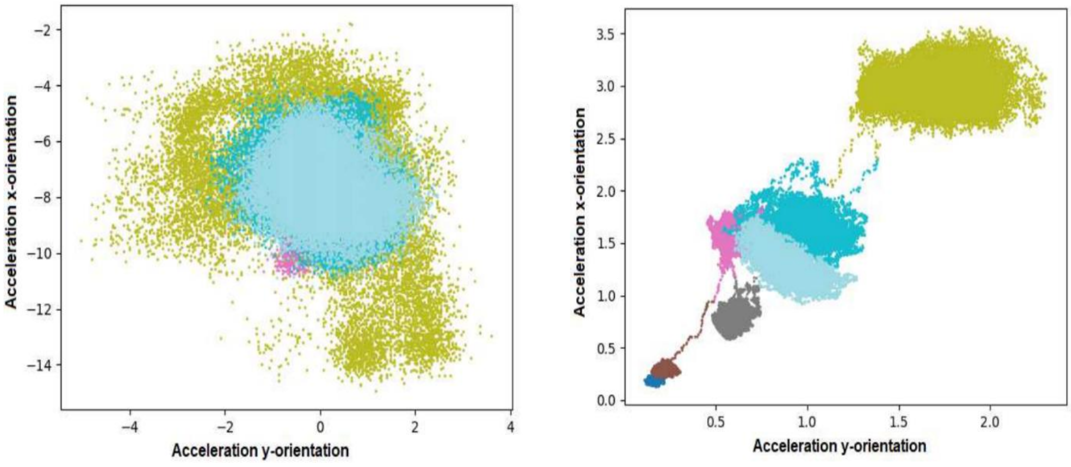
In both the processing methods, an additional smoothing can be carried out by calculating the moving average according to equation (3).

$$X_i = \frac{1}{G} * \sum_{i-g}^{i+g} X_i \qquad (3)$$

The parameter G specifies the degree of smoothing. The parameter g is the difference between the indices between the instantaneous value X_i and the maximum value $X_{g\pm i}$ specified by the degree of smoothing. Thus g depends on the degree of smoothing G and can be determined according to (4).

$$g = \frac{G-1}{2} \qquad (4)$$

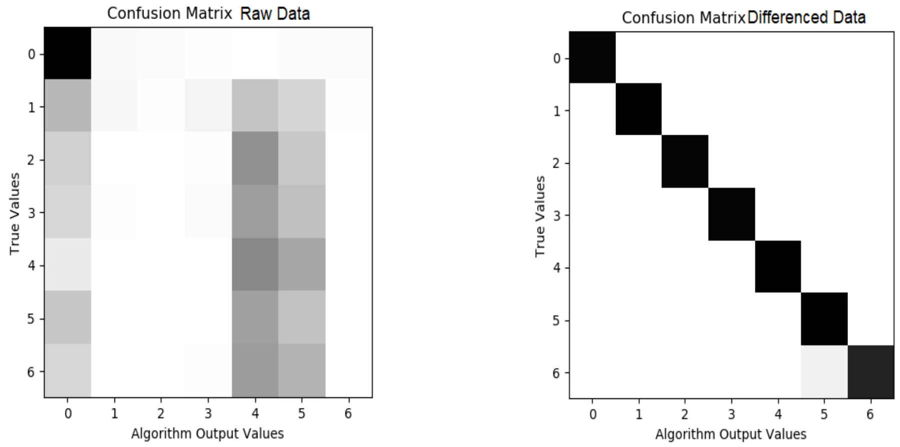
With the degree of smoothing G, first optimizations regarding the classification of the measured sensor data can be carried out [7]. **Error! Reference source not found.** shows the effect of processing by means of differencing compared to the unprocessed raw data. The coloring in the pictures illustrates the different class assignments. **Error! Reference source not found.**(a) shows the acquired raw data of an acceleration sensor in the x-orientation. It can be seen that a delimitation regarding the classes is not clear. For example, the acceleration in the direction of the x-axis at -8 m/s^2 is not unique and can in principle be assigned to any class. In **Error! Reference source not found.**(b) the classes are more delimited after the differencing and smoothing and thus a class assignment is clearer. For example, the value 0.7 can be clearly assigned to the class shown in gray. Figure 5(a) shows the raw data from the measurements in x- and y-orientation. A classification is clearly not possible due to the overlapping point clouds. Figure 5(b), on the other hand, shows the data prepared after the differencing. It can be seen that the point clouds are now clearly distinguishable, making visual and algorithmic class assignment easier.



(a)

Figure 5. Comparison of raw data (a) and data prepared by differencing (b) in two axes.

5. Evaluation



(a)

Figure 6. Comparison of the prediction results of the focus cluster algorithm with raw data (a) and data prepared by differencing (b).

The evaluation of the ML algorithms with regard to the respective sensors and the data processing was divided into a training and a test phase. In the training phase, the data records were divided evenly by feeding every tenth data value of the respective training method to the ML algorithm. As a result, the respective ML algorithm was trained with 10% of the data. The complete data set was then evaluated in the test phase. Several ML algorithms were considered for the recorded data sets. These included decision trees [8][9], the gradient boost method [8][10] a focus cluster algorithm [11] and artificial neural networks (ANN) [12]. The investigations revealed that ANNs are less suitable for data sets with low attribute numbers due to the long duration in the training phase. Therefore, only the decision trees, the gradient boost method and the focus cluster algorithm were used for the further experiments [7]. Figure 6 shows two confusion matrices [13] for the focus cluster algorithm, which show the distribution between the actual class and the class determined by the algorithm. The numbers on the axes correspond to the seven defined classes in which the data records have been categorized. The darker an area, the more often the ML algorithm has assigned data records to a class. A correct assignment is obtained if the assigned class corresponds to the actual class. Ideally, you would get a black diagonal from top left to bottom right. Figure 6(b) shows the result for the data sets prepared after differencing and smoothing. It can be seen that the majority of the data records were assigned to the actual classes, the hit rate here was over 98%. On the other hand, it can be seen in Figure 6(a) that a significantly lower hit rate has been achieved for the unprepared data sets.

6. USING STATISTICAL METHODS FOR PREDICTION

The preceding test was performed by measuring the vibration of a stand-alone motor on a workbench, which was not build into a working machine. Therefore this data cannot be used to make a prediction for maintenance, but the feasibility of categorizing a motor by its vibration and magnetic field was studied. To get a better picture of the real working conditions of such a motor a larger data set was collected by mounting three multifunction sensors (GY 521, BNO 055, MPU 92.65) on a thermal system which is used in day to day operations. These data sets were then used to build a statistical model based on auto regression and moving average (ARMA[14],[15]) of the vibration. The statistical models were created for every sensor orientation separately to get an optimal result for each time series. As the metric to compare the different models was chosen the maximum relative deviation (MRD) according to (5).

$$MRD = \max_i \left(\left| \frac{X_i - Y_i}{X_i} \right| \right) \qquad (5)$$

In (5) "X_i" are the measured sensor values used to build the model and "Y_i" are the values predicted by the model at this time-step.

7. RESULTS

To compare the results of the ML algorithms for the respective sensors, a matrix with the relevant properties was created for each combination of ML algorithm and sensor:

- Classification accuracy (performance) of the algorithms
- Computing time for the training and testing phase of the algorithms
- Smoothing factor G

Table 2 shows the comparison of the processing methods examined using the combination of the multifunction sensor GY 521 and the gradient boost method.

Table 2. Comparison of the data preparation methods with the combination of the multifunction sensor GY521 and the gradient boost method.

GY521 with Gradient Boost method	**Performance (%)**		**Computing Time Training (s)**		**Computing Time Testing (s)**	
	G=0	G=99	G=0	G=99	G=0	G=99
Raw Data	98.51	83.97	4.85	6.88	1.11	0.90
Integration	82.84	85.15	5.45	3.82	1.07	0.66
Differencing	79.76	85.1	5.04	3.83	1.50	0.61

It can be seen that the highest performance is achieved when using the raw data. In the training and test phases the integrated and the differenced data are slightly faster. A comparison was made for each sensor and ML algorithm combination. The best performing data processing method was then selected for each combination.

The comparison tables of the sensors which have achieved the best results of the sensor types examined are listed in Tables 3 to 5. These were the ADXL 345 (accelerometer), the MPU 60.50 (gyroscope) and the GY 521 (accelerometer, gyroscope and magnetic field). The processing method with the highest performance for the respective algorithm is shown for each of the sensors.

Table 3: Comparison of the ML algorithms on the combination ADXL 345 (accelerometer) and the best performing data processing method.

ADXL345	**Decision Tree**	**Gradient Boost Method**	**Focus Cluster Algorithm**
Conditioning method	Differencing G=99	Differencing G=99	Differencing G=99
Performance	93.79%	96.6%	98.98%
Computing time training	1.78s	5.88s	0.32s
Computing time testing	31.99s	1.44s	34.88s

The result of the examinations according to Table 3 is that all acceleration sensors achieved the greatest performance with smoothing (G = 99) and data prepared by differencing. The focus cluster algorithm achieved the highest performance.

Table 4. Comparison of the ML algorithms on the combination MPU 60.50 (accelerometer And gyroskope) and the best performing data processing method.

MPU60.50	**Decision Tree**	**Gradient Boost Method**	**Focus Cluster Algorithm**
Conditioning method	Raw data, G=0	Raw data, G=0	Differencing G=99
Performance	91.46%	95.24%	90.89%
Computing time training	0.09s	5.54s	0.91s
Computing time testing	0.28s	1.46s	31.99s

The result according to Table 4 is that the highest performance was achieved with unprocessed and unsmoothed data with the gyroscopes. The gradient boost process achieved the highest performance.

Table 5. Comparison of the ML algorithms on the combination GY 521 (Multisensor) and the best performing data processing method.

GY521	**Decision Tree**	**Gradient Boost Method**	**Focus Cluster Algorithm**
Conditioning method	Raw data, G=0	Raw data, G=0	Differencing G=99
Performance	95.7%	98.51%	99.89%
Computing time training	0.09s	4.85s	1.27s
Computing time testing	0.28s	1.11s	37.63s

In the case of the multifunction sensors with acceleration, magnetic field sensors and gyroscope, it can be seen from Table 5 that the highest performance was achieved with smoothing (G = 99) and differenced data using the focus cluster algorithm.

Error! Reference source not found.shows the lowest MRDs for each separate sensor orientation with their number of auto regressive (p) and moving average (q) terms.

Table 6: Lowest MRD for each sensor orientation with their corresponding p and q numbers.

	p	**q**	**MRD (%)**
Acc X	5	5	20.3
Acc Y	3	3	24.4
Acc Z	2	5	10.1
Gyro X	7	5	22.4
Gyro Y	4	3	24.7
Gyro Z	7	6	28.3
Mag X	4	5	10.6
Mag Y	3	5	22.6
Mag Z	7	7	28.2

These results show a maximum deviation up 28.3% with the gyroscope values in z orientation. The lowest deviation was reached with acceleration in z orientation and the magnetic field in x orientation with only 10.1% and 10.6% respectively. The difference in accuracy of the models is rather high and therefore an ARMA model can only be used to model two of the nine time series measured. For the remaining seven there should be used other means of modeling.

8. CONCLUSION

It has been found that the processing of the raw data in the form of smoothing and differencing in combination with the focus cluster algorithm gave the best results for acceleration sensors. The gyroscopes examined showed that the unprocessed raw data without smoothing in combination with the gradient boost method achieved the highest classifiability. The multisensors examined gave the best results when using the focus cluster algorithm in combination with smoothed and differenced data. In addition it was found, that an ARMA model could be used to predict the acceleration in z orientation and the magnetic field in x orientation.

9. OUTLOOK

Based on these results, the combination of detection measure, data processing method and ML algorithm can in the next step be used for a PMA strategy. For a complete PMA, further detection measures have to be examined. For that purpose, this procedure is continued with further sensor types in order to find an optimal combination for all necessary detection measures. In the future, a prediction model is to be developed on the basis of these results, with which predictions can be made about the degree of wear of system components of a thermal system under automation. Formal aging and error models of the respective system components must also be created in order to map the aging process of components. These models can then be used to make probabilistic statements about the failure probabilities of the individual assemblies. Such models could be based on Dynamic Bayesian Networks (DBN) [16] auto regression and moving average (ARMA) [17] or, as the focus cluster algorithm has yielded such an high performance, a multi dimensional focus trajectory. In addition to that, the statistical models used to predict the motor vibration in day to day operations could be extended to auto regression, integrated, moving average to get a better result for all sensor orientations.

REFERENCES

[1] O. H. H. R. a. P.-M. S. S. Feldmann, Predictive Maintenance - Service der Zukunft, Roland Berger GmbH, 2017.

[2] „DIN EN 13306: Instandhaltung – Begriffe der Instandhaltung; Dreisprachige Fassung EN 13306:2017 (Deutsche Norm),“ Beuth Verlag, Berlin, 2018.

[3] B. Ebert, Prozessoptimierung bei Industrie 4.0 durch Risikoanalysen, Berlin, Heidelberg: Springer-Verlag, 2018.

[4] R. B. S. G. Frank Duscheck, „https://www.bearingpoint.com,“ BearingPoint GmbH, 18 01 2018. [Online]. Available: https://www.bearingpoint.com/files/BearingPoint_Studie_Maintenance_.pdf. [Zugriff am 07 12 2019].

[5] H. M. Hashemian, „State-of-the-Art Predictive Maintenance Techniques,“ in IEEE Transactions On Instrumentation And Measurement Vol. 60, 2011.

[6] M. Küstner, „Auswahl und Bewertung eines Sensorkonzepts für die Implementierung einer Predictive Maintenance zur Temperaturbehandlung unter Automation,“ Weikersheim, Künzelsau, 2019.

[7] M. Gaust, „Entwurf und Implementierung von maschinellen Lernalgorithmen zur Klassifikation von Maschinendaten für eine vorausschauende Wartung von Industrie-Thermosystemen,“ Weikersheim, Künzelsau, 2019.

[8] Pedregosa, „JMLR 12 Scikit-learn: Machine Learning in Python,“ Journal of Machine Learning Research, 2011.

[9] P. V. A. E. S. Thomas T., Applications of Decision Trees. In: Machine Learning Approaches in Cyber Security Analytics, Singapore: Springer, 2020.

[10] N. D. a. S. P. A. Binding, „Machine Learning Predictive Maintenance on Data in the Wild," in IEEE 5th World Forum on Internet of Things (WF-IoT), Limerick, Ireland, IEEE, 2019, pp. 507-512.

[11] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems, Boston: O'Reilly UK Ltd., 2017.

[12] T. Rashid, Neuronale Netze selbst programmieren: Ein verständlicher Einstieg mit Python, Boston: O'Reilly, 2017.

[13] M. Füllsack, „Systems sciences at ISIS," Institute for Systems Sciences, Innovation and Sustainability Research at the Karl-Franzens University Graz, Austria, 2013. [Online]. Available: http://systems-sciences.uni-graz.at/etextbook/bigdata/confusionmatrix.html. [Zugriff am 28 11 2019].

[14] C.-Y. Lin, Y.-M. Hsieh, F.-T. Cheng, H.-C. Huang und M. Adnan, „Time Series Prediction Algorithm for Intelligent Predictive Maintenance," in IEEE Robotics and Automation Letters, IEEE, 219, pp. 2807 - 2814.

[15] H. Akaike, „Maximum likelihood identification of gaussian autoregressive moving average models," in Biometrika, vol. 60, 1973, pp. 255-265.

[16] A. G. u. E. H. P. Dagum, Dynamic Network Models for Forecasting, San Francisco, CA, United States: Morgan Kaufmann Publishers Inc., 1992.

[17] Y. L. Y. Z. R. X. a. F. Z. Jinjiang Wang, „An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples," in Renewable Energy Vol. 145, 2020, pp. 642 - 650.

LFO2: An Enhanced Version of Learning-From-OPT Caching Algorithm

Yipkei Kwok and David L. Sullivan

Department of Computer Science, Mathematics and Physics,
Missouri Western State University, Saint Joseph, Missouri, USA

Abstract

Recent machine learning-based caching algorithm have shown promise. Among them, Learning-From-OPT (LFO) is the state-of-the-art supervised learning caching algorithm. LFO has a parameter named Window Size, which defines how often the algorithm generates a new machine-learning model. While using a small window size allows the algorithm to be more adaptive to changes in request behaviors, experimenting with LFO revealed that the performance of LFO suffers dramatically with small window sizes. This paper proposes LFO2, an improved LFO algorithm, which achieves high object hit ratios (OHR) with small window sizes. This results show a 9% OHR increase with LFO2. As the next step, the machine-learning parameters will be investigated for tuning opportunities to further enhance performance.

Keywords

Content Delivery Network, Cache, Machine Learning, Supervised learning

1. Introduction

When accessing Internet objects such as web pages and videos, end users experience access latency. To reduce the access latency, rather than delivering objects to their distant end users, it is increasingly common for content providers (e.g., web sites and video streaming service providers) to deliver their objects through content delivery networks [1, 2], where a content delivery network (CDN) is a geographically distributed network of cache servers. When requesting an object, an end user sends a request to a nearby cache server. If the object is available on the server, the server delivers the object directly to the end user. Because of the geographical proximity, the access latency is reduced. Otherwise, the server forwards the request to the content provider and waits for the requested object to arrive from the provider. Upon the arrival, the cache server (1) forwards the object to the end user and (2) optionally cache the object into its local storage for future accesses. While a server may blindly cache each object that it receives from content providers, admission decisions can be made by a *admission algorithm* [1]. When caching a new object, if the cache is already full, the server picks an object to evict. An *eviction algorithm* [3] decides which object(s) to evict. Therefore, the factors that decide the cache content (i.e., which objects are being cached) are (1) requests generated by end-users, (2) the admission algorithm, (3) the eviction algorithm, and (4) the cache size. Ideally, we would want as many requested objects being serviced by a cache server as is possible. The effectiveness of a cache server is quantified by its *object hit rate* (OHR) [4]. If, on average, 30 requests out of 50 are serviced directly at the cache, the OHR is 60%.

This paper addresses admission and eviction algorithms, collectively, as caching algorithms. There are two categories of caching algorithms: heuristics-based algorithms and adaption-based algorithms. Heuristics-based algorithms, such as GDSF [5], make decisions based on specific object properties, or combinations of them. Adaption-based algorithms constantly monitors request behaviors and self-adjust accordingly. This algorithm category consists of three sub-catgories: machine learning-based algorithms (e.g., LRB [6]), and

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 61-68, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111806

non-machine learning-based algorithms (e.g., AdaptSize [1]). Existing caching algorithms will be discussed more thoroughly in the later part of this paper. Among machine learning-based techniques, LFO [7] demonstrates performance that exceeds several latest (representative) techniques in both categories of heuristics-based and adaptive-based techniques. Nonetheless, when experimenting with LFO, the experiment results revealed an opportunity to modify the algorithm, which led to significant improvement in its performance.

The structure of this paper consists of Section 2 that discusses the related work in the literature, Section 3 that provides an overview of LFO, Section 4 that discusses the proposed modification, Section 5 that illustrates the evaluations of LFO2 and results, and Section 6 that concludes this paper.

2. Related Work

Existing techniques for achieving high OHRs on CDN servers can be divided into two categories: heuristic-based technique and adapation-based techniques.

Heuristic-based techniques. These techniques make caching decisions based on assumptions related to object access properties. Examples of these properties are recency and frequency. Well-known recency-based techniques are LAMA [8], 2Q [9], and LRU-K [10], where LFU-K [11] is a well-known frequency-based technique. There are more sophisticated heuristics-based techniques that consider both recency and frequency. An example of these techniques is LHD [12]. These techniques impose a weight, either as a user-defined constant or as an adaptive variable, to decide which of the two properties is more decisive in making the caching decisions. Experiment results show that these techniques are capable of reaching the OHR of nearly 60%. While being relatively easy to implement, a drawback of heuristic-based technique is that, when their underlying assumption no longer holds, their performance suffers. Therefore, their performance may not be robust for request sequences that have varying request behaviors over time.

Adaptation-based techniques. This category of techniques continuously monitor the behavior of the request sequence and the system, and adjust their internal parameters accordingly. Among them, there are two sub-categories: machine learning-based techniques and non-machine learning-based techniques. Some machine-learning based techniques, such as Harvesting Randomness [13], are based on reinforcement learning, while some, such as LFO [7] and Pecc [14], are based on supervised learning. LFO is the state-of-the-art supervised-learning caching algorithm. Some non-machine learning-based techniques are based on Hill Climbing [15, 1, 16], where some are based on mathematical prediction [17, 18, 19]. A hill-climbing algorithm usually runs multiple simulations simultaneously, where each simulates the effect of a parameter candidate. After the simulation, the algorithm re-configures itself to adopt the best parameter candidate.

3. LFO Overview

LFO is an iterative algorithm, where *window size*, a user-specified constant, specifies the length of an iteration. The default value of window size is 1 million, which implies that a new window starts after each sequence of 1 million requests. Figure 1 describes the mechanism of LFO on a high level. During each window, for each incoming request, LFO collects information of (1) timestamp, (2) object ID, (3) object size, and (4) remaining caching space at the request arrival. These information constitutes the features of the request. At the end of each window, LFO deduces a machine-learning model to guide caching decisions during the next window. For each request, the model takes its features and decides if to cache the requested object. It is worth noting that if the model has decided to not cache an object that has just generated a hit, LFO evicts the object anyway. As a result, LFO may consume only a fraction of the cache space. LFO uses lightweight boosted decision trees based on the LGBM library [20] to create such a model using the input of (1) the features of each request arrived during the window, and (2) the best decision that it could have made for each of those requests. To deduce the latter, LFO first budget the cache resources in a window in terms of cache size (i.e., space) and window size (i.e., time). Then, LFO attempts to fit as many requests into the budget as is possible, where requests that fit into the budget are marked as *to-cache*. To maximize OHR, LFO prioritizes requests that consume little resource in terms of cache space and time, where time refers to how long the requested object stayed in the cache before generating a hit. LFO marks requests that do not fit into the budget as *not-to-cache*. This paper calls the resultant sequence of decisions the *optimal caching decisions* and the process described above that produces the sequence *Optimal Caching*. Practically, such a model maps the feature of each request to the best decision that could have been made. In the next window, given the features of each request, the model predicts the best caching decision for the request aiming at maximizing the OHR of the next window.

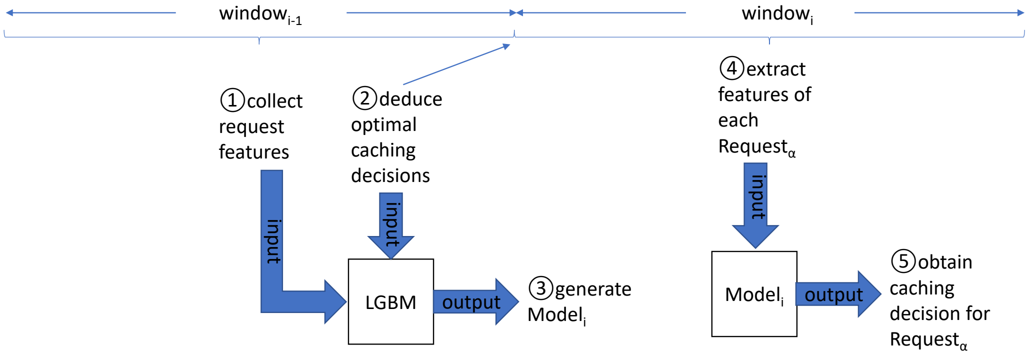

window$_{i-1}$
window$_i$
①collect request features
②deduce optimal caching decisions
input
input
LGBM
output
③generate Model$_i$
④extract features of each Request$_\alpha$
input
Model$_i$
output
⑤obtain caching decision for Request$_\alpha$

Figure 1: Using the request features collected throughout a window, LFO deduces the optimal caching decisions and generates a model for making caching decisions in the next window. [7]

4. The Effects of Window Size on Performance and LFO2

When experimenting with LFO, experiment results revealed that Window Size has a significant impact on the performance of LFO in terms of OHR. Section 4.1 presents the performance results of LFO with different window sizes and the analysis of the performance differences. While the results show that LFO performs better with large window sizes, using small window sizes allows LFO to be more adaptive to changes in request behaviors. This observation agrees with the common understanding in the literature [21, 22, 23]. Based on this performance analysis, This paper proposes LFO2, a revised LFO design, which allows LFO to achieve high OHRs using small window sizes. Section 4.2 presents LFO2 in detail.

4.1 Window Size Influences on The Performance

The default window size of LFO is 1 million. LFO, which was implemented on WebCacheSim2 [24], was evaluated with window sizes of 1, 2, 4, and 8 millions using the CDN1 trace [25]. Table 1 summarizes the results. As Window Size increases, OHR increases. Achieving high OHRs requires simultaneously meeting two necessary conditions. First, optimal caching must produce a sequence of caching decisions that leads to high OHR. Second, the machine-learning models produce caching decisions that closely assimilate the optimal caching decisions. To quantify the quality of caching decisions produced by the models, this paper introduces a metric named *Decision Accuracy*, which measures the percentages of caching decisions generated by the models that match the optimal caching decisions.

Table 1: Effects of Window Size on OHR, Decision Accuracy, and the number of False Negatives.

window size (million)	OHR	decision accuracy (%)	false negatives (%)
1	48.72%	84.12%	39.96%
2	58.08%	82.22%	33.49%
4	67.29%	81.60%	26.85%
8	74.76%	82.75%	21.35%

As shown in Table 1, the general trend is that decision accuracy decreases as window size increases with the exception of the window size of 8M requests, which slightly exceeds the decision accuracy when the window size of 2M requests by approximately 0.5%. As window size is doubled, the amount of data used for training is doubled accordingly, which enhances the accuracy of the model [26]. This suggests why the 8 million-requests window size achieves a decision accuracy higher than when the window sizes are 2 millions and 4 millions requests.

However, it is commonly known that request behaviors, including reuse distance and object size, vary over time [21, 22, 27]. As a result, a model generated at the beginning of a window may no longer accurately reflect the request behaviors during the later part in the window. Therefore, this explains why the decision accuracy decreases as the window size increases. Despite of the 8 million-request window size being an exception, the window size of 1 million request still yields the highest decision accuracy among all window sizes.

While the decision accuracy values suggest that our models are fairly accurate, OHRs remained low with small window sizes. Given the two necessary conditions for achieving high OHRs discussed above, the only plausible explanation is that Optimal Caching is unable to produce caching decisions that lead to high OHRs when windows are small. An in-depth analysis shows that Optimal Caching misses many caching opportunities, where it should mark the requests as to-cache as those requests may yield hits with relatively little resources, in terms of space and time. These missed opportunities happen at the final request to each object that is accessed during a window. In the absence of future access information, Optimal Caching assumes that the requested object will not be accessed again and marks request as not-to-cache. While reasonable, those requests may potentially lead to hits if they were, otherwise, marked as to-cache. This paper calls those requests *false negatives*. Figure 2 illustrate how reducing the window size increases the number of false negatives. Given the full knowledge of object access in the trace, we would ideally mark each of the first 8 requests, except the second request to Object 3, as to-cache. Given the window size of 8 requests, without knowing that Objects 1, 2, and 4 will soon be accessed at the beginning of the next window, Optimal Caching marks the last 3 requests in the window as not-to-cache. However, marking them as to-cache would yields 3 additional hits in the next window. Therefore, the last 3 requests in the window are false negatives. As shown on the lower part of the figure, as the window size is reduced by half to 4, the number of false negatives increases to 6. The smaller the window size, the more false negatives there are in the optimal caching decisions. This increases the likeliness that the machine-learning model decides to not cache the accessed objects. This explains why OHR increases as the window size increases.

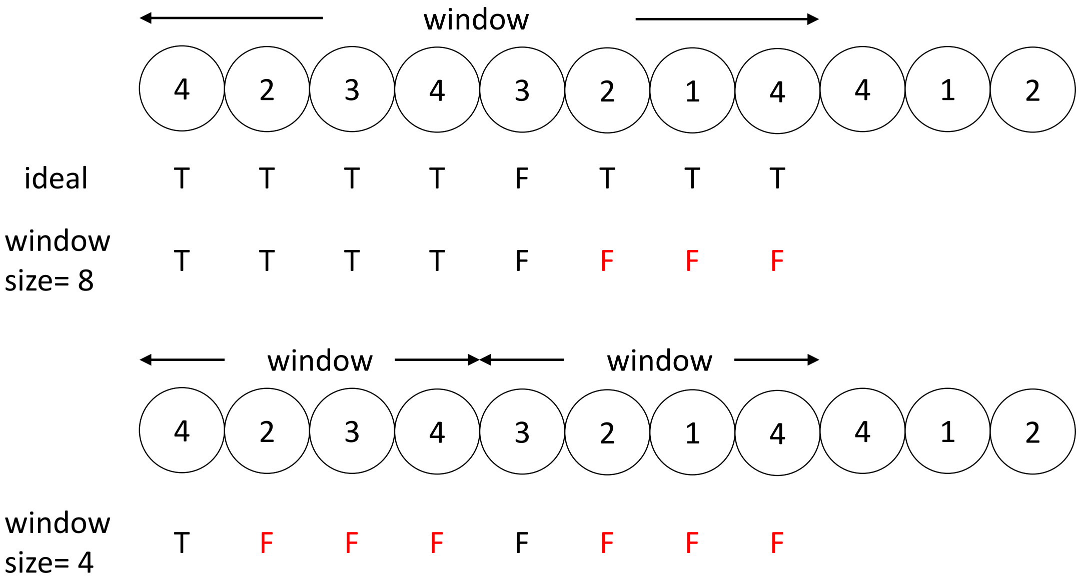


Figure 2: The smaller the window size, the more request are marked as not-to-cache. F's denote false negatives.

To count the number of false negatives in the window sizes of 1, 2, 4, and 8 millions, an ideal sequence of caching decisions, as shown in Figure 2, is needed. Since the number of false negatives decreases as the window size increases, the window size that equals to the trace length yields the sequence of caching decisions containing the fewest false negatives possible. This is because the caching decisions are made with the full knowledge of object accesses. This paper calls such sequence the *ultimate caching decisions*, where the process for deducing the sequence *Ultimate Caching*.

The sequences of optimal and ultimate caching decisions were compared to obtain the percentage of false negatives for each window size, which is presented in Table 1. By increasing the window size from 1 million to 8 millions, OHR increases by over 25% while the percentage of false negatives drops from 19%. This observation hints us that if we manage to reduce false negatives by a certain amount with the window size of 1 million, we should be increase the OHR by a similar amount.

4.2 LFO2 Algorithm

This is how LFO2 reduces the number of false negatives. When its Optimal Caching encounters the final request to each object accessed during a window, rather than marking it as not-to-cache, it gives the request a ``second chance," by consulting the model for the likelihood that the requested object will be accessed again. If the likelihood is high enough, LFO2 marks it as to-cache. Since this is the second chance, it is reasonable for the request to be subject to higher requirement, in terms of the likelihood. LFO2 marks the request as to-cache only if its likelihood is higher than a threshold, known as *Likelihood Threshold* in this paper. The idea of giving objects that are about to be evicted a second chance has been shown to be enhance performance in areas such as memory paging [28]. Note that it is inevitable that LFO2 mistakenly marks a request as to-cache when trying to give the request a second chance. In the actual request arrival sequence, those requests may actually consume much cache resource in order to generate hits and, therefore, should not be cached. This paper calls those requests *false positives*. As for false negatives, false positives are identified by comparing the sequences of ultimate and optimal caching decisions. Section 5 presents the performance results of LFO2 and the effects of Likelihood threshold on the numbers of false negatives and positives.

5. Evaluation

This section evaluates the effectiveness of LFO2 and compares its performance with that of LFO. Section 5.1 discusses the experiment setup that we used for the evaluation, while Section 5.2 discusses and analyzes the evaluation results.

5.1 Experiment Methodology

We implemented LFO2 on WebCacheSim2 and evaluated it with the CDN1 trace that we used in Section 4.1. To ensure that none of the caching algorithm may achieve high OHRs by naively caching every object requested for future use, the cache size was set to 256GB, which is only one tenth of the total size of objects requested by the cache. As mentioned in Section 4, LFO attempts to correct the false negatives among the optimal caching decisions. To quantify the optimality of the optimal caching decisions generated by LFO2, we measured the percentage of optimal caching decisions that matched the ultimate caching decisions. Among the mismatches, there are false negatives and false positives, where the latter are those requests that Ultimate Caching recommends not to cache while LFO2 mistakenly suggested to cache. To understand how LFO2 affect the cache space consumption, we reported the maximum amount of cache space consumed in each experiment. The literature recognizes that small windows all scheduling algorithms to be more adaptive to changes in workload characteristics through more frequent self-adjusting. To reaffirm this common belief, we measured the *Decision Correctness*, the percentage of decisions made by the model during a window that matches the optimal caching decisions deduced at the end of the window.

5.2 Evaluation Results

To evaluated the claim that LFO2 is capable of achieving high OHRs even with small windows, LFO2 was experimented with the window size of 2 million requests. Table 2 presented the results. The table also includes the LFO results for window sizes of 2 and 8 millions for reference. Note that, among the 4 window sizes experimented with LFO, it performed the best with the window size of 8 millions. LFO2 outperformed LFO with the window size of 8 million requests by nearly 9%. It achieved so by dramatically reducing the number of false negatives, by up to 33%, as shown in the table. The results show that, as Likelihood Threshold decreases, the number of false negatives decrease accordingly. This is because, the lower the threshold is, the more likely Optimal Caching marks the final request to each object begin accessed during a window as to-cache.

Table 2: The performance (OHR) of LFO2 with window size of 2 million requests. Data of false negatives and positives and maximum cache occupancy are presented to help understanding performance. LFO results of window sizes of 2 and 8 million requests are presented for readers' reference.

	window size (million)	Likelihood threshold	OHR	false negatives (%)	false positives (%)	maximum cache occupancy (GB)
LFO	2	N/A	58.08%	33.48%	0.00%	58
LFO	8	N/A	74.76%	74.76%	0.00%	177
LFO2	2	0.5	83.64%	0.53%	3.73%	254
	2	0.6	83.65%	0.74%	3.72%	254
	2	0.7	83.31%	1.01%	3.70%	254
	2	0.8	81.42%	4.60%	3.49%	254
	2	0.9	58.10%	33.44%	0.00%	58

However, lowering the likelihood threshold also increases the possibility that Optimal caching mistakenly marks requests as to-cache. This explains why, as Likelihood Threshold decreases, the percentage of false positives increased, though slightly. Nonetheless, the percentage of false positives never exceeded 4%.

By reducing false negatives in optimal caching decisions, LFO2 creates models that have higher tendency to cache objects. This is reflected at the max cache occupancy (i.e., the largest amount of cache occupied throughout an experiment). For LFO, as the window size increases from 2 to 8 million requests, the the models become more inclined to cache objects. As a result, the maximum cache occupancy increases from 58 to 177GB. Similarly, in LFO2, with Likelihood Threshold below 0.9, the models identified more requests, whose accessed objects, if cached, would generate hits to enhance the OHR. Therefore, the maximum cache occupancy further increases to 254GB, which almost filled up the cache. This paper argues that the increase in occupancy is justifiable based on the OHR improvements by LFO2. Indeed, in the literature, high resource utilization is a desirable property.

6. Conclusion

Achieving high OHRs on content delivery networks is crucial to end-user experience when accessing Internet content. Among machine learning-based techniques for achieving high OHR, LFO is one of the state-of-the-art algorithms. This paper presented LFO2, an improved version of LFO. LFO2's machine-learning models are more accurate in identifying objects that increase the OHR, if cached. We evaluated LFO2 using a trace of requests captured in a production environment. LFO2 achieves a 9% OHR improvement over LFO. While LFO2 almost fully occupied the entire cache in order to achieve such a performance improvement, we see it as LFO2's advantage as it better utilizes cache resources. In the future, the possibility of machine-learning parameter tuning for enhancing caching performance will be investigated.

References

[1] D. S. Berger, R. K. Sitaraman, and M. Harchol-Balter, ``Adaptsize: Orchestrating the hot object memory cache in a content delivery network," in *14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17)*, 2017, pp. 483--498. *Cited on page(s):* 1, 2

[2] A. Blankstein, S. Sen, and M. J. Freedman, ``Hyperbolic caching: Flexible caching for web applications," in *2017 USENIX Annual Technical Conference (USENIX ATC 17)*. Santa Clara, CA: USENIX Association, 2017, pp. 499--511. [Online]. Available: https://www.usenix.org/conference/atc17/technical-sessions/presentation/blankstein *Cited on page(s):* 1

[3] M. Bilal and S.-G. Kang, ``A cache management scheme for efficient content eviction and replication in cache networks," *IEEE Access*, vol. 5, pp. 1692--1701, 2017. *Cited on page(s):* 1

[4] D. Menasce, ``Scaling web sites through caching," *IEEE Internet Computing*, vol. 7, no. 4, pp. 86--89, 2003. *Cited on page(s):* 1

[5] Z. Zhao, Y. Ma, and Q. Cong, ``Gdsf-based low access latency web proxy caching replacement algorithm," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, ser. CSAI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 232–236. [Online]. Available: https://doi.org/10.1145/3297156.3297237 *Cited on page(s):* 1

[6] Z. Song, D. S. Berger, K. Li, and W. Lloyd, ``Learning relaxed belady for content distribution network caching," in *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*. Santa Clara, CA: USENIX Association, Feb. 2020, pp. 529--544. [Online]. Available: https://www.usenix.org/conference/nsdi20/presentation/song *Cited on page(s):* 1

[7] D. S. Berger, ``Towards lightweight and robust machine learning for cdn caching," in *Proceedings of the 17th ACM Workshop on Hot Topics in Networks*, 2018, pp. 134--140. *Cited on page(s):* 2, 3

[8] X. Hu, X. Wang, Y. Li, L. Zhou, Y. Luo, C. Ding, S. Jiang, and Z. Wang, ``LAMA: Optimized locality-aware memory allocation for key-value cache," in *2015 {USENIX} Annual Technical Conference ({USENIX}{ATC} 15)*, 2015, pp. 57--69. *Cited on page(s):* 2

[9] T. Johnson, D. Shasha *et al.*, ``2q: a low overhead high performance bu er management replacement algorithm," in *Proceedings of the 20th International Conference on Very Large Data Bases*. Citeseer, 1994, pp. 439--450. *Cited on page(s):* 2

[10] E. J. O'neil, P. E. O'neil, and G. Weikum, ``The lru-k page replacement algorithm for database disk buffering," *Acm Sigmod Record*, vol. 22, no. 2, pp. 297--306, 1993. *Cited on page(s):* 2

[11] L. B. Sokolinsky, ``Lfu-k: An effective buffer management replacement algorithm," in *Database Systems for Advanced Applications*, Y. Lee, J. Li, K.-Y. Whang, and D. Lee, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 670--681. *Cited on page(s):* 2

[12] N. Beckmann, H. Chen, and A. Cidon, ``LHD: Improving cache hit rate by maximizing hit density," in *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*. Renton, WA: USENIX Association, Apr. 2018, pp. 389--403. [Online]. Available: https://www.usenix.org/conference/nsdi18/presentation/beckmann *Cited on page(s):* 2

[13] M. Lecuyer, J. Lockerman, L. Nelson, S. Sen, A. Sharma, and A. Slivkins, ``Harvesting randomness to optimize distributed systems," in *Proceedings of the 16th ACM Workshop on Hot Topics in Networks*, 2017, pp. 178--184. *Cited on page(s):* 2

[14] A. Bhardwaj and V. Janardhan, ``Pecc: Prediction-error correcting cache," in *Workshop on ML for Systems at NeurIPS*, vol. 32, 2018. *Cited on page(s):* 2

[15] S. Bansal and D. S. Modha, ``Car: Clock with adaptive replacement." in *FAST*, vol. 4, 2004, pp. 187--200. *Cited on page(s):* 2

[16] A. Cidon, A. Eisenman, M. Alizadeh, and S. Katti, ``Cliffhanger: Scaling performance cliffs in web memory caches," in *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, 2016, pp. 379--392. *Cited on page(s):* 2

[17] G. Almási, C. Caşcaval, and D. A. Padua, ``Calculating stack distances efficiently," in *Proceedings of the 2002 workshop on Memory system performance*, 2002, pp. 37--43. *Cited on page(s):* 2

[18] X. Hu, X. Wang, L. Zhou, Y. Luo, Z. Wang, C. Ding, and C. Ye, ``Fast miss ratio curve modeling for storage cache," *ACM Transactions on Storage (TOS)*, vol. 14, no. 2, pp. 1--34, 2018. *Cited on page(s):* 2

[19] C. A. Waldspurger, N. Park, A. Garthwaite, and I. Ahmad, ``Efficient {MRC} construction with {SHARDS}," in *13th {USENIX} Conference on File and Storage Technologies ({FAST} 15)*, 2015, pp. 95--110. *Cited on page(s):* 2

[20] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, ``Lightgbm: A highly efficient gradient boosting decision tree,'' *Advances in neural information processing systems*, vol. 30, pp. 3146--3154, 2017. *Cited on page(s):* 2

[21] S. Arunagiri, Y. Kwok, P. J. Teller, R. Portillo, and S. R. Seelam, ``A heuristic for selecting the cycle length for FAIRIO,'' Department of Computer Science, The University of Texas at El Paso, Tech. Rep. UTEP-CS-11-53, September 2011. *Cited on page(s):* 3, 4

[22] S. Arunagiri, Y. Kwok, P. Teller, R. Portillo, and S. Seelam, ``Fairio: An algorithm for differentiated i/o performance,'' in *Computer Architecture and High Performance Computing (SBAC-PAD), 2011 23rd International Symposium on*, oct. 2011, pp. 88 --95. *Cited on page(s):* 3, 4

[23] S. Arunagiri, Y. Kwok, P. J. Teller, R. A. Portillo, and S. R. Seelam, ``Fairio: A throughput-oriented algorithm for differentiated i/o performance,'' *International Journal of Parallel Programming*, vol. 42, no. 1, pp. 165--197, Feb 2014. [Online]. Available: http://dx.doi.org/10.1007/s10766-012-0217-6 *Cited on page(s):* 3

[24] ``Webcachesim2 - a simulator for cdn caching and web caching policies.'' https://github.com/sunnyszy/lrb, accessed: 2021-07-30. *Cited on page(s):* 3

[25] ``Data: publicly available cdn request trace (from a cdn server in san francisco serving wikipedia production traffic) used in the evaluation.'' https://www.cs.cmu.edu/~dberger1/data_and_software.html, accessed: 2021-07-30. *Cited on page(s):* 3

[26] ``Lgbm parameter tuning for better accuracy.'' https://lightgbm.readthedocs.io/en/latest/Parameters-Tuning.html?highlight=parameter%20tuning, accessed: 2021-07-30. *Cited on page(s):* 3

[27] S. Arunagiri, Y. Kwok, P. Teller, R. Portillo, and S. Seelam, ``FAIRIO: a throughput-oriented algorithm for differentiated I/O performance,'' *Int. J. of Parallel Programming*, pp. 1--33, 2012. *Cited on page(s):* 4

[28] K. M. Saleem, M. Iqbal, H. Saadi, F. Qazi, and D.-e.-S. Agha, ``Second chance page replacement algorithm with optimal (scao),'' in *2019 International Conference on Information Science and Communication Technology (ICISCT)*, 2019, pp. 1--5. *Cited on page(s):* 5

SURVEY ON SOME OPTIMIZATION POSSIBILITIES FOR DATA PLANE APPLICATIONS

Gereltsetseg Altangerel, Tejfel Máté

Department of Programming Languages and Compilers, ELTE,
Eötvös Loránd University, Budapest, Hungary

ABSTRACT

By programming both the data plane and the control plane, network operators can customize their networks based on their needs, regardless of the hardware manufacturer. Control plane programming, a major component of the SDN (Software Defined Network) concept, has been developed for more than 10 years and successfully implemented in real networks. Efforts to develop reconfigurable data planes and high-level network programming languages make it truly possible to program data planes. Therefore, the programmable data planes and SDNs offer great flexibility in network customization, allowing many innovations to be introduced on the network. The general focus of research on the data plane is data-plane abstractions, languages and compilers, data plane algorithms, and applications. This paper outlines some emerging applications on the data plane and offers opportunities for further improvement and optimization.

KEYWORDS

Data plane, load balancing, in-network caching, in-network computing, in-network data aggregation, INT.

1. INTRODUCTION

Efforts to create fully programmable networks have been going on for a very long time. SDN technology is part of this long history. Many factors in business and technology have led to the creation of SDNs and fully programmable networks. In the case of a technological view, the main difficulty of traditional networks is a proprietary system, so the introduction of new technologies and protocols is very slow, and network management is complicated. From a business perspective, network operators needed to reduce investment and operation costs, and take full control of their networks: defining their own control plane and data plane algorithms.

To meet these requirements, the SDN defines two important features: first, separate the data plane and control panel, and second, the controller platforms can control multiple forwarding elements using a well-defined API (Application Programming Interface), such as OpenFlow, one of the successful protocols. Not only do they simplify network management, but they also open up development gateways to each control plane and data plane, allowing for many network innovations.

So far, SDN has made significant strides in the industry and has conducted extensive research on the control plane. For instance, commercial switches have supported OpenFlow, and a variety of

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 69-81, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111807

controller platforms have emerged, based on which several control plane applications have developed and implemented in the major data centers such as Google.

A fully programmable network has two pillars: a programmable data plane and a programmable control plane. The data plane programming began to be discussed in the 2000s with the advent of the merchant chip, but it became a reality in 2015, and research in this area is in great demand today [1].

In recent years, researches on developing data plane programming languages, creating programmable switching architecture, and improving the performance of programmable switches have become more mature. Thanks to the success of these fundamental studies and implementations, data plane applications are evolving, and some are entering production and beginning to bear fruit.

The general focus of research on the data plane is data-plane abstractions, languages and compilers, data plane algorithms, and applications [2]. This paper summarizes the research of some data plane applications and provides some ideas in detail on how to improve and optimize these applications.

The rest of this survey paper is organized as follows. Section 2 provides background information on data plane programming, Section 3 describes the data plane applications and future optimization ideas, and the final section presents conclusions and future work.

2. BACKGROUND

2.1. Data plane programmability

Network devices process packets with the help of control plane and data plane algorithms. Data plane algorithms define the forwarding behaviour of a network device (packet processing stages, tables, and so on), while control plane algorithms define rules for manipulating a packet in the data plane, sense network, detect network failures, and update packet processing rules [3]. In the SDN network, the control plane algorithms running on the controller platform (e.g., server) manage the data plane. For example, routing algorithms in the control plane define packet forwarding rules based on the destination IP address. These rules are installed in the routing table of the data plane via API.

As a result of many years of research and improvement on SDN, the control plane can be flexibly programmed by the end-user (network operator). The data plane also needed to be programmed by the end-user to introduce innovations quickly in the network. In a traditional network, the implementation process of a new feature [4] or protocol goes through many stages, from software developer to standard organization and chip designer. For example, as a result of this process, it took 4.5 years for the VxLAN protocol to the network from the first proposal [5]. This can be seen as one of the critical reasons why internet architecture has not changed for many years. If the data plane is flexible enough, both the end-user and hardware vendor can quickly deploy the new features [3].

The following research works have been taken to make the data plane programmable:

1. **Developing data plane programming languages**: Domain-specific programming languages for defining data plane algorithms and functionalities (forwarding behaviour) are being developed. Examples include FAST [6], Domino [7], Protocol-Oblivious Forwarding [8], and NetKAT [9], P4 [10], with P4 being the most successful.

2. **Creating data plane architectures**: To map the data plane algorithm defined by domain-specific language to the switch ASIC hardware, the hardware vendor must provide the data plane architecture (programmable building blocks and data plane interface between them)[11]. This architecture is also called a data plane model or hardware abstraction in some literature. For example, architecture for Tofino programmable switches is protocol independent switching architecture (PISA) and based on this architecture, data plane algorithms can be defined in P4.
3. **Improving performance of programmable switch**: The belief that the performance of a programmable switch cannot reach the performance of a fixed-function switch is obsolete, and thanks to numerous studies and technologies, the performance of a programmable switch has approached/same as that of a fixed-function switch [12].
4. **Defining API**: Providing an interface for connecting the control plane and the programmable data plane. For example, the P4 compiler creates an API that connects the data plane to the control plane [11].

2.2. P4 language

P4 is a domain-specific programming language for defining packet processing algorithms on the data plane of a programmable network device (target)[11]. This section briefly describes the P4 language, its advantages, programmable packet processing components that can be defined in P4, and how to run P4 code on the target.

P4 is currently one of the most popular and well-defined languages and has two main advantages: target and protocol-independency, so a lot of data plane applications currently in development are on P4. Target-independency means that P4 program can run on any type of target. To ensure this feature, the hardware vendor must implement a generic architecture and a compiler backend fora given target: provide them to the P4 developer [13] and so, the P4 program is easily mapped to the target with help of these. Protocol-independency means that P4 developers can define their rich set of protocols and data plane behaviour/functionalities.

According to the architecture, the main blocks that can be programmed on P4 are packet parser, match-action units (one or more), and deparser. The general data plane architectures of P4 for research purposes are V1 and Portable Switch Architecture (PSA) [14] and, optimization of the P4 data plane application has been doing based on these architectures. Figure 1 describes basic pipeline in V1 model architecture. Parser recognizes incoming packets and extracts headers and fields from the packet. After this, the match-action pipeline processes extracted packet headers. A match-action unit contains one or more tables and actions. For example, the IPv4 routing table showed in Figure 2 can be created here, and the match key is the destination IP address and based on which, corresponding actions such as drop or forward are performed. In this stage, the header can be added, subtracted, and modified. The deparser builds the outgoing packet by assembling the processed headers and the original packet payload [10]. In the case of PSA architecture, it is possible to define more detailed pipelines with more than one pair of parser and deparser for ingress and egress. Also, the match-action tables and external functions can be determined between parser and deparser.

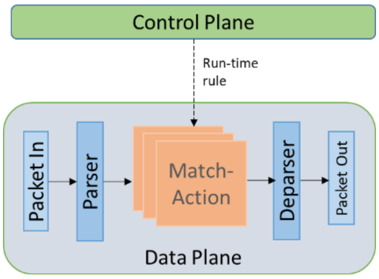
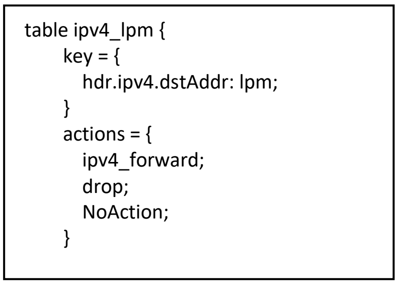


Figure 1. Abstract packet forwarding in P4



```
table ipv4_lpm {
    key = {
       hdr.ipv4.dstAddr: lpm;
    }
    actions = {
       ipv4_forward;
       drop;
       NoAction;
    }
```

Figure 2. IPv4 table example

3. Data Plane Applications and Optimizations

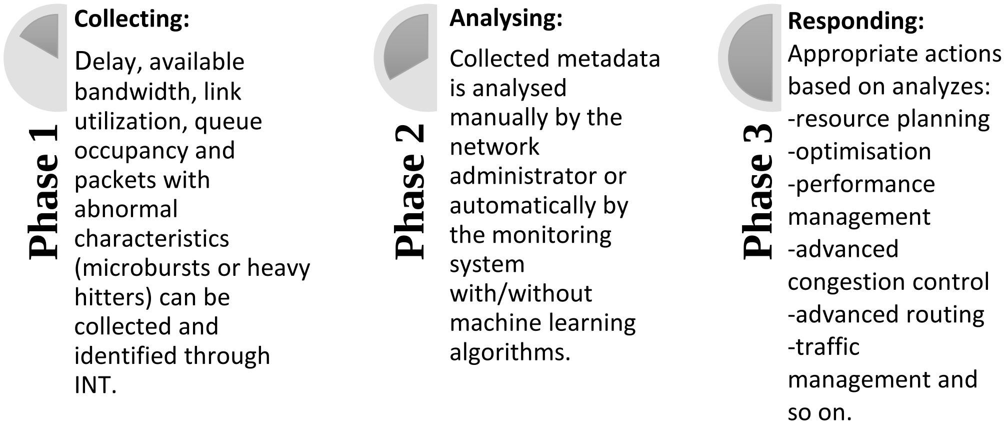
Developing novel and optimal applications on the data plane are one of the interesting areas of data plane research. The general directions of applications are in-band telemetry, load-balancing, in-network computation, deployment of consensus protocols, congestion control, queue management and traffic management [2]. This section describes some of these applications, their motivation, approach, challenges, and future improvement and optimization possibilities.

3.1. In-band Network Telemetry (INT)

One of the foundations for effective network management and control is network measurement. More accurate, precise, real-time measurements are considered good. Traditional measurement and monitoring methods are active methods (ping, traceroute), passive methods based on traffic mirrors, and hybrid method -a combination of these [15]. These are simple to deploy, but the downside is that they put extra load on the network during monitoring, so they can't be much precise in some cases. In other words, the process of measuring the network itself can affect the state of the network.

With the advent of programmable data planes, the In-band Network Telemetry framework, a more direct network measurement, is originated on a data plane without the involvement of a control plane. The basic idea is to collect the status of network devices (metadata) using a normal packet or probe packet (INT packet) that is transmitted over the network. Intermediary devices embed their own metadata into the INT packet. Therefore, it does not create a much more additional load on the network compared with traditional measurement. Also, it is more detailed, accurate and near-real time. One disadvantage is that metadata is limited by packet’s maximum transmission unit (MTU). INT instructions (header) on what to collect from the devices are added to packet at the source INT node and then that packet is transmitted through network for collecting device’s state. The metadata and INT header is removed from packet on the edge device (INT sink node). The sink node then performs the appropriate monitoring or actions, for example, it forwards the collected report to another external device or server for further monitoring [16].

INT operation can be divided into 3 phases:



Collecting:

Delay, available bandwidth, link utilization, queue occupancy and packets with abnormal characteristics (microbursts or heavy hitters) can be collected and identified through INT.

Analysing:

Collected metadata is analysed manually by the network administrator or automatically by the monitoring system with/without machine learning algorithms.

Responding:

Appropriate actions based on analyzes:
-resource planning
-optimisation
-performance management
-advanced congestion control
-advanced routing
-traffic management and so on.

Figure 3. INT phases

3.1.1. Recent optimization works around INT

Optimization on phase 1: The most important thing to consider when collecting network status via INT is not to compromise the performance of the network, intermediary devices, and the monitoring servers. To fulfill this requirement, it is important to determine the proper size and structure of the INT packet within the MTU, minimize the number of flows/packets for telemetry, filter unimportant telemetry information, and choose optimal collection mechanisms. Some studies around these are:

- **Optimizing telemetry data:** It can be optimizing number of monitored network elements, not including the INT header in all flow packets [17], sampling packets for monitoring, and using threshold to identify in-band monitored flow. For example, P4-based INT doesn't support sampling; therefore, adding an INT header to all incoming packets will create high network overhead in a large scale network. One of the works on this [18] suggests sampling strategies based on rate and event. In the rate based strategy, the INT source node inserts the INT header into every Rth packet, where R is a configurable parameter. Another work [18] bound the amount of information added to each packet.
- **Intelligent trigger:** Fault detection platform with event-based and policy-based trigger is considered intelligent mechanism. The event can be detected by data plane or monitoring server. This solution [17] offloads event detection from monitoring server to an in-network P4 application and it reduces network overhead and monitoring server load. KeySight [19] suggested event-triggered fault detection platform based on Bloom filter. PAINT [20] offers policy based detection (by monitoring system): network operators use Service Provisioning Language (SPL) to define and deploy in-band network telemetry services. PAINT automatically parses service policies.

Since the network device and the monitoring node (server, sink, and analyser) both have limited processing capabilities, it is important to determine whether it is optimal to detect the event with either a network device or a monitoring server. In addition, event-based and policy-based detection algorithms themselves are one thing to optimize.

Optimization on phase 2: To make monitoring more effective, it is being combined with machine learning methods. The machine learning part of the monitoring system is called the knowledge plane in some works [21] or the knowledge-defined network. An example of these solutions is Barefoot Deep Insight [22], which is the world's first commercial-level packet-by-packet status monitoring system. Combined with machine learning technologies, Barefoot Deep Insight can realize automatic abnormality monitoring of network performance at nanosecond time, including microburst detection, abnormal packet loss detection, abnormal queue detection, and so on.

Optimization on phase 3: The applications such as congestion control and advanced routing based on INT should be efficient. For example, the efficiency of congestion control is evaluated based on the congestion detection, and the resolving time. There are congestion control mechanisms on different INT characteristics such as link load based [23], rate-based [24], and queuing and processing delay based congestion control [25]. The efficiency of these solutions should be determined by performance evaluation.

Therefore, in addition to the optimization ideas mentioned at each of the above INT stages, there are opportunities for further research to develop solutions for other types of networks, such as adapting or expanding the in-band telemetry system in a wireless sensor network (WSN) and Internet of Things (IoT) data network. For example, IoT packets are too small, making it difficult to identify abnormal behaviour of packet [26].

3.2. In-Network Computing

Traditional network devices often focus on achieving high throughput, so processing is limited on transmitted data. With the advent of flexible programmable networking devices, it has become possible to perform more computations on network devices (in-network computations) without reducing packet processing rates. In other words, it means that a set of computing operations on an end-server and middlebox can be offloaded to a network device. This has the following advantages.

- Higher layer functionalities, such as transport and application layers, are processed at line rate, reducing latency and increasing throughput.
- Reduces traffic, thereby reducing network congestion, which is one of the factors degrading application performance.
- Saving energy by running servers [27].

First of all, it is important to determine what type of computation operations are most optimal to run in-network. According to the studies, the most feasible applications are in-network packet aggregation, in-network caching and applications alleviating control plane load [1]. Since data centres are early adopters of the SDN network, most of these applications are currently designed for data centre networks.

3.2.1. In-Network Packet Aggregation

The group of applications with partition/aggregation patterns in the data centre network includes search, query processing, dataflow computing, graph processing, and stream processing, and deep learning frameworks. During the partitioning phase, job requests are divided into sub-tasks, which are executed in parallel on different worker servers and each worker server produces partial results. In order to obtain the final result, the partial results are collected and aggregated at the aggregation stage. During the aggregation phase, data (e.g partial results) must be transmitted between a large numbers of workers, which puts a heavy load on the network. For example, a trace of Facebook's data center shows that 46 percent of all traffic is generated during the aggregation phase. Furthermore, it leads to network bottleneck [28].

Data aggregation functionality is usually performed at the application layer. If it is done on the network path, traffic load can be reduced. Other reasons for in-network aggregations are that behind these functions are usually simple arithmetic logic operations, so placing them on a switch is simple, and since these algorithms are communicative associative functions, there is no need to pay attention to the packet sequence. DAIET [27] which is built in P4 is an in-network aggregation prototype solution for machine learning and graph analytics applications. However, the solution is generic enough and can be used in various partition/aggregation data centre application.

The next effective segment to use in-network aggregation is to combine small-sized and large numbers of packets. The idea has existed for a very long time, but there is no real implementation. This kind of packet aggregation/disaggregation has many important benefits. For example, aggregating multiple, and small-sized IoT packets into one transmission unit can reduce the additional overhead associated with each transfer. Wang et al. [29] introduce proof-of-concept designs and implementations on IoT packet aggregation and disaggregation purely in P4 pipelines of the switching ASIC.

3.2.2. Applications alleviating control plane load

It is now technically possible to offload most of the tasks on a control plane to a data plane. The main benefit of this is that it can accelerate the control plane. However, some tasks are not optimal to run on data plane because they require a lot of resources. Therefore, research on what tasks should be offloaded on the data plane is one of the interesting topics in the future.

The case study on [30] suggested how to perform key control plane tasks such as failure detection, and notification, connectivity retrieval, and computation on a data plane, and implemented the proposed algorithms in BMv2 P4 software switch. It also discussed the advantages and disadvantages of implementing control plane tasks on a data plan. Another case study in this topic is the implementation of Time-synchronization Protocol (DPTP) on the Tofino programmable switch with P4 pipeline [30]. Global time-synchronization on the data-plane is very much necessary for supporting distributed applications. The key research questions around this are, first, to determine what types of control plane tasks can be optimally deployed on the data plane and how to overcome hardware constraints in deployment.

3.2.3. In-network caching

Modern network services such as search engines, social networking e-commerce are used by billions of users and generate huge amounts of traffic on the network. To view a single web page, you may need to access hundreds or thousands of storage servers in the background [31]. One mechanism to deliver these services to users with high throughput and low latency is caching, which is a crucial way to improve the performance of a storage system. The idea is that to retrieve items on the storage system more quickly, high-access items (hot items) must be temporarily stored in the cache and the cache should be updated regularly based on hot items. The hot items can be changed abruptly, and most users like to access that hot items, which can lead to an imbalance in network traffic. For example, 60-70 percent of Facebook users access 10 percent of the total content [32]. Therefore, when building a caching system, these issues need to be considered.

Traditional networks use flash-based caches, disk-based caches, and server-based caches, and data plane programming provides new opportunity to create in-network caching. This means that it is possible to create a cache on a programmable network device. Because network devices naturally placed on the path between the client and server, creating a cache on the path can further reduce latency. The key-value store data structure is used to build the database in the cache because it is general and widely used by applications. It is used as a basic interface to the caching[33].

Netcache [34] is new key-value store architecture by leveraging flexibility, and the power of a modern programmable switch to handle queries on hot items of the storage server. It is built on top of rack (ToR) switch in the data centre network. Therefore, ToR switch plays important role and has 3 main modules: L2/L3 routing, on-path caching for key-value items, and query statistics. The Query statistic module identifies the hot items, and based on these statistics, the controller updates the cache. The core of Netcache is packet-processing pipeline which detect, index, store and serve key-value items. For example match-action table classify key on packet header and values are stored in register array, on-chip memory in programmable switch. One ToR switch can cache items on a storage server only connected to it, and cannot work with other ToR switches in a coherent way.

IncCache [35] is another in-network, key-value store system built in a programmable data plane. What distinguishes it from Netcache is that it is implemented in the core, aggregation and ToR switch of the data centre network, as well as end-host server, and maintains the cache coherence using a directory-based cache coherence protocol.

These works are good start for in-network caching and both reduce latency by a certain percentage. The Netcache architecture was created on a Tofino and commodity server-based switch with a P4 pipeline, while IncCache was developed on Cavium XPliant switch and the forwarding plane was defined in a proprietary language.

According to the discussion on those works, the following questions can be open in the future: Mostly network requests (read) are processed from the cache. So, can write/delete requests be processed from cache? Do you need compression to reduce the cache size? , and so on.

3.2.4. Consensus protocols (in-network)

Running some application-level protocols on the data plane is another interesting topic: for example, the implementation of consensus protocols for distributed networks in the data plane. Paxos is popular consensus protocol used in fault-tolerant networks and is commonly used in data center applications. Implementing this in the data plane will improve the performance of the protocol itself and the performance of applications based on this protocol service [36]. Data plane programmability allows for tight integration between the application and the network but, the developers should always consider how network-level optimization affects the top level.

3.3. Load Balancing Applications

The main purpose of the load balancer is to efficiently distribute the load over multiple pieces of network infrastructure in order to maximize throughput, minimize response time, and prevent overloading of single resource. The data centre network has redundant resources, so load balancers play an important role in the optimal use of these resources. Data centre networks perform load balancing in more than one way. The L3 load balancer(s) selects one of the many equal-cost paths that can route the packet, while the L4 load balancer(s) chooses the one of serving instances (servers) for the incoming service request [37].

Layer 3 load balancing mechanisms in the Data Centre network and Internet try to choose the congestion-free and optimal path from the multiple paths, so that bisection bandwidth can be used more efficiently. Layer 3 load balancing mechanisms are usually implemented on the data plane.

The most commonly implemented method is the Equal-Cost Multi-path Routing(ECMP) and the per-flow based load balancing mechanism, which randomly assigns one of the equal cost paths to each flow. Because the flow is distributed randomly, performance may be reduced if two elephant flows are allocated in the same path [38]. In addition, it is the congestion-oblivious mechanism that does not track the over-utilized path.

Conga [39] was improved ECMP, and it is a congestion-aware mechanism and maintains the congestion status of each path on the leaf/spine switch in the data centre network. However, due to the limited memory of the switch (leaf), it is not possible to scale this mechanism as the network grows. In addition, because Conga is implemented on custom hardware, it is costly to redesign (requires modification on chip architecture), which means that network operators cannot change the mechanism to suit their network.

With the advent of the programmable data plane, it became possible to develop a customizable load balancing mechanism on the programmable data plane. HULA [40], programmed in P4 is the first load balancing mechanism explicitly designed for the programmable switch architecture and it is scalable and congestion-aware. Conga centralizes the congestion track at one point (leaf switch), while HULA does it in a distributed manner. Each HULA switch maintains only the congestion state for the best next hop to reach the destination, not the congestion state for all paths to the destination. Therefore, each HULA switch makes a local decision when selecting a path, while the CONGA depends on the leaf switch. By tracking congestion in this distributed way, scalability is better than CONGA. In addition, it can automatically detect network failures. Therefore, this work inspires network operators to create a more optimal load balancing mechanism for their network in a programmable data plane.

The layer 4 load balancers could be hardware, cluster of servers and commodity server, and they are usually implemented on the commodity server in data center. It is also called software load balancers (SLB). Thanks to data plane programmability, they can also be developed on the switching ASIC. When designing a Layer 4 load balancer, the following two are important. First, the incoming connections to the servers must be very well-tuned to the bisection bandwidth of the physical network (uniform load distribution of the incoming connections across the network and servers). Second, providing per connection consistency (PCC): the ability to map packets belonging to the same connection to the same server, even if there are presence changes to the active servers and load balancers. But, meeting both these requirements at the same time has been an elusive goal [41].

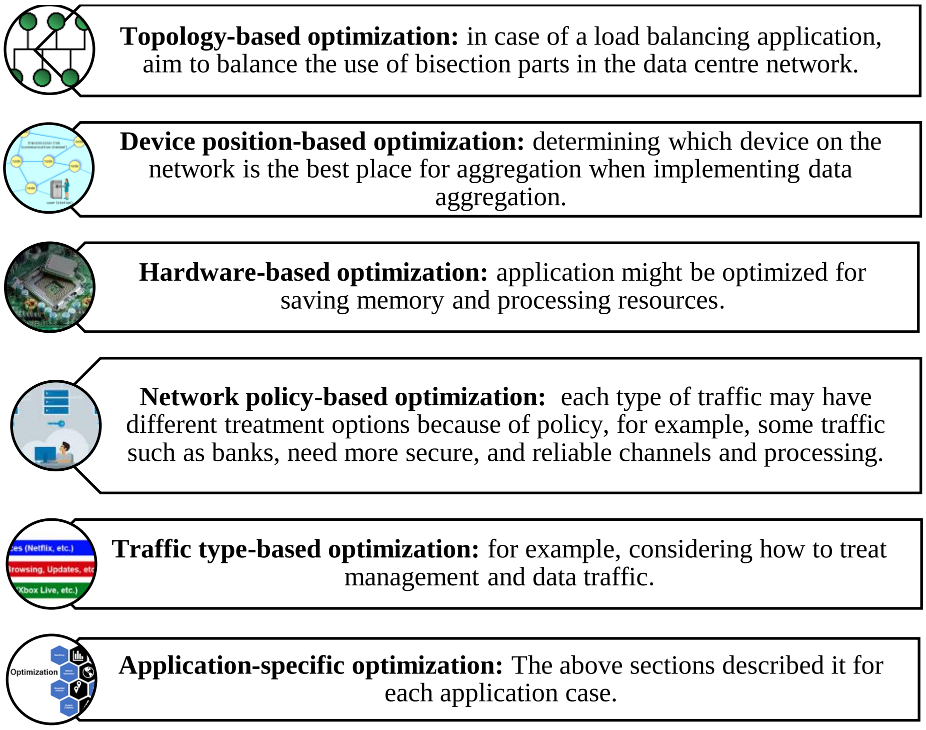
It was not easy to ensure the PCC because the switching ASIC doesn't have enough memory to store a large number of connection states. However, with the continuous increase in memory size, it is possible to implement the L4 load balancer on the switch. The main advantage of implementing it on switch is that there is no additional software load balancer in between application traffic and application server. This allows balancing load at line rate. SilkRoad [35] was proposed as a load balancer on a programmable switching ASIC and implemented using 400 lines of P4 code. The performance measurement on SilkRoad show that it can balance 10 million connections at line speed.

SHELL [42] tried to implement a stateless load balancer on P4-NetFPGA programmable switch, and it is easier to deploy on a network device than a stateful solution. Moreover, SHELL is application-agnostic and load-awareness.

The above descriptions provide examples of implementing L3 and L4 load balancers on a programmable switch. Network traffic is constantly changing, so load balancing mechanisms need to be congestion-aware, dynamic, and with low latency. The results of empirical analysis of these implementations seem reasonable. In the future, the researchers can do an analytical study in terms of optimality and scalability on these in order to look for opportunities improving dynamic nature.

3.4. General factors for optimizing the data plane applications

Each application use case has its own optimization factors, which are described above. In this section, we described the general factors of optimization. What application will be on the data plane, and what kind of application optimization is needed generally depend on the type of network, such as local area network (LAN), wide area network (WAN), data centre network, industrial network, time-sensitive network, etc. For example, load balancing, traffic management, and congestion control applications play an important role in a data centre, but may not apply to other types of networks. Subsequent optimization factors may include network topology, traffic types, and so on. These are explained in a little more detail with the specific example as in the followings:



Topology-based optimization: in case of a load balancing application, aim to balance the use of bisection parts in the data centre network.

Device position-based optimization: determining which device on the network is the best place for aggregation when implementing data aggregation.

Hardware-based optimization: application might be optimized for saving memory and processing resources.

Network policy-based optimization: each type of traffic may have different treatment options because of policy, for example, some traffic such as banks, need more secure, and reliable channels and processing.

Traffic type-based optimization: for example, considering how to treat management and data traffic.

Application-specific optimization: The above sections described it for each application case.

Figure 4. Optimization factors

4. Conclusions and Future Works

With the advent of data plane programming, applications such as network monitoring, traffic aggregation, caching, and load balancing are being redesigned on the data plane. A lot of applications are developed on P4. Some solutions implemented on a BMv2 P4 prototype switch are not a guarantee that the solution will work effectively on real equipment in the production network, but it is a good start to design and promote innovation. Some are implemented on hardware switches such as FPGA and Tofino but have not been fully tested in the actual network. So there is a lot of work to be done to improve and optimize these solutions. However, it should be noted that the P4 application working group has developed some cases that can be used in the production network.

Many of these solutions are currently for data centers and promise to significantly improve the data center network. Similarly, there are many opportunities to develop new application cases in the data plane for other types of networks in the future. For example, the main requirement of an industrial network is reliability, low and predictable latency. Data plane programmability helps to reduce latency in the industrial network, and some prototype implementations such as in-network sensor monitoring, data caching for industrial automation, and in-network robotic control applications are made in BMv2 switch [43][44]. In addition, some solutions are emerging as the prototype for time-sensitive networks, such as in-network time synchronization. Furthermore, these kinds of studies can be well developed and optimized in the real network.

Also, it is possible to combine these applications and develop effective solutions. Therefore, the researchers can determine which combination of applications is the most optimal. For example, according to the INT specification, network troubleshooting, advanced congestion control, advanced routing, and network data plane verification can be made based on INT monitoring.

Other topics to consider about optimization are that developer needs to think how network optimization is related (or irrelevant) to transport-level optimization [45]. Current in-network works are mainly focused on optimizing the network layer. However, transport protocols will affect the performance of any in-network solution. In addition, programmable switches do not support floating point calculations used in more complex operations, such as artificial intelligence (AI) and machine learning (ML) algorithms. For example, AI-enabled analysis can be used to understand network problems caused by managing the complexity, scale, and dynamics of modern networks [46].

In the future, the development and optimization of a data plane program should take into account the general factors, and specific factors identified for each case used in our paper. This paper not only gave a better understanding of the data plane applications but also offered specific ideas about what can be done in the future.

ACKNOWLEDGEMENTS

The research has been supported by the project "Application Domain Specific Highly Reliable IT Solutions" implemented with the support of the NRDI Fund of Hungary, financed under the Thematic Excellence Programme TKP2020-NKA-06 (National Challenges Sub programme) funding scheme.

REFERENCES

[1] “IEEE ICC 2018: Keynote by Nick McKeown, Programmable Forwarding Planes Are Here To Stay.” [Online]. Available: https://www.youtube.com/watch?v=8ie0FcsN07U&list=PLCn7fbhhPRgRGjLTn4ISi6v3hZu1AHT3T&index=14. [Accessed: 22-Apr-2021].

[2] R. Bifulco and G. Retvari, “A survey on the programmable data plane: Abstractions, architectures, and open problems,” IEEE Int. Conf. High Perform. Switch. Routing, HPSR, vol. 2018-June, 2018, doi: 10.1109/HPSR.2018.8850761.

[3] F. Hauser et al., “A Survey on Data Plane Programming with P4: Fundamentals, Advances, and Applied Research,” 2021.

[4] G. Altangerel, E. Tsogbaatar, and D. Yamkhin, “Performance analysis on IPv6 transition technologies and transition method,” in Proceedings - 2016 11th International Forum on Strategic Technology, IFOST 2016, 2017, doi: 10.1109/IFOST.2016.7884155.

[5] M. Budiu and C. Dodd, “The P416 programming language,” Oper. Syst. Rev., vol. 51, no. 1, pp. 5–14, 2017, doi: 10.1145/3139645.3139648.

[6] M. Moshref, A. Bhargava, A. Gupta, M. Yu, and R. Govindan, “Flow-level state transition as a new switch primitive for SDN,” ACM SIGCOMM Comput. Commun. Rev., vol. 44, no. 4, pp. 377–378, 2015, doi: 10.1145/2740070.2631439.

[7] A. Sivaraman, A. Cheung, M. Budiu, C. Kim, M. Alizadeh, and H. Balakrishnan, “Packet Transactions : High-Level Programming for Line-Rate Switches.”

[8] H. Song, “Protocol-oblivious forwarding: Unleash the power of SDN through a future-proof forwarding plane,” HotSDN 2013 - Proc. 2013 ACM SIGCOMM Work. Hot Top. Softw. Defin. Netw., pp. 127–132, 2013, doi: 10.1145/2491185.2491190.

[9] C. J. Anderson et al., “NetKAT: Semantic foundations for networks,” ACM SIGPLAN Not., vol. 49, no. 1, pp. 113–126, 2014, doi: 10.1145/2578855.2535862.

[10] P. Bosshart et al., “P4: Programming protocol-independent packet processors,” Comput. Commun. Rev., vol. 44, no. 3, pp. 87–95, 2014, doi: 10.1145/2656877.2656890.

[11] The P4 Language Consortium, “P4 16 Language Specification version 1.2.1,” pp. 1–163, 2020.

[12] G. Antichi1, T. Benson, N. Foster, F. M. V. Ramos, and and J. Sherry, “Programmable Network Data Planes (Dagstuhl Seminar 19141),” 2019, doi: 10.4230/DagRep.9.3.178.

[13] P. Voros, D. Horpacsi, R. Kitlei, D. Lesko, M. Tejfel, and S. Laki, “T4P4S: A target-independent compiler for protocol-independent packet processors,” IEEE Int. Conf. High Perform. Switch. Routing, HPSR, vol. 2018-June, no. August, 2018, doi: 10.1109/HPSR.2018.8850752.

[14] The P4.org Architecture Working Group, “P4 16 Portable Switch Architecture (PSA) The P4.org Architecture Working Group,” pp. 1–68, 2018.

[15] L. Tan et al., “In-band Network Telemetry: A Survey,” Comput. Networks, vol. 186, no. August 2020, 2021, doi: 10.1016/j.comnet.2020.107763.

[16] C. Systems, “In-band Network Telemetry (INT) Dataplane Specification,” pp. 1–55, 2020.

[17] J. Vestin, A. Kassler, D. Bhamare, K. J. Grinnemo, J. O. Andersson, and G. Pongracz, “Programmable event detection for in-band network telemetry,” arXiv. arXiv, 26-Sep-2019.

[18] R. Ben Basat, G. Antichi, and M. Mitzenmacher, “PINT : Probabilistic In-band Network Telemetry,” vol. 4.

[19] Z. Xia et al., “KeySight : A Scalable Troubleshooting Platform Based on Network Telemetry,” pp. 2–3, 2018.

[20] Y. Tang, Y. Wu, G. Cheng, and Z. Xu, “Intelligence enabled SDN fault localization via programmable in-band network telemetry,” IEEE Int. Conf. High Perform. Switch. Routing, HPSR, vol. 2019-May, pp. 1–6, 2019, doi: 10.1109/HPSR.2019.8808121.

[21] J. Hyun and J. W. K. Hong, “Knowledge-defined networking using in-band network telemetry,” 19th Asia-Pacific Netw. Oper. Manag. Symp. Manag. a World Things, APNOMS 2017, pp. 54–57, 2017, doi: 10.1109/APNOMS.2017.8094178.

[22] “Barefoot Deep Insight.” [Online]. Available: https://www.intel.com/content/www/us/en/products/network-io/programmable-ethernet-switch.html. [Accessed: 22-Apr-2021].

[23] Y. Li et al., “HPCC: High precision congestion control,” SIGCOMM 2019 - Proc. 2019 Conf. ACM Spec. Interes. Gr. Data Commun., pp. 44–58, 2019, doi: 10.1145/3341302.3342085.

[24] V. Jeyakumar, M. Alizadeh, C. Kim, and D. Mazières, “Tiny packet programs for low-latency network control and monitoring,” Proc. 12th ACM Work. Hot Top. Networks, HotNets 2013, 2013, doi: 10.1145/2535771.2535780.

[25] B. Turkovic, F. Kuipers, N. Van Adrichem, and K. Langendoen, “Fast network congestion detection and avoidance using P4,” NEAT 2018 - Proc. 2018 Work. Netw. Emerg. Appl. Technol. Part SIGCOMM 2018, pp. 45–51, 2018, doi: 10.1145/3229574.3229581.

[26] E. Tsogbaatar et al., “DeL-IoT: A deep ensemble learning approach to uncover anomalies in IoT,” Internet of Things, vol. 14, no. March, p. 100391, 2021, doi: 10.1016/j.iot.2021.100391.

[27] A. Sapio, I. Abdelaziz, A. Aldilaijan, M. Canini, and P. Kalnis, “In-Network Computation is a Dumb Idea Whose Time Has Come.”

[28] P. Costa et al., “NetAgg : Using Middleboxes for Application-specific On-path Aggregation in Data Centres,” pp. 249–261.

[29] S. Wang, C. Wu, Y. Lin, and C. Huang, “High-Speed Data-Plane Packet Aggregation and Disaggregation by P4 Switches,” vol. 4, 2019.

[30] P. G. Kannan, R. Joshi, and M. C. Chan, “Precise Time-synchronization in the Data-Plane using Programmable Switching ASICs,” no. 2, pp. 8–20.

[31] R. Nishtala et al., “Scaling memcache at facebook,” Proc. 10th USENIX Symp. Networked Syst. Des. Implementation, NSDI 2013, pp. 385–398, 2013.

[32] B. Atikoglu, Y. Xu, E. Frachtenberg, S. Jiang, and M. Paleczny, “Workload analysis of a large-scale key-value store,” Perform. Eval. Rev., vol. 40, no. 1 SPEC. ISS., pp. 53–64, 2012, doi: 10.1145/2254756.2254766.

[33] “Key–value database.” [Online]. Available: https://en.wikipedia.org/wiki/Key–value_database. [Accessed: 22-Apr-2021].

[34] X. Jin et al., “NetCache : Balancing Key-Value Stores with Fast In-Network Caching,” no. Figure 1, pp. 121–136.

[35] J. Nelson and L. Ceze, “IncBricks : Toward In-Network Computation with an In-Network Cache.”

[36] H. Tu, D. Daniele, M. Canini, F. Pedone, and R. Soul, “NetPaxos : Consensus at Network Speed.”

[37] A. Aghdai, C. Y. Chu, Y. Xu, D. H. Dai, J. Xu, and H. Jonathan Chao, “Spotlight: Scalable transport layer load balancing for data center networks,” arXiv, 2018.

[38] J. L. Ye, C. Chen, and Y. Huang Chu, “A Weighted ECMP Load Balancing Scheme for Data Centers Using P4 Switches,” Proc. 2018 IEEE 7th Int. Conf. Cloud Networking, CloudNet 2018, pp. 1–4, 2018, doi: 10.1109/CloudNet.2018.8549549.

[39] M. Alizadeh et al., “CONGA : Distributed Congestion-Aware Load Balancing for Datacenters,” pp. 503–514.

[40] N. Katta, M. Hira, C. Kim, A. Sivaraman, and J. Rexford, “HULA : Scalable Load Balancing Using Programmable Data Planes,” 2016.

[41] T. Barbette et al., “A high-speed load-balancer design with guaranteed per-connection-consistency,” Proc. 17th USENIX Symp. Networked Syst. Des. Implementation, NSDI 2020, pp. 667–683, 2020.

[42] B. Pit-Claudel, Y. Desmouceaux, P. Pfister, M. Townsley, and T. Clausen, “Stateless Load-Aware Load Balancing in P4,” Proc. - Int. Conf. Netw. Protoc. ICNP, vol. 2018-Septe, pp. 418–423, 2018, doi: 10.1109/ICNP.2018.00058.

[43] F. E. R. Cesen, L. Csikor, C. Recalde, C. E. Rothenberg, and G. Pongracz, “Towards low latency industrial robot control in programmable data planes,” Proc. 2020 IEEE Conf. Netw. Softwarization Bridg. Gap Between AI Netw. Softwarization, NetSoft 2020, pp. 165–169, 2020, doi: 10.1109/NetSoft48620.2020.9165531.

[44] J. Vestin, A. Kassler, and J. Akerberg, “FastReact: In-Network Control and Caching for Industrial Control Networks using Programmable Data Planes,” IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA, vol. 2018-Septe, pp. 219–226, 2018, doi: 10.1109/ETFA.2018.8502456.

[45] “In-Network Data-Center Computing.” [Online]. Available: https://tools.ietf.org/id/draft-he-coin-datacenter-00.html. [Accessed: 22-Apr-2021].

[46] “Networking Technology Trends Report - Download - Cisco.” [Online]. Available: https://www.cisco.com/c/en/us/solutions/enterprise-networks/networking-technology-trends.html#~trends. [Accessed: 22-Apr-2021].

AUTHORS

Tejfel Máté received his B.Sc., M.Sc. and Ph.D. Degree in Computer Science, from ELTE, Budapest Hungary. He is currently working as an Associate Professor in the Department of Programming Languages and Compilers, ELTE. His research interest includes programming languages, correctness check, Software Defined Networks, and network optimization. For more information, visit his database at 0000-0001-8982-1398 orchid-id.

Altangerel Gereltsetseg received her B.Sc. & M.Sc. Degree in Information technology, from Mongolian University of Science and Technology (MUST). She is currently a Ph.D. student in the Department of Programming languages and compilers, ELTE under the supervision of Professor Tejfel Máté. Her research interest includes Software-defined Networks, deeply programmable network, and network optimization. For more information, visit her database at 0000-0002-1594-8158 orchid-id.

GPF: A GREEN POWER FORWARDING TECHNIQUE FOR ENERGY-EFFICIENT NETWORK OPERATIONS

Rahil Gandotra[1] and Levi Perigo[2]

[1]Interdisciplinary Telecom Program, University of Colorado Boulder, USA
[2]Department of Computer Science, University of Colorado Boulder, USA

ABSTRACT

The energy consumption of network infrastructures is increasing; therefore, research efforts designed to diminish this growing carbon footprint are necessary. Building on prior work, which determined a difference in the energy consumption of network hardware based on their forwarding configurations and developed a real-time network energy monitoring tool, this research proposes a novel technique to incorporate individual device energy efficiency into network routing decisions. A new routing metric and algorithm are presented to select the lowest-power, least-congested paths between destinations, known as Green Power Forwarding (GPF). In addition, a network dial is developed to enhance GPF by allowing network administrators to tune the network to optimally operate between energy savings and network performance. To ensure the scope of this research for industry adoption, implementation details for different generations of networking infrastructure (past, present, and future) are also discussed. The experiment results indicate that significant energy and, in turn, cost savings can be achieved by employing the proposed GPF technique without a reduction in network performance. The future directions for this research include developing dynamically-tuning network dial modes and extending the principles to inter-domain routing.

KEYWORDS

Energy efficiency, intent-based networking, network optimization dial, SDN, programmable control plane, OpenFlow, programmable data plane, P4.

1. INTRODUCTION

The energy consumption of the information and communications technology (ICT) infrastructure has been rising steadily - from 1815 TWh or 8% of the global energy use in 2012 to 2,547 TWh or 10% of the global energy use in 2017 [1]. These high energy consumptions translate into large carbon footprints since coal, the most carbon-intensive approach to producing energy, still remains the single largest source of power generation globally (36.4%) [2]. Researchers have predicted the global energy demand of the ICT sector by 2030 to be 8,265 TWh or 21% of the global energy use, with the worst-case prediction being 30,715 TWh or 51% of global energy use [3]. The network infrastructure alone is predicted to consume 2,641 TWh by 2030 making it a considerable portion of the total ICT energy consumption (32%). Therefore, research efforts to attempt to reduce it are warranted.

While there have been significant improvements in the energy efficiency of data centers and networks, they have mostly been overwhelmed by the growth of data traffic caused by increased

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 83-97, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111808

consumption of services [4, 5]. For instance, Van Heddeghem et al. estimated that the growth of electricity consumed by networks (at 10.4% per year) was much faster than the global electricity demand itself (at 3% per year) [6]. While compute equipment is designed to be reasonably energy proportional, i.e. their energy consumption is proportional to their load, networking equipment is not energy proportional because it is a shared resource and is expected to always be available [7]. The numerous efforts put into compute energy management and cooling technologies have resulted in networks becoming a bigger fraction of the total ICT power budget [8].

This paper builds on prior research conducted on comparing the energy efficiencies of networking hardware [9] and developing a real-time network power monitoring framework [10], with the objective to incorporate network energy efficiency into the network-wide forwarding decisions, and to study the tradeoffs between energy savings and network performance. [9] presented a detailed energy efficiency study of a variety of network devices and analyzed the power consumption when traffic flows through individual devices as a function of their configuration. The test results utilized an in-line power meter and indicated that significant differences in energy efficiency of devices were prevalent based on their forwarding configurations, such as the number or type of the forwarding entries, or whether the forwarding entries were stored in the hardware or the software table. In order to enable the non-intrusive collection of power information from network devices, [10] presented a framework employing information models - both standardized IETF and non-standard customized models – to demonstrate the feasibility of real-time data gathering for effective network energy management.

The motivation behind the research in this study is to build on previous work by developing energy-efficient forwarding mechanisms that incorporate individual device efficiencies into network-wide forwarding decisions with minimal impact to network performance. The results from this work could be used to enhance the body of knowledge in the vital field of network energy management. The rest of the paper is organized as follows: Section 2 discusses related work and highlights the novelty of this work; Section 3 describes the research methodology – both the algorithm and its different mechanisms; Section 4 discusses the research results; Section 5 present the conclusions.

2. RELATED WORK

The existing work on incorporating energy awareness into networks can be categorized as five distinct techniques: sleeping of network elements, link rate adaptation, proxying, store and forward, and network traffic aggregation.

Sleeping of network elements was based on the intuition that because computer networks are provisioned for peak load and are generally underutilized, parts of the network or individual devices can be put to sleep during off-peak hours [11]. The work involved in this approach was two-fold: predicting low-utilization periods suitable for energy savings, and reconfiguring the network to put network elements to sleep [10]. The target network element to be put to sleep could either be the physical device(s) [13-15], or individual components of one device [16-19]. Since networking devices typically provide limited support for putting the system or components to sleep, the techniques in this category face the challenges of waking up a sleeping element and understanding the impact of sleeping on protocols which require network presence.

Link rate adaptation was identified as a potential approach after research showed that high link rates (1/10 Gbps) consumed more energy than low link rates (10/100 Mbps) [20]. The work in this approach involved identifying periods of low traffic wherein the links could be operated at slower rates, and reconfiguring the device link speeds network wide [21]. Both reactive approaches, based on matching the link speeds with the current utilization [22, 23], and proactive approaches, predicting future traffic patterns from historical utilization data [17, 24], were proposed. However, since several routing and switching protocols operate based on link speed metrics, the interoperability of such protocols with the link rate adaptation techniques remains to be examined.

Proxying is based on the hypothesis that periodic network chatter traffic does not allow end hosts to sleep effectively and that network elements could be employed as proxies to enable infrastructure-level savings [20]. The techniques proposed in this category attempted to counter the requirements necessitating networked end-hosts to be connected and responsive even while sleeping, i.e. always-on should not mean always powered on [25]. Both on-system [21, 26] and off-system [27, 28] proxies were proposed to offload some of the end-host responsibilities. The impact of proxying on session-based protocols and the additional energy consumption of a proxy as compared to the amount of energy saved are some of the open problems in this category.

The observation that many web applications exhibit bursty traffic with high inter-arrival traffic times led to the development of store and forward techniques [29]. While the traditional practice in network engineering is to avoid bursts using congestion avoidance and QoS queues, the insight behind the techniques in this area was to shape traffic in bursts in order to allow network elements to be put in low power/sleep mode in between these bursts [17]. The techniques in this category included in-network bunch-and-burst [30, 31] and a parallel network implementing time-based scheduling in conjunction with the Internet [32, 33]. However, practical implementations have demonstrated these techniques to be suitable only for specific network topologies due to the cost of an added end-to-end delay and the problem of synchronizing bursts across the network to enable efficient sleeping.

Network traffic aggregation techniques aimed to proactively find the minimum-power subsets of the infrastructure that can provide the desired level of performance while enabling energy savings [23]. The work in this category involved collecting traffic statistics such as the traffic matrix and the desired fault-tolerance levels, optimizing the problem while satisfying all constraints, and reconfiguring the devices to put them in low-power/sleep modes [34, 35]. The major open problems in this category include finding methods to reduce the computing time that allows for on-the-fly solutions and the general lack of support for putting network hardware to sleep or wake them back up.

While only a few of the proposed techniques have seen industry adoption, for example link mode adaptation was standardized as IEEE 802.3az, a majority of them suffer from critical flaws. Firstly, most networking hardware deployed has limited support for putting a device or its individual components into sleep or low-power modes. Since network devices are expected to be always on and connected, the vendors typically provide minimal energy management capabilities. Secondly, waking up a sleeping element is a non-trivial problem. Unless the device is connected to a ‘smart’ networked power strip, a device which is powered off becomes laborious to connect back. Thirdly, the impact of frequent device/component-level changes on cost-based switching and routing protocols would need to be minimized. Frequent and incorrect changes could in fact lead to an increase in the energy consumption because powering on a sleeping element or transitioning to another mode consumes additional energy and time.

This paper presents a sixth distinct and novel technique for saving energy in networks – energy-efficient forwarding - by utilizing individual device energy efficiencies for the network forwarding decisions. In other words, the aim of this technique was to deliver traffic between endpoints in a network while attempting to consume the minimum amount of energy as possible. The proposed technique does not make any assumptions about the underlying hardware support for sleep/low-power modes, and instead employs the differences in energy efficiency of different devices while processing traffic. An approach to provide tradeoffs between energy savings and network performance is also presented. The proposed technique includes both an algorithm for making forwarding decisions as well as a mechanism for implementing it over the current and the next generation of networking paradigms. Since the technique aims to be transparent to applications and to reduce operational energy consumption, it could be used in conjunction with any of the five abovementioned techniques to allow for further energy savings.

3. RESEARCH METHODOLOGY

The metricidentified in previous work to compare energy efficiency between devices was Energy Consumption Rating (ECR), which is the aggregated energy consumption normalized to the actual throughput. A lower metric implied the device consumed less energy while processing the same amount of traffic. While this metric is helpful in determining which device processes traffic more energy efficiently, it is not useful for making forwarding decisions and suffers from two major problems: the congestion problem, and the sleeping giant problem.

The congestion problem is presented as: consider two similar network devices A and B. If at any instance, device A is experiencing much higher amount of traffic as compared to device B, its ECR value would be much lower than device B’s. However, a low value of ECR does not imply that additional traffic could be forwarded through device A, as it is already heavily congested.

The sleeping giant problem is presented as: Consider two dissimilar network devices A and B. At any instance, both devices are experiencing similar amounts of bandwidth, however, since device B is a bigger, more feature-rich box, it is consuming higher amounts of power which makes its ECR value higher than device A’s. Therefore, even though device B has the potential to support higher amounts of bandwidth, it will not be preferred at this instance because of its current lower ECR value.

So, a more useful metric instead for making forwarding decisions would be one that helps in determining the lowest-power and least-congested paths. In other words, devices that consume less power and have high available bandwidth (Avail_Bandwidth) should be preferred over those consuming high power and having low available bandwidth. The metric Energy Efficiency Potential (EEP) in introduced and used in this research, which is the aggregated energy consumption normalized to the available bandwidth of the device.

The technique presented in this paper, called Green Power Forwarding (GPF), consists of both a GPF algorithm for making decisions about finding the most energy-efficient paths in the network, and mechanisms for implementing the GPF algorithm over different varieties of network infrastructures such as traditional, programmable-control, and programmable-data networks. The GPF algorithm is based on Dijkstra's Shortest Path First (SPF) algorithm but instead of using link speed as the routing metric, Watt/Avail_Mbps is used to compute least-power, least-congested paths.

Algorithm 1. GPF algorithm

```
Given→
nodes: vertices in the graph; denoted by u, v
ee: energy efficiency values of each node; denoted by $ee_N$
weights: weighted edges connecting two nodes; denoted by w(u,v)
Initialize→
s: source node
dist: an array such that dist(s) = 0, and for all other nodes v, dist (v) = $\infty$
Q: a queue of all nodes in the graph
S: an empty set
Run→
for each node s in nodes:
while Q is not empty:
   pop node v from Q that is not in S with the smallest dist(v)
   add node v to S
   update dist such that for all adjacent nodes u of v (from w(u,v)):
if dist(u) >dist(v) + ee(u):
dist(u) = dist(v) + ee(u)
      else:
         no change
```

To enhance GPF and provide tradeoffs between energy savings and network performance, a network dial is designed with the capability to switch between different modes. The goal behind this network dial is to translate the intent of the network administrator to operate the network in a certain energy mode into a distributed systems problem and solving it by fetching the required energy and performance parameters, computing routing metrics as per the dial mode set, running the GPF algorithm, and reconfiguring the network accordingly. The energy savings are determined by one parameter – the current power consumption of the device in Watts, and network performance is determined by three parameters – the available bandwidth in Mbps, the total delay in ms, and the total packet loss in %. Therefore, the network dial should allow tradeoffs between these four parameters – power, bandwidth, delay, and loss. The final weighted formula that combines these four parameters for the purposes of comparing routes is defined in Eq. (1). The final routing metric (ee) uses four constants – c1 through c4, to act as multipliers in the final routing calculation, and a lower metric would be preferred when making forwarding decisions.

$$ee = c1.Power + \frac{c2.10^4}{Bandwidth} + c3.Delay + c4.Loss \qquad (1)$$

Here, the energy savings metric is defined as the first term (c1.Power), while the network performance metric is defined as the remaining part of the equation (the addition of the second, third, and fourth terms). The four constants – c1, c2, c3 and c4 – would need to be manipulated over a range such as 0 to 1 in order to implement the different network dial modes.Dividing 104 by the available bandwidth helps differentiate between higher amounts of bandwidth. Otherwise, all available bandwidth values above 1 Mbps would yield a value between 0 and 1, thereby not impacting the final routing metric significantly. The 104 term helps in differentiating between bandwidths smaller than 10000 Mbps (which would yield a value greater than 1) and greater than 10000 Mbps (which would yield a value smaller than 1).Next, to understand the behaviour of how these constants impact the final routing metric ee, statistical analysis was performed to determine for a given pair of values of power, bandwidth, and delay for two devices A and B, which device would have a lower metric when c1, c2, c3 vary from 0 to 1. The loss value and the c4 constant was ignored in this analysis in order to allow visualizing plots in three dimensions.

Also, network performance is considered adequately by including the bandwidth and delay parameters. A sample plot is shown in Fig. 1, wherein black points indicate that the routing metric of device A was lower than B, the yellow points indicate the opposite, and red points indicate that both devices had the same ee value for those particular values of c1, c2 and c3.

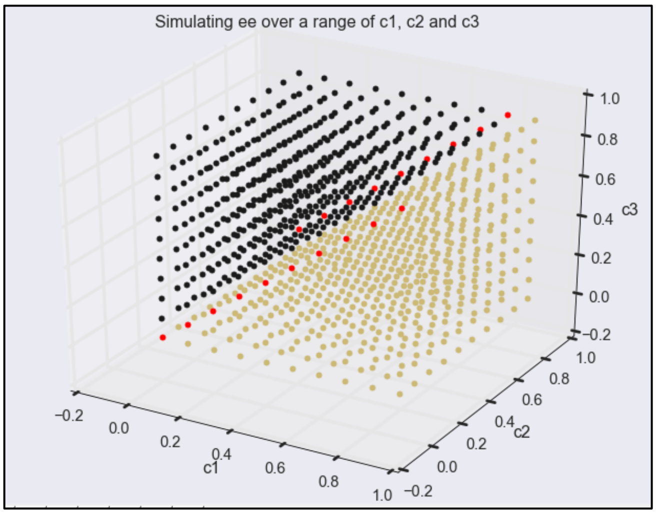


Figure 1. Scatter plot depicting ee values over a range of c1, c2, c3

Simulating over a wide range of power, bandwidth and delay values helped identify a trend of how the constants' values need to be modified in order to get the desired result. Furthermore, the different network dial modes defined are shown in Table I. Since in many instances energy savings come at the cost of network performance, three modes are defined to cover the entire range of operations for the network administrator to select. Although just three dial modes have been defined in this work, the same principles could be used to develop as many modes as required. Implementing these tradeoffs allows for utilizing the opportunity costs associated with energy savings at the expense of network performance, and vice versa.

Table 1. Network dial modes

Dial mode	**Description**
1	Forwarding decisions are based on energy efficiency only; network performance parameters are not considered.
2 (default)	Forwarding decisions are based on both energy efficiency and network performance parameters.
3	Forwarding decisions are based on network performance parameters only; energy efficiency is not considered.

The constants' values chosen for each mode are depicted in Fig. 2. For dial mode 1, c1 is set to 1, and c2, c3, and c4 are set to 0 to ensure that the ee value is calculated by the power consumption term only, and the three network performance parameters (bandwidth, delay, and loss) are not considered. For dial mode 3, c1 is set to 0, and c2, c3, and c4 are set to 1, so that ee is calculated based on network performance only and energy efficiency is not considered. For dial mode 2, c1 is set to 0.75, and c2, c3, and c4 are each set to 0.25. The rationale behind these values is that since network performance is calculated based on three parameters and energy savings is calculated by just one parameter, the weightage for the four constants should be proportional. Therefore, c1 is set to three times the values of c2, c3 and c4. So, between dial modes 1 and 3, c1 decreases from 1 to 0, and c2, c3, and c4 increase from 0 to 1. These values depicted are one instance of the dial implementation and any values can be set to the constants to create intermediary modes according to specific requirements.

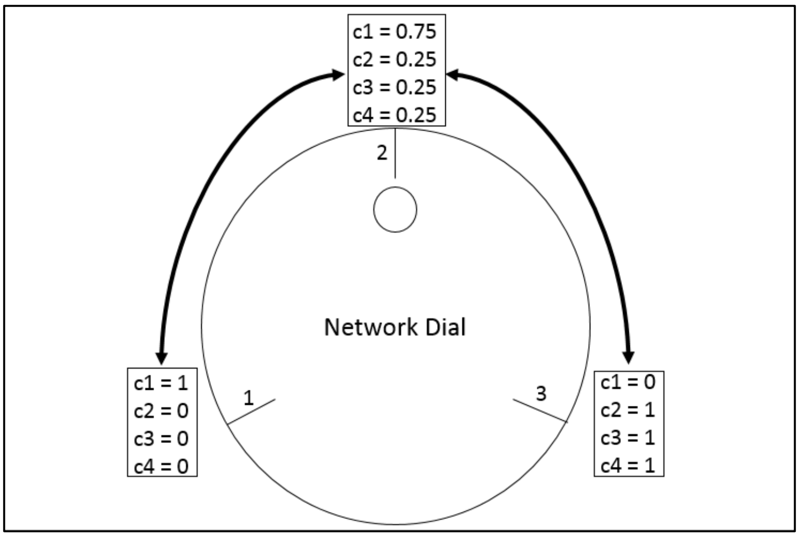


Figure 2. Constants' values in different dial modes for ee metric calculation

The mechanisms to implement GPF provide the implementation details on the technologies employed when working with the different generations of networks. The mechanism details for three types of networking paradigms are provided – traditional, programmable control, and programmable data. These three types represent how networks have evolved over the past decades in an effort to introduce programmability by disaggregating the control and data planes, logically centralizing the control plane, and developing programmable data pipelines.Table 2 provides an overview of the implementation details for each generation.

Table 2. Implementation overview of each networking generation

Networking generation	**Simulated devices**	**Forwarding mechanism**	**Control mechanism**
Traditional	Cisco routers	Static/dynamic routing (OSPF/EIGRP)	SSH
Programmable-control	Open vSwitch	OpenFlow match-action pipelines	OpenFlow
Programmable-data	Stratum Bmv2	P4 match-action pipelines	P4Runtime

The network topology used for the experiments is shown in Fig. 3. It consists of a two-tier Clos architecture with every leaf switch connected to each of the spine switches, and the two leaf switches connected to one host each. Even though the devices used are switches, they are operated as routers with every interface being on a different network in order to allow manipulation of IP traffic. The five spine switches represent the devices used in previous research so that actual power consumption values could be used for the tests [9]. Every switch is connected to an out-of-band network controller which is responsible for making the forwarding decisions for the traffic flowing between the two hosts. The control connections shown are different for the different types of networks tested.

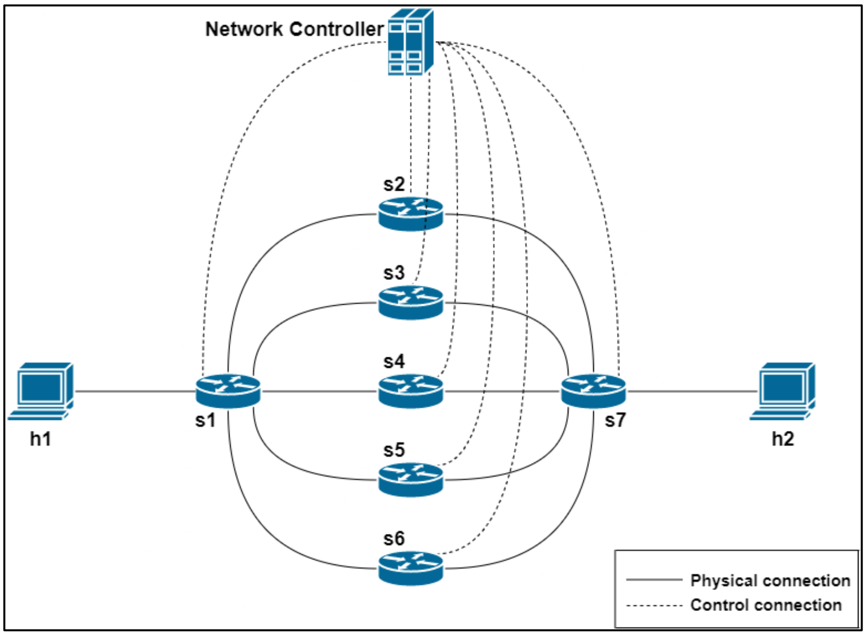


Figure 3. Network topology used in the experiments

The high-level operation of GPF is shown in Fig. 4. The network monitoring tool developed in Chapter III gathers the power, throughput, delay, and packet loss values from the network infrastructure and exposes this information as REST endpoints. The network controller fetches this information using REST endpoints, and also accepts an input on the network dial mode from the user. Next, it runs the GPF algorithm and computes the next hop information for each destination network for each source node. Once the algorithm run is complete, the outputs have to be translated by the control connection module in the specific control connection format – SSH/OpenFlow/P4Runtime. This process is run continuously so that real- time network changes can be transposed onto the network operating behavior. Since the network controller is running on a separate high-CPU server, its processing requirements do not impact the network devices and their ability to process and forward traffic.

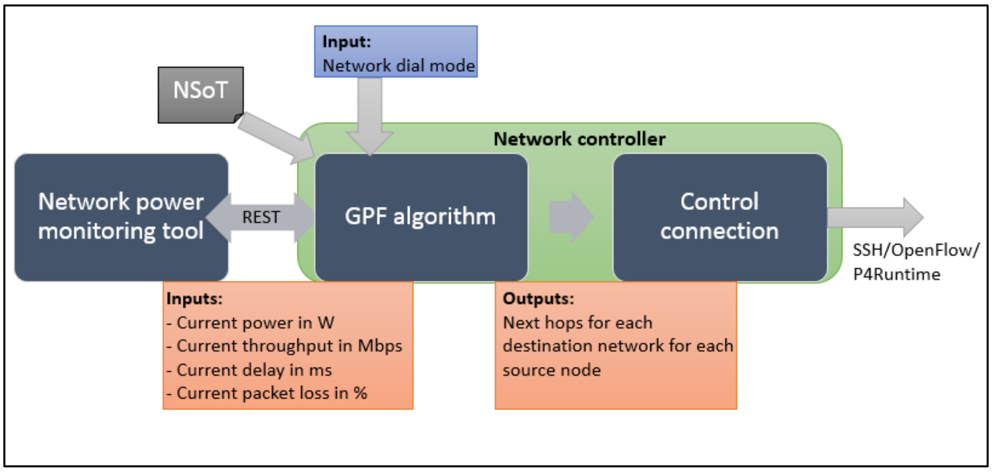


Figure 4. High-level overview of GPF operation

3.1. Traditional Network

The traditional network consists of devices with coupled control and data planes, making control decisions locally in a distributed manner, operating protocols supported by the fixed-function data pipelines. The network devices used in the experiment were virtual Cisco 7200 routers simulated in GNS3 [37]. The forwarding mechanisms tested included both static and dynamic routing scenarios; OSPF and EIGRP were employed for dynamic routing testing.

The network controller first reads a NSoT file to discover the devices' information which include their management IP and port. This information is used to send REST calls to the network power monitoring tool to obtain the current power, throughput, delay and loss values. Also, the controller accepts the network dial mode input from the user. Next, the routing metric is computed based on the network dial mode input, and the GPF algorithm is run and its output is a routing table constructed for all source and destination nodes with next hops indicating the path selected based on the dial mode.

SSH was used as the control protocol to configure the appropriate routes onto the devices. The control connection module used a Python library called Netmiko to initiate SSH connections and send routing commands onto the devices [38]. In the case of static routing scenario, the network controller configures static routes on all devices based on the next hops information computed by the GPF algorithm. In the case of dynamic routing with OSPF, the network controller overrides the local routing information computed by each device running Dijkstra's SPF algorithm independently by configuring static routes which have lower administrative distances. In the case of dynamic routing with EIGRP which supports setting a preferred route by influencing its metrics, the network controller uses offset lists to prefer one path over the other.

This process is run continuously with the network controller fetching the current power and network performance parameters every five seconds and reconfiguring the network if required based on the output of the GPF algorithm. A change in the network dial mode also triggers a rerun of this process in order to reconfigure the network based on the new dial mode.

3.2. Programmable-control network

Programmable-control networks consist of devices with decoupled control and data planes, running a logically centralized controller, and operating protocols supported by their fixed-function data pipelines. The network devices used in the experiment were Open vSwitch (OvS) devices simulated in GNS3 [39]. The forwarding mechanism tested was the OpenFlow match-action pipelines.

The network controller first reads a NSoT file to discover the devices' information which include their OpenFlow datapath ID (DPID), management IP and port. This information is used to send REST calls to the network power monitoring tool to fetch the current power, throughput, delay and loss values. Also, the controller accepts the network dial mode input from the user. Next, the routing metric is computed based on the network dial mode input, and the GPF algorithm is run and its output is a routing table constructed for all source and destination nodes with next hops indicating the path selected based on the dial mode.

OpenFlow was used as the control protocol to configure the appropriate flow entries onto the devices. The control connection module ran an instance of the RYU SDN controller to enable northbound and southbound interfaces [36]. rest_router application was run on Ryu which exposes a convenient set of API's for setting addresses and routes on the OvS devices. The network controller uses REST API calls to push the appropriate flow entries onto the devices using OpenFlow Flow_Mod messages matching on the destination network with the action to forward out to its next hop computed from the GPF algorithm.

This process is run continuously on the controller with the network controller fetching the real-time power and network performance parameters every five seconds and reconfiguring the network if required based on the output of the GPF algorithm. A change in the network dial mode also triggers a rerun of this process on the network controller in order to reconfigure the network based on the new dial mode.

3.3. Programmable-data network

Until now, the real-time power and network performance parameters were fetched by the network controller from the network power monitoring tool. This information was collected and used to make forwarding decisions in a centralized manner. With the advent of programmable packet parsers that support custom headers or protocols, we tested the exchange of the power and network performance parameters in a distributed manner to test if the devices could make energy-efficient forwarding decisions independently.

The programmable-data network enables the devices to expose programmable parsers along with the match-action tables which allow for the design and development of customized protocols. The network devices used in the experiment were Stratum BMv2 switches [40] simulated using the network emulation software Mininet [41]. Stratum is an open-source next-generation switch operating system that exposes a set of SDN interfaces including P4Runtime [42]. BMv2 is a reference P4 software switch that implements the packet- processing behavior specified by a P4 program [43]. The forwarding mechanism tested was the P4 match-action tables. Two types of approaches were developed – in-band and out-of-band.

The in-band approach involved developing a custom P4 data-plane program with a custom header – Real-time Network Power Protocol (RNPP) – added in between the Ethernet and IPv4 headers in order to transmit the required information between network devices in band. The RNPP header is developed as below –

```
header rnpp {
   bit<16>proto_id;
   bit<16>dst_id;
   bit<16>dpid;
   bit<10>power;
   bit<20>bw;
   bit<10>delay;
   bit<8>loss;
}
```

Forwarding was still based on the IPv4_LPM table while the power and network performance parameters exchanged were stored in a custom table – RNPP_table. The network controller employed P4Runtime in order to – *(i)* fetch the current RNPP values from each device's RNPP_table, and *(ii)* run the GPF algorithm based on the collected values and populate the IPv4_LPM table in order to implement energy-efficient forwarding.

The out-of-band approach involved using P4 Packet_Out and Packet_In messages to allow the exchange of the power and network performance parameters out-of-band. A custom P4 data-plane code was developed that supported adding customized header fields to the Packet_Out and Packet_In messages. The flow of operation is as follows: the network controller, using P4Runtime, first generates and sends Packet_Out messages with the customized header bits initialized but containing null values from each switch. The switches are configured to broadcast these messages out each port after populating the header fields with their respective power and network performance parameters and stripping the Packet_Out specific bits. Next, switches receiving these custom packets are configured using P4Runtime to forward these packets to the network controller via Packet_In messages after adding the Packet_In specific bits. This allows for the exchange and computation of the energy-efficient information in a distributed manner while ensuring this traffic remains out-of-band from actual data traffic.

4. RESULTS AND ANALYSIS

The results from the working of the GPF algorithm and the network dial are presented and analyzed in this section. The experiments were based on power consumption data collected from hardware in the prior work [9]. The first set of tests involved comparing a cost-based SPF with GPF (see Fig. 5). The y-axis of the graph indicates the end-to-end Watt/Mbps value of the selected path by each algorithm. As noted from the figure, since all link costs were set to 1 Gbps in the experiments, the cost-based SPF, by design, randomly selected one of the five paths, while GPF always selected the lowest-power, least-congested path.Watt/Mbps helps in normalizing the energy efficiency of different paths and indicates the energy cost per Mbps of traffic.

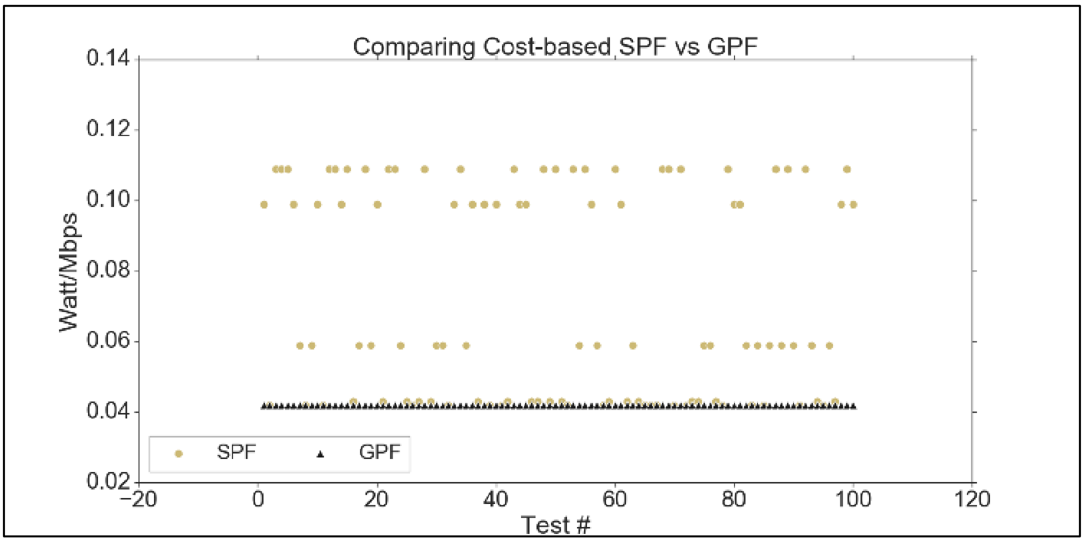


Figure 5. Comparing Watt/Mbps - cost-based SPF vs GPF

Table 3 shows the basic statistical differences between the two algorithms. Both mean and standard deviation values for GPF are lower than that of SPF implying that not only did GPF select low-power paths, but the amount of variation in the power consumption of selected paths was also smaller than that of SPF.

Table 3. Statistical comparison – Watt/Mbps – SPF vs GPF

Statistic descriptor	SPF – Watt/Mbps	GPF – Watt/Mbps
Num. of tests	100	100
Mean	0.0718	0.0398
Median	0.059	0.040
Standard deviation	0.0277	0.0007
Min.	0.039	0.039
Max.	0.109	0.041

While the differences in the Watt/Mbps values might seem insignificant, when scaled up to the size of production networks handling terabits of traffic, the differences become substantial. Fig. 6 compares the operational energy expenditures in US dollars for 500 devices running cost-based SPF and GPF over different durations. The final dollar values are calculated based on an average electricity cost of $0.137 per kWh in the USA, and do not include other operational costs such as cooling [44].In addition, 1 Watt-hour of network equipment energy savings result in an additional 1.59 Watt-hour of facility-level energy savings [45].

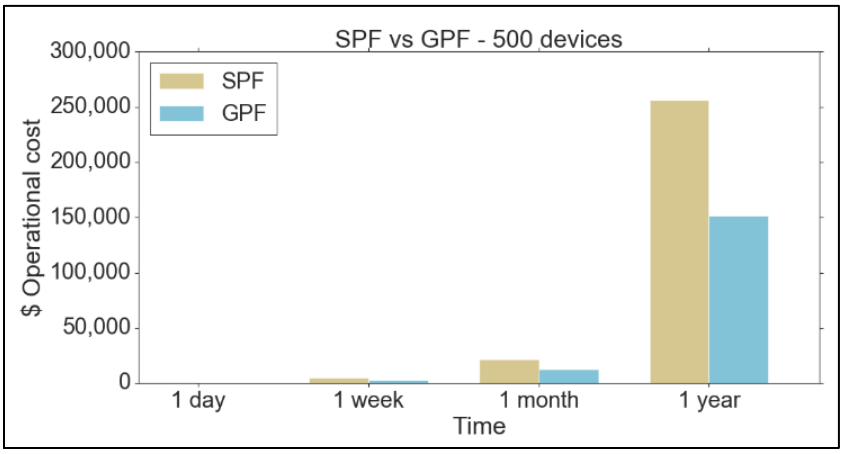


Figure 6. Comparing $ operational cost – SPF vs GPF

To test the network dial implementation, multiple tests were conducted in each of the three modes to check the energy savings and the network performance metrics for the selected path (see Fig. 7). For clarity, a low value for both the energy savings and network performance metrics implies more energy savings and better network performance. For mode 1, low values of the energy savings metric and high values of the network performance metric are noted. For mode 3, high values of the energy savings metric and low values of the network performance metric are noted. For mode 2, intermediate values of the energy savings and the network performance metrics are noted. A broader impact of this dial is that it promotes national security by enabling the government and defense networks to work for longer durations in the event of a disaster, by allowing operations in the high energy savings, low network performance mode.

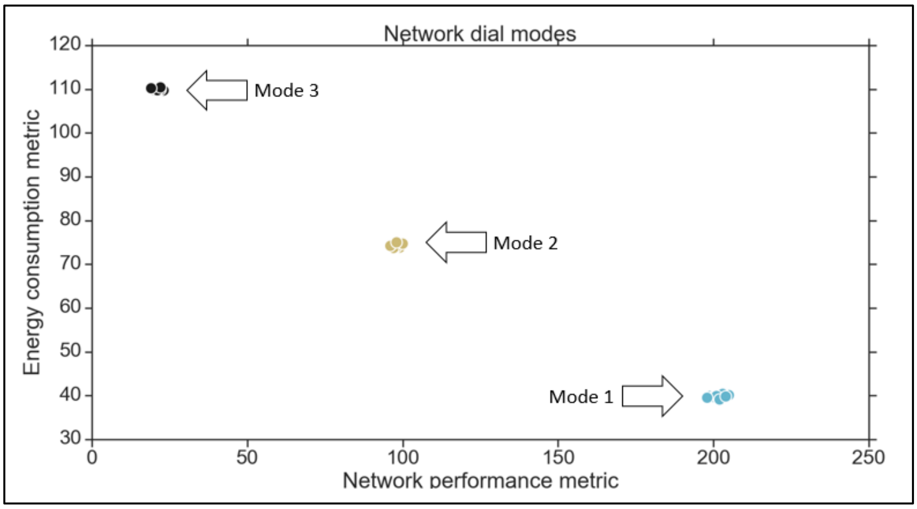


Figure 7. Energy consumption and network performance metrics in different network dial modes

5. CONCLUSIONS

A novel technique to incorporate energy efficiency into the network forwarding decisions and to provide tradeoffs between energy savings and network performance was presented in this paper. The GPF algorithm developed from this research computes lowest-power, least-congested paths for all network destinations. The experiment results indicate that significant energy and cost savings could be achieved by prioritizing devices that process traffic more energy efficiently. In addition, a network dial with multiple modes is described that provides a convenient way to tradeoff between energy savings and network performance.The results from the network dial implementation confirm the feasibility of operating the network in different energy modes.The paper also describes implementation mechanisms for different generations of networks – traditional, programmable-control, and programmable-data -in order to demonstrate the potential for industry adoption.

The future directions for this work include employing machine-learning algorithms to enable the dynamic optimization of the dial modes, developing a hardware testbed to implement the algorithm on and test with different network devices, and extending the principles of this research to inter-domain routing.

REFERENCES

[1] P. Corcoran and A. Andrae, “Emerging trends in electricity consumption for consumer ICT,” *National University of Ireland, Galway, Connacht, Ireland, Tech. Rep.,* Jul. 2013.

[2] BP, “Statistical review of world energy,” vol. 69, 2020. [online] Available at: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-full-report.pdf.

[3] A. S. G. Andrae and T. Edler, “On global electricity usage of communication technology: Trends to 2030,” *Challenges*, vol. 6, pp. 117-157, Jun. 2015.

[4] C. Preist and P. Shabajee, “Energy use in the media cloud: Behaviour change, or technofix?” in *Proceedings of the IEEE 2nd International Conference on Cloud Computing Technology and Science*, pp. 581-586, Nov. 2010.

[5] C. Preist, D. Schien, and E. Blevis, “Understanding and mitigating the effects of devices and cloud service design decisions on the environmental footprint of digital infrastructure,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1324-1337, May 2016.

[6] W. V. Heddeghem et al., “Trends in worldwide ICT electricity consumption from 2007 to 2012,” *Computer Communications*, vol. 50, pp. 64-76, Sep. 2014.

[7] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, “A power benchmarking framework for network devices,” in *Proceedings of IFIP Networking*, pp. 795-808, May 2009.

[8] P. Mahadevan, S. Banerjee, and P. Sharma, “Energy proportionality of an enterprise network,” in *Proceedings of the 1st ACM SIGCOMM workshop on Green networking*, pp. 53-60, Aug. 2010.

[9] R. Gandotra and L. Perigo, “Comparing energy efficiencies of SDN hardware based on forwarding configurations,” in *Proceedings of the 29th International Conference on Computer Communications and Networks (ICCCN),* Honolulu, USA, Aug. 2020.

[10] R. Gandotra and L. Perigo, "We've got the power: A framework for real-time network power monitoring," *Journal of Computer and Communications*, vol. 8, pp. 75-88, May 2020.

[11] M. Gupta and S. Singh, “Greening of the Internet,” in *Proceedings of ACM SIGCOMM*, pp. 19-26, Aug. 2003.

[12] M. Gupta, S. Grover, and S. Singh, “A feasibility study for power management in LAN switches,” in *Proceedings of the 12th IEEE International Conference on Network Protocols (ICNP),* Berlin, Germany, pp. 361-371, Oct. 2004.

[13] L. Chiaraviglio, M. Mellia, and F. Neri, “Energy-aware backbone networks: A case study,” in Proceedings of the 1st International Workshop on Green Communications (GreenComm) in conjunction with the IEEE International Conference on Communications, Dresden, Germany, Jun. 2009.

[14] K.-H. Ho and C.-C. Cheung, “Green distributed routing protocol for sleep coordination in wired core networks,” in *Proceedings of 6th International Conference on Networked Computing*, pp. 1-6, May 2010.

[15] K. Xie et al., “E^{3}MC: Improving energy efficiency via elastic multi-controller SDN in data center networks,” *IEEE Access*, vol. 4, pp. 6780-6791, Oct. 2016.

[16] M. Gupta and S. Singh, “Using low-power modes for energy conservation in Ethernet LANs,” in *Proceedings of the 26th Annual IEEE Conference on Computer Communications (INFOCOM 2007),* Anchorage, Alaska, pp. 2451 – 2455, May 2007.

[17] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, “Reducing network energy consumption via sleeping and rate adaptation,” in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation (NDSI),* San Francisco, USA, pp. 323-336, Apr. 2008.

[18] G. Ananthanarayanan and R. H. Katz, “Greening the Switch,” in Proceedings of the USENIX Workshop on Power Aware Computing and Systems (HotPower), held at the Symposium on Operating Systems Design and Implementation (OSDI 2008), San Diego, USA, Dec. 2008.

[19] K. Christensen, P. Reviriego, B. Nordman, M. Mostowfi, and J. A. Maestro, “IEEE 802.3az: The road to energy efficient Ethernet,” *IEEE Communications. Magazine*, vol. 48, no. 11, pp. 50–56, Nov. 2010.

[20] K. Christensen, B. Nordman, and R. Brown, “Power management in networked devices,” *IEEE Computer*, vol. 37, pp. 91–93, Aug. 2004.

[21] C. Gunaratne, K. Christensen, and B. Nordman, “Managing energy consumption costs in desktop PCs and LAN switches with proxying, split TCP connections and scaling of link speed,” *International Journal of Network Management*, vol. 15, pp. 297–310, Sep. 2005.

[22] C. Gunaratne, K. Christensen, and S. W. Suen, “Ethernet Adaptive Link Rate (ALR): Analysis of a buffer threshold policy,” in *Proceedings of the IEEE Global Communications Conference (GLOBECOM 2006),* San Francisco, USA, pp. 533-534, Nov. 2006.

[23] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, “Energy aware network operations,” in *Proceedings of the IEEE Global Internet Symposium*, Rio de Janeiro, Brazil, pp. 1-6, Apr. 2009.

[24] M. d. S. Conterato, T. C. Ferreto, F. Rossi, W. d. S. Marques, and P. S. S. d. Souza, “Reducing energy consumption in SDN-based data center networks through flow consolidation strategies,” in *Proceedings of 34th ACM/SIGAPP Symposium on Applied Computing*, pp. 1384-1391, Apr. 2019.

[25] K. J. Christensen, C. Gunaratne, B. Nordman, and A. D. George, “The next frontier for communications networks: Power management,” *Computer Communications*, vol. 27, pp. 1758–1770, Dec. 2004.

[26] M. Jimeno and K. Christensen, “A Prototype Power Management Proxy for Gnutella Peer-to-Peer File Sharing,” in *Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN 2007),* Oct. 2007.

[27] M. Allman, K. Christensen, B. Nordman, and V. Paxson, “Enabling an energy-efficient future Internet through selectively connected end systems,” *Proc. ACM SIGCOMM HotNets Workshop (HotNets 07),* Atlanta, GA, Nov. 2007.

[28] S. Nedevschi et al., “Skilled in the art of being idle: Reducing energy waste in networked systems,” in *Proceedings of the. 6th ACM/USENIX Symposium on Networked Systems Design and Implementation (NSDI),* Boston, USA, pp. 381-394, Apr. 2009.

[29] A. Akella, R. Balan and S. Seshan.. "Protocols for low power," *SIGCOMM CCR*, Jan. 2002.

[30] M. Rodríguez-Pérez, S. Herrería-Alonso, M. Fernández-Veiga, and C. López-García, “Improved opportunistic sleeping algorithms for LAN switches,” in *Proceedings of the IEEE Globecom,* Honolulu, USA, pp. 1-6, Nov. 2009.

[31] P. Reviriego, J. A. Maestro, J. A. Hernandez, and D. Larrabeiti, “Burst transmission for Energy-Efficient Ethernet,” *IEEE Internet Comput.*, vol. 14, no. 4, pp. 50–57, Jul. 2010.

[32] C.-S. Li, Y. Ofek, and M. Yung, “Time-driven priority flow control for real-time heterogeneous Internetworking,” *IEEE INFOCOM’96*, Mar. 1996.

[33] M. Baldi and Y. Ofek, “Time for a “Greener” Internet,” in Proceedings of the 1st International Workshop on Green Communications (GreenComm) in conjunction with the IEEE International Conference on Communications, Dresden, Germany, Jun. 2009.

[34] B. Heller et al., “ElasticTree: Saving energy in data center networks,” in *Proc. of USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pp. 249-264, Apr. 2010.

[35] M. Zhang, C. Yi, B. Liu and B. Zhang, “GreenTE: Power-aware traffic engineering,” in *Proceedings of the 18th IEEE International Conference on Network Protocols*, pp. 21-30, Oct. 2010.

[36] Github.com. Ryu – Component-based software defined networking framework. [Online] Available: https://github.com/faucetsdn/ryu.

[37] Gns3.com. GNS3 | The software that empowers network professionals. [Online] Available: https://www.gns3.com/.

[38] Github.com. Netmiko – Multi-vendor library to simplify Paramiko SSH connections to network devices. [Online] Available: https://github.com/ktbyers/netmiko.

[39] B. Pfaff et al., “The design and implementation of Open vSwitch,” in Proceedings of the 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI), pp. 117-130, May 2015.

[40] Github.com. Stratum - Enabling the era of next-generation SDN. [Online] Available: https://github.com/stratum/stratum.

[41] Github.com. Mininet – Rapid prototyping for software defined networks. [Online] Available: https://github.com/mininet/mininet.

[42] P4.org API Working Group, “P4Runtime specification.” [Online] Available: https://p4.org/p4runtime/spec/v1.2.0/P4Runtime-Spec.pdf.

[43] P. Bosshart et al., “P4: Programming Protocol-Independent Packet Processors,” ACM SIGCOMM Computer Communication Review, vol. 44, no. 3, pp. 87-95, Jul. 2014.

[44] U.S. Bureau of Labor Statistics, “Average energy prices for the United States.” [Online] Available: https://www.bls.gov/regions/midwest/data/AverageEnergyPrices_SelectedAreas_ Table.htm.

[45] R. Ascierto and A. Lawrence, “Global data center survey,” Uptime Institute, Jul. 2020.

SECURITY IN AGILE DEVELOPMENT, USE CASE IN TYPEFORM

Pau Julià, David Salvador and Marc Peña

Security Team, Typeform, Barcelona, Spain

ABSTRACT

Software development methodologies have evolved during the last years to reduce the time to market to the minimum possible. Agile is one of the most common and used methodologies for rapid application development. As the agile manifesto defines in its 12 principles, one of its main goals is to satisfy the customer needs through early and continuous delivery of valuable software. Significantly, that none of the principles refers to security. In this paper, we will explain how Typeform integrates security activities into the whole development process, reducing at the same time the phases on the S-SDLC to reduce friction and improve delivery while maintaining the security level.

KEYWORDS

Security, S-SDLC, SDLC, AGILE, Development, Methodology, Use Case.

1. INTRODUCTION

Software development methodologies have evolved during the last years to reduce the time to market to the minimum possible. Methodologies like waterfall cannot adapt to these goals, so new methodologies arose to gain flexibility and allow rapid development, commonly called RAD [1] (Rapid Application Development), but at the same time unwittingly pushing the security checks and reviews out of the development and delivery flow. Agile [2] is one of the most common and used methodologies for rapid application development. As the agile manifesto defines in its 12 principles, one of its main goals is to satisfy the customer needs through early and continuous delivery of valuable software. Significantly, none of the principles refers to security.

2. CURRENT CONTEXT

In a waterfall development model, organizations usually performed security-related activities during the non-functional testing phase [3]. During this phase security engineers reviewed, tested, and analyzed the implemented solution, or new release of the software, before going live to production. This situation contributed to the creation of friction between the security and the engineering teams. This friction was a consequence of the security team's output during that testing phase: bugs to be fixed, new requirements, etc. Basically adding more workload to the development team after their implementation phase. This new workload also meant a blocking situation for the implementation to go live, affecting business goals and needs. On top of that, the security team's perception of the organization could be affected negatively by being seen as a "blocker" rather than helping the organization.

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 99-109, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111809

Most organizations had, and still have, a low ratio between security team members and engineering team members. This situation adds extra complexity to the task of improving security organization-wide. On many occasions, the security team will be the bottleneck of the organization. A simple way of trying to solve this problem would be to add a security team member to every engineering team. However, most organizations will not be able to do so due to costs, resources, and scalability reasons. Nowadays, the ratio between security and engineering headcount in Typeform is 1 to 30.

2.1. From Blockers to Facilitators

Security bugs will be found when the security team tests the implemented software. Those security bugs are added to the engineering team's backlog which may also include functional, UI, performance, and other types of bugs. This addition of more work to the team's backlog might cause fatigue and disengagement.

Unsurprisingly the security team has been perceived as the "NO department"; the gatekeeper that decides what can and what cannot be done with the information and with the technology; the team that blocks projects, tasks, ideas and introduces a delay in the delivery [4]. As a side effect of this, interaction with the security team is avoided as much as possible and not considered even for consultation in companies where security is not the core of the business.

Certain security policies put in place by the security team due to compliance or legal requirements tend to forbid or limit the usage of tools, libraries, and other software. In addition to that, certain security policies might force the organization members to go through slow and exhaustive processes to reach certain ends. This kind of situations push organization members to find ways to avoid those policies so, instead of promoting security within the organization, they lower it.

The goal and the mission of the security department must be just the opposite: identify critical risks affecting the critical assets and treat them with effective controls and be flexible with the less important ones.

Reducing friction with the rest of the teams and becoming an active facilitator team, will help improve the security across the company. On top of that, engineering teams will see security as a status to achieve and not as an obstacle. The only way to achieve that is to refocus security in a way that is non-intrusive yet effective.

2.2. Shifting Security to the Left

Systems Sciences Institute at IBM reported [5] that it costs six times more to fix a bug found during implementation than one identified during design. In addition, it is reported that the cost to fix bugs found during the testing phase could be 15 times more than the cost of fixing those found during design.

A paradigm called "shift security to the left" [6] was created aimed to identify security bugs at the earliest development stage possible. One of the key principles is that the responsibility of security should be shared with the engineering team. For this "shift" to be successful the security team must provide efficient tools and processes to the engineering team that will allow them to be as productive as before with the least amount of friction. For example, providing an IDE plugin to detect vulnerable dependencies as they are added.

As new tools and processes are provided to the engineering team new frictions in the development workflow can appear. For example, in an organization where every committed code is scanned for third-party vulnerabilities, giving the full information of the vulnerable component to the engineer can block the development. These details are useful for the security team but they can be too domain-specific. Following the previous example, a better way would be to inform the engineer that a dependency should be upgraded and provide an easy way to do it. The role of the security team is to provide tools and processes that are developer-friendly.

3. S-SDLC AT TYPEFORM

In this section, we explain how Typeform implemented an ad hoc Secure Software Development Life Cycle (S-SDLC) [7], its benefits, challenges, and problems found during its implementation. The S-SDLC described in this section was designed having in mind an organization with an agile development methodology and a short time to market need.

Typeform is an online Software-as-a-Service (SaaS) company that specialises in online form building and online surveys. Its main software creates dynamic forms based on user needs. Due to that, Typeform is dealing with a huge volume of information from its clients and the respondents of their forms and this implies a big responsibility to guarantee the confidentiality, integrity and availability of that information.

At a software engineering level, it follows an event-based microservice architecture approach. Every microservice is owned by one of the 20 engineering teams that exist in Typeform. The average deployment rate of a new version of a microservice is 0,5 hours (around 50 per day).

From a security perspective, an organization that continuously deploys new versions of its microservices is a double-edged sword. On the bright side, the same agility that allows new features to be fastly released, allows for security fixes to find their way into production in a quick way. So, attack windows will be shorter than in other less agile organizations. As a drawback, it is not feasible for the security team to keep track of every change or addition made to the production environment. Instead, the security team should provide tools, processes, and controls to avoid bringing flawed implementations into production.

3.1. Goals

Typeform’s S-SDLC has been designed to achieve the following goals:

Goal 1 – Developments securely achieve requirements and objectives from design.
Goal 2 – Security risks are identified and treated as soon as possible, ideally before they appear.
Goal 3 – The S-SDLC activities have a minimal impact on the development’s time to market.
Goal 4 – Security team’s perception is improved throughout the organization.

3.2. Phases

Given Typeform’s SDLC, agile and with a high rate of deployments, we classified the security related activities added to the SDLC in three steps: build, deploy, and run (Figure 1).

This classification follows an approach similar to a DevOps culture one: multidisciplinary teams with a high degree of autonomy and accountability of their developments. These teams prioritize mean time to mitigate (MTTM) and mean time to remediate (MTTR). Therefore, the workflow of these teams is heavily based on deploying code to production in a fast way. To not be disruptive to the organization, the S-SDLC has to be built around keeping these teams' objectives.

We consider as an S-SDLC activity any task or process, automated or not, that evaluates the security of the implementation and produces an output. This output can be, for example, in the form of a code fix, a design recommendation, or a vulnerability report. Next, we will describe every S-SDLC activity added in every phase.

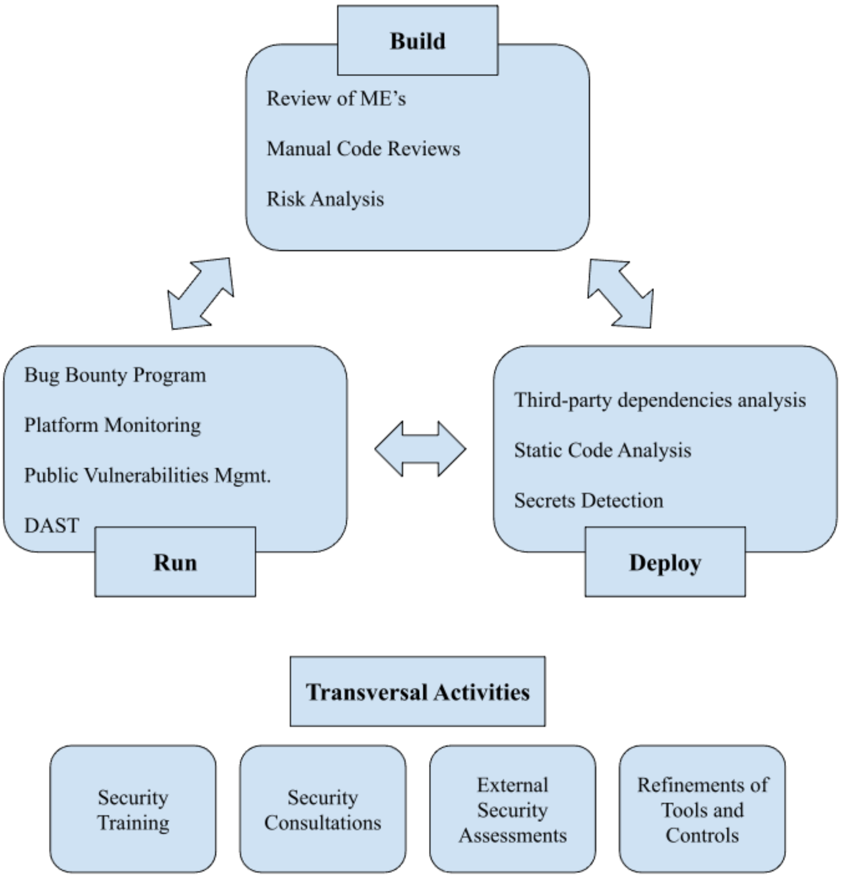


Figure 1. S-SDLC Phases and their security activities

3.2.1. Build

We consider this phase to start when a new development is conceived and to finish when the first implementation of it is done. That includes designs, spikes, or any other activity performed before there is an implementation complete enough to be deployed to production. Functionality is the factor that marks the suitability for its deployment in production. It also includes code that is deployed to testing environments.

The S-SDLC activities included in this phase are:

3.2.1.1. Review of Meaningful Experiments

In Typeform, a Meaningful Experiment (ME) is any A/B test [8], feature or technical work implemented in the product application. This concept has been created to empower engineering teams to experiment, learn from the outcomes, and consistently deliver quality and improvements.

Every ME is started by one or multiple engineering team members creating a Wiki-style page with definitions, potential problems, architecture changes, and any other useful information. The ME is then reviewed by different organization stakeholders: data analytics team, infrastructure team, architecture among others.

The security team is also one of the stakeholders who will review the ME. Before the review process, the authors of the ME have to fill out a small questionnaire with security questions. The answers will help the security team have an overview of what will be modified and how. The engineering team can start working on the ME implementation but before moving to the next phase, deploy, the ME needs to be approved by all stakeholders, including the security team. If appropriate, the ME can be labeled as "Security Review" meaning that further actions need to be taken by the security team. Those can be manually reviewing the source code, usually in the form of a Pull Request, dynamically testing the implementation when it reaches production or any other follow-up activity during the implementation's lifecycle.

Leveraging the MEs the security team is aware of every new implementation and can detect security problems at a very early stage. Given that the MEs reviews are made asynchronously, the security team will not be blocking their development. For the cases where new requirements are arisen by the security team, those will not create a heavy workload on the engineering team as they are still in the early development stages.

As a sample we took a period of 12 weeks, where 384 MEs were created (Figure 2); an average of 32 per week. Most of those MEs had no security relevance, for example, a ME about a color scheme redesign. From the MEs which had security relevance and needed further analysis, 49 of them were canceled during the Build phase, 43 of them needed further information, design, and documentation work, and 84 of them were accepted to move forward. Finally, from the MEs moving forward only 13 of them, a 3% of the total, were tagged for further Security Review.

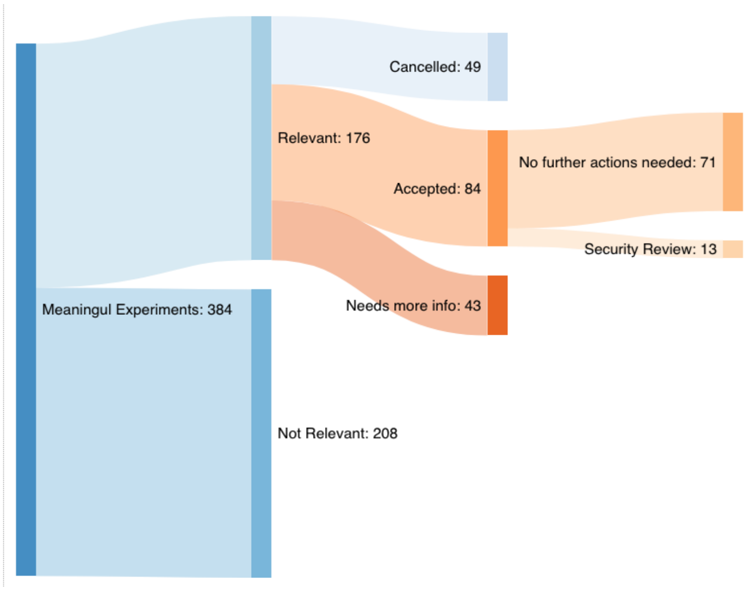


Figure 2. Meaningful Experiments review for 12 weeks.

3.2.1.2. Manual Code Reviews

Certain implementations related to sensitive features require a code revision by the security team: some examples are authentication and authorization, changes to core architecture, integrations with third parties, user-provided webhooks, etc. Those reviews are usually performed through one, or multiple, Pull Requests where the security team is added as a mandatory reviewer. As previously mentioned, this situation happens when a ME has been labeled as "Security Review". This situation allows the security team to have a good overview of sensitive features and their modifications over time. However, we found that in some cases this can be blocked: it is not until a security member has reviewed the code that the implementation can move forward. This blocking situation depends on the workload of the security team and its prioritization.

3.2.1.3. Risk Analysis

As with Manual Code Reviews, certain MEs labeled as "Security Review" will go through extra analysis during the Build phase. In this case, an ad hoc risk analysis will be performed by the security team.

The objective of this risk analysis is to evaluate the effects that the requirements defined in the ME will have against the organization's security in terms of information confidentiality, integrity, availability, and auditability. The outcome of the analysis will come in the form of action to be taken to accept, mitigate, avoid or reject the identified risks.

The risk analysis also includes a threat modeling process. This process aims to: identify security objectives, break down the application's architecture to identify potential security attacks, identify and categorize potential security attacks, and estimate the risk.

3.2.2. Deploy

This is a short-timed but crucial phase. This one includes the time range between a code implementation is ready to be deployed to production to when it is finally live in production.

All of the activities in this phase fall under the category of Static Application Security Testing (SAST), where the source code is scanned for security issues. The output of these scans is then used to decide if the implementation is moving forward and being deployed to production or if it needs to be fixed.

These activities reinforce the "Shifting Security to the Left" paradigm by providing scan results and fix suggestions in an early phase where the code has not reached production.

The S-SDLC activities included in this phase are:

3.2.2.1. Third-party dependencies vulnerability analysis

The third-party dependencies included in every source code repository are monitored automatically by a tool with a vulnerability database that is continuously updated.

When a vulnerability is detected in one of the third-party dependencies, and a fix exists, a Pull Request is created in the affected repositories. This Pull Request includes a fix for that vulnerability which is usually a version upgrade. The Pull Request also includes an automatically generated digest of the vulnerability and its severity. This flow helps the security team become a facilitator, as a solution is already being provided (the fix Pull Request) instead of just opening an issue in the organization's ticketing system, reducing the friction between the engineering and the security team.

The analysis of third-party dependencies is also made on every Pull Request made to the repository by the engineering team. The output of the scan will inform the Pull Request author whether a component with vulnerabilities is being introduced through that Pull Request or not. This flow helps to shift security to the left as the engineering team is responsible for that Pull Request.

3.2.2.2. Static Code Analysis

The most critical and sensitive source code repositories are integrated with a commercial Static Code Analysis tool.

For every new commit made to the main branch of the repository, a package with the new code is generated and sent to the Static Code Analysis tool. The tool scans the new code for vulnerabilities and stores the results. Afterward, a notification is sent to the security team if new vulnerabilities have been detected. If so, a security team member reviews the results and, if appropriate, opens an issue in the organization's ticketing system describing the vulnerability and suggesting a fix.

The Static Code Analysis scan result does not block any development, it is an asynchronous process that keeps an updated vulnerability status of the source code. The development can proceed and the security team reviews the results when convenient or needed.
It's been challenging to find a balance when choosing the number of repositories to scan and not introduce a lot of noise by notifications and excessive workload.

3.2.2.3. Secrets Detection

Unfortunately, uploading by mistake credentials, tokens, keys, or any other type of secret to the source code repository is a common issue. The impact of these incidents can be fatal, especially if the repository is publicly available; an attacker can use those leaked secrets as a starting point for a much bigger attack such as data exfiltration, defacement, advanced persistent threat, etc.

To avoid this we included what we call a "secrets detection" control in every Pull Request and commit made to the source code repository. The added code is scanned looking for patterns that might match with tokens, credentials, or any other type of secret. If any of the patterns match the engineering team member submitting that code is notified and redirected to a Wiki-style page with the steps to follow: revoke the secret and remove it from the source control versioning system.

In the early versions of our secrets detection control, the detection patterns needed several tuning as the rate of false positives was high. After some iterations, we managed to lower that rate while still covering a wide range of potential cases.

3.2.3. Run

This phase can be better described as a status. It refers to the moment where the code is live in production and remains there. However, that running code might be available only to a part of the users through the usage of "Feature Flags": a software practice of gating functionalities. A functionality, or feature, can be turned on and off via a management service allowing certain users, for example, the ones geographically located in Europe, to have access to that feature.

The S-SDLC activities included in this phase are:

3.2.3.1. Bug Bounty Program

Bug Bounty programs adoption is becoming very spread among organizations. This type of program offers recognition and usually an economical compensation to individuals who report bugs of certain software or service, especially bugs related to security vulnerabilities.

Every bug reported to the program is first triaged by an external security team part of the bug bounty program. If the bug is confirmed then Typeform's security team comes into action by manually reviewing the issue and triaging it. This manual intervention depends on the bug's severity and impact. Depending on those factors a further investigation to identify the root cause and if that same bug exists somewhere else in our product is performed. On multiple occasions we have encountered bugs that existed in different parts of the code too and, thanks to the internal expertise of our security team, we were able to identify them before being reported again in the Bug Bounty program.

Having a Bug Bounty program in place helped to close the gap between the amount of engineering team members and security team members. This gap is closed with the addition of both the individuals reporting bugs to the program and the external security team performing a first triage of the bugs.

Managing a Bug Bounty program has its challenges too; triaging timelines need to be met which can result sometimes in excessive workload to a small security team. Plus, certain disagreements with the bug, its severity, and impact sometimes happen between the reporting individual and both the external and Typeform's security team.

3.2.3.2. Platform Monitoring

It is key to our platform, the SaaS offered by Typeform, to have a security monitoring program. This is achieved by collecting all the data produced by the infrastructure, applications, and tools that power our platform. Once collected it can be analyzed and exploited allowing the security team to identify in early stages any suspicious activity or potential security vulnerability. This early identification helps to block those suspicious activities potentially before harm can be done. Other benefits include audit compliance, service level functional monitoring, performance measuring, liability limitation, and capacity planning.

Given the dimension of our platform, a fixed number of security team members need to be assigned to this task.

3.2.3.3. Public Vulnerabilities Management

Typeform's platform infrastructure is continuously scanned for public vulnerabilities affecting any of the components such as operating systems, virtualization software, networking components, etc. The scan is performed through a set of tools and services that inform the security team whenever a vulnerability is detected in any of the monitored components.

As with the Platform Monitoring activity, a fixed number of security team members need to be assigned to the management and refinement of this activity. Usually the same members given the shared knowledge and expertise on those matters.

3.2.3.4. Dynamic Application Security Testing (DAST)

Once the code has reached production and is running, it is tested for common web application vulnerabilities. This is done through a commercial software tool that tests for injection flaws, like SQL injection and Cross-Site Scripting, broken authentication, business logic flaws and other types of web application vulnerabilities. These tests are performed periodically and the output of the tests is sent to the security team for further review if needed.

Whenever a Meaningful Experiment has been labelled as “Security Review” these types of tests will also be carried out by a security team member manually and tool-assisted. To reduce the risk of having sensitive untested code running in production we use “Feature Flags”, which allow the untested feature or implementation to be available only to the security team for its review.

3.2.4. Transversal Activities

The security team performs other activities which contribute to increasing the level of security awareness, meeting standards and being compliant, and, in general, increasing the security in the organization. Some of those activities are performed regularly and some others on demand but we consider them as transversal, as multiple teams and stakeholders benefit from them.

3.2.4.1. Security Training

Historically, security training sessions at Typeform were performed as a one-hour long workshop. However, last year the organization went from an on-site working approach to an almost fully remote one. This new working situation, plus the fact of not having enough security team members to handle all the workload that comes with these training sessions, made us take a different approach: asynchronous training sessions.

We split the training material and topics into small, focused, and short asynchronous training sessions. To do that we used our product: online forms.

This format, online forms, allows the engineering team members to access the training materials whenever they find them suitable. Also, with an average of 5 minutes to go through all the slides, the online forms can be consumed in a more direct and distraction-free way.

Periodically, the security team adds new training material but, once the material is prepared, there is no need to invest a regular amount of time on it: opposed to what was being done with on-site training.

3.2.4.2. Security Consultations

The security team is always available to help with any matter related to the design, implementation, deployment, testing, or any other aspect where security is considered. Same as with any doubt regarding data privacy, GDPR or any other regulation or policy. These consultations are part of the day-to-day work in the team and help promote an organization-wide security culture.

3.2.4.3. External Security Assessments

The whole Typeform application and infrastructure goes through periodic third party security assessments. Having an independent entity performing the assessments increases the level of confidence and trust in Typeform’s organization. Usually, these assessments are performed every 6 months.

3.2.4.4. Refinement of Tools and Controls

Another day-to-day task of the security team is to manage all the tools and processes offered by security. This includes the secrets detection software, the application to monitor third party dependencies, the tool to scan source code, etc. These management tasks may include adding new features, fixing possible bugs, tuning configurations or managing its users.

4. RESULTS

The four goals of the designed S-SDLC (Section 4.1) have been achieved.

The first goal fulfillment has been demonstrated by external audits. Two independents external penetration tests were performed last year, and the outcomes of both were successful: no major issues were found.

The second goal is fulfilled by leveraging Meaningful Experiments. Security is considered from a very early stage by making the security team one of its stakeholders.

As for the third goal, the S-SDLC activities have a minimal impact in the development. Meeting this objective was not an easy task. For every activity, the security team needed to find a good balance between adding security and keeping the development process agile. This balance was not found instantly, and several iterations and refinements of the activity were needed. For example, when first adding Secrets Detection to the Deploy phase many false positive detections were found. Multiple refinements of the detection rules were, and still are on some occasions, needed.

As for the fourth goal, measuring the improvement of the security team’s perception throughout the organization is difficult. However, we consider that multiple factors indicate an improvement: submission of security tools improvements by the engineering team, increase in the number of security consultations, engineering teams taking ownership of certain activities like third-party dependencies analysis, etc.

5. FUTURE WORK

Typeform’s S-SDLC, as any other organizations, can still be improved. For that reason, the security team has some new activities to add to the S-SDLC and improvements to be made in certain phases or activities.

Shift security even more to the left. Review which of the activities performed during the “Deploy” phase could be partially or completely moved to the “Build” phase. For example, by adding an IDE plugin that statically analyses the code as the engineers create it and raises a warning if any potential issues are detected. These types of changes need to be reviewed with the engineering team to avoid adding unnecessary friction as it might complicate the software coding process and not even translate into enough security benefits.

Study the creation of a security champion program. This program suggests the creation of a new role in the engineering teams. A security champion is someone who has enough security knowledge to help with common and basic security challenges faced during implementation, design or any other development process. People having this role require some previous knowledge which, usually, is given by the security team. The creation of this role would be a good way of closing the gap between security and engineering team members. However, providing the necessary knowledge to these people would also require an investment of many hours by the security team. So, its viability needs to be first studied and reviewed.

6. CONCLUSIONS

This paper presents the details of a Secure Software Development Life Cycle implemented in a Software-as-a-Service organization, Typeform.

After reviewing the organization's context, four goals are set for the S-SDLC to achieve. The goals are set to be aligned with the organization's goals of a fast and frequent delivery of new software features. Then, the organization's SDLC is split in three phases: build, deploy, and run. For every phase, multiple security activities are detailed. These activities might be tasks or processes that evaluate the security of the software at a certain point of its development and produce an output.

Afterwards, how the S-SDLC goals have been achieved are detailed. We consider that our S-SDLC implementation met all the four defined goals by providing different indicators.

Finally, some improvements and future work related to the S-SDLC are discussed.

REFERENCES

[1] Martin, James (1991). *Rapid Application Development*. Macmillan. pp. 81–90.

[2] Kent Beck; James Grenning; Robert C. Martin; Mike Beedle; Jim Highsmith; Steve Mellor; Arie van Bennekum; Andrew Hunt; Ken Schwaber; Alistair Cockburn; Ron Jeffries; Jeff Sutherland; Ward Cunningham; Jon Kern; Dave Thomas; Martin Fowler; Brian Marick (2001). "Manifesto for Agile Software Development". Agile Alliance. Retrieved 14 June2010.

[3] https://www.toolsqa.com/software-testing/istqb/functional-and-non-functional-testing/. Retrieved 14 June 2021.

[4] Aaron Rinehart, Kelly Shortridge (2020). Security Chaos Engineering, Chapter 3 - SCE versus Security Theater— Getting Drama out of Security.

[5] https://www.researchgate.net/figure/IBM-System-Science-Institute-Relative-Cost-of-Fixing-Defects_fig1_255965523

[6] Donald Firesmith (23 March 2015). "Four Types of Shift Left Testing". Archived from the original on 2015-09-05. Retrieved 27 March 2015.

[7] Voitova, Anastasiia (2 November 2020). "Using SSDLC to Prepare for Security Incidents". *DZone*.

[8] Young, Scott W. H. (August 2014). "Improving Library User Experience with A/B Testing: Principles and Process". *Weave: Journal of Library User Experience*.

Automated Testing of Data Survivability and Restorability

Sylvain Muller and Ciarán Bryce

University of Applied Sciences (HES-SO),
Geneva, Switzerland

Abstract. Regular data backups are fundamental for protection against cyber-attacks and damage to infrastructure. To ensure a successful restoration, backed up data must be tested regularly for restorability to the company's current environment. Cloud providers generally test their backed-up data, but a testing framework is also required for locally stored files and databases. The paper proposes an automated test framework that validates the continued usability of backed up data for target restoration environments. The framework tests backups of Excel files, MySQL and Postgres databases, PDF documents and flat files.

Keywords Security, backup, automation, testing, infrastructure-as-code.

1 Introduction

Data is a company's most valuable asset. However, cyber-security attacks like ransomware or physical events like damage, loss or theft are serious threats to data in all companies. Effecting data backups is fundamental to companies' security and survivability since this is the only way company data can survive a physical or cyber-emergency [4].

Many companies have a regrettable tendency to forget that backup is only half of the matter. Restoration must be an efficient and well practiced process since, contrary to backups, restorations are initiated in emergency conditions and failure to restore correctly undoes the benefit of doing the backup in the first place.

Two conditions must be met for a data restoration to succeed. First, the backed up data, or *snapshot*, must survive until the restoration, and be readable during the operation. Second, the environment in which the snapshot is restored must be able to accommodate the restored data. The data may be unreadable in a different environment to the one in which the snapshot is made due to incorrect software versions. This can happen if a company moves location after an emergency or if there is a significant change in the environment between backup and restoration.

Companies have been moving to the cloud over the past decade, partly to delegate data backup and restoration issues. However, not all SMEs trust or are ready to move to the cloud. They prefer to manage data locally, and consequently, require a means to ensure backup and restoration for their files and databases.

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 111-126, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111810

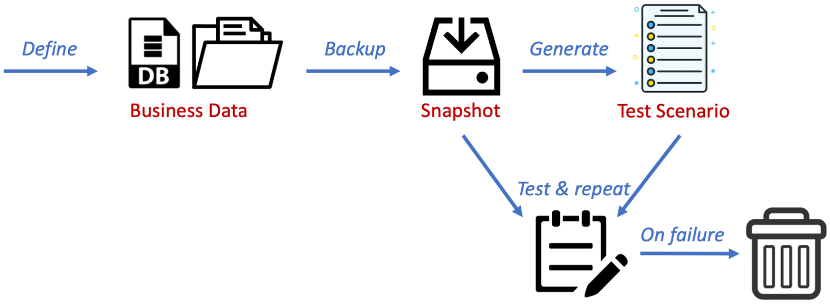


Fig. 1. Overview of Approach.

This paper proposes a framework where snapshots can be automatically tested for their usability, and for restorability with respect to target environments. The framework currently works on snapshots of Excel documents, Postgres and MySQL databases and flat files (standard directories). An overview of our approach is shown in Figure 1. The key features are the following:

- The business defines the crucial data files and databases that must survive.
- When a snapshot is created, a test scenario is generated for the backup data. An example test for an Excel file could verify the number of sheets in the document or the values of specific cells. The test for a database can verify the number of tables, column names and the values of certain database entries.
- A snapshot's usability is verified by loading the snapshot into a folder and running the tests. In the case of MySQL and Postgres, where the snapshot must be processed before running the tests, the snapshot is loaded into a Docker environment. The environment can be thrown away after running the test suite. The snapshot is considered usable only if all tests pass.
- We test the restorability of the snapshot by tweaking different properties of the Docker environment, e.g., by replacing Ubuntu 18 by Ubuntu 20 to test for restorability to a more recent Ubuntu environment.

Our aim is a simple automated framework that can be used as easily as possible. This means minimal competence to use the tool. For instance, the tests generated can mostly be generated automatically, even though the data manager can extend the suite with further tests. The approach is inspired by automated testing frameworks used in software development environments, e.g., [5, 8].

The remainder of this paper is organized as follows. We delve more deeply into the backup challenge in Section 2. We present the data testing model in Section 3 and its implementation in Section 4. Related work is described in Section 5. Conclusions and future work are presented in Section 6.

2 Backup and Restoration Challenges

We begin by clarifying the terminology used in the remainder of this paper:

- **Business data** refers to company data under normal operation. This data must survive a physical or cyber-emergency, so needs to get backed up and safely restored. Business data is only and precisely that data of value to a company in files and databases, so generally is a subset of the all data owned.
- A **snapshot** is a copy of business data for restoration. A restoration is made from a snapshot, so the snapshot must be stored separately from business data. In this way, damage to business data in an emergency does not impact snapshots. For instance, a ransomware that infects business data must not infect snapshots.
- A **backup** is a verb capturing the act of creating a snapshot from business data.
- A **restoration** is the act of replacing business data with a snapshot.
- An **archive** designates historical or non-current data. We do not consider archives here. Archives are generally read-only, stored on different media types and perhaps in different formats. The role of an archive is to index some historical event in the company's history, such as an old tax declaration, and not to provide data for use in a restoration following an incident.

Backup operations are universally considered essential to operational security in every company [13]. However, backups are only part of the story. The real challenge is to execute a successful restoration. Notably, backups are done under normal company operation. If a backup fails, it can simply be restarted. In contrast, a restoration is done in an emergency, and if it fails, the backup operation was all for nothing [17]. Companies therefore require simple frameworks to support restoration.

When a company manages its backups, restorations might fail for three reasons.

R1 The backup operation fails, so there is no usable data to restore.
R2 The snapshot becomes unusable.
R3 The environment in which the restoration is made cannot accommodate the snapshot.

Reason **R1** is the failure of the backup operation. This can be due to a configuration error (related to permissions or access paths), a storage or network error, write during copy, etc [1]. Another common reason for backup failure is human error, or where a company forgets or postpones the backup [20].

Concerning Reason **R2**, a snapshot is unusable if it cannot be read during a restoration. This happens if the snapshot data gets infected by a cyber-attack, if the storage media gets lost or damaged, or if the credentials to access a backup service where the snapshot is stored get lost or expire.

Definition Snapshot *survivability* is the property that the snapshot is usable for a subsequent restoration operation.

Some infrastructures test survivability by simply creating a message digest of a newly created snapshot; verifying that the snapshot digest remains unchanged is considered sufficient for survivability. However, this approach does not catch all errors. For instance, the message digest is not available on a cloud service that stores the backups. Also, the initial snapshot might be erroneous so the message digest is in fact a digest of invalid data. Finally, the approach requires us to be able to securely store the message digest.

Reason **R3** for restoration failure is that the environment in which the restoration is made cannot accommodate the backup. An extreme example of this is a restoration to a different environment to the one in which the backup was created, e.g., a company forced to move after an incident at its main office, and where the backup office has a different network or software infrastructure. Automating tests in these scenarios is hard. Nonetheless, we can test for restoration environments where software versions have changed since the backup environment.

Definition Snapshot *restorability* with respect to an environment E is the property that the snapshot can be restored within E.

A backup and restoration model for SMEs must be applicable to a variety of small company environments. Their infrastructures can be basic but diverse. The storage devices used include cloud boxes, hard disks, USB keys and external drives, as well as Cloud services (e.g., Dropbox, GoogleDrive, etc.). The applications that generate the data to be backed up include classical office applications (e.g., MS Excel and Word), databases (e.g., MySQL and MongoDB in cases where the database is part of a full-stack solution), flat files and folders, as well as custom business applications such as a payroll application or applications linked to company equipment (e.g., a dentist's radiography machine's application). The implication of this is that snapshots should be file based (rather than server images), that space considerations are important, and that snapshot testing must be neutral with respect to the storage media chosen for snapshots.

There are also emerging financial motivations for companies to achieve survivability and restorability. The first is the EU General Data Protection Regulation, which requires companies holding personal data belonging to EU citizens to implement secure backups. A second motivation is cyber-insurance. Following a cyber or physical incident, cyber-protection insurance contracts can pay for restorations. Therefore, it is important to make the restoration as reliable as possible to reduce costs for insurers and reduce insurance premiums for companies.

3 Model

Our goal is to automate the testing of business data. Implicit in this approach is the idea that the *data manager* – the person in the company responsible for the data – first identifies that data of high value to the company, and whose survival needs to be assured, c.f., Figure 1. After snapshot creation, a *test scenario* is generated by the data manager.

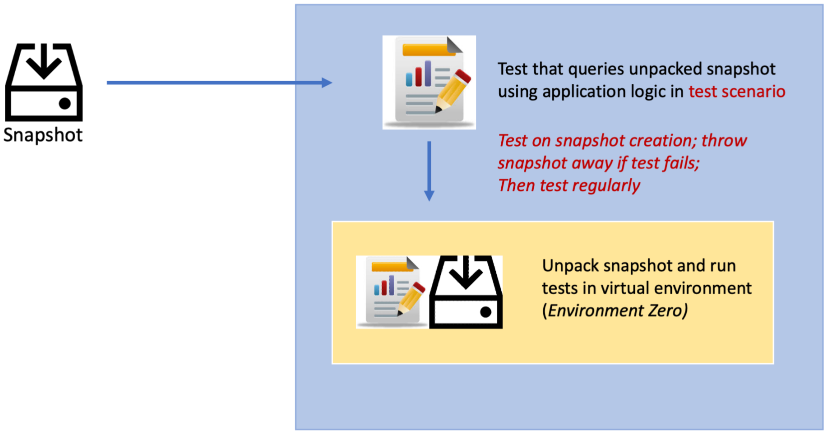

Snapshot
Test that queries unpacked snapshot using application logic in test scenario
Test on snapshot creation; throw snapshot away if test fails; Then test regularly
Unpack snapshot and run tests in virtual environment (Environment Zero)

Fig. 2. Overview of Approach.

The snapshot can be tested for survivability just after it is created, to ensure that the backup operation was successful, and at regular intervals to ensure that the snapshot is still accessible from its repository. Restoration tests, where we test that data is usable in changed environments, can be done at regular intervals.

3.1 Survivability Testing

The natural way to test snapshot survivability is simply to restore the snapshot to a temporary environment and to examine the restored contents. This approach is taken here, where automated tests check the contents of restored data, *c.f.*, Figure 2.

A shared folder is created in which the snapshot is copied and the tests are run. For MySQL and Postgres databases, an initialization phase transforms the snapshot into a queryable data structure. For this reason, the tests are run in a Docker-based virtual environment with the snapshot folder being shared with Docker.

The environment created for the survivability test is defined in a Hashicorp Language based *scenario* file, and an example is given in Figure 3. The locally_data_dir element defines where the snapshot being tested is stored.

```
module = "world"

version = "0.1.0"

local_data_dir = "backup"

environment database {
        resource mysql {
                image = "mysql:5.7.33"
                envs = {
                        MYSQL_ROOT_PASSWORD = "root"
                        MYSQL_DATABASE = "world"
                }
                ports = {
                        mysql = 3306
                }
              }
            }

```

Fig. 3. Test Scenario Description in HCL.

The image clause specifies the application image loaded into the Docker environment, which is a MySQL database in this example. The default Docker image for MySQL uses Debian Buster-Slim and PostgreSQL uses either Alpine Linux or Debian. Our default Docker container runs an Alpine Linux operating system.

The test framework is inspired by automated test frameworks used in modern software development environments like Java and Ruby/Rails, e.g., JUnit [2], RSpec [8], etc. The framework currently permits tests to be written that verify the contents of MS Excel and its OpenOffice cousin, as well as PDF documents, MySQL and Postgres dumps, as well as flat files.

The basic test operators of our framework are given in the Tables 1, 2, 3 and 4. A *test scenario* is written using these operators – and the expected contents of the snapshot. An example of tests written for an SQL snapshot is shown in Figure 4; a PDF file test is in Figure 5. These simply examine contents of the file (e.g., database entries) or examine structure (e.g., author names, names of columns or tables, etc.).

3.2 Restoration Testing

The survivability tests for databases restore data to a minimal environment, created within a Docker virtual container running with an Alpine or Buster Debian Linux. As we reject newly created snapshots that do not pass survivability tests, we know

Table 1. Operation list for MySQL

Test	Description
AssertColumnsExist(table, column_names)	Verify columns in table
AssertColumnMatch(table, column_name, regex)	Verify column values match regex
AssertRowCount(table, min, max)	Verify row count in range
Query(query, return_values)	Query table for specified values

Table 2. Operation list for Excel

Test	Description
AssertSheetsExist(sheet_names)	Check for sheet names
AssertCellsMatch(sheet, range, expr, type)	range can be A1;B5, type can be value, formula or link
AssertRowCount(sheet, min, max)	Verify that row count is in range
AssertColumnCount(sheet, min, max)	Verify that column count is in range

Table 3. Operation list for PDF files

Test	Description
AssertPageCount(min, max)	Number of document pages
AssertImagePageCount(page, min, max)	Number of images on page
AssertHasWatermark()	Document watermark
AssertTitleMatch(regex)	Document title
AssertAuthorMatch(regex)	Document author
AssertContentMatch(page, regex)	Content match to regex on page

Table 4. Operation list for folders

Test	Description
AssertFileCount(min, max)	Count the number of pages in the document
AssertFileNameMatch(regex)	Match file names to regex expressions
AssertFileExists(name)	File exists?
AssertFileSize(name, min, max)	File size
AssertDirExists(name)	Directory exists?

```
// Valid the structure and consistency of city table.
export function testCityTable() {
  mysql.assertColumnsExist(this, 'city', ['ID', 'Name', 'CountryCode', 'District', 'Population'])
  mysql.assertRowsCount(this, 'city', 4000, 4500)
  mysql.assertColumnMatch(this, 'city', 'Population', '[0-9]+')
}

// Valid the structure and consistency of country table.
export function testCountryTable() {
  const columns = [
    'Code', 'Name', 'Continent', 'Region',
    'SurfaceArea', 'IndepYear', 'Population', 'LifeExpectancy',
    'GNP', 'GNPOld', 'LocalName', 'GovernmentForm',
    'HeadOfState', 'Capital', 'Code2'
  ]
  mysql.assertColumnsExist(this, 'country', columns)
  mysql.assertRowsCount(this, 'country', 239, 239)
  mysql.assertColumnMatch(this, 'country', 'Code', '^[A-Z]{3}$')
}

// Valid the structure and consistency of language table.
export function testLanguageTable() {
  mysql.assertColumnsExist(this, 'countrylanguage', ['CountryCode', 'Language', 'IsOfficial', 'Percentage'])
  mysql.assertRowsCount(this, 'countrylanguage', 980, 1000)
  mysql.assertColumnMatch(this, 'countrylanguage', 'IsOfficial', '^[^TF]*(T|F){1}[^TF]*$')
}
```

Fig. 4. Example Tests for a MySQL database.

```
export function setup() {
  document = new pdf.Open('paper.pdf')
}

export function testPageCount() {
  document.assertPagesCount(this, 10, 15)
}

export function testWatermark() {
  document.assertHasWatermarks(this)
}

export function testAuthorMatch() {
  document.assertAuthorMatch(this, 'Sylvain Muller and Ciaran Bryce')
}

export function testTitleMatch() {
  document.assertTitleMatch(this, 'Automated Testing of Data Survivability and Restorability')
}

export function testImageCount() {
  document.assertImagesCount(this, 2, 1, 1)
}
```

Fig. 5. Example Tests for a PDF file

that there does exist a minimum, *Environment Zero*, in which data can be restored. Further, this environment is reproducible over time since it is created from the same set of configuration parameters.

Restoration testing is really just a special case of survivability testing. The key feature with restoration testing is to test with different software environments and versions, e.g., testing restoration to an older or future version of the database, or to a different operating system. This is done by defining a different HCL scenario file with a different Docker image.

Concretely, a Docker image file is downloaded from hub.docker.com using the Dockerfile that specifies the environment. For instance, the Dockerfile could specify "FROM mysql:8.0.18", as in the Dockerfile of Figure 6, meaning that the restoration test is made with version 8.0.18 of MySQL. The test can be replicated with different scenarios for different software environments. For instance, if MySQL 10.0 is made available, a new agent can be created with a Dockerfile with "FROM mysql:10.0".

The Docker framework can test restoration with different environmental settings. This is done by creating a shared folder between the calling environment and Docker in which system files can be transferred. This allows us to test the restoration against these new system files. For instance, consider the command: *docker run –name=psql -U softpeel -d -v /pgdata:/var/lib/pgsql/data -p 5000:3306 psql*. This runs the Postgres server in a Docker environment with the particularity that the

```
FROM golang:alpine as builder

WORKDIR /app

RUN mkdir -p /data
RUN CGO_ENABLED=0 GOOS=linux GOARCH=amd64 go build github.com/tigerwill90/
    replicas/cmd/mysql

FROM mysql:8.0.18

ENTRYPOINT ["/mysql"]
EXPOSE 5000
```

Fig. 6. An edited Dockerfile importing an Ubuntu OS image.

configuration folder in the Docker environment is replaced by the psgdata folder in the host environment.

3.3 Discussion

We believe that the framework presented here is quite flexible with respect to how it integrates into a company's data backup process.

When should snapshots be made? We do not specifically address this question in our framework. Our aim is simply to make it possible to test the snapshots. Issues such as the frequency of snapshot creation, conservation duration, number of snapshots, etc., is a business decision that company owners must make for themselves based on the value of their data and calculated risks to data. On the other hand, the framework does not impose any particular rhythm on the backup operations.

Who writes tests, and when? A test suite for a snapshot should be generated any time after the snapshot is created. A good rule of thumb is to write and run the test suite just after snapshot creation since this permits detection of snapshots corrupted during creation.

As suggested in the examples in Figures 4 and 5, our aim is to support the automatic generation of tests from business data. This is because most tests simply examine random content from the business data, or verify structure like the number of sheets in an Excel file, the names of tables and columns in a Postgres database, etc. On the other hand, a data manager is able to extend the test suite with his own tests. This approach is now common practice in automated software testing [8].

What should be done when a test suite fails? A test suite failure means that the snapshot cannot be restored in the environment described in the Docker-based environment. If this is the current working environment of the organization, then this means that the snapshot is no longer usable – unless some environment can be found and tested that permits restoration.

In reality, a test failure underlies the advantage of our model, since it pinpoints snapshots that cannot be restored before an emergency where a real restoration from the snapshot is needed.

How does the testing framework integrate in the backup process? Snapshot testing is an independent brick in the process of saving and restoring business data. Tests can be launched manually or automatically via a cron job. Tests can also be triggered via REST calls to our test motor's HTTP server. The latter allows us to integrate testing into an automated backup process based on the open-source Restic framework (see *Restic.net*). Restic automates the copy of file repositories that include both local folders as well as Cloud providers (e.g., Amazon, Google Cloud, MicroSoft Azure), FTP servers and REST-enabled Web servers.

What is particular about restoration tests? Fundamentally, there is no real difference between a survivability and restoration test from the perspective of using the framework. Only the restoration environment described in the scenario file changes.

An ideal situation is where we can test that snapshots can be restored in a wholly different environment, as can happen when a company foresees the need to change location after a major incident or when a software migration is being prepared for at the local site. Testing for major environmental changes is a challenge, since there are so many parameters to consider: machine hardware and network configurations, software licenses, user permissions, etc. This is beyond the scope of our current framework, but does merit further research.

4 Implementation

The framework is a *motor* written in Golang that runs on Linux, Mac OS X and Windows. The motor is accessed via a REST API that also includes encapsulates a command line interface. The only external dependency of the tool is with Docker[1] which provides the virtual containers in which survivability and restorability tests are run on database snapshots. This has the advantage that the tests do not have a side effect on the rest of the system. Further, the snapshot itself does not get modified in the real environment; only in the virtual environment. In this way, the snapshot is immutable for its whole lifetime.

The tests are written in JavaScript; the template is shown in Figure 7. Test programmers simply add the tests they require to the test scenario. Added test functions simply need to begin with "test" and respect Camel casing.

A test can be run from the command line and specific tests can also be executed:

[1] www.docker.com

```
import { sleep } from "scenario";

// Setup allow to do extra setup before testing like recover a database
export function setup() {
        console.log(’recovering date’)
        sleep(30)
}

// Each test name must begin by the suffix ’test’.
export function testShouldFail() {
        this.fail(’failed’)
}

// Teardown allow you to perform clean up operation
export function teardown() {
        // some clean up
}
```

Fig. 7. JavaScript test template.

- scenario run – run all test methods in the suite
- scenario testShould[A-z]+$ – run all test methods that match the argument.

The setup method reads the data from the folder where the snapshot is stored. The example of Figure 8 comes from the test of a MySQL database.

The framework is written in just over 10 thousand lines of Golang code. The kernel of the framework implements assertions for database and other content types, as well as the link to Docker for starting the test environment.

From a performance perspective, the majority of the execution time is on loading the data into the Docker environment. For instance, the MySQL example of Figure 4 takes 30 seconds on a 400 kB database. This includes the time to create the Docker container with the required image, unzip the database, load it into the container and then run the tests. The time is comparable to that needed to start a database server in a production environment.

5 Related Work

All standard operating systems come with tools for backup and restoration [14]. Windows for instance allows for incremental and differential backups of the file system. Files and folders possess an archive flag that is set whenever the file is modified, and in need of backup. This bit is reset when an incremental or full backup is made. The idea behind this approach is to facilitate a full system restoration. In contrast, the goal of our approach is to permit an SME to identify a critical set of business data that needs to be restorable, and then to handle the backup and restoration of just that data. This allows the SME to optimize backup storage and to focus on the company-critical data. Further, our method supports restoration in environments different to those in which backups are done.

```
import { sleep, retry } from "scenario"
import archive from "scenario/archive"
//import { unzip } from "scenario/archive"
import docker from "scenario/docker"
import sql from "scenario/assert/sql"

const config = {
  user: ’softpeel’,
  password: db.postgres.envs.POSTGRES_PASSWORD,
  db_name: db.postgres.envs.POSTGRES_DB,
  port: db.postgres.getPort(’postgres’),
  ip: db.postgres.getIp()
}

const postgres = new sql.Client(’postgres’, config)

export function setup() {
  // Exponential backoff retry that ping world database it is ready
  retry(function () {
    return postgres.ping()
  }, 30)

  const zip = archive.Zip(’softpeel.sql.zip’, {overwrite: true})
  zip.extract()
  const f = open(’softpeel.sql’)

  // MySQL cli config to recover sql dump.
  const config = {
    cmd: [’psql’, ’-U’, ’softpeel’, ’-d’, db.postgres.envs.POSTGRES_DB],
    envs: [‘POSTGRES_PASSWORD=${db.postgres.envs.POSTGRES_PASSWORD}‘],
    stdin: f,
  }

  // Execute command on mysql resource
  const code = docker.exec(db.postgres, config)
  if (code != 0) {
    throw (’recovery fail’)
  }
}
```

Fig. 8. The setup method creates reads the data for the enclosing Docker environment

With the advent of the cloud, an emerging approach to handle backups is to save a snapshot of each virtual machine's current image, e.g., [19]. This contrasts with the file-based backup because all application code and data is backed up. Most cloud providers implement this technique. There are several challenges nonetheless. First, the image is not certain to have been taken at a moment when the data is consistent; for instance, a transaction may have been running on the database. Second, the solution requires a lot of storage space, which makes indexing and deduplication harder, e.g., [7]. In contrast, the file-based approach such as the one taken in this paper encourages users to identify business critical data, is more economical in storage space, and snapshots can be restored to a wider range of environments.

Recent research in data storage generally looks at issues like storage area networks design, performance and security, e.g., [6, 1], backup techniques, e.g., [9], or specific angles of backups like minimizing service outage during backups [18, 11] or deduplication [12]. However, the issues of survivability and restorability are less frequently treated. This needs to change as regulations increasingly require companies to conserve data and comply to external requests for data conservation.

Frameworks with scripts that automate backups have appeared, e.g., *restic.net*, and are accessible to SMEs. This is a positive development since automating the backup and restoration procedure reduces the margin of error in the operations. We have used our test framework in Restic scripts where the tests are invoked in the script after the snapshot is created., but it can be added to any automated backup system.

A related field is *infrastructure-as-code* (IaC), e.g., [3, 15, 10]. IaC is the DevOps [16] idea where developers and operators use automated scripts for the provision of software systems in their company. IaC greatly facilitates software system updates since the process is made as efficient as possible. Importantly, it is easier to do a *rollback* if the system is discovered to have post-deployment bugs since the steps taken by human operators are removed or reduced. IaC was invented for code deployment but the principles are just as applicable to data backups and restorations. Further, it is natural to be able to add a test framework such as the one presented in this paper to these IaC scripts since data, like code and services, are fast-class entities in company infrastructures.

6 Conclusions

This paper has presented a framework for the methodical testing of file-based data backups. The file-based approach optimizes backup storage space and allows companies to focus on their critical business data. The approach can also be used by companies that run virtual servers on the cloud, even though the cloud provider creates backup snapshots of the virtual machines. Storing data in the cloud does not remove responsibility from companies to control and test their backed up data.

Automated testing works for MicroSoft and Open Office (spread-sheet and document) files, MySQL exports, flat files and PDF documents. Future work will extend test coverage to image files and encrypted files. To test encrypted snapshots, we need to extend the framework to run the tests in a sandboxed environment, such as that supported by Linux AppArmor, and for which Docker can be run [21].

References

1. George Amvrosiadis and Medha Bhadkamkar. Getting back up: Understanding how enterprise data backups fail. In *2016 USENIX Annual Technical Conference, USENIX ATC 2016, Denver, CO, USA, June 22-24, 2016.*, pages 479–492, 2016.
2. Yoonsik Cheon and Gary T. Leavens. A simple and practical approach to unit testing: The JML and junit way. In *ECOOP 2002 - Object-Oriented Programming, 16th European Conference, Malaga, Spain, June 10-14, 2002, Proceedings*, pages 231–255, 2002.
3. Clauirton de Siebra, Rosberg Lacerda, Italo Cerqueira, Jonysberg P. Quintino, Fabiana Florentin, Fabio Q. B. da Silva, and André L. M. Santos. From theory to practice: The challenges of a devops infrastructure as code implementation. In *Proceedings of the 13th International Conference on Software Technologies, ICSOFT 2018, Porto, Portugal, July 26-28, 2018.*, pages 461–470, 2018.
4. Vasiliki Diamantopoulou, Aggeliki Tsohou, and Maria Karyda. From ISO/IEC 27002: 2013 information security controls to personal data protection controls: Guidelines for GDPR compliance. In Sokratis K. Katsikas, Frédéric Cuppens, Nora Cuppens, Costas Lambrinoudakis, Christos Kalloniatis, John Mylopoulos, Annie I. Antón, Stefanos Gritzalis, Frank Pallas, Jörg Pohle, M. Angela Sasse, Weizhi Meng, Steven Furnell, and Joaquín García-Alfaro, editors, *Computer Security - ESORICS 2019 International Workshops, CyberICPS, SECPRE, SPOSE, and ADIoT, Luxembourg City, Luxembourg, September 26-27, 2019 Revised Selected Papers*, volume 11980 of *Lecture Notes in Computer Science*, pages 238–257. Springer, 2019.
5. Thomas Fehlmann and Eberhard Kranich. A framework for automated testing. In Murat Yilmaz, Jörg Niemann, Paul M. Clarke, and Richard Messnarz, editors, *Systems, Software and Services Process Improvement - 27th European Conference, EuroSPI 2020, Düsseldorf, Germany, September 9-11, 2020, Proceedings*, volume 1251 of *Communications in Computer and Information Science*, pages 275–288. Springer, 2020.
6. Kazuo Goda. Storage area network. In *Encyclopedia of Database Systems, Second Edition.* 2018.
7. Jianxin Li, Yangyang Zhang, Jingsheng Zheng, Hanqing Liu, Bo Li, and Jinpeng Huai. Towards an efficient snapshot approach for virtual machines in clouds. *Inf. Sci.*, 379:3–22, 2017.
8. Myron Marston and Ian Dees. *Effective Testing with RSpec 3: Building Ruby Apps with Confidence.* The Pragmatic Bookshelf, 2017.
9. Donghee Min, Taegye Hwang, Joonhyouk Jang, Yookun Cho, and Jiman Hong. An efficient backup-recovery technique to process large data in distributed key-value store. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing, Salamanca, Spain, April 13-17, 2015*, pages 2072–2074, 2015.
10. Kief Morris. *Infrastructure as Code.* O'Reilly, 2016.
11. Pratik Mukherjee and Valentina Salapura. Challenges of DB2 restore in a distributed systems environment and engineered solutions. In *48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, DSN Workshops 2018, Luxembourg, June 25-28, 2018*, pages 38–42, 2018.
12. J. K. Periasamy and B. Latha. Efficient hash function-based duplication detection algorithm for data deduplication deduction and reduction. *Concurr. Comput. Pract. Exp.*, 33(3), 2021.

13. Antònia Mas Picahaco, Antoni Lluís Mesquida, Esperança Amengual Alcover, and Bartomeu Fluxà. ISO/IEC 15504 best practices to facilitate ISO/IEC 27000 implementation. In *ENASE 2010 - Proceedings of the Fifth International Conference on Evaluation of Novel Approaches to Software Engineering, Athens, Greece, July 22-24, 2010*, pages 192–198, 2010.
14. W. Curtis Preston. *Backup and recovery - inexpensive backup solutions for open systems: covers Windows, Linux, Unix, and OS X.* O'Reilly, 2007.
15. Akond Rahman, Rezvan Mahdavi-Hezaveh, and Laurie Williams. A systematic mapping study of infrastructure as code research. *Information & Software Technology*, 108:65–77, 2019.
16. James Roche. Adopting devops practices in quality assurance. *Commun. ACM*, 56(11):38–43, 2013.
17. Jibran Saleem, Bamidele Adebisi, Ruth Ande, and Mohammad Hammoudeh. A state of the art survey - impact of cyber attacks on sme's. In *Proceedings of the International Conference on Future Networks and Distributed Systems, ICFNDS 2017, Cambridge, United Kingdom, July 19-20, 2017*, page 52, 2017.
18. Xiaoyan Yin, Javier Alonso, Fumio Machida, Ermeson C. Andrade, and Kishor S. Trivedi. Availability modeling and analysis for data backup and restore operations. In *IEEE 31st Symposium on Reliable Distributed Systems, SRDS 2012, Irvine, CA, USA, October 8-11, 2012*, pages 141–150, 2012.
19. Lingfang Zeng, Shijie Xu, and Yang Wang. Vmbackup: an efficient framework for online virtual machine image backup and recovery. *Concurr. Comput. Pract. Exp.*, 28(9):2630–2643, 2016.
20. Lingfang Zeng, Shijie Xu, and Yang Wang. Data backup: Do business want to measure recovery potential? *Issues in Information Systems*, 19(1):20–28, 2018.
21. Hui Zhu and Christian Gehrmann. Lic-sec: an enhanced apparmor docker security profile generator. *IACR Cryptol. ePrint Arch.*, 2020:1147, 2020.

DATA PROTECTION THROUGH DATA SECURITY-AS-A-SERVICE USING BLOCKCHAIN ENABLED PLATFORM

Dr Magesh Kasthuri, Hitarshi Buch,
Krishna Moorthy and Vinod Panicker

Wipro Limited, India

ABSTRACT

Data access is inevitable in today's world and it is prone to threat attacks and hence data security is utmost important for any enterprise to handle industrial solutions. The economics of data being used across the industries rapidly growing in current digital world so the potential data related threats is also rapidly growing. Data security is an integrated solution component for any Enterprise solution but with the growing demand on data security and potential threat handling, Data Security as a Service (DSaaS)f is a new model widely accepted in modern age architecture in Blockchain and Big Data world combining the power of cloud based security services, decentralized network in Blockchain and tamper-proof ledger management. Any Enterprise Security architecture comprises of how data is handled in a secured way and how integration between services (consumers/producers or API interaction or any middleware services) handles data between them. Hence it is inevitable to that future technology adoption should include Data Security-as-a-service for zero-trust solution design complying with compliance and security standards for industry.

KEYWORDS

Data Security, Blockchain, Decentralized Ledger, DSaaS, Data Loss Prevention (DLP), User and Entity Behaviour Analytics (UEBA), Cloud Access Service Broker (CASB), Certificate Management, Key Management.

1. INTRODUCTION

Data has become the easy target for the cyber attackers due to its spread and availability. The data explosion or data breach has become a new criminal activity impacting business, government, and individuals. Protecting sensitive data is the key priority for the digital enterprises as innovative ways are being found to compromise the systems, attack the security, and steal data. Having a cloud-based offering is not a solution considering its centralized systems, interfaces and accessibility still gives a vulnerable target. Also, some of the companies are not ready to accept their business data or personal data is stored in the cloud or third party provider (SaaS). Hence, handling of data securely is the key in digital transformation is the need of this hour.

Data security is a practice and technique to safeguard the sensitive data in the digital ecosystem. It includes protecting the data from unwanted actions through unauthorized access, data theft and corruptions throughout the life cycle of a data. It aims to maintain the data confidentiality, reliability, availability, and data integrity. It encompasses the organizational policies, country specific policies and procedures to protect the data from destructive forces.

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 127-143, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111811

In Cloud platforms, security and compliance feature for Federal Risk and Authorization Management Program (FedRAMP), US Department of Defence Architecture Framework (DODAF) and UK's Ministry of Defence Architecture Framework (MODAF) are supported by native services and during application design, supported cloud services for these compliance framework to be chosen for data protection activities.

This article provides foundational blocks of designing a DSaaS solution, possible DSaaS requirements at various states of data, DSaaS capabilities required, use cases and possible reference architecture for DSaaS by leveraging Blockchain solution. The use of Blockchain technology does not fully remove the inherent risks of data security as it needs to be pro-actively managed.

2. LITERATURE REVIEW

Subashini et al [1] explains the real-time common threats to data applications in cloud platforms like AWS or Azure and how it can be approached in different architectural models like Multi-cloud or hybrid cloud solution design. They have discussed how data at rest and data in motion is important in architectural decision for cloud application design.

Deyan et al. [2] in their survey paper explains how data privacy is important in various security and compliance requirements including country restrictions like regulatory policies and how it can be addressed in cloud platforms like Azure, Amazon and Google Cloud in both backend database platforms and in front-end application services. The author explains some key problems in data security including public access restrictions and data management issues.

Vladimir et al. [3] talks about some of the common data security issues in cloud architecture in a given domain solutions like Healthcare and Financial services and how common data security policies like General Data Protection Regulations (GDPR) helps in customer personal data security policies and adoption of customer identity access management (CIAM) solutions. The authors explains a data protection approach for Data Loss Prevention (DLP)

Ravi et al. [4] proposed a solution approach for integrating security services for data protection in cloud platforms including application, platform and infrastructure security services and how to address non-functional requirements for application design. The authors proposes an efficient User and Entity Behaviour Analytics (UEBA) framework for cloud data protection.

Eman et al. [5] has done detailed survey for data protection techniques in cloud platforms and how Software as a Service (SaaS) solutions like Microsoft Active Directory (AD) or Amazon Cognito along with Key vault and certification manager helps in data protection in two-way handshake during application integration services.

Vijay et al. [6] proposed a new-age service model in cloud platforms for enhanced application protection and how cloud services can handle data security in backend data store and integration services. They have also explained Five pillars of Well-architected framework covers data security as a pillar in enterprise application design in cloud platforms. The authors introduces Cloud Access Service Broker (CASB) method for cloud data security.

Mohammed et al. [7] introduced a new-age solution called SECaaS which means security-as-a-service framework which can be applied on cloud-based application design for data protection and how this can be an cloud agnostic solution to be used for any hyperscalers in cloud application design.

3. KEY COMPONENTS OF DSAAS

In this context, conceptualizing Data Security as a Service (DSaaS), that provides various data security requirements, data security capabilities, services, policies, procedures, and associated use cases forms an important design consideration for any Enterprise. Key Components of DSaaS are shown in below Figure-1.

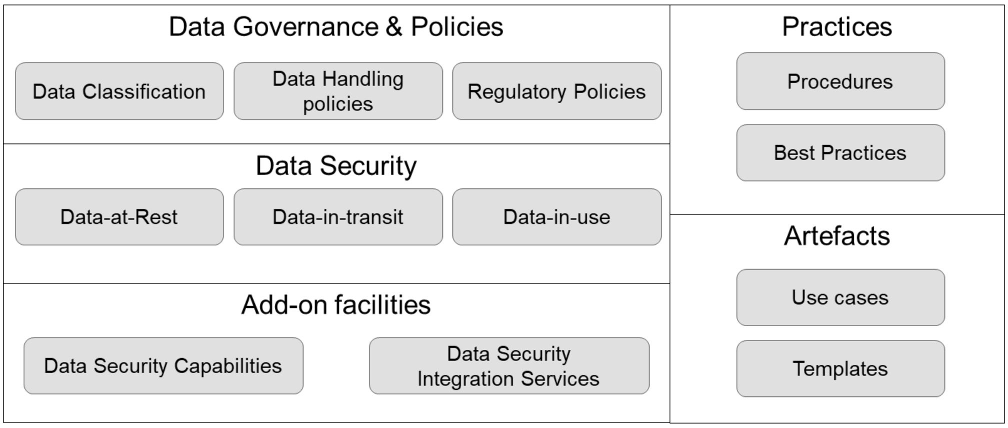


Figure 1. Key components of DSaaS

Blockchain platform can act as a main enabler for DSaaS in which a Distributed Ledger Technology providing an increased cyber resiliency and maintains ledger integrity because of its decentralized architecture, implementation of enhanced security frameworks for tamper-proof transaction, access patterns with no single point of failure (SPoF). The data is stored in blocks and connected with chain of blocks; thus, attacking a specific block does not affect the other blocks and the attacker needs to tamper all the blocks, but then detection is evident. The encryption and cryptography solution that Blockchain applications use to manage the data or transactions blocks protects individual transactions or records and the entire ledger. Thus, Blockchain proves to be a holistic capability to serve DSaaS requirements.

4. DATA SECURITY REQUIREMENTS

With external attacks account for the majority of data breaches suffered, Attackers take advantage of the fact that firms are interconnected and reliant on many other components of a broader business ecosystem. Emerging Data Protection regulations and Data privacy and protection compliance like General Data Protection Regulation (GDPR), General Data Protection Regulation (CCPA), Health Insurance Portability and Accountability Act (HIPAA) and Payment Card Industry Data Security Standard (PCI DSS) makes tough for organizations to identify what data falls under the umbrella of such regulations.

CXOs across different industries indicated that Personally Identifiable Information (PII) and Intellectual Property (IP) as the top two data types compromised in a breach followed by theft of payment and credit card data. Malicious internal incidents are on the rise and attackers are compromising employee credentials to data or tamper with it. Masquerade as insiders (with legitimate access privileges) in their efforts to access sensitive data causes potential data threats. Lot has changed with enterprises embracing cloud and cloud based services and increasing risk in data protection in such an environment. DSaaS (Data Security as a Service) like encryption as a service and certificate management as a service are now used by organization to improve their security posture in the cloud.

Considering the current exposure to the threats, the data security requirements are explained using its three states of data in a digital ecosystem. The three states of data are: **Data-at-Rest, Data-in-Transit or Data-in-Motion and Data-in-Use.** Any data can be exposed to threats when it is at rest, or in-transit or in-use. It requires a protection layer and robust data security solution to enable this requirement. There are multiple approaches to secure the data at various juncture and encryption plays a major role in enabling this.

4.1. Data-at-Rest

Data-at-Rest is generally termed as one of the states of digital data, where data is not moving, and the data is not being transferred or accessed. It is a stable state and physically stored when compared to the other states and it can be stored in cloud, end points(computers, PCs, mobile devices) file servers, tape drive, hard drive, computers, any physical devices, or archive storage/document management systems, etc. in any form.

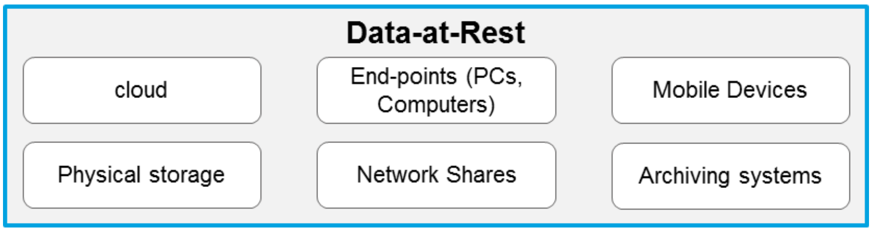


Figure 2. Data-at-Rest services

The data-at-Rest scenarios are increasing concern to organisations, business and government. So focus is to provide data security requirements for data-at-Rest to avoid malicious attacks and theft of physical storage media. In general, these data storage areas are highly protected with various defensive layers including security zones, firewalls, anti-virus layers and physical securities. However, these are not impassable.

The data security requirements start from where these data gets stored i.e., storage location, storage type, storage media and access requirements. Additionally, another critical point is to understand how it is classified based on the data classification requirements. So, key considerations on data security at rest are:

- **Secured, physical location:** Physical location where data is stored is one of the critical considerations whilst adopting data security at rest.
- **Size of the network:** Size of the network plays a major role here if a Blockchain network is not too large or well distributed and it becomes a potential risk for the attack.
- **Secured, physical location with classification of data storage at classified locations:** Ideal to consider multiple locations for multiple types of data based on the data calcification outcome. But it is also a challenge considering multiple copies of data can be available in PCs, Storage Devices, and Mobile devices. With the distributed architecture of Blockchain, it is feasible to store data at chunks across the locations.
- **Storage:** Data Storage on mobile devices is a key challenge considering use of these mobile devices are very common nowadays. Hence data protection that is stored within these mobile devices is another challenge to consider.
- **Infrastructure set-up:** A key parameter that influences the Data-at-Rest is its infrastructure set-up and right-levels of secured server farms with right patches up-to-date. It includes redundant, highly available distributed architecture to have a reliable storage. In addition to this, having data controls on cloud storage is good, however the actual encryption keys are

owned by the storage provider and hence control is not with the company. Applying right sets of protection layers for the storage media also a critical requirement to avoid the potential risks.

- **Access Policies:** Considerations to be applied with strong access policies to prevent these risks, as longer the data remains unused in storage, the more likely it might be at risk. Additionally, admins of permissioned Blockchain networks can hamper the blocks e.g., rewriting blocks history, delete resources etc.
- **Blockchain users backing up the Private keys in physical media:** Theft of these private keys is another big challenge as these keys are used for critical operations in the Blockchain ecosystem. So, having a secured key governance practices are key for this.
- **Automated approach to evaluate the security requirements:** Automated process to evaluate the security risks considering automated data classification framework and data protection evaluation to identify potential risks.

To prevent these data being accessed or stolen, companies apply security protection measures with additional layers of defence such as data encryption, password protection or both. The security options used for this type of data are broadly referred to as Data-at-Rest Protection (DARP). The companies should consider right data security requirements for the data-at-rest as part of the preventive Data Loss Plan along with country specific data protection regulations as applicable.

4.2. Data-in-Transit or Data-in-Motion

Data-in-Transit or Data-in-Motion or Data-in-Flight is termed as data which is travelling from one point to another through public or private communication channels such as messaging, mailing, sharing, chats, cloud, collaboration tools (eg. Microsoft Teams, Zoom meeting), applications etc. It is data moving in a network i.e. opposite in meaning for Data-at-Rest. In the current digital world, data is shared across in multiple ways to collaborate each other. Since data is moving in a network and it contains many nodes, where different access points are connected to the same network, hence data-in-motion needs to be protected.

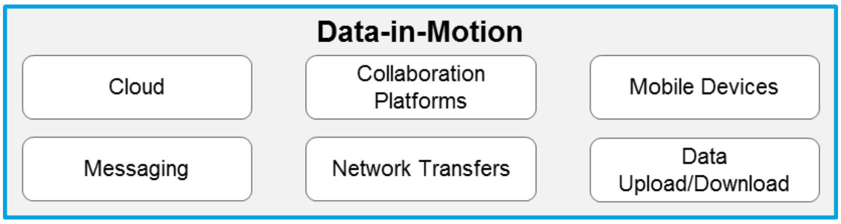


Figure 3. Data-in-motion services

Key considerations for Data in motion are:

- **Need for a Transport layer security:** Transport layer security to be considered for all the sensitive data that is being traversed in the network using SSL/TLS protocols. However, the challenge is, there are infinite number of mechanisms and means of channels for data sharing. Examples include email encryption using PKI (Public Key Infrastructure), File Transfers using Secured File Transfer Protocols (sFTPs) and HTTPS.
- **Prevent all:** Prevent accessing the sensitive information even from the root users/administration users.

- **Data transfer/sharing mechanisms:** Moving sensitive data from one location to another using USB drives, uploading to cloud storage, web contents etc.
- **Secured access from individual workstations vs End-user attacks:** It is common mechanism to leverage end-user access points for malicious practice as it is an entry point. The attacker will gather user credentials to infiltrate the network
- **Data Leak Prevention:** Identification of the purpose of retrieving sensitive data whilst it-in-transit is a challenge. Also, unable to identify potential risks when data is accessed by the endpoint or at receiving end as control of transport layer is no more applicable.
- **Private Key management:** The recent cyberattacks have proved that they are achieved through stealing end-user keys and then using those to enter the network, instead attacking the server farm directly through malware/virus/physical assault. So, the private key management is the key consideration for data protection.
- **Automated approach to evaluate the security requirements:** Automated process to evaluate the security risks using suspicious activities in the network, diagnose potential threats and proactively improve the security.

4.3. Data-in-Use

Data-in-Use can also be called as active data in the context of being worked in a database or accessed by an application. In general, any data is opened by either application or users, for its consumption or any treatment, then we can consider the state as Data-in-Use. In this state, data is more vulnerable considering data is decrypted whilst processing the data. The requirement is to have additional controls prior to access the data. The examples include working/accessing the Applications (on premise, cloud, and mobile) and Database.

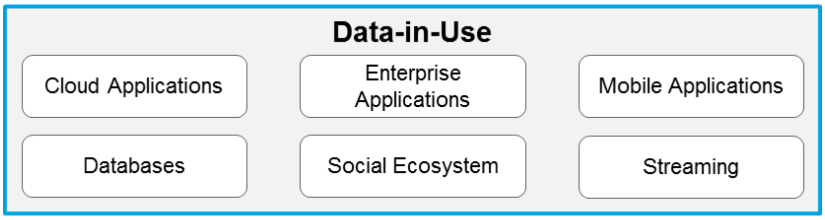


Figure 4. Data-in-use services

Key considerations for Data-in-use are:

- **Authentication/Identity Requirements:** To ensure right user is identified and authenticated to access the data/platform.
- **Authorization/Role based access:** On authentication, user/program should consider right persona and sufficient privileges to process/access/operate the data leading to focus on data visibility and it is usability.
- **Difficult to control the data after it is being accessed:** After users are authenticated and authorized; the user can take a photo and use for malicious purposes. These controls are difficult to put in place.
- **User code-of-conduct:** It is a challenge to apply user code of conduct considering human element in the public domain/digital ecosystem and remote usage environments.
- **Encrypting Memory:** Encrypting memory to be considered as it prevents data accessibility from malicious users but with the trade-off of performance due to additional activity of encryption and decryption.
- **Data Sharing:** Rights to download the documents or sharing, forwarding etc is being widely used and needs to be considered as part of data security requirement.

- **Automated approach to evaluate the security requirements:** Automated process to evaluate the security risks using suspicious activities in the network, diagnose potential threats and proactively improve the security.

Compromising of data-in-Use enables access to encrypted data-at-Rest and data-in-Motion, through accessing the private leys of data-at-Rest or data-at-Motion and thus manipulating the original content will lead into data security risks.

5. RELEVANT USE CASES

Relevant use case have been identified where data security service is traditionally used to prevent data leak. Since data security implementation have move forward from financial services to Digital document management, Digital Authority and Signature, Digital Identity solutions, Smart contract solutions for Supplychain services, Digital record management using Distributed ledger services, Decentralized network services for workflow management and security services, Data security as a service adoption to various industrial usecases have been emerging for the last five years across Healthcare, Manufacturing, Retail and Non-banking domain in Financial services like Claim processing and verification in Insurance, Investment management in Capital Markets to name a few.

There are various usecases in Healthcare and Medical service Industry being explored to use Blockchain platform. In addition, post COVID19 situation, Gartner research says that there are some key reasons for Blockchain based Healthcare application being evolved where data security is the primary concern as explained below:

- With disruption in Financial services, urgent transaction in cash liquidity is important to expedite business transactions such as purchase of medical devices, realization of purchase order etc.,
- Improve trust in supply chain processing and track assets with high transparency.
- Digital Document signing and asset management to be suitable for Insurance Claims processing.
- Rapidly processing Supply chain services in invoicing, purchase order approval, Zero touch payments and trustless party handling
- Using digital currency and crypto-currency for Supply chain financial activities.

Some of the popular usecase scenario where Data security as a Service can potential create higher impact to business risk management are discussed below.

5.1. Use case – Electronic Medical Record (EMR) Services

Electronic Medical Record (EMR) or Electronic Health Record (EHR) are very important to be handled safely and data operations with EMR is a costly operation and important for the Healthcare Industry. When a patient is admitted for medical emergency, understanding the patient’s medical history is very important and prima-focus step before starting the treatment and hence data handling with utmost speed with data privacy handling is always a demanding scenario in Healthcare industry.

Growing nations are handling these Medical Records in central repository which can be accessed across different hospitals. These kind of EMR has popularly demanding as it can be used for Hospitals for patient history and Insurance companies for claim verification.

Hence Data Security as a Service (DSaaS) is one of the most sought out solution approach for EMR with Blockchain based solution, where we can implement Smart contract based solution to enable alerts based on fluctuations in these readings so that hospitals can remotely take active measures to handle patients efficiently.

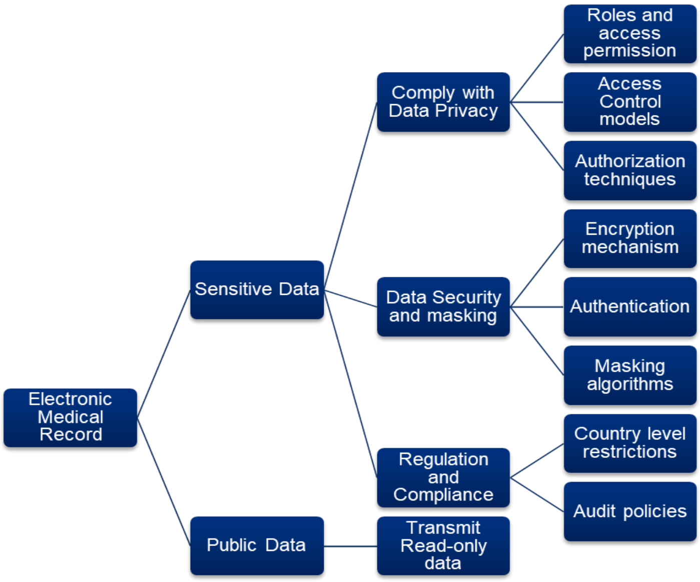


Figure 5. Data Security path for EMR

This can be achieved with a Blockchain based solution, where IoT transmission between hospitals across different locations can trigger events like search patient records, patient treatment history, doctor's advice during previous ailment etc., Based on these events, actions can be associated for automating the approval process in order to ensure the entire EMR transmission is automated completely without manual intervention and at the same time without compromising security in openly transmitting records to unauthorized parties using Blockchain security services.

5.2. Use case – Retail Supplychain Services

Implementing a Retail Supplychain solution involves handling people (governance), process (activities and information flow) and technology (using blockchain smart contracts to accelerate the workflow or using faster data encryption and decryption mechanism for both Data at rest and data in transit).

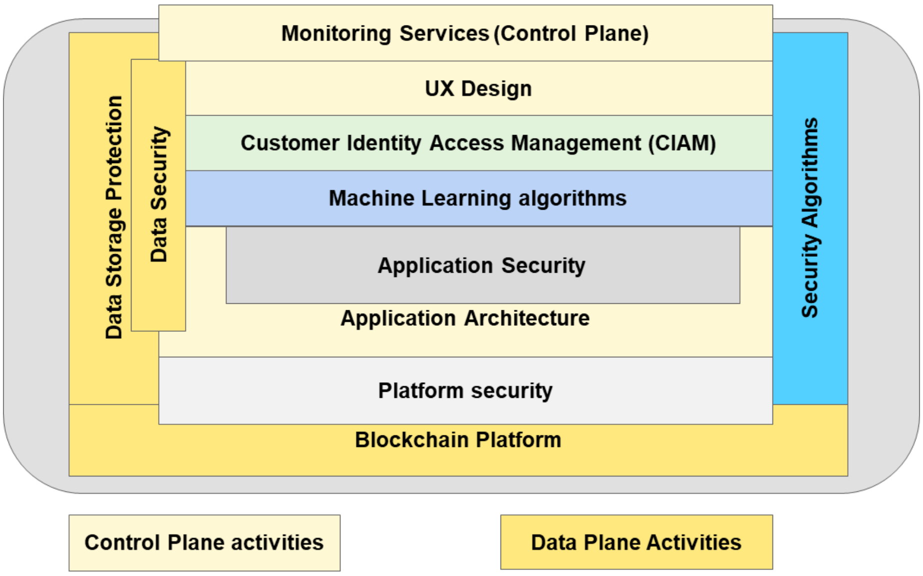


Figure 6. Data Security integrated in Retail Supplychain platform

Operational activities in a Blockchain platform are typically divided as Control Plane and Data Plane as shown in Figure 6. Control Plane is the driver which can be used to create and manage any cloud resources. Data plane is the operational activity which handles the capabilities of the resources created/managed by Control plane. These kind of resource management through Control plane and the Control plane helps to control and manage security services, auditing, policy driven activities, activity logs and resource hierarchy services.

During Retail Supplychain process, multiple parties are involved in purchase order, invoicing, payment approval, payment transfer etc., and hence in a traditional approach it takes considerable time to execute the entire Supplychain process. Using a Blockchain platform, helps to accelerate the process by involving multiple parties (node approver and owner) to execute the entire workflow operation. But Data Security could be a roadblock to address in such case. Hence implementing a Data Security as a Service platform integration with private Blockchain platform can help to develop a zero-trust security integration for the entire Supplychain process.

6. DSaaS Capabilities

Data security or protection can be provided by ensuring that a specific set of technology services and patterns are used in conjunction with blockchain to ensure that data breaches are prevented. The factors based on which this can be ascertained are:

- Data and documents are the most important asset that any enterprise application needs to protect. The level of protection to be enforced must be determined based on how confidential and sensitive the data asset is to the organization.
- Another factor that needs to be considered is the stage at which data security needs to be implemented. As described earlier data security mechanisms for data-at-rest will be little different from techniques required to protect data-in-use.
- Finally, irrespective of the data security mechanism used effective monitoring for proactively detecting data breach will always be required

In order to fulfill the enterprise grade data security requirements, DSaaS will require specific set of components and services as explained below.

6.1. Data Storage Protection Services

Data and document storage requirements require techniques that protect data-at-rest. Blockchain enabled data storage techniques would comprise of:

i) **Blockchain for Data Storage:** Data that does not contain confidential or personal information and is smaller in size (<10KB) can be directly stored on blockchain. The append-only ledger of blockchain ensures that data cannot be altered unless it undergoes the validation and consensus in-built into the blockchain network. Depending on the blockchain platform being used; there can be limitations on how data can be queried and may have to be mitigated by using an offchain read-only storage.

ii) **Blockchain for Proof of Existence:** Depending on the blockchain platform being used there can be limitations on how data can be queried. When data and/or documents are of larger size then traditional centralized storage mechanisms such as databases and filesystems must be used, which are susceptible to breaches as the content can be changed without getting detected. In such scenarios, storing the digital fingerprint of data asset as a one-way hash in blockchain is the recommended approach. This ensures that digital representation is always available on blockchain which can act as ir-refutable timestamped proof of the state of the data asset. A verification service will be provided so that the actual data or physical document can be verified against its proof recorded on blockchain. Additional metadata such as ownership, reference to the physical storage and access rights will also be maintained in blockchain for easy tracking of transaction.

iii) **Data Sharding:** For highly sensitive data / documents storing it at one central location can pose a high security risk. Therefore, sharding of data and documents into multiple parts and storing it on distributed nodes is recommended. Technology components like IPFS (Inter Planetary File System) provide a protocol and peer-to-peer network for storing data in a distributed file system. Each file is assigned with a unique identifier in a global namespace and then stored in a distributed manner, which can be reassembled on demand. The fragmented parts ensure that the document cannot be tampered with to protect the data.

iv) **Data Encryption:** Data-at-rest in production-grade systems is maintained in encrypted format. Symmetric encryption using industry standard algorithms (AES, DES) will be used for storing data on blockchain and when using sharding techniques to further strengthen data security. Asymmetric encryption will only be used for data storage when the data stored is intended to be consumed by specific stakeholder.

6.2. Data Usage Protection Services

There are several proven techniques for securing data-at-rest and there are several layers of security that can be implemented above the data / file storage layer. But protecting data-in-use is more complicated considering that its usage is spread across all the components of a solution. Some of the protection services recommended for DSaaS solution are:

i) **Secured Multi-Party Computation:** SMPC is a technique of distributed computing and cryptography which enables entities (individuals, applications or devices) to work with data while ensuring that the data and/or encryption keys are kept in a protected state. Multiple entities can participate in handling these confidential data. SMPC provides a new model for protecting data-in-use by strengthening the traditional security mechanisms. SMPC also helps in mitigating data residency issues by eliminating the single-point-of-failure risk because of the ability to “split” the confidential data or cryptographic key into multiple parts that can be re-assembled on-demand at runtime when a data transaction is executed. SMPC and blockchain are complementary technologies and it can be applied to all data lifecycle stages.

ii) **Decentralized PKI:** When data is being exchanged or used between different parties or systems, then its security is heavily reliant on PKI (Public Key Infrastructure) usage for authentication and asymmetric encryption of data. For production-grade systems, the public and private keys are supposed to be procured from a trusted CA (Certificate Authority). But these centralized CA are also susceptible to breaches which can lead to key theft and impersonation. Therefore, usage of Decentralized PKI on blockchain based framework like Web of Trust model is recommended approach for complete transparency and security. The different capabilities of the DPKI system that will be available on blockchain will include:

***ii-a) Identity Registration* –** Smart contract-based registration of UUID (user's unique identifier) and their public key. The rules governing registration and renewal of identifiers will be transparently maintained on blockchain.

***ii-b) User-controlled key generation* –** For the registered identity, the generation of public-private keypair will be initiated by the user. Private keys must be generated in a decentralized manner under user's control. User may authorize an agent to manage keys on their behalf.

ii-c) Master key and sub-keys – Each user or entity whose identity is registered will be assigned a master private key and sub-keys from the master key. The sub-keys will be maintained as that identity's metadata and used for transactions signing related to that identifier, whereas the master key controlled by the user will be used to manage the identity's metadata. To ensure that master key is secured and can be recovered, SMPC (Secured Multiparty Computation) will be used to shard and store parts of the master key.

ii-d) Public key availability – All ledger participants will get automatic access to the identities public key which can be used for asymmetric data encryption

ii-e) Identity Revocation – Identity can be deleted or revoked only by the identity holder or via a workflow process that required multiple entities to provide their approval

iii) **Blockchain based ACL:** Access Control Lists or registries managed via smart contracts on blockchain will provide security rules to be enforced by last-mile security agents. The advantage of maintaining these rules on blockchain is that any change to the access levels for data or transaction type will always be available on blockchain as an audit trail. This technique also ensures that data leakage is prevented at the blockchain layer because smart contract-based authorization controls who can read or write to blockchain.

iv) **Multi-Signature Accounts**: Critical, high value transactions will be protected by ensuring that multiple signatures are required before the transaction is accepted and processed further in blockchain. This technique can be used for both onchain and offchain data. In case of offchain data, the smart contract events will be triggered when the required signatures are in place based on which the offchain execution can be controlled.

v) **Smart Contract Oracles**: Since smart contracts are not allowed to directly communicate with systems external to blockchain which is required to ensure that the smart contract output is deterministic. Therefore, smart contract oracles are required to supply the external data to blockchain. Data security rules implemented on blockchain can require validations from an external system. This can be enabled by initiating the Oracles component that listens to smart contract events and automatically fetches external data and supplies it to smart contract. This pattern ensures that smart contracts can obtain external data on-demand and make informed decisions.

vi) **Blockchain based UEBA**: User and entity behaviour analytics or UEBA arose out of the malicious behaviour by users and other entities. Self-sovereign Identity (SSI) based Blockchain trust networks have a critical role in ensuring that UEBA can be done responsibly keeping user privacy as top priority. UEBA uses machine learning and algorithms to strengthen security by monitoring users and other entities, detecting anomalies in behaviour patterns that could be indicative of a threat. A SSI based Blockchain network can be used to present proof of normal behaviour of user rather than collect raw user data in a UEBA. This is a proactive approach to security and gaining visibility into user and entity behaviour without compromising on user's privacy at any point in time.

7. Reference Architecture

DSaaS (Data Security as a Service) architecture will comprise of all the components and services mentioned in previous section in a layered architecture. The services can be leveraged by any kind of solution irrespective of whether blockchain is used not. As shown in Figure-1, the DSaaS Architecture comprises of well-defined microservices and underlying components.

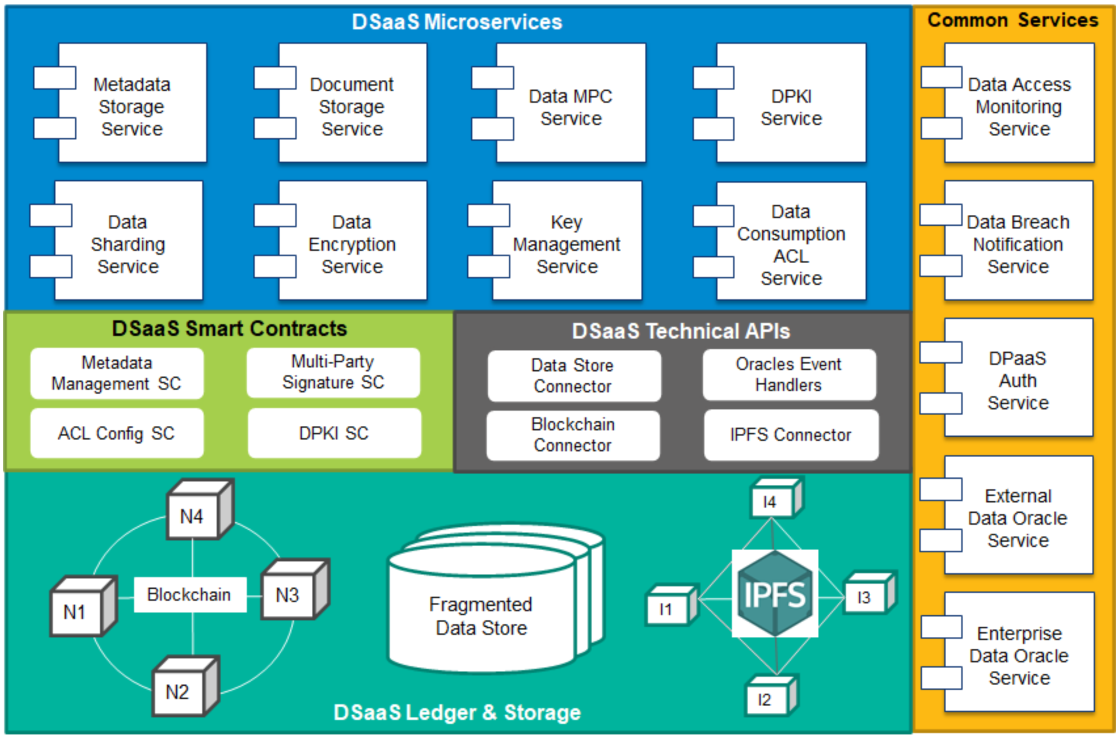


Figure 7. DSaaS Reference Architecture

The reference architecture comprises of the following of key building blocks that are segregated into different buckets.

7.1. DSaaS Micro Services

Microservices architecture will be followed to build APIs with the right granularity and cohesive functionality and the components in the architecture are:

Metadata Storage Services: These are generic APIs which will provide ability to the client application to store metadata related to their assets and process in blockchain so that data protected at rest with complete provenance and audit trail. This can be used to store small sized data, or it can be used to store the reference and digital representation of large sized data or document

Document Management Service: For protecting large sized documents this API can be used to store the physical document on IPFS and its metadata stored on blockchain. Documents can either be uploaded as an attachment or a shared folder be specified for upload of very large documents.

Data Sharding Service: Highly sensitive large sized data and/or digital assets will be protected by using this API that performs data fragmentation and stores in an encrypted token form in the file system. The metadata and its associated access rights are maintained on blockchain using which the data is reconstructed by decoding and combining all fragments as shown in Figure-8 below.

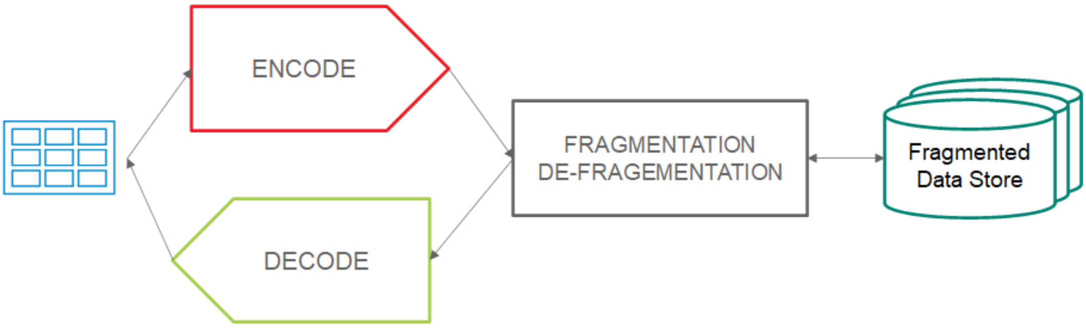


Figure 8. Data Sharding / Fragmentation

Data Encryption Service: This API will provide options of using symmetric encryption algorithms, which can be used to encrypt / decrypt data before storing it on blockchain or in the fragmented file store.

Data Multi-party Commutation (MPC) Service: This set of APIs enable the usage of multi-party computation capabilities ranging from enrolment of entities that would participate in the computation process to low level cryptographic functions as well as interactive and non-interactive functions required for effectively using MPC to protect the data-in-use.

DPKI Service: This set of APIs will help client applications to leverage the Decentralized PKI capabilities which involve identity registration, key generation, revocation etc. by leveraging the corresponding smart contracts deployed on blockchain.

Key Management Service: Generic APIs which will allows other microservices and client applications to be able to securely generate public-private key infrastructure using blockchain.

Data Consumption ACL Service: This API will allow setting up of registry with roles and permissions on blockchain that would be used for authorizing access to different types of data based on its sensitivity, confidentiality etc.

Common Services: APIs which will cover the cross-cutting concerns that cover monitoring, notifications, authentication / authorization as well as Oracles functionality to connect to external and enterprise data sources.

7.2. DSaaS Smart Contracts

The smart contacts layer represents the storage and logic deployed on blockchain which is required to provide functionality for data storage functionality and advanced functionality related to SMPC, DPKI microservices. Each smart contract will comprise of multiple functions with the appropriate role-based authorization checks.

7.3. DSaaS Technical APIs

This set of APIs are fine-grained technical APIs which are required for interacting with blockchain, IPFS and other data stores. This will also include the smart contract event handlers which will trigger the retrieval of external data from Oracles.

7.4. Blockchain

This is a distributed ledger network that forms the core of the DSaaS offering. A minimum viable number of nodes will be required to maintain this blockchain network to ensure high availability. A permissioned blockchain platform will be used so that it provides the required smart contract functionality as well as higher performance.

7.5. IPFS (Inter Planetary File System)

This component will provide a decentralized network to store and maintain the physical documents. Each document stored in IPFS will be allocated a unique hash identifier which will be maintained on blockchain.

7.6. Fragmented Data Store

This component will comprise of multiple logical and physical partitions in which the sharded data will be stored in encrypted format.

8. DSAAS ENABLED IMPLEMENTATIONS

The previous section describes the reference architecture comprising of the building blocks of a DSaaS framework. It is pertinent to outline how this reference architecture can be used to fulfil the data security requirements of the use cases described above. The actual realization of the architecture may vary based on the functional architecture, so the best fit DSaaS component is outlined in this section.

In the Electronic Medical Records (EMR) user scenario, some of the key data protection services required for this use case are:

Metadata Storage Service: Protecting the PII (Personally Identifiable Information) data of patients is one of the key requirements so health records against the patient's UUID on blockchain. For highest level of data security, the data fragmentation services to shard and store patient's health records can be leveraged.

DPKI Service: Identities related to patients, health care providers so that public key infrastructure is generated and distributed securely to each user.

Data Consumption ACL Service: Each actor in the EMR use case will have specific data access rights. Therefore, a default ACL limiting access to ledger data will be setup. Additionally, dynamic update of ACL will also be enabled for scenarios where patient provides their consent and agrees to share their data with a particular health care provider.

Document Management Service: The detailed medical comprising of pathology tests, X-ray, CT-scan reports etc. will be maintained by the health care provider's in electronic format. If physical access to the document needs to be made available, then the document storage services can be used for maintaining such documents in IPFS. If the documents are centrally managed, then their unique digital fingerprint will be maintained on blockchain via the metadata storage services.

In this use case, we design a solution which is PII compliant and uses Data storage protection services leveraging blockchain to provide DSaaS capabilities with zero-trust security platform for integrated solution.

In the Retail Supply chain User scenario, Supply chain is a vast area and hence the data security requirements also vary depending on the functionality being addressed. Some of the key data protection services required for this use case are:

DPKI & ACL Services: The identity and access management requirements for this retail-centric use case will require these components.

Data MPC Service: Large orders and high value payments will leverage the MPC services to ensure that the transaction signing key is sharded and secured to prevent fraudulent transactions by obtaining multi-party confirmation.

Data Encryption Service: Confidential data transactions in supply chain such as trading agreement, purchase order management, invoice generation and settlement etc. will leverage the encryption services to ensure that the data can be read and processed by the authorized recipient only.

Data Oracle Services: Supply chain logistics and financial transactions may require data for validation or processing purposes from external sources such as retrieving the latest forex rates etc. Similar order and product management will require validations from data quality perspective from the enterprise business applications. This can be achieved by using data oracles services to interact with external and/or enterprise data sources.

In this use case, we develop a solution which uses Customer Identity Access Management (CIAM) like PingIdentity or Gigya and other pluggable interface components of DSaaS reference architecture like microservices, connectors and API services for integrated Retail Supplychain solution design.

9. FUTURE SCOPE OF WORK

Data security as a Service can be enhanced in future as cloud agnostic solution and blockchain agnostic approach to be used for any cloud platforms and using any blockchain platforms to enhance its features with cloud native service integration for better solution approach for enterprise application design.

Also, these features can be developed in future using Microservices architecture along with polyglot database for better flexibility to integrate with multiple application and reuse the asynchronous communication services for internal and external application integration.

10. CONCLUSIONS

Data Security is the key to strengthen the Enterprise architecture when handling workflow operation involving multiple parties. For various industries like Fintech (Insurance, Payment, Investment banking) and Healthcare (EMR/EHR, Medical Retail Supplychain operations), it is important that an efficient pluggable DSaaS integration is incorporated which can help in business agility, cost efficiency and improved Governance and security compliance.

As shown in the reference architecture of DSaaS solution, there are many pluggable components which can be used to integrate a business agile solution for integrated Enterprise security to enable industry level compliance like FedRamp, PII, TOSCA, PCI or HIPAA compliance service functions.

REFERENCES

[1] Subashini, Subashini, and Veeraruna Kavitha. "A survey on security issues in service delivery models of cloud computing." Journal of network and computer applications 34.1 (2011): 1-11.

[2] Chen, Deyan, and Hong Zhao. "Data security and privacy protection issues in cloud computing." 2012 International Conference on Computer Science and Electronics Engineering. Vol. 1. IEEE, 2012.

[3] Getov, Vladimir. "Security as a service in smart clouds--opportunities and concerns." 2012 IEEE 36th Annual Computer Software and Applications Conference. IEEE, 2012.

[4] Kumar, P. Ravi, P. Herbert Raj, and P. Jelciana. "Exploring data security issues and solutions in cloud computing." Procedia Computer Science 125 (2018): 691-697.

[5] Mohamed, Eman M., Hatem S. Abdelkader, and Sherif El-Etriby. "Enhanced data security model for cloud computing." 2012 8th International Conference on Informatics and Systems (INFOS). IEEE, 2012.

[6] Varadharajan, Vijay, and Udaya Tupakula. "Security as a service model for cloud environment." IEEE Transactions on network and Service management 11.1 (2014): 60-75.

[7] Hussain, Mohammed, and Hanady Abdulsalam. "SECaaS: security as a service for cloud-based applications." Proceedings of the Second Kuwait Conference on e-Services and e-Systems. 2011.

AUTHORS

Dr. Magesh is a Distinguished Member of Technical Staff at Wipro. Magesh holds a Ph.D in Deep Learning and Genetic Algorithms. He is a senior member of IEEE and has published more than 50 articles in OpenSource For You, PC Quest, Cutter Business IT Journal and other notable international journals. He has also published around 480 thought leadership articles on AIML, Blockchain, and Cloud on LinkedIn with the hashtag #shorticle.

Vinod Panicker is a Chief Architect & Distinguished Member of Technical Staff with over 19+ years of software development experience. Vinod currently leads the Blockchain initiative for the Cybersecurity & Risk Group at Wipro. He is an expert in open source and crowd sourcing platforms. He was the Lead architect for Wipro's inner sourcing platform, Open connect and helped it scale seamlessly to over 30K users.

Hitarshi has 20+ years of experience in IT architecture, consulting, design and implementation using blockchain, API, SOA, BPM and Java/J2EE technologies. He has experience in IT transformation and modernization initiatives and has provided enterprise-wide SOA-based solutions.

In his current role, at Wipro Technologies as a Chief Architect in Service Transformation at Wipro, he leads the Center of Excellence initiatives as part of the Blockchain practice. His charter involves applied research and building technology assets around blockchain protocols such as Hyperledger, Quorum, Besu, Corda, Multichain, Hedera etc.

Krishna Mty is a Lead Architect and Distinguished Member of Technical Staff in Wipro and part of Integrated Digital Engineering and Application Services.

A NOVEL REGIONAL FUSION NETWORK FOR 3D OBJECT DETECTION BASED ON RGB IMAGES AND POINT CLOUDS

Hung-Hao Chen[1], Chia-Hung Wang[1], Hsueh-Wei Chen[1],
Pei-Yung Hsiao[2], Li-Chen Fu[1] and Yi-Feng Su[3]

[1]Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
[2]Department of Electrical Engineering,
National University of Kaohsiung, Kaohsiung, Taiwan
[3]Research and Development Division,
Automotive Research and Testing Center (ARTC), Changhua, Taiwan

ABSTRACT

The current fusion-based methods transform LiDAR data into bird's eye view (BEV) representations or 3D voxel, leading to information loss and heavy computation cost of 3D convolution. In contrast, we directly consume raw point clouds and perform fusion between two modalities. We employ the concept of region proposal network to generate proposals from two streams, respectively. In order to make two sensors compensate the weakness of each other, we utilize the calibration parameters to project proposals from one stream onto the other. With the proposed multi-scale feature aggregation module, we are able to combine the extracted region-of-interest-level (RoI-level) features of RGB stream from different receptive fields, resulting in fertilizing feature richness. Experiments on KITTI dataset show that our proposed network outperforms other fusion-based methods with meaningful improvements as compared to 3D object detection methods under challenging setting.

KEYWORDS

Machine Learning, 3D Object Detection, Data Fusion, Autonomous Driving.

1. INTRODUCTION

Owing to the speedy development of computer vision technologies, more and more companies have started to invest in and invent intelligent vehicles. Therefore, autonomous driving has become a popular issue nowadays. The most essential property of autonomous driving is to perceive the surroundings of vehicles and provide safety for drivers. Accordingly, the key to this property is object detection. Over the past few years, there has been many successful 2D object detection approaches proposed, such as Faster R-CNN [1] and RetineNet [2]. However, 2D object detection is unable to provide sufficient ability of perceptions in comparison with 3D object detection because 2D object detection lacks the information of depth and the knowledge of orientation. The depth can hint that the distance of the object is too close, and the orientation is capable of knowing whether the object is in the same direction as the vehicle. With the help of 3D detection, the intelligent vehicles are able to make precise decisions under different situations. In order to detect on-road objects, most of the intelligent vehicles are equipped with multiple

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 145-160, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111812

sensors such as RGB cameras and LiDARs. Thus, various 3D object detectors based on these sensors are proposed.

Some image-based approaches were presented to utilize monocular [3,4] or stereo images [5,6] to better obtain 3D information of objects. RGB images are good at providing color information and detailed contours of front view. Nevertheless, they still suffer from the limitation of insufficient depth information.

On the contrary, LiDAR-based methods were also proposed to explore the use of 3D LiDAR points. In comparison with RGB images, LiDAR points offer accurate depth information that can be leveraged to localize the objects in the 3D space. Some works [7,8,9] transformed 3D point clouds into 2D bird's eye view (BEV) images or 2D front view images and performed typical convolutional operations to obtain the latent features. Other methods [10] voxelized the 3D point clouds and applied 3D convolution on the generated voxels. However, LiDAR-based methods suffer from sparse observations especially at long range.

To compensate the disadvantages of two sensors, we present Regional Fusion network for 3D object detection (RF3D), which is a fusion-based framework that leverages both cameras and LiDARs jointly. We generate region proposals from both streams, respectively, rather than from LiDAR stream only. Thus, proposals can be projected from one stream to the other stream mutually. In this way, we can fuse the features in deeper layers for better refinement, taking advantage of two sensors and predicting accurate estimations. Furthermore, the presented multi-scale feature aggregation module makes use of different levels of RGB features to obtain low-level contents and high-level semantic meanings simultaneously. With the help of proposed regional fusion layer, the fusion between two streams of feature maps from different sensors is conducted in RoI-level, avoiding cascading redundant feature-level fusion. To verify the effectiveness of RF3D, we conduct several experiment on KITTI Vision Benchmark [11]. The experimental results manifest that our network outperforms other methods under hard difficulty in 3D detection.

In this paper, we design a Multi-Scale Feature Aggregation module (MSFA) with upsampling and downsampling layers to aggregate features from different receptive fields. For two stream fusion, we propose Regional Fusion Layer to fuse point clouds and RGB images based on the RoI estimated in the first stage. Based on above methods, we present a novel two-stream deep architecture for 3D detection, Regional Fusion Network (RF3D), that simultaneously conducts on both point clouds and RGB images in a fusion way for autonomous driving.

The rest of the paper is organized as follows. The overview including image-based, LiDAR-based and fusion-based approaches are introduced in section 2. Then, we define the problem formulation in section 3. In section 4, we present the overall architecture and details in our proposed RF3D. The experimental results on KITTI dataset are shown in section 5. In section 6, we conclude the paper and give directions for future improvement.

2. Related Work

3D object detection is very necessary for intelligent transportation systems. Recently, many works on this topic have gradually emerged. After reviewing the existing approaches of 3D object detection, we categorize these approaches into three groups, namely, image-based approaches, LiDAR-based approaches and fusion-based approaches. To sum up, they are divided according to the inputs.

2.1. Image-based Approaches

Using RGB images to infer accurate 2D bounding boxes of objects is no longer difficult for many state-of-the-art methods since RGB images can provide texture and color information in the form of pixel-wise intensity. Also, there are many works that utilize RGB images to predict 3D bounding boxes of objects. MonoFENet [3] used monocular image to additionally generate the disparity map to enhance the extracted features. D4LCN [4] firstly generated the depth map using the monocular image, and then the depth-guided filtering module was utilized to fuse features of image stream and depth stream. DSGN [5] detected 3D objects on a differentiable volumetric representation that effectively encoded 3D geometric structure for 3D regular space. Disp R-CNN [6] predicted disparity only for pixels on objects of interest and learned a category-specific shape prior for more accurate disparity estimation. However, these image-based methods suffer from the inherent difficulties of estimating depth from images and as a result perform poorly in 3D localization.

2.2. LiDAR-based Approaches

Unlike RGB images, point clouds collected by LiDARs are unordered and discrete. As a result, raw point clouds cannot serve as the inputs of the convolutional layer. Pixor++ [7] and Pointpillars [9] firstly transformed the 3D point clouds into the 2D BEV images, and utilized a 2D CNN to learn the point cloud features for the 3D bounding boxes generation. VoxelNet [12] grouped the point clouds into the voxels and used a 3D CNN to learn the features of the voxels to generate the 3D bounding boxes. However, the BEV projection and voxelization process suffered from the information loss due to the data quantization. Moreover, the 3D CNN was both memory and computation inefficient. On the other hand, PointRCNN [13] directly learned point-wise features and generated 3D bounding boxes from raw point clouds and utilized ground-truth augmentation to gain significant improvements. TANet [14] jointly used channel-wise, voxel-wise, and point-wise attention to alleviate the impact of noisy points. Although depth measurements provided by LiDARs are useful for localizing the 3D bounding boxes of objects, the observations are usually sparse especially at long range.

2.3. Fusion-based Approaches

Since the fusion mechanism between RGB images and LiDAR point clouds remains an open problem nowadays, there are only few approaches that take both RGB images and LiDAR point clouds as inputs. AVOD-FPN [15] applied a 2D convolutional network on both RGB images and LiDAR BEV representations, and fused them at the intermediate region-wise convolutional feature map via feature concatenation. Frustum PointNet [16] utilized mature 2D object detection to firstly generate the 2D region proposals based on the RGB images, and lifted the proposals to the 3D frustums. Then, the points inside the 3D frustums were used to infer the 3D bounding boxes. However, the 2D object detection was the bottleneck. PointPainting [17] designed a painted version of PointRCNN [13] by appending the class score from image to each point. PI-RCNN [18] proposed an Attentive Cont-conv Fusion (PACF) module to fuse point and image features. MMF [19] used a joint model to do four tasks, and each task could benefit from other tasks. ContFuse [20] performed one-way fusion to fuse the feature maps of the RGB images to the BEV feature maps, and CrossFusion [21] utilized the spatial relationship between the BEV features and RGB features to perform two-way fusion. Both ContFuse and CrossFusion applied hierarchical feature-level fusion, which was time-consuming and redundant. Besides, none of the aforementioned methods directly use raw point clouds to perform fusion. Consequently, the information may be lost during the process of data quantization.

In this work, we aim to propose a fusion-based 3D object detection network that exploits the use of raw point clouds and RGB images. In addition, our presented network generates rough proposals from two streams, respectively, and the network fuses two inputs in RoI-level, which avoids extra computation cost on the fusion of non-interest regions. To increase the richness of RGB features, our presented multi-scale feature aggregation module further provides the RGB features with richer information from different features that are with various receptive fields.

3. PROBLEM FORMULATION

We present a deep learning network that aims to solve the task of 3D object detection consuming the inputs of RGB images and LiDAR point clouds. Firstly, an RGB image can be regarded as a set of integer pixel values I_{RGB}, where $I_{RGB} = \{v_{ij} | 1 \leq i \leq W, 1 \leq j \leq H\}$, W denotes the width and H symbolizes the height of the image. Each element v_{ij} in the image is an integer with the range of $[0, 255]$. On the contrary, a LiDAR point cloud can be represented as a set of discrete points I_{LiDAR}, where $I_{LiDAR} = \{P_s | s = 1,2, \ldots, N\}$ and N stands for the number of points in a point cloud. Note that N might vary among different collected frames. Additionally, each point P_s can be parameterized into a four-dimensional tensor (x_s, y_s, z_s, r_s), where $(x_s, y_s, z_s)$is the coordinate with regard to the origin of coordinate system and $r_s$denotes the reflectiveness of the point P_s.

Given RGB images I_{RGB} and LiDAR point clouds I_{LiDAR}, our goal is to predict accurate 3D detection that contains both localization and classification information. In general, the outputs of the 3D object detection are represented as a set of 3D bounding boxes O_{box}, where $O_{box} = \{B_k | k = 1,2, \ldots M\}$ and M symbolizes the number of predicted 3D bounding boxes. Furthermore, each 3D bounding box B_k is composed of an eight-dimensional tensor $(x_k, y_k, z_k, w_k, h_k, l_k, \theta_k, cls_k)$, where $(x_k, y_k, z_k)$is the localization information that denotes the center coordinate of the bounding box with respect to the coordinate of the LiDAR and (w_k, h_k, l_k)represents the size of the bounding box. In the typical 3D on-road object detection, there only exists yaw rotation along with the axis perpendicular to the ground which is denoted as θ_k. Last but not the least, the classification information is represented as cls_k, indicating the category that the bounding box belongs to.

To sum up, the entire formula for the 3D object detection task T_{det} can be denoted as

$$T_{det}(I_{RGB}, I_{LiDAR}) = O_{box} = \{B_k | k = 1,2, \ldots M\} \quad (1)$$

The goal is to propose a 3D detection network that can generate accurate 3D bounding boxes O_{box}based on RGB images I_{RGB} and LiDAR point clouds I_{LiDAR}.

4. REGIONAL FUSION NETWORK

The architecture of RF3D is shown as Figure 1. Our proposed method is composed of five major components including (1) backbone for retrieving latent features, (2) mutual projection for projecting proposals from LiDAR to RGB stream and the reverse, (3) multi-scale feature aggregation module for generating rich RGB features in different scales, and (4) regional fusion layer for performing RoI-level fusion between two input sources.

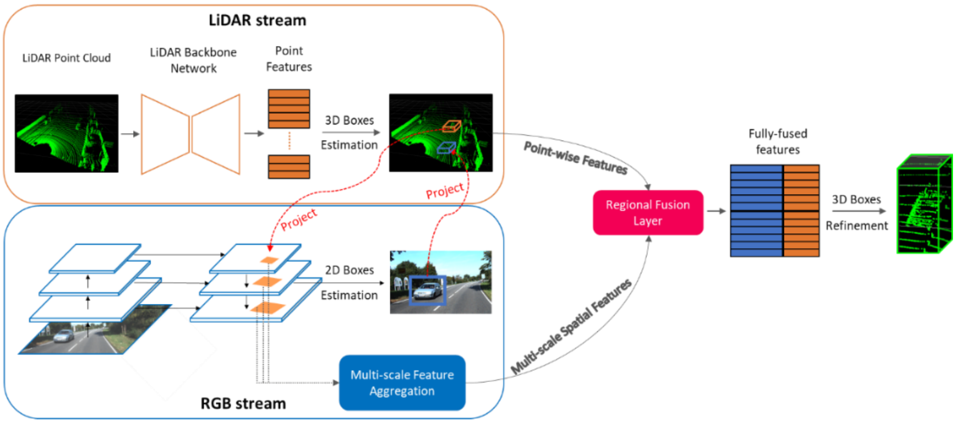


Figure 1. Overview of the proposed Regional Fusion Network

4.1. Backbone Network

The backbone networks aim to obtain discriminative features and generate 2D proposals from RGB images and 3D proposals from LiDAR point clouds, respectively. In order to perform fusion between RGB images and LiDAR point clouds, there exists two streams in our network. One stream is for RGB images and the other is for LiDAR point clouds. However, the discrete and unordered data format of point clouds is very different from pure images that we are not able to apply conventional convolutional operation on the point clouds. Consequently, we utilize separate backbone networks for RGB stream and LiDAR stream.

4.1.1. LiDAR Stream

We utilize PointNet++ [22] as the backbone network, as shown in Figure 2, for the LiDAR stream due to its capabilities of handling unordered issue and learning point-wise features of point clouds. Specifically, we employ four sets of abstraction layers as well as multi-scale grouping that are utilized to subsample original 16,384 points into regions with sizes of 4096, 1024, 256 and 64, respectively. Then, the feature propagation layer is used to obtain the point-wise features for the 3D proposal generation and fusion.

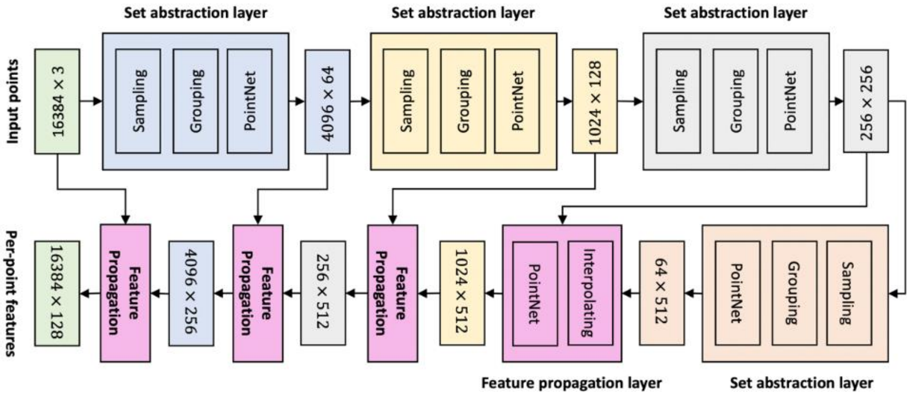


Figure 2. The LiDAR stream backbone PointNet++

4.1.2. RGB Stream

We apply ResNet-50 [23] combined with a feature pyramid network (FPN) [24] as shown in Figure 3. It augments a standard convolutional network using lateral connections and atop-down pathway so as to obtain rich multi-scale feature maps from a single resolution input image. We exploit the feature maps C_2, C_3 and C_4 of ResNet-50 having scales of $1/4$, $1/8$ and $1/16$ to build the feature pyramid. Consequently, the resultant feature pyramid is leveraged to generate 2D proposals from RGB images and provides multi-scale feature maps for multi-scale feature aggregation module.

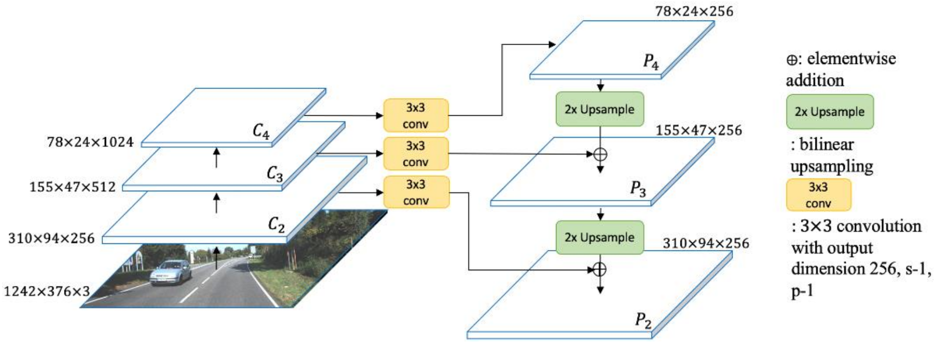


Figure 3. The RGB stream backbone ResNet-50-FPN

4.2. Mutual Projection

As aforementioned, LiDARs and RGB cameras have their own disadvantages. LiDARs possess sparse observations at long range while RGB cameras have limited usage in nighttime, cloudy and rainy weather. Some objects might be detected in one stream while they cannot be captured in the other stream. In order to perform regional fusion and make two sensors benefit each other, we have to obtain an object in both LiDAR stream and RGB stream. Therefore, we project the proposals, which are estimated in the backbone networks, from one stream onto the other. To be more specific, 2D proposals from the RGB stream are projected onto the 3D LiDAR coordinate system and 3D proposals from the LiDAR stream are projected onto the 2D image coordinate system as well, as depicted in Figure 4.

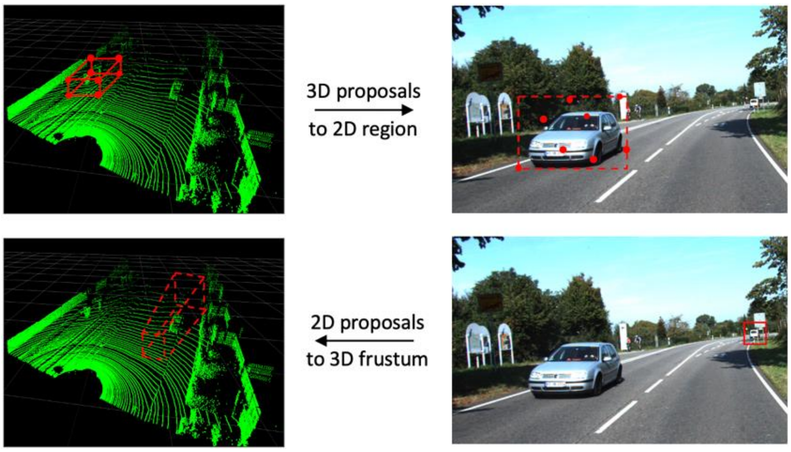


Figure 4. Illustration of mutual projection. The first row showing a 3D bounding box generated from the LiDAR stream is projected onto the RGB image. On the other hand, the second row demonstrating a 2D bounding box predicted from the RGB stream is lifted to a 3D frustum with near and far planes.

4.2.1. 3D Proposals to Images

Given a 3D proposal whose coordinates of eight corners are $\{C_n^{LiDAR} | n = 1,2,\ldots,8\}$, where $C_n^{LiDAR} = (x^{LiDAR}, y^{LiDAR}, z^{LiDAR})$ and x^{LiDAR}, y^{LiDAR}, z^{LiDAR} represent the coordinates in the LiDAR coordinate system, we can utilize the calibration matrix to project each point C^{LiDAR} to the image coordinate system and generate corresponding eight points $\{(u_n^{RGB}, v_n^{RGB}) | n = 1,2,\ldots,8\}$, where $C_n^{RGB} = (u_n^{RGB}, v_n^{RGB})$ and u^{RGB}, v^{RGB} symbolize the coordinates in the image coordinate system. The calibration matrix is pre-determined, and the entire projecting process can be performed through matrix multiplication.

After obtaining eight corners in the image view, we find the tightest 2D bounding box that can bound all eight corners as the corresponding projected 2D proposals. Hence, RGB features inside 2D bounding boxes are utilized to conduct RoI-level fusion in proposed regional fusion layer.

4.2.2. 2D Proposals to Point Clouds

A 2D proposal which is in the image coordinate system can be lifted to a frustum. A frustum is constructed with two planes which are near-plane and far-plane in the LiDAR coordinate system as shown in Figure 5. The near one is generated with smaller predefined depth d_{near} and the far one is obtained from predefined larger depth d_{far}. As a result, a 3D frustum is generated through connecting these two planes. However, there might be some points that do not belong to the object detected from the RGB stream. Inspired by Frustum PointNet [16], we only select the 3D points whose confidence scores generated in the backbone are greater than the predefined threshold in the frustum. Therefore, those selected points and original 2D proposals are fed into the presented regional fusion layer to conduct RoI-level fusion so as to make two sensors benefit each other.

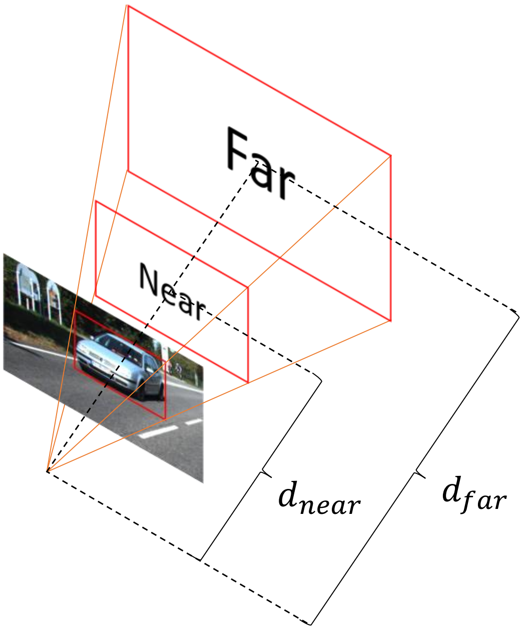


Figure 5. Back projection from the 2D proposals to the 3D space

4.3. Multi-scale Feature Aggregation Module

Features are the most demanding components for the network to generate high-quality predictions. In general, a CNN comprises a number of convolutional layers to extract discriminative features of images. In addition, convolutional layers that are located in different levels can generate various kinds of features. Low-level features are more content descriptive. Besides, the receptive field of the low-level layer is relatively small so that the information of small-size objects can be preserved well. On the other hand, deep high-level layers usually generate class-specific features having more semantic meanings. Since the receptive field of the high-level layer is large, some knowledge of small-sized objects might lose, leaving only global information. In order to perform RoI-level fusion and localize the objects precisely, we have to keep information from multiple receptive fields together as shown in Figure 6. Therefore, we rely not only on low-level features that indicate the appearances of objects but also high-level features that give the semantic meanings of objects.

Accordingly, the 2D proposals from the RGB stream and projected 2D proposals from the LiDAR stream are generated through the backbone networks. Both of them are 2D bounding boxes in nature. In order to generate high-quality RGB features that contain high-level and low-level information simultaneously, we aggregate features from multiple receptive fields. Inspired by Mask R-CNN [25], given the bounding box and multi-scale feature maps P_2, P_3 and P_4, we apply RoIAlign to extract the corresponding features and pool the feature maps into the sizes of 16×16, 8×8 and 4×4 based on different receptive fields, respectively. After that, we utilize upsampling operation as well as downsampling operation to resize the features and aggregate them together. Hence, the multi-scale spatial features containing both high-level semantic meanings and low-level geometric information are generated. Then, these features representing potential foreground objects are used to perform RoI-level with the point-wise features in the proposed regional fusion layer.

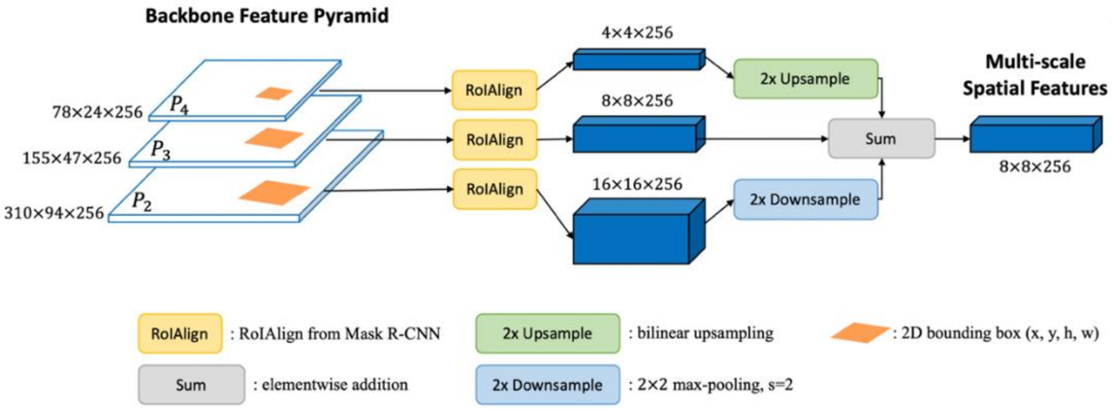

Backbone Feature Pyramid
78×24×256
P_4
155×47×256
P_3
310×94×256
P_2
RoIAlign
RoIAlign
RoIAlign
4×4×256
8×8×256
16×16×256
2x Upsample
2x Downsample
Sum
Multi-scale Spatial Features
8×8×256
RoIAlign : RoIAlign from Mask R-CNN
2x Upsample : bilinear upsampling
: 2D bounding box (x, y, h, w)
Sum : elementwise addition
2x Downsample : 2×2 max-pooling, s=2

Figure 6. Architecture of multi-scale feature aggregation module

4.4. Regional Fusion Layer

Generally, RGB images provide rich color information of objects while LiDAR point clouds have fine-grained 3D structures. Each kind of data has its own superiority. In order to obtain high-quality detection results, fusion between RGB images and LiDAR point clouds is inevitable. Besides, the spatial relationship between RGB images and LiDAR point clouds is necessary to reason the fusion process. Without utilizing the spatial relationship between two sources, the fusion process may result in severe errors and lack the ability to learn representative fused features. With a known calibration projection matrix, the projection from a point cloud to an RGB image can be completed. Each point of point clouds in the 3D space is related to a pixel in an image. This one-to-one correspondence can be utilized to fuse the data and supply each point feature with additional information from the RGB stream.

As illustrated in Figure 7, the proposed regional fusion layer leverages spatial features from the RGB stream and the point-wise features from the LiDAR stream to conduct the data fusion between two sources. For each proposal generated in the backbone network, our main purpose is to associate its point-wise features with pixel-wise RGB features so as to increase the feature richness of the LiDAR features for the box refinement. As a matter of fact, we choose to enrich LiDAR features because they are more suitable for performing 3D object detection than the RGB features.

At the first step of regional fusion layer, we apply 1×1 convolution on the spatial features and resize the spatial features along with the height and width dimension on the RGB feature map F_{RGB}, where $F_{RGB} \in \mathbb{R}^{H \times W \times C}$. The transformed resized RGB features are denoted as F'_{RGB}, where $F'_{RGB} \in \mathbb{R}^{HW \times C}$, and HW stands for the number of pixels and C represents the number of channels. Then, we apply attention mechanism to find the correspondence between RGB features F'_{RGB} and LiDAR point-wise features F_{LiDAR}, where $F_{LiDAR} \in \mathbb{R}^{N \times C}$, and N denotes the number of sampled points in the proposal and C symbolizes the channels. In our experiments, the number of sampled points in each 3D proposal is set 512. In the procedure of attention mechanism, we first calculate the attention scores M, whose formula is defined as

$$M = Softmax\left(F_{LiDAR} \times F'^{T}_{RGB}\right) \tag{2}$$

where the superscript T represents the transpose matrix and $M \in \mathbb{R}^{N \times HW}$. In addition, the softmax function is applied along each row in the matrix. As a result, each row vector in M, representing the importance scores of pixels contributing to each LiDAR point, is set as the size of $1 \times HW$. After obtaining the attention scores M, we use the matrix M to calculate the weighted summation of pixel-level RGB features with respect to each LiDAR point, whose formula is defined as

$$F_{attenRGB} = M \times F'_{RGB} \tag{3}$$

where $F_{attenRGB} \in \mathbb{R}^{N \times C}$, and $F_{attenRGB}$ serves as the additional RGB information for the point-wise features. Finally, we concatenate F_{LiDAR} and $F_{attenRGB}$ together and generate fully fused features, which can be utilized for the box refinement.

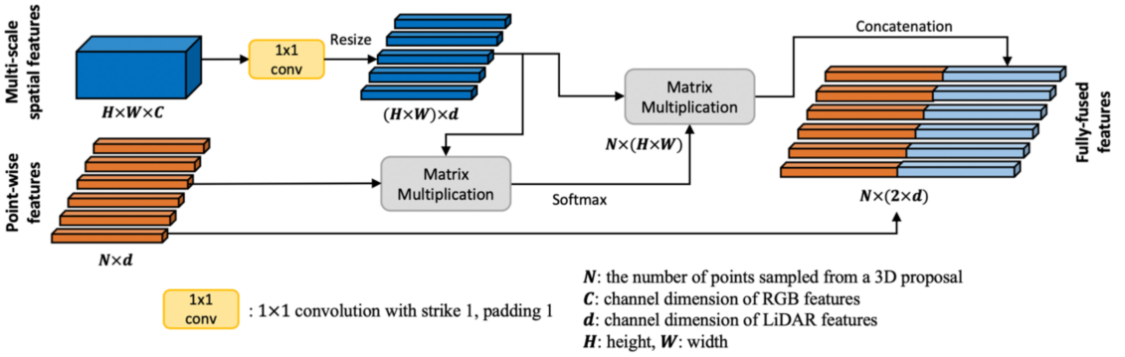

Multi-scale spatial features
H×W×C
1x1 conv
Resize
(H×W)×d
Concatenation
Matrix Multiplication
N×(H×W)
Fully-fused features
Point-wise features
Matrix Multiplication
Softmax
N×(2×d)
N×d
1x1 conv
: 1×1 convolution with strike 1, padding 1
N: the number of points sampled from a 3D proposal
C: channel dimension of RGB features
d: channel dimension of LiDAR features
H: height, W: width

Figure 7. Operational steps of regional fusion layer

4.5. Loss Function

In our network, we use a multi-task loss to train our network. To be more specific, we define the total loss function as the summation of regression loss and classification loss. Since large regression targets are not good for training a detector, we normalize the center and the size of each ground-truth as well as anchor box. The center of each ground-truth and anchor box is normalized as

$$\Delta x = \frac{x^g - x}{w^g}, \Delta y = \frac{y^g - y}{h^g}, \Delta z = \frac{z^g - z}{l^g} \tag{4}$$

where g stands for the ground-truth. In contrast, the size of each ground-truth and anchor box is normalized as

$$\Delta w = \frac{w^g}{w}, \Delta h = \frac{h^g}{h}, \Delta l = \frac{l^g}{l} \tag{5}$$

As for the orientation of each ground-truth and anchor box, it is defined as

$$\Delta\theta = \theta^g - \theta \tag{6}$$

By normalizing the anchor box and the ground-truth, we can obtain a regression tensor T for each of them, where $T = (\Delta x, \Delta y, \Delta z, \Delta w, \Delta h, \Delta l, \Delta\theta)$. To calculate box regression loss L_{box}, we apply the common smooth L1 loss being represented as

$$L_{box}(T^g, T^a) = \sum_{\substack{j\in\{\Delta x,\Delta y,\Delta z,\\ \Delta w,\Delta h,\Delta l,\Delta\theta\}}} smooth_{L1}\left(T_j^g, T_j^a\right) \tag{7}$$

in which

$$smooth_{L1}\left(T_j^g, T_j^a\right) = \begin{cases} 0.5(T_j^g - T_j^a)^2 & , if\left|T_j^g - T_j^a\right| < 1 \\ \left|T_j^g - T_j^a\right| - 0.5 & , otherwise \end{cases} \tag{8}$$

where a denotes the anchor box. On the other hand, we utilize simply binary cross-entropy loss as our classification loss L_{cls}, which can be expressed as

$$L_{cls}(cls^g, cl^a) = -[cls^g \log(cls^a) + (1 - cls^g)\log(1 - cls^a)] \tag{9}$$

After all, the multi-task loss we use to train our model is a weighted sum of the box regression loss L_{box} and the classification loss L_{cls}, which can be expressed as

$$L = \alpha\frac{1}{N}\sum_{i=1}^{N} L_{cls}\left(cls_i^g, cls_i^a\right) + \beta\frac{1}{N_{pos}}\sum_{i\in positive} L_{box}\left(T_i^g, T_i^a\right) \tag{10}$$

where N symbolizes the total number of positive and negative samples, that is to say, $N = N_{pos} + N_{neg}$, and α,β are the hyperparameters controlling the ratio of these two losses.

5. EXPERIMENT

5.1. Experimental Data

In this paper, we choose the task of 3D object detection in KITTI Vision Benchmark to validate our proposed RF3D. In the task of 3D object detection of the benchmark, there are 7,481 training data and 7,518 testing data, and each of them comprises an RGB image, a LiDAR point cloud as well as a calibration file. There are three object categories annotated in the dataset, including car, pedestrian and cyclist. Besides, the category of car has the most sufficient training samples in the dataset. As a consequence, we choose the category of car to evaluate the testing set performance of our approaches as other methods selected. Following the KITTI setting, we accomplish evaluations on three difficulty regimes, namely easy, moderate and hard, which is decided occlusion level, truncated level and distance of the object.

5.2. Evaluation Metric

The predicted results of 3D detection are verified by submitting to KITTI official testing server. The Average Precision (AP) with 40 points is adopted as the evaluation metric for both 3D and BEV detection. In the class of car, the threshold of Intersection over Union (IoU) is set as 0.7 to determine whether the prediction belongs to true positive or false positive.

5.3. Implementation Details

For the LiDAR stream, we only preserve points belonging to the image view via calibration parameters. We subsample 16,384 points from each frame as inputs. For those frames with the number of points fewer than 16,384, we randomly choose points until retrieving 16,384 points. For the RGB stream, we resize the image to the size of 1242×376. The number of points inside 3D proposals for fusion is set 512. We do not apply data augmentation in our experiments because mismatch problems usually occur between point clouds and images.

We implement our network on single GPU GTX 1080 Ti with Pytorch [26]. Two stages of RF3D are trained separately. First stage network, which is utilized to generate proposals, is trained with batch size 8, and second stage network, which is exploited to refine 3D boxes, is trained with batch size 3. Adam [27] is used for optimization with weight decay of 0.001. The learning rate is initialized as 0.001 and decay with a factor of 0.5 at 100, 150, 180 and 200 epochs, respectively.

5.4. Experiment Result of 3D Detection Benchmark

The 3D detection results of the class car on KITTI testing dataset is shown in Table 1. The task of 3D detection is more challenging than that of BEV detection, because 3D detection requires the involvement of height information. Our RF3D outperforms other published state-of-the-art methods with respect to AP under all difficulty regimes in 3D detection except for MMF [19], which is the state-of-the-art fusion-based method in easy and moderate difficulties. In our experimental results, we observe that our network surpasses other methods by a large margin under the hard case. This situation represents that directly utilizing raw point clouds as inputs can preserve the 3D geometric information of those highly occluded or truncated objects. In addition, observing the additional information from RGB stream and the proposed regional fusion layer, the network is able to predict high-quality 3D bounding boxes.

Table 1. Comparison of results on KITTI 3D detection benchmark testing split (car), where PC denotes point clouds.

Method	Types of Input	3D AP of car (in %)		
		Easy	Moderate	Hard
Disp R-CNN [6]	Image	59.58	39.34	31.99
VoxelNet [12]	PC	81.97	65.46	62.85
PointPillars [9]	PC	79.05	74.99	68.30
TANet [14]	PC	83.81	75.38	68.32
F-PointNet [16]	PC+Image	81.20	70.39	62.19
ContFuse [20]	PC+Image	82.54	66.22	64.04
AVOD-FPN [15]	PC+Image	81.94	71.88	66.38
CrossFusion [21]	PC+Image	83.20	74.50	67.01
PointPainting [17]	PC+Image	82.11	71.70	67.08
PI-RCNN [18]	PC+Image	84.37	74.82	70.03
MMF [19]	PC+Image	86.81	76.75	68.41
ours	PC+Image	85.18	75.76	70.99

5.5. Ablation Study on Components

Since KITTI official testing server has limited submissions per month, we use the validation set to conduct our ablation studies and several experiments. We follow the rule proposed in [28] to split the training data into training set and validation set. As a consequence, there are total 3,712 training frames and 3,769 validation frames, respectively.

There are two components presented to reason the fusion between two input sources, including the multi-scale feature aggregation module and the regional fusion layer. The multi-scale feature aggregation module enriches the feature maps of RGB stream by combining feature maps from different receptive fields with upsampling and downsampling layers. The regional fusion layer utilizes the proposals from one stream and their projected proposals from the other to perform the RoI-level fusion so as to fertilize the LiDAR features with additional RGB information. Other than the proposed two modules, we also exploit the proposals generated from both streams to make two sensors compensate with each other. In order to validate the effectiveness of these methods, we conduct several ablation studies on the validation set of class car as well. The experimental results are shown in Table 2.

In the beginning, we simply utilize the LiDAR data to perform 3D object detection without any RGB images as listed in the first row. There is no fusion between two input sources. Secondly, we utilize two input sources simultaneously without the multi-scale feature aggregation module as presented in the second row. Meanwhile, we only leverage the 3D proposals generated from the LiDAR stream and their corresponding projected ones from the RGB stream to perform fusion. As a consequence, there is no 2D region proposal generated from the RGB stream. Besides, we exploit the multi-scale feature aggregation module alone to validate its effectiveness for improving the fusion as illustrated in the third row. After that, we enable the network to generate 2D and 3D proposals simultaneously and project proposals from one stream onto the other stream as shown in the fourth row. In this way, we demonstrate the importance of 2D estimations generated from the RGB images. Finally, we combine all the properties together to use as the last derived model. It is obvious that the performance is more profitable than the others. Therefore, we choose the last one as our final model and comparison with other methods.

Table 2. Ablation studies of each component on KITTI validation split of 3D detection (car), where the RF layer stands for regional fusion layer and the MSFA module indicates multi-scale feature aggregation module.

2D proposals	RF layer	MSFA module	3D AP of car (in %)		
			Easy	Moderate	Hard
✗	✗	✗	83.78	74.34	73.67
✗	✓	✗	86.12	76.89	75.54
✗	✓	✓	88.21	78.42	76.82
✓	✓	✗	87.84	78.36	77.10
✓	✓	✓	89.54	79.22	78.37

5.6. Qualitative Results

We visualize several predicted results from KITTI dataset as illustrated in Figure 8. It is observed that some objects are very difficult to be captured through only RGB images due to serious occlusion and truncation. However, with the help of 3D proposals generated from LiDAR point clouds, these highly occluded and truncated objects can be easily detected since raw point clouds do not suffer from these issues. We also find that several objects have limited points collected in point clouds, resulting in poor performance of 3D proposals generation. Since we simultaneously utilize RGB images to generate 2D proposals, it is verified that the RGB images can compensate the weaknesses of LiDAR point clouds.

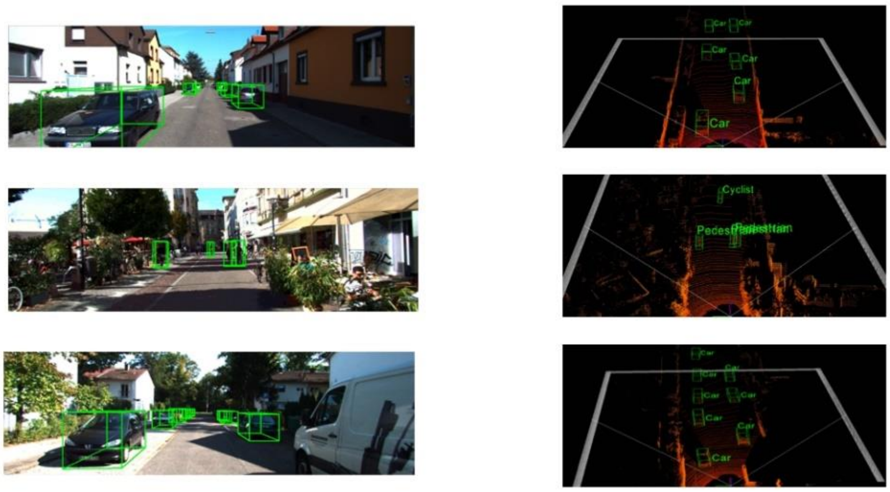


Figure 8. Visualization of the prediction results of RF3D on KITTI dataset

6. Conclusions

In this paper, we propose Regional Fusion Network for 3D on-road object detection. Our network directly consumes raw point clouds as inputs to perform data fusion. To the best of our knowledge, we are the first to integrate raw point clouds and RGB images to conduct 3D object detection. We are able to compensate the weaknesses of two sensors through projecting the proposals from one stream to the other. Additionally, our proposed multi-scale feature aggregation module can combine features from different receptive fields to enrich the RGB features and improve overall detection results. Moreover, the presented regional fusion layer is able to fuse two inputs based on their corresponding RoIs and provide additional RGB information for LiDAR features. The experimental results on KITTI Vision Benchmark show that our model outperforms other methods in 3D detection especially under challenge setting. However, in order to obtain satisfying detection results, our proposed RF3D has longer inference time. The future research emphasizes on designing an efficient and lightweight proposed RF3D to reduce inference time. Besides, data augmentation techniques for both point clouds and images can be developed to improve the performance of the presented Regional Fusion Network.

ACKNOWLEDGEMENTS

This work was in cooperation with Automotive Research and Testing Center under the grant number of 110-EC-17-A-25-1588 in the project of Department of Industrial Technology of Ministry of Economic Affairs, Taiwan ROC. This research was also supported by the Joint Research Center for AI Technology and All Vista Healthcare under Ministry of Science and Technology of Taiwan, and Center for Artificial Intelligence & Advanced Robotics, National Taiwan University, under the grant numbers of 110-2634-F-002-042 and 110-2634-F-002-016.

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.

[2] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.

[3] W. Bao, B. Xu, and Z. Chen, "Monofenet: Monocular 3d object detection with feature enhancement networks," *IEEE Transactions on Image Processing,* vol. 29, pp. 2753-2765, 2019.

[4] M. Ding *et al.*, "Learning depth-guided convolutions for monocular 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 1000-1001.

[5] Y. Chen, S. Liu, X. Shen, and J. Jia, "Dsgn: Deep stereo geometry network for 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12536-12545.

[6] J. Sun *et al.*, "Disp R-CNN: Stereo 3D Object Detection via Shape Prior Guided Instance Disparity Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10548-10557.

[7] B. Yang, M. Liang, and R. Urtasun, "Hdnet: Exploiting hd maps for 3d object detection," in *Conference on Robot Learning*, 2018, pp. 146-155.

[8] B. Yang, W. Luo, and R. Urtasun, "Pixor: Real-time 3d object detection from point clouds," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 7652-7660.

[9] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12697-12705.

[10] Z. Wu *et al.*, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1912-1920.

[11] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354-3361.

[12] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4490-4499.

[13] S. Shi, X. Wang, and H. Li, "Pointrcnn: 3d object proposal generation and detection from point cloud," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 770-779.

[14] Z. Liu, X. Zhao, T. Huang, R. Hu, Y. Zhou, and X. Bai, "TANet: Robust 3D Object Detection from Point Clouds with Triple Attention," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 34, pp. 11677-11684, 04/03 2020.

[15] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3d proposal generation and object detection from view aggregation," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1-8.

[16] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum pointnets for 3d object detection from rgb-d data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 918-927.

[17] S. Vora, A. H. Lang, B. Helou, and O. Beijbom, "PointPainting: Sequential Fusion for 3D Object Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13-19 June 2020 2020, pp. 4603-4611.

[18] L. Xie *et al.*, "Pi-rcnn: An efficient multi-sensor 3d object detector with point-based attentive cont-conv fusion module," *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 34, pp. 12460-12467, 2020.

[19] M. Liang, B. Yang, Y. Chen, R. Hu, and R. Urtasun, "Multi-Task Multi-Sensor Fusion for 3D Object Detection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15-20 June 2019 2019, pp. 7337-7345.

[20] M. Liang, B. Yang, S. Wang, and R. Urtasun, "Deep continuous fusion for multi-sensor 3d object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 641-656.

[21] D.-S. Hong, H.-H. Chen, P.-Y. Hsiao, L.-C. Fu, and S.-M. Siao, "CrossFusion net: Deep 3D object detection based on RGB images and point clouds in autonomous driving," *Image and Vision Computing,* vol. 100, p. 103955, 2020.

[22] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in neural information processing systems*, 2017, pp. 5099-5108.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[24] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.

[25] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.

[26] A. Paszke *et al.*, "Automatic differentiation in pytorch," 2017.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the 3rd International Conference on Learning Representations (ICLR),* 2014.

[28] X. Chen *et al.*, "3d object proposals for accurate object class detection," in *Advances in Neural Information Processing Systems*, 2015, pp. 424-432.

AUTHORS

Hung-Hao Chen received the B.S. degree in Department of Computer Science and Engineering from National Chen Kung University, Tainan, Taiwan in 2018 and the M.S. degree with the Department of Computer Science and Engineering in Department of Computer Science and Engineering from National Taiwan University, Taipei, Taiwan in 2020. His research interests include deep learning and computer vision.

Chia-Hung Wang received the B.S. degree in Department of Electrical Engineering, College of Electrical and Computer Science from National Taiwan University of Science and Technology, Taipei, Taiwan in 2018 and the M.S. degree with the Department of Computer Science and Engineering in Department of Computer Science and Engineering from National Taiwan University, Taipei, Taiwan in 2021. His research interests include deep learning, computer vision and sensor fusion.

Hsueh-Wei Chen received the B.S degree in Department of Computer Science and Information Engineering from National Chung Cheng University, Chiayi, Taiwan in 2020. He is currently pursuing the M.S degree with the Department of Computer Science and Engineering in Department of Computer Science and Engineering from National Taiwan University, Taipei, Taiwan. His research interests include deep learning and object detection.

Pei-Yung Hsiao received the B.S. degree in chemical engineering from Tung Hai University, in 1980 and the M.S. and Ph.D. degrees in electrical engineering from the National Taiwan University, in 1987 and 1990, respectively. In 1990, he was an Associate Professor in the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan. In 1998, he was the CEO of Aetex Biometric Corporation. He is currently a Professor in the Department of Electrical Engineering, National Univ. of Kaohsiung. His research interests and industrial experiences include VLSI/CAD, image processing, fingerprint recognition, visual detection, embedded systems, and FPGA rapid prototyping.

Li-Chen Fu received the B.S. degree from National Taiwan University in 1981, and the M.S. and Ph.D. degrees from the University of California, Berkeley, in 1985 and 1987, respectively. Since 1987, he has been on the faculty of and currently is a professor in both the Department of Electrical Engineering and Department of Computer Science & Information Engineering of National Taiwan University. He is now a senior member of both the Robotics and Automation Society and Automatic Control Society of IEEE, and he became an IEEE Fellow (F) in 2004. His areas of research interest include robotics, FMS scheduling, shop floor control, home automation, visual detection and tracking, E-commerce, and control theory & applications.

Yi-Feng Su received the M.S. degree in Electrical Engineering from National Changhua University of Education (NCUE), Changhua, Taiwan, in 2005. Presently, he works as an engineer in the Automotive Research & Testing Center (ARTC), Taiwan. His research areas include image processing, machine vision, algorithm development, and applications of embedded systems.

K-NEAREST NEIGHBOUR AND DYNAMIC TIME WARPING FOR ONLINE SIGNATURE VERIFICATION

Mohammad Saleem and BenceKovari

Department of Automation and Applied Informatics,
Budapest University of Technology and Economics, Budapest, Hungary

ABSTRACT

Online signatures are one of the most commonly used biometrics. Several verification systems and public databases were presented in this field. This paper presents a combination of k-nearest neighbor and dynamic time warping algorithms as a verification system using the recently published DeepSignDB database. Our algorithm was applied on both finger and stylus input signatures which represent both office and mobile scenarios. The system was first tested on the development set of the database. It achieved an error rate of 6.04% for the stylus input signatures, 5.20% for the finger input signatures, and 6.00% for a combination of both types. The system was also applied to the evaluation set of the database and achieved very promising results, especially for finger input signatures.

KEYWORDS

Online signature verification, k-nearest neighbor, dynamic time warping.

1. INTRODUCTION

Biometrics are used for authentication and identification purposes; signatures are widely used in bank checks, documents, payments, and many other fields. Signatures are classified into online and offline signatures based on the input methods. In online signatures, special devices are used to acquire the signatures, such as tablets and digital pens. These devices can capture several features like position, pressure, azimuth, and altitude as a function of time (see table 1). Online signatures have more features than offline signatures, where signatures are acquired using a regular pen and paper, then scanned and processed as an image. Thus, it is harder to forge an online signature compared to an offline signature. Nowadays, online signatures are used more frequently than before due to digital devices' development and the need for quick actions and smart methods to keep up withthe digital evolution.

Data acquisition, preprocessing, feature extraction, and verification are the main stages of any online signature verification system. These stages together form the process of classifying a signature as genuine or forged. Several methods and algorithms can be applied for each step. The effect of these different algorithms on the accuracy of the verification system varies for other systems.

Many databases are publicly available for researchers, such as the database of Signature Verification Competition2004 (SVC2004) [1], the Spanish Ministry of Science and Technology database (MCYT-100) [2], the Dutch and Chinese subsets of the Signature Verification

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 161-168, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111813

Competition 2011 database (SigComp'11) [3], BiosecureID [4][5], and the German database of the Signature Verification Competition 2015 (SigComp'15) [6]. In this work, a recently published database is used, the DeepSignDB [7]. These databases varyby the signees number, signature number, type of forgery, the input device used, and some other features.

The same signee provides very similar (yet not exactly the same) signatures. These signatures might vary by size, position, pressure, or any other feature. To reduce these inner-class differences, several preprocessing methods can be applied. Scaling and translation algorithms, where the signature is scaled and shifted to a specific range of points, are commonly used algorithms in data preprocessing. Some other methods can be applied, such as zero pressure removal, rotation, or resampling.

The last step of the verification is the classification phase, where several similarity measurements and classification algorithms can be applied to decide whetherthe questioned signature is genuine or forged. Then the system accuracy is evaluated using specific evaluation methods. All the previous steps will be discussed in detail throughout the paper.

In the following section, the related work from the literatureis briefly presented. Afterward, we describe our work methodology and present the experimental results. Finally, the results are evaluated, and the paper is concluded.

2. Related Work

Dynamic time warping (DTW) is widely used for verification and similarity measurement. DTW finds the minimum distance between two-time series, which may vary in length [8]. It is one of the most commonly usedalgorithms measuring the similarities of time series. It has also shownpromising results in the field of online signature verification.

K-nearest neighbor (k-NN) algorithm is applied to calculate upper and lower thresholds of the proposed algorithms, which are then used for signature classification. k-NN is a one-class classification algorithm and was previously used in some verification systems [9] [10].

DTW and k-NN were applied in the state of the art of online signature verification field, but not together. Feng et al. proposed a warping technique for DTW [11]. Faundez-Zanuy proposed a method using a combination of vector-quantization and DTW [12], and Parziale et al. used stability-modulated DTW [13].

There are several published signature verification competition that compares different system on the same database such as the signature verification competition 2004 (SVC2004) [1], the international conference on document analysis and recognition ICDAR competitions (ICDAR 2009 [14]), (ICDAR 2013 [15]) and (ICDAR 2015 [6]), the Signature Competition 2011 (SigComp2011 [3]). These signature verification competitionsprovide fair comparisons between the verification systems applied on the same database under similar circumstances.

3. Methodology

DeepSignDB database is used in this work. It is a combination of the most commonly used databases in the field. More than 70,000 signatures were acquired in this database from 1526 signees [7]; see figure 1. It contains both stylus and finger inputs using eight different devices. The DeepSignDB is divided into two sets.

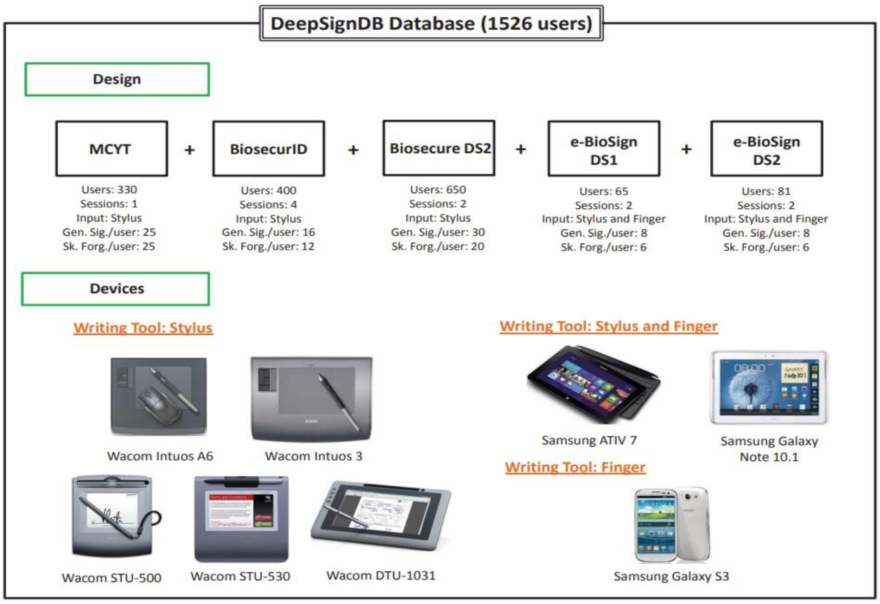


Figure 1. DeepSignDB database [7].

The development set is used for training purposes, and the evaluation set, which helps researchers to evaluate the efficiency and accuracy of their proposed systems. The development and evaluation sets represent 70% and 30% of the database, respectively.

Preprocessing phase is very important to enhance the accuracy of the similarity measurement and reduce the inner class effect. The signature points were filtered in this work by removing the points with zero or shallow pressure values except for the finger input signatures, as theyhave zero pressure values already for all their points.

Previously, we studied the effect of some preprocessing methods on verification accuracy [16]. Scaling and translation have shown a strong impact on accuracy; thus, both were applied to all the signatures to reduce the inner-class differences that occurred since the same signee might provide very similar signatures but still with different scaling and starting points. Thus, the following algorithms were applied to scale the signature to the [0,1] range and shift the center of gravity to the origin.

$$\hat{x}(i) = x_{newMin} + \frac{x(i) - x_{oldMin}}{x_{oldMax} - x_{oldMin}} * (x_{newMax} - x_{newMin}) \quad (1)$$

$$\hat{x}(i) = x(i) - \mu_{\mathrm{x}} \qquad (2)$$

Where x_{oldMin} and x_{oldMin} are the new min-max points range, and$\mu_{\mathrm{x}}$is the points mean.

Although many features can be extracted from online signatures, the combination of the horizontal position (X), vertical position (Y), and pressure (P)are used for the classification purpose, see Table 1. All the preprocessing and similarity measurements were applied to the XYP combination. It showed more accurate results compared to the individuallyused features.

Table 1. Feature's combination

#	**Feature**
1	x-coordinate: X
2	y-coordinate: Y
3	Pressure: P
Combination	**XYP**

The proposed system uses DTW and k-NN algorithms in the classification phase. DTW is used as a distance measurement between the signatures. The k-NN algorithm is used to select the reference signatures and calculate upper and lower threshold, which plays a significant part in calculating the prediction of the tested signature. The k-NN algorithm is applying using the following formula:

$$d(S, S_{nn}) < \theta \frac{1}{K} \sum_{k=1}^{K} d(S_{nn}, S_{knn}) \quad (3)$$

where d is the distance between the questioned signature (S) and its nearest neighbor (S_{nn}), S_{knn}represents the k-nearest neighbor, and θ is a threshold used for the classification calculations.

In the classification phase, the distance d between the questioned signature and the reference signature is calculated using DTW. The distance is used to calculate the prediction for the questioned signature P_q using a calculated forgery threshold F_{th}, genuine threshold G_{th} and a scaling parameter s as follows:

$$P_q = \frac{s * F_{th} - d_s}{s * F_{th} - G_{th}} \, (4)$$

Where P_q is the prediction value, s is the scale, F_{th} and G_{th} are the forged and genuine thresholds, respectively.

The prediction values are between zero and one, where zero represents a genuine signature, while one represents a forged signature. Furthermore, a threshold is assigned to classify the signature as genuine or forged based on its prediction value. Both false acceptance rate (FAR) and false rejection rate (FRR) were considered in the accuracy evaluation. Several thresholds can be applied, and the best result is chosen using the equal error rate (EER) where FAR and FRR cross.

4. EXPERIMENTAL RESULTS

As mentioned in the previous sections, DeepSignDB is divided into a development set and evaluation set. It contains both mobile and office scenarios, random and skilled forgery. In the development phase, the comparisons were provided using two strategies, 1vs1 where only one signature is available as a reference, and 4vs1, where four reference signatures are available. In the evaluation phase, only 1vs1 comparisons were available. The proposed system showed strong performance, especially when using 4vs1 comparisons. For the development set of the DeepSignDB, the results achieved were as following:

- **Task1**

In Task1, only stylus input signatures were available. It represents the office scenarios. Our system achieved 6.04% EER (see figure 2).

- **Task2**

In Task2, only mobile scenario signatures using finger input were considered. The system achieved an EER of 5.20%, see figure 3.

- **Task3**

In Task3, both mobile and office scenario signatures were considered. The system achieved an EER equal to 6.00%, see figure4.

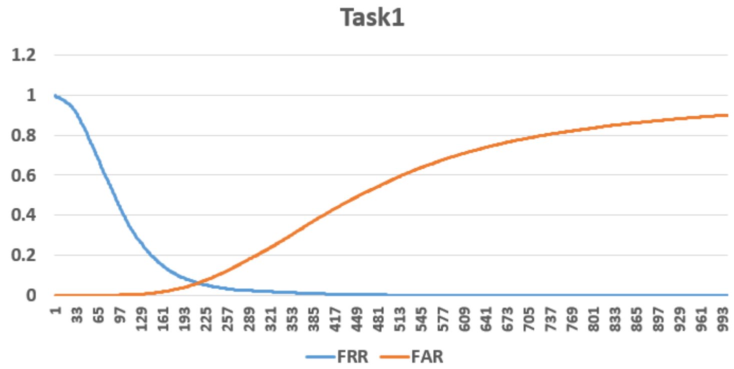


Figure 2. Task1 results.

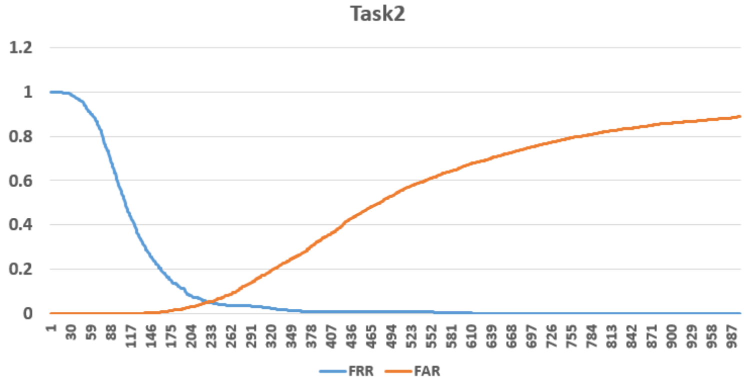


Figure 3. Task2 results.

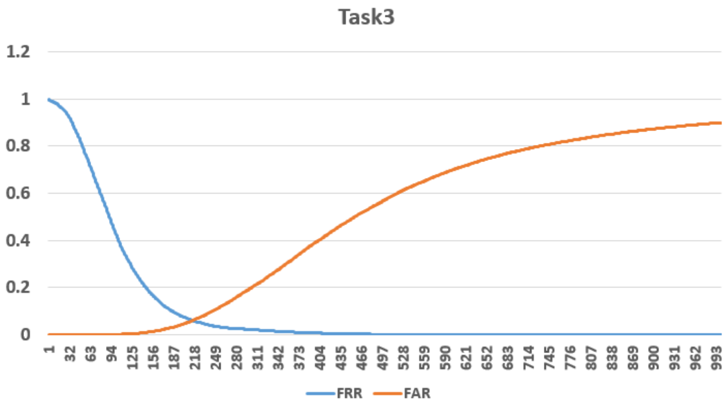


Figure 4. Task3 results.

- **Evaluation set**

For the evaluation set, the EER achieved was 13.28%; for this set, only 1vs1 scenarios were considered. This result is promising, and we believe it can be improved, which is currently under process.

The DeepSignDB database was published recently for the competition. Thus, there are not many published results using it. Table 2 shows a comparison with some other verification systems using different databases.

Table 2. A comparison between verification system for finger input signatures

Study	Databse	EER
Tolosana et al. (2021) [7]	DeepSignDB	13.8%
Lai and Jin (2018) [17]	Mobisig	10.9%
Li et al. (2019) [18]	Mobisig	16.1%
Proposed	**DeepSignDB**	**13.28%**

5. Conclusion

A combination of the k-nearest neighbour and dynamic time warping algorithms is presented in this work as an online signature verification system. DeepSignDB was used for system development and evaluation, a combination of several databases with different input methods. The database is divided into development and evaluation sets where both office and mobile scenarios of signatures were used. The system achieved 6.04% EER when using stylus input signatures, 5.20% for finger input signatures, and 6.00% when using a combination of both types. The system also achieved 13.28% EER on the evaluation set of the database for finger input signatures. These promising results also showed that the system could be improved to adopt more scenarios and achieve higher accuracy.

Acknowledgments

The work presented in this paper has been carried out in the frame of project no. 2019-1.1.1-PIACI-KFI-2019-00263, which has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2019-1.1. funding scheme.

REFERENCES

[1] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "Svc2004: First international signature verification competition," in International conference on biometric authentication, pp. 16–22, Springer, 2004.

[2] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, et al., "MCYT baseline corpus: a bimodal biometric database," IEE Proceedings-Vision, Image and Signal Processing, vol. 150, no. 6, pp. 395–401, 2003.

[3] M. Liwicki, M. I. Malik, C. E. Van Den Heuvel, X. Chen, C. Berger, R. Stoel, M. Blumenstein, and B. Found, "Signature verification competition for online and offline skilled forgeries (sigcomp2011)," in 2011 International Conference on Document Analysis and Recognition, pp. 1480–1484, IEEE, 2011.

[4] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, et al., "BiosecurID: a multimodal biometric database," Pattern Analysis and Applications, vol. 13, no. 2, pp. 235–246, 2010.

[5] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, et al., "The multiscenario multi environment biosecure multimodal database (BMDB)," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 6, pp. 1097–1111, 2009.

[6] M. I. Malik, S. Ahmed, A. Marcelli, U. Pal, M. Blumenstein, L. Alewijns, and M. Liwicki, "Icdar2015 competition on signature verification and writer identification for on-and off-line skilled forgeries (sigwicomp2015)," in 2015 13th International Conference on Document Analysis and Recognition (IC-DAR), pp. 1186–1190, IEEE, 2015.

[7] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Deep-sign: Deep online signature verification," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 2, pp. 229–239, 2021.

[8] P. H. Franses and T. Wiemann, "Intertemporal similarity of economic time series: An application of dynamic time warping," Computational Economics, vol. 56, no. 1, pp. 59–75, 2020.

[9] L. Nanni, "Experimental comparison of one-class classifiers for online signature verification," Neurocomputing, vol. 69, no. 7-9, pp. 869–873, 2006.

[10] S. S. Khan and A. Ahmad, "Relationship between variants of one-class nearest neighbours and creating their accurate ensembles," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 9, pp. 1796–1809, 2018.

[11] H. Feng and C. C. Wah, "Online signature verification using a new extreme points warping technique," Pattern Recognition Letters, vol. 24, no. 16,pp. 2943–2951, 2003.

[12] M. Faundez-Zanuy, "Online signature recognition based on VQ-DTW," Pattern Recognition, vol. 40, no. 3, pp. 981–992, 2007.

[13] A. Parziale, M. Diaz, M. A. Ferrer, and A. Marcelli, "SM-DTW: Stability modulated dynamic time warping for signature verification," Pattern Recognition Letters, vol. 121, pp. 113–122, 2019.

[14] V. L. Blankers, C. E. van den Heuvel, K. Y. Franke, and L. G. Vuurpijl, "ICDAR 2009 signature verification competition," in 2009 10th International Conference on Document Analysis and Recognition, pp. 1403–1407, IEEE, 2009.

[15] M. I. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein, and B. Found, "ICDAR 2013 competitions on signature verification and writer identification for on-and offline skilled forgeries (SigWiComp 2013)," in 2013 12th International Conference on Document Analysis and Recognition, pp.1477–1483, IEEE, 2013.

[16] Saleem, Mohammad, and Bence Kovari. "Preprocessing approaches in DTW based online signature verification." Pollack Periodica 15.1 (2020): 148-157.

[17] S. Lai and L. Jin, "Recurrent Adaptation Networks for Online Signature Verification," IEEE Trans. on Information Forensics and Security, vol. 14, no. 6, pp. 1624–1637, 2018.

[18] C. Li, X. Zhang, F. Lin, Z. Wang, J. Liu, R. Zhang and H. Wang, "A Stroke-based RNN for Writer-Independent Online Signature Verification," in Proc. International Conference on Document Analysis and Recognition (ICDAR), 2019.

AUTHORS

Mohamad Saleem is a Ph.D. candidate in software engineering at Budapest University of Technology and Economics, Hungary, at the Department of Automation and Applied Informatics. His research interests include online signature verification. He worked as a researcher and TA at Yarmouk University, Jordan, and Budapest University of Technology and Economics. He is a member of the Jordanian engineers' association.

Bence Kovari received his Ph.D. degree in software engineering from Budapest University of Technology and Economics, Hungary, in 2013. Since then, he has been working as a researcher and teacher currently as an associate professor at the Department of Automation and Applied Informatics. His research interests include software engineering and automated verification of handwritten signatures. He has over 50 publications in the field. He is a member of the Hungarian Association for Image Processing and Pattern Recognition, a member of the John von Neumann Computer Society.

USING AI TO LEARN INDUSTRY SPECIFIC BIG DATA FOR BUSINESS OPERATION AND CRISIS MANAGEMENT

Yew Kee Wong

School of Information Engineering, HuangHuai University, Henan, China

ABSTRACT

Artificial intelligence has been a buzz word that is impacting every industry in the world. With the rise of such advanced technology, there will be always a question regarding its impact on our social life, environment and economy thus impacting all efforts exerted towards sustainable development. In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets for different industries and business operations. Numerous use cases have shown that AI can ensure an effective supply of information to citizens, users and customers in times of crisis. This paper aims to analyse some of the different methods and scenario which can be applied to AI and big data, as well as the opportunities provided by the application in various business operations and crisis management domains.

KEYWORDS

Artificial Intelligence, Big Data, Business Operations, Crisis Management.

1. INTRODUCTION

Artificial intelligence (AI) is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. AI is accomplished by studying how human brain thinks, and how people learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems [1]. AI is a science and innovation based on disciplines such as Computer Science, Biology, Psychology, Linguistics, Mathematics, and Engineering. A major thrust of AI is in the development of computer functions associated with human intelligence, for example, reasoning, learning, and problem solving. Out of the following areas, one or multiple areas can contribute to build an intelligent system [2]. This paper aims to analyse some of the use of big data for the AI development and its applications in various business operations and crisis management.

2. WHAT IS BIG DATA

The Big data refers to significant volumes of data that cannot be processed effectively with the traditional applications that are currently used. The processing of big data begins with raw data that isn't aggregated and is most often impossible to store in the memory of a single computer. A buzzword that is used to describe immense volumes of data, unstructured, structured and semi-

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 169-176, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111814

structured, big data can inundate a business on a day-to-day basis. Big data is used to analyse insights, which can lead to better decisions and strategic business moves [3]. The definition of big data: "Big data is high-volume, and high-velocity or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." The characteristics of Big Data are commonly referred to as the four Vs:

Volume of Big Data

The volume of data refers to the size of the data sets that need to be analysed and processed, which are now frequently larger than terabytes and petabytes. The sheer volume of the data requires distinct and different processing technologies than traditional storage and processing capabilities. In other words, this means that the data sets in Big Data are too large to process with a regular laptop or desktop processor. An example of a high-volume data set would be all credit card transactions on a day within Asia.

Velocity of Big Data

Velocity refers to the speed with which data is generated. High velocity data is generated with such a pace that it requires distinct (distributed) processing techniques. An example of a data that is generated with high velocity would be Instagram messages or Wechat posts.

Variety of Big Data

Variety makes Big Data really big. Big Data comes from a great variety of sources and generally is one out of three types: structured, semi structured and unstructured data. The variety in data types frequently requires distinct processing capabilities and specialist algorithms. An example of high variety data sets would be the CCTV audio and video files that are generated at various locations in a city.

Veracity of Big Data

Veracity refers to the quality of the data that is being analysed. High veracity data has many records that are valuable to analyse and that contribute in a meaningful way to the overall results. Low veracity data, on the other hand, contains a high percentage of meaningless data. The non-valuable in these data sets is referred to as noise. An example of a high veracity data set would be data from a medical experiment or trial.

Data that is high volume, high velocity and high variety must be processed with advanced tools (analytics and algorithms) to reveal meaningful information. Because of these characteristics of the data, the knowledge domain that deals with the storage, processing, and analysis of these data sets has been labelled Big Data [4].

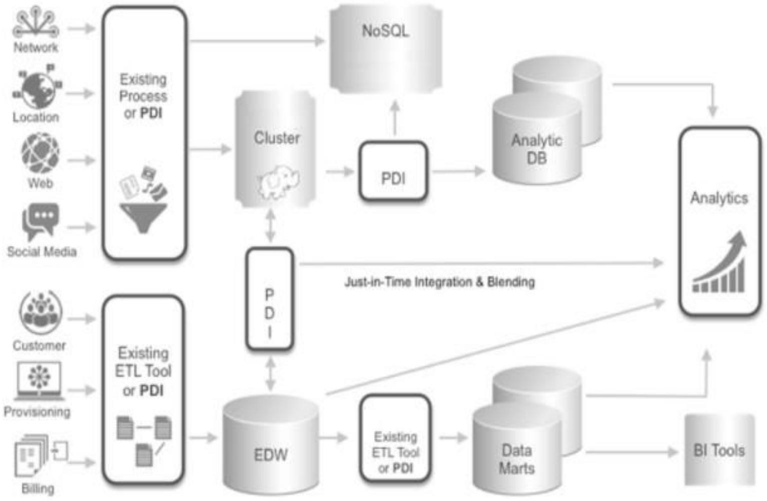


Figure 1. Big Data Architecture. (arccil.com)

2.1. Types of Big Data

There are 3 types of big data; unstructured data, structured data and semi-structured data.

2.1.1. Unstructured data

Any data with unknown form or the structure is classified as unstructured data.

2.1.2. Structured data

Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data.

2.1.3. Semi-structured data

Semi-structured data can contain both the forms of data.
Dealing with unstructured and structured data, data science is a field that comprises everything that is related to data cleansing, preparation, and analysis. Data science is the combination of statistics, mathematics, programming, problem-solving, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing, and aligning data [5]. This umbrella term includes various techniques that are used when extracting insights and information from data.

2.1.4. Big data benefits

- Big data makes it possible for you to gain more complete answers because you have more information.
- More complete answers mean more confidence in the data, which means a completely different approach to tackling problems.

2.2. What is Big Data Analytics

Data analytics involves applying an algorithmic or mechanical process to derive insights and running through several data sets to look for meaningful correlations. It is used in several industries, which enables organizations and data analytics companies to make more informed decisions, as well as verify and disprove existing theories or models [6] [7]. The focus of data analytics lies in inference, which is the process of deriving conclusions that are solely based on what the researcher already knows.

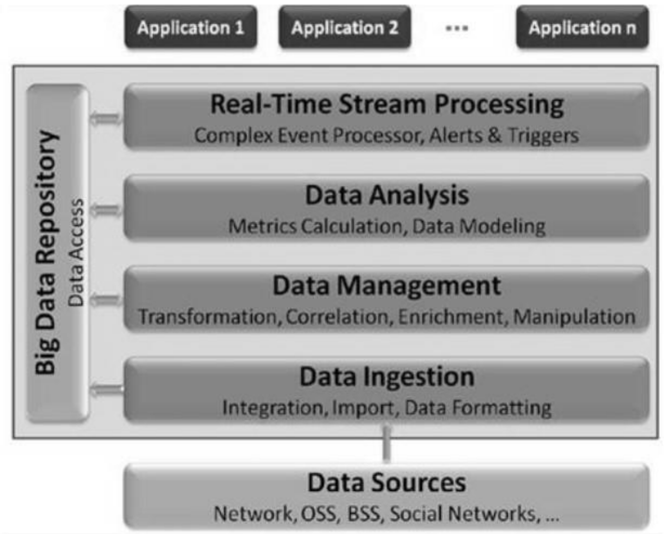

Application 1
Application 2
...
Application n
Big Data Repository
Data Access
Real-Time Stream Processing
Complex Event Processor, Alerts & Triggers
Data Analysis
Metrics Calculation, Data Modeling
Data Management
Transformation, Correlation, Enrichment, Manipulation
Data Ingestion
Integration, Import, Data Formatting
Data Sources
Network, OSS, BSS, Social Networks, ...

Figure 2. Big Data Analytics Architecture.

3. USING AI IN SENSITIVE BUSINESS OPERATIONS

The artificial intelligence rules define the way the online learning system assigned learning materials and exercises for the learner to follow [8]. These are the basic rules which we have carry out in our experiments, in which we find it effective in improving the learners understanding.

3.1. Financial Industry

Artificial intelligence (AI), along with other financial technology (fintech) innovations, are significantly changing the ways that financial business are being run, especially in the fields like trading and insurance, leading the traditional financial industry into a new era [9].
Robots replacing humans

Back in 2000, Goldman Sach's New York headquarters employed 600 traders, buying and selling stock on the orders of the investment bank's clients. Today there are just two equity traders left, as automated trading programs have taken over the rest of the work. Meanwhile, BlackRock, the world's biggest money manager, also cut more than 40 jobs earlier this year, replacing some of its human portfolio managers with artificially intelligent, computerized stocktrading algorithms. Those two big companies are not the only financial institutions replacing human jobs with robots. By 2025, AI technologies will reduce employees in the capital markets by 230,000 people worldwide, according to a report by the financial services consultancy Opimas [10].

Big new frontiers are only just beginning to opening up in fintech from AI, block chain and robotics to biometrics, augmented reality and cybersecurity. Among all the fintech innovations, the prospect of the block chain has the highest expectation. The block chain will change the way people store information, which is real, spreading fast and cross-border, and its 'de-centric' feature will allow everyone to know what other people are doing. The application of block chain in finance will once again bring about a revolutionary impact on the industry, just like AI does.

3.2. Health Industry

The Artificial intelligence (AI) is reshaping operations across industries. Arguably, healthcare is where these changes are poised to make the biggest impact – optimizing uptime and availability of the treatment solutions.Using AI-powered tools capable of processing large amounts of data and making real-time recommendations, healthcare organizations are learning they can reduce administrative waste in a number of areas, from medical equipment maintenance to hospital bed assignments [11].

Artificial intelligence is reinventing and reinvigorating modern healthcare through technologies that can predict, comprehend, learn and act. The ability of AI to transform clinical care has received widespread attention, but the technology’s potential extends beyond patient care to processes across the spectrum of healthcare operations.In healthcare and other industries that depend on reliable equipment performance, few things are more disruptive than unexpected outages. These unplanned stops create costly emergency situations, such as extended downtime, rush delivery of parts and overtime to repair the equipment.

Facing pressure to improve profitability and efficiency, many healthcare organizations are turning to emerging technologies like AI and big data analytics to improve upon existing maintenance operations. Until recently, maintenance typically involved either reacting to an unexpected problem or adhering to a preventive maintenance schedule, which can sometimes result in unnecessary maintenance. Line.

3.3. Manufacturing Industry

AI is core to manufacturing's real-time future. Real-time monitoring provides many benefits, including troubleshooting production bottlenecks, tracking scrap rates, meeting customer delivery dates, and more. It's an excellent source of contextually relevant data that can be used for training machine learning models. Supervised and unsupervised machine learning algorithms can interpret multiple production shifts' real-time data in seconds and discover previously unknown processes, products, and workflow patterns [12].

The manufacturing industry has exploited the use of AI technology, and in particular knowledgebased systems, throughout the manufacturing lifecycle. This has been motivated by the competitive challenge of improving quality while at the same time decreasing costs and reducing design and production time. Just-in-time manufacturing and simultaneous engineering have further required companies to focus on exploiting technology to improve manufacture planning and coordination, and on providing more intelligent processing in all aspects of manufacturing. The objective is to improve quality, to reduce costs, and to speed up the design and manufacturing process.

4. USING AI IN CRISIS MANAGEMENT

4.1. Extreme Weather Forecast

According to the UN Office for the Coordination of Human Affairs, in 2016 over 100 million people were affected by natural disasters including earthquakes, hurricanes and floods. Technology has a vital role to play in providing the appropriate situational awareness that then shapes practical, life-saving decisions for effective crisis management. These decisions may involve the evacuation of the most dangerous areas after an earthquake, or explore tactical options about how and where to position critical resources like medicine, food, clean water and shelter. Through utilising the data tweeted and texted by citizens in a crisis zone, rescuers have access to the knowledge needed to devise a strategy for immediate rescue attempts and for longer term help [13].

Issues can arise, however, due to the volume of available data, and high-quality filtering systems are needed to avoid using inaccurate data that could misdirect humanitarian aid, potentially wasting time, resources, and human trust in the system. Humanitarian responders may, understandably, question the specificity of information, therefore, building their trust and encouraging uptake of AI technology is a socially meaningful endeavour; without this, a system is unlikely to be adopted in the field. Machine learning, understood as the refinement of how AI 'learns' to use algorithms and other data, offers a solution to detecting key information taken from social media messages. Hence, researchers are focusing efforts on improving how the millions of messages are sifted by algorithms to overcome inaccuracy, ensuring that only the most important data is identified and shared.

4.2. Man-Made Environmental Disaster

The case of BP oil spill in 2010 provides an important example for understanding how these principles are valued by public opinion in a crisis situation, and how the communication actions by a corporation in this type of circumstances might have long-term effect on the brand image of the organization. On April 20, 2010, a BP's Deepwater Horizon oil rig exploded, causing what has been called the worst environmental disaster in U.S. history and taking the lives of 11 rig workers. For 87 straight days, oil and methane gas spewed from an uncapped well-head, 1 mile below the surface of the ocean. The federal government estimated 4.2 million barrels of oil spilled into the Gulf of Mexico [14].

The accumulation of unsafe supervisory action had resulted in risk levels substantially increasing. Not only were risks increasing, but they were also incrementally becoming more aggressive in nature. For instance, one of the first acts of unsafe supervision is illustrated when BP neglected its responsibility of ensuring safety protocols were carried out after the completion of the Macondo Well. This was a major mistake on BP's part, violating safety protocols which may have identified the issues present with the cementing of the well. Should these issues have been identified sooner, the likeliness of the crisis happening would potentially be slim. In addition to this, there was also very little supervision during and after works were carried out. This can be attributed to the aforementioned organisational restructuring which created much confusion regarding who was accountable for the assurance of safety [15].

4.3. Natural Disaster

Researchers have found that AI can be used to predict natural disasters. With enormous amounts of good quality datasets, AI can predict the occurrence of numerous natural disasters, which can be the difference between life and death for thousands of people [16]. Some of the natural disasters that can be predicted by AI are:

4.3.1. Earthquakes

AI systems can be trained with the help of seismic data to analyse the magnitude and patterns of earthquakes and predict the location of earthquakes and aftershocks.

4.3.2. Floods

Various researchers and technology experts are developing AI-based applications with the help of rainfall records and flood simulations to predict and monitor flooding.

4.3.3. Volcanic eruptions:

AI-powered systems can accurately predict volcanic eruptions with the help of seismic data and geological information.

4.3.4. Hurricanes:

AI can use satellite to predict and monitor the path and intensity of hurricanes and tornadoes.

5. CONCLUSIONS

The study is assessing new frameworks for effective prevention measures and how AI can fit in and foster the early warning process. So further experiments and understanding the interrelation between AI and big data, what frameworks and systems that worked, and how AI can impact on different business operations whether by introducing new innovations that foster crisis management learning process and early prevention measures. The study from various reviews show promising results in using AI to learn specific industry big data and further evaluation and research is in progress.

REFERENCES

[1] M. K.Kakhani, S. Kakhani and S. R.Biradar, (2015). Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8), pp.228232.

[2] A. Gandomi and M. Haider, (2015). Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 35(2), pp.137-144.

[3] C. Lynch, (2008). Big data: How do your data grow?, Nature, 455, pp.28-29.

[4] X. Jin, B. W.Wah, X. Cheng and Y. Wang, (2015). Significance and challenges of big data research, Big Data Research, 2(2), pp.59-64.

[5] R. Kitchin, (2014). Big Data, new epistemologies and paradigm shifts, Big Data Society, 1(1).

[6] C. L. Philip, Q. Chen and C. Y. Zhang, (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data, Information Sciences, 275, pp.314-347.

[7] K. Kambatla, G. Kollias, V. Kumar and A. Gram, (2014). Trends in big data analytics, Journal of Parallel and Distributed Computing, 74(7), pp.2561-2573.

[8] S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, (2014). On the use of mapreduce for imbalanced big data using random forest, Information Sciences, 285, pp.112-137.

[9] MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, (2014). Health big data analytics: current perspectives, challenges and potential solutions, International Journal of Big Data Intelligence, 1, pp.114-126.

[10] Xinhua China Daily, (18-Sep-2017). How is AI disrupting financial industry.
http://www.chinadaily.com.cn/business/2017-09/18/content_32147126.htm

[11] Focus Elekta's Online Magazine, (2019). How AI is revolutionizing healthcare operations, https://focus.elekta.com/2019/10/how-ai-is-revolutionizing-healthcare-operations/

[12] Louis Columbus, (18-May-2020). 10 Ways AI is improving manufacturing in 2020, Forbes. https://www.forbes.com/sites/louiscolumbus/2020/05/18/10-ways-ai-is-improvingmanufacturing-in-2020/?sh=3530e5d1e85a

[13] Kejriwal M. & Zhou P., (2019). SAVIZ: Interactive Exploration and Visualization of Situation Labeling Classifiers over Crisis Social Media Data, *International Conference on Advances in Social Networks Analysis and Mining, Vancouver*, Aug 27-30, pp705-708.

[14] National Commission, (2011). The Gulf Oil Disaster and the Future of Offshore Drilling.

[15] Dhaimaan Mahmud, (2019). Crisis Management Analysis of the BP Oil Spill, Birmingham Business School.

[16] Naveen Joshi, (15-Mar-2019). How AI can and will predict disasters, Forbes,https://www.forbes.com/sites/cognitiveworld/2019/03/15/how-ai-can-and-will predictdisasters/?sh=7cddf40d5be2

AUTHOR

Prof. Yew Kee Wong (Eric) is a Professor of Artificial Intelligence (AI) & Advanced Learning Technology at the HuangHuai University in Henan, China. He obtained his BSc (Hons) undergraduate degree in Computing Systems and a Ph.D. in AI from The Nottingham Trent University in Nottingham, U.K. He was the Senior Programme Director at The University of Hong Kong (HKU) from 2001 to 2016. Prior to joining the education sector, he has worked in international technology companies, HewlettPackard (HP) and Unisys as an AI consultant. His research interests include AI, online learning, big data analytics, machine learning, Internet of Things (IOT) and blockchain technology.

BURNOUTWORDS - DETECTING BURNOUT FOR A CLINICAL SETTING

Sukanya Nath and Mascha Kurpicz-Briki

Institute for Data Applications and Security IDAS,
Bern University of Applied Sciences, Biel/Bienne, Switzerland

ABSTRACT

Burnout, a syndrome conceptualized as resulting from major workplace stress that has not been successfully managed, is a major problem of today's society, in particular in crisis times such as a global pandemic situation. Burnout detection is hard, because the symptoms often overlap with other diseases and syndromes. Typical clinical approaches are using inventories to assess burnout for their patients, even though free-text approaches are considered promising. In research of natural language processing (NLP) applied to mental health, often data from social media is used and not real patient data, which leads to some limitations for the application in clinical use cases.

In this paper, we fill the gap and provide a dataset using extracts from interviews with burnout patients containing 216 records. We train a support vector machine (SVM) classifier to detect burnout in text snippets with an accuracy of around 80%, which is clearly higher than the random baseline of our setup. This provides the foundation for a next generation of clinical methods based on NLP.

KEYWORDS

Natural Language Processing, Psychology, Burnout, Machine Learning.

1. INTRODUCTION

Stress causes several syndromes or diseases and is a major problem of today's society. The *Stress in America Report 2019* from the American Psychological Association has shown that Americans believe that a healthy stress level is on average 3.8 (scale ranging from 1 to 10, where 10 is "a great deal of stress" and 1 is "little or no stress"), however, they report to have experienced in average a stress level of 4.9 [1]. Nearly 8 in 10 Americans say that the coronavirus pandemic is a significant source of stress in their lives [2].

Sometimes, this stress can lead to a burnout syndrome. Last year, the WHO has included burnout in the 11th Revision of the International Classification of Diseases (ICD-11) as a syndrome[1]. In ICD-11, burnout is defined as follows:

"Burnout is a syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed. It is characterised by three dimensions: 1) feelings of energy

[1] https://icd.who.int/browse11/l-m/en\#/http://id.who.int/icd/entity/129180281

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 177-191, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111815

depletion or exhaustion; 2) increased mental distance from one's job, or feelings of negativism or cynicism related to one's job; and 3) a sense of ineffectiveness and lack of accomplishment." [2]

In the global pandemic crisis around COVID-19, research has shown an increase of burnout, in particular on frontline personnel in the health sector [3] [4].

Burnout identification is complex, because it overlaps with other syndromes [5] and multiple definitions exist [6]. For example, fatigue is a common symptom for both depression and burnout [7]. To identify burnout in clinical intervention, inventories are used. Inventories are psychological tests, where the person concerned fills out a questionnaire. The currently used metric, in both practice and most studies, measures burnout with self-test inventories, and has been criticized in the literature [8] [9].

Inventories with scaling questions, even though often used in practice, have major limitations. In personality inventories, people tend to fake their results [10]. This risk might be increased with a delicate topic such as burnout. Furthermore, extreme response bias (ERB), i.e., the tendency of some respondents to choose the highest or lowest option, is a well-known issue [11]. Other research also reports defensiveness (denying symptoms) and social desirability (to show an exaggeratedly positive image) in inventories [12].

Literature agrees that the Maslach Burnout Inventory (MBI) [13] [14] and the Tedium Measure [15], later called Burnout Measure, are most commonly used in practice and research [16] [9]. The MBI is an introspective psychological inventory and consists of 22 items in three dimensions: emotional exhaustion, depersonalization and personal accomplishment. The Tedium Measure [15] is often used as an alternative to MBI. This inventory consists of 21 items, and each item has to be classified by frequency.

Apart the methods using inventories, burnout components can also be identified by independent judges in interview extracts [17] [9]. The drawback of interviews or free-text questions is that they result in large overhead (interviews, transcription and interpretation) and therefore promising approaches are often not further explored [9].

In this paper, we present BurnoutWords, a dataset based on extracts from conversations with burnout patients, a control group and experts. We provide insights into the wording of burnout patients, extract features from the dataset and allow further research to develop new approaches to enable new clinical methods with text-based burnout identification. We train a burnout classifier using Support Vector Machine (SVM) models and reach an accuracy with a clear improvement over the random baseline.

Whereas most related work in the field investigates social media content, our dataset includes extracts from interviews with burnout patients, aggregated from different previous work and pre-processed for automated evaluation. Using interview extracts from real burnout patients reduces the noise as compared to social media data. This allows to fill the gap between natural language processing (NLP) and the application of the new technologies to develop new methods for application in clinical psychology.

Our paper contributed to the current state-of-the-art by providing a new type of burnout detection dataset, which is based on interview extracts instead of social media data. We furthermore demonstrate how such a dataset can be used to develop new technologies for clinical psychology.

[2] https://icd.who.int/browse11/l-m/en\#/http://id.who.int/icd/entity/129180281

Therefore, our work creates the foundation for future work in the field and new methods for clinical burnout detection.

In section 2, we described the related work with regard to the application of NLP methods and existing datasets for burnout and depression detection. Then, in section 3, we describe the BurnoutWords dataset and how it has been assembled. In section 4 we describe our experimental setup. We then describe (section 5) and discuss (section 6) the results before concluding the article in section 7.

2. Related Work

2.1. Natural Language Processing for Burnout/Depression Detection

Few works exist for burnout detection in general, and as far as of our knowledge, no comparable dataset exists. Burnout is not a disease, but considered as a syndrome, making its detection often hard, because its symptoms overlap with other syndromes or diseases [5] in particular depression. Therefore, we consider additionally the related work for depression detection with the help of NLP.

In particular, we focus on text-based data and therefore do not further discuss approaches based on biomarkers, vocal data (e.g., [18]) or image-based approaches (e.g., [19]). It has been shown that in clinical psychology, written language often plays a central role in diagnosis [20]. The authors summarize the linguistic and social indicators that have been applied to automatic depression detection in different contexts, e.g., narratives written by college students with or without depression [21].

The work focusing on data from social media is the majority of available research in the field, but also work about online forums exist, e.g., [22]. In general, in the related work, the concerned users have been identified by using a screening survey, their public sharing of a diagnosis, or their membership in an online forum [23].

Whereas depression is often diagnosed as being present or absent (e.g., presence of depression symptoms on facebook posts for college students [24]), a model based on survey answers and the language used in facebook posts assesses the severity of depression [25] using the depression facet scores of the Big 5 item pool [26]. Changes in depression across seasons are observed which confirms results from clinical psychology. It has been shown that explicit depression references are rare, but when they appear, they are strong indicators for a real-life depression [27]. Other research targets particular types of depression: a study analyzing Twitter data provides an approach to identify mothers at risk for postpartum depression [28] (complemented later with shared facebook data [29]).

Different technologies are used for the detection of depression on social media, e.g., based on the linguistic inquiry word count (LIWC) lexicon [30] containing multiple psychological constructs. Some more recent approaches use deep learning architectures to achieve better results, which requires a large amount of data. Research has shown how a neural network can be designed to detect depression with limited data and without any exhaustive feature engineering [31], presenting a neural network architecture that optimizes word embeddings. SenseMood [32] is applying a CNN-based classifier and Google's Bert model [33] on posted images and tweets from users with or without depression, combining visual and textual features. They are using a dataset previously presented in research [34], which contains a set of users with anchor tweets matching the strict pattern that they have been diagnosed depression.

In an enterprise context, studies have used the Valence-Arousal-Dominance (VAD) model [35] to study productivity or risk for burnout in data such as issues and comments from a software development management tool [36]. Burnout risk is measured as low valence and dominance, and high arousal.

Whereas most work focusses on social media data only, it has been shown that the language from Facebook posts can be used to predict depression for consenting individuals recorded in medical records [37]. Another study has examined how language patterns on social media change prior to emergency department visits [38].

When mental health and in particular depression is studied from social media data only, the user's mental state is reflected from published posts and thoughts. This can lead to limitations in the data sets. In our work, we train our model based on text data transcribed from interviews with confirmed burnout patients. This provides the basis for new approaches to be applied not only on new data from social media, but also as a tool support for professionals in clinical intervention.

2.2. Existing Datasets for Burnout/Depression Detection

Social Media: The dataset from ReachOut triage shared task [39] consists of 65'024 forum posts that were manually labelled by expert judges by their severity. This dataset addresses different types of mental health crisis, not focusing particularly on a specific diagnosis or syndrome like depression or burnout. The dataset from the CLPsych 2015 shared task on depression detection was constructed using tweets from users that have stated explicitly that they have been diagnosed with depression or PTSD [40] [41]. For each user up to 3000 recent public tweets were added to the dataset. Other work has collected data from Twitter, creating three datasets: Depression, Non-Depression, and Candidate-Depression [34]. Based on the user's information and current tweet, an anchor tweet [41] to determine the mental health was selected. To simulate observation over time as in a clinical setup, tweets following in the next month after the anchor tweet were also considered for the selected users.

Clinical Interviews/Crisis Counselor: The Crisis Text Line dataset [42] is based on the data collected by a 24/7 text-based crisis support hotline. It can be made available for researchers upon application and contains labeled data concerning different topics including depression. The Distress Analysis Interview Corpus (DAIC) [43] [44] contains clinical interviews conducted by humans and agents in English language. The semi-structured clinical interviews provide a contribution to detect psychological distress conditions such as anxiety, depression or post-traumatic stress disorder. Each interview does also include a depression score from the PHQ-8 inventory [45]. The data has been transcribed and annotated for a variety of features[3]. The AVEC 2014 challenge about depression detection [46] included also interviews in German language which are available as video and audio traces. The original challenge did not provide transcripts. However, other work has used automatic speech recognition technology to transcribe interviews from this corpus and make it available upon request [47].

Disease Information from Clinical Notes: The SemEval-2014 Task 7 dataset [48] includes clinical notes which are annotated with disease/disorder mentions. The dataset is based on the Shared Annotated Resources (ShARe) project[4]. Recent research has provided a large public dataset for clinical motivated symptom extraction from clinical notes [49]. Such approaches are helpful to automatically process and aggregate clinical data. In the dataset presented in this paper,

[3] https://dcapswoz.ict.usc.edu/

[4] http://share.healthnlp.org

we go one step further: in addition to the clinical symptoms, thoughts and sentiments expressed by the patients are included in the dataset and thus in the classifier.

Based on our literature research, we conclude that there is currently no publicly available dataset that provides labeled patient interview data in text form for burnout detection, neither for the English nor the German language. In this paper, we therefore introduce a first version of the BurnoutWords dataset for the German language to the research community to enable future research in the field.

3. THE BURNOUTWORDS DATASET

3.1. Dataset Content and Origin

This paper presents the BurnoutWords dataset, containing extracts from interviews in the German language. The interviews have been conducted in the context of different previous work where extracts have been published, and have been collected and aggregated in our research. The dataset contains texts and corresponding labels, whether the texts describe the current burnout situation (label *burnout*), or the view of a person that has no (more) burnout, or potential measures and prevention of burnout (label *noburnout*). Table 1 shows an extract from the dataset to illustrate the format.

Table 1. Example: one record of the BurnoutWords dataset. Label 1 indicates burnout (label 0 would indicate the class noburnout). Translation to English: (The) doctor made the diagnosis. (The) doctor is a good colleague of mine and takes a lot of time for me. I only realized it ca. 2 months later. It needs time until one can accept it. I informed myself about the disease. The information on the internet did not help me. The doctor could give me advice.

Text	**Label**
Arzt stellte Diagnose. Arzt ist guter Kollege von mir und nimmt sich viel Zeit für mich. Ich realisierte es aber erst ca. zwei Monate später. Braucht Zeit bis man es akzeptieren kann. Informierte mich über die Krankheit. Die Informationen im Internet haben mir nicht geholfen. Der Arzt konnte mir Auskunft geben.	1

A part of the extracts has been published in a thesis [50] and, in agreement with the author of the thesis, the extracts have been re-used in this research. The thesis studies burnout in medical doctors. The author conducted interviews with confirmed burnout patients (i.e., the medical doctors), and with a control group of medical doctors not having burnout. We collected and pre-processed the relevant data from the thesis and labeled it into the following classes: burnout patients (*burnout*), and control group (*noburnout*).

More interview extracts have been collected from a book investigating on burnout in an enterprise context [51]. The author conducted interviews with seven different patients that previously had a burnout, working in different domains (i.e., IT, marketing, engineering, health care, coaching and journalism). We identified the questions concerning descriptions of their burnout, including thoughts and symptoms of that time, and included their answers to the class *burnout* of our dataset. Furthermore, we included answers to questions discussing their current life after burnout, and hints to avoid or handle burnout in the data labeled as *noburnout*.

The author [51] also conducted interviews with experts, including medical doctors, psychologists and coaches. We selected answers describing the symptoms and emotions of burnout patients and labeled them as *burnout*. We identified questions concerning the handling and prevention of burnout and included the answers in the class *noburnout*.

The original data provides some demographic information about the participants in the interviews; however, it is partially modified for privacy reasons. Since only extracts from the interviews are available (the entire interviews cannot be used for further research due to data protection law limitations), the text content is not equally distributed between the participants. Due to this and the partial anonymization of demographic data, it is not considered in the dataset. The dataset contains an overall number of 293 records, 216 in the class *burnout* (73.72%) and 77 in the class *noburnout* (26.28%). Due to the fact that only extracts of the interviews were published, a majority of those extracts comes from the burnout patients and not the control group.

3.2. Data Ethics

The local ethics commission has approved to use this publicly available data for our research. Technology itself is neither good nor bad. However, the ethical aspects of the application of such technologies must be discussed. In our research, we focus on providing the required technology to allow future clinical methods to assess mental health. Such methods should be based on the voluntariness of the involved individuals and provide benefits for individuals and the society. When such technologies are misused as an instrument of power, for example using them in a company, to automatically assess written communication of collaborators without their explicit consent or with negative consequences in case of not giving their consent, this would be a major ethical issue. We therefore provide access to our dataset only upon request for future research in an academic or clinical context.

4. Experimental Setup

4.1. The Words of Burnout Patients

Preprocessing: First of all, characters such as brackets or quotation marks have been removed. Then, double whitespaces have been removed. Finally, the character *ß* is replaced with *ss*, since this character is not common in all the German speaking areas. The pre-processed data is stored in an additional column of the data table. The original data is also available for further research. Responses with less than 200 characters were excluded, since we assume that they do not contain enough information.

In a first step, we examine the words that are being used by the burnout patients and the control group respectively. We therefore consider the ten most used nouns (lemmatized) for each group, excluding the words that appear in both groups. We assume that those words are conversational words that do not add value for the resolution of the question whether there is a burnout or not.

4.2. Model Training

We split the data into a training and a test set. We are using around 70% of the data for feature extraction and training of the model, and around 30% of the data to evaluate the model against completely new data that it was not trained with. Figure 1 shows the experimental setup as explained in detail in the following subsections.

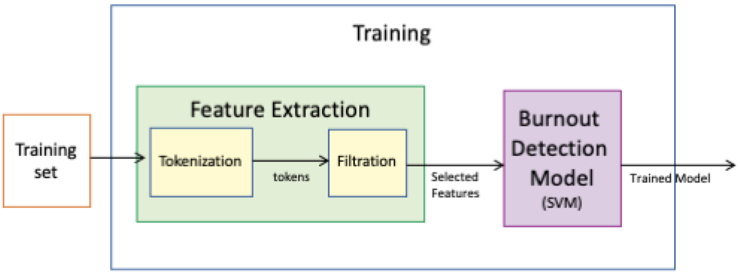


Figure 1. The experimental setup of our training including feature extraction and the burnout detection model.

4.2.1. Feature Extraction

The first part of the training process is the feature extraction, which can be further divided into two phases. The first is the tokenization phase. At the end of this phase, we produced a list of all the tokens which could be potential features. The second phase is the feature filtration stage. In this phase, evaluated the goodness of the features. The features which passed the filtration phase are considered as selected features.

Tokenization: In order to generate features, we divided the training set into the two classes - *burnout* and *noburnout* - and aggregated the contents of each class. At the end of this first step, we had one large file containing the text from all the *burnout* cases, and another large file of text from all the *noburnout* cases. The *burnout* text file was larger than the *noburnout* text file, as we had a higher number of burnout cases. We fed both text files to a function which created a list of tokens (using SpaCy[5]). We filtered the stop-words as they contain little information about the topic. Similarly, we filtered out the punctuations as they also contain little information about the topic and in particular because the interview excerpts are a form of notes and not directly written by the interviewee. We produced a list of tokens for the *burnout* class and another list of tokens for the *noburnout* class.

Feature Filtration: In this phase, our aim is to select those features which can help us achieve a clear boundary between the two classes during the model training. As such, we are interested in those features which occur with a much higher frequency in one class but not in the other class. At first, the lists of tokens of both classes were counted for frequency of occurrence, and a feature dictionary was created such that the key is the token/feature and the value is the frequency of this token/feature. As mentioned before, we have more *burnout* cases than *noburnout* cases. To reduce the advantage of length for the *burnout* class, we further normalized each of the values of this feature dictionary with the total number of tokens in the respective class text file. We then computed the difference between the frequencies of both classes: we took all the features present in the *burnout* and *noburnout* feature dictionaries and note the difference of their normalized frequencies:

$$\text{diff} = \text{burnout}_{\text{norm.freq}} - \text{noburnout}_{\text{norm.freq}}$$

We created a new dataset (dataframe) containing the following columns: feature, burnout frequency normalized, noburnout frequency normalized, difference. We sorted using the

[5] https://spacy.io

frequency difference. Note that if we have a positive difference, the *burnout normalized frequency* is greater than the one for *noburnout*. In other words, the feature has occurred relatively more frequently in the *burnout* text rather than *noburnout* text. In the same way, if we have a negative difference, the feature has occurred more frequently in the *noburnout* text than in the *burnout* text.

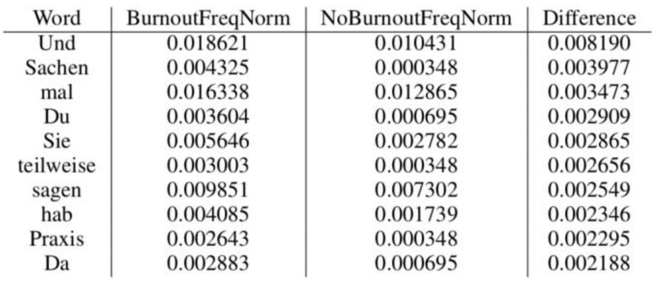
For a given feature set size, say N (where N is an even number), we selected N/2 features which had the greatest positive difference and another N/2 features with the greatest negative difference. In this way, we selected those features which had the greatest (positive or negative) difference between the two classes in the training set. We did not consider features which occur relatively commonly in both classes as they are not very discriminatory.

4.2.2. Burnout Detection Model

This section describes the second part of the training phase. In this phase, we transform the training and test sets into their corresponding feature vector matrix according to the previously selected features. For the purpose of training our burnout detection model, only the training set is used, while the test set is used for model evaluation.

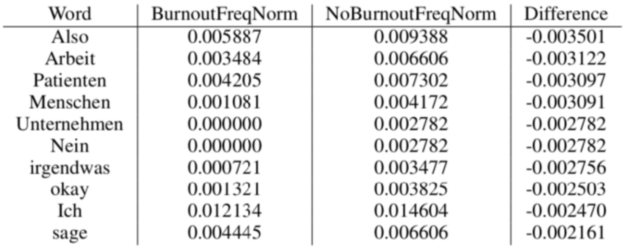
During training, we loop over the different features identified in the previous experiment, using different feature set sizes (N=10,20,...,100). For each iteration we created Support Vector Machine (SVM) models using different configurations. Support Vector Machine models [52] [53] are supervised learning models that attempt to find a hyperplane separating the classes in N dimensions where N is the number of features. A set of data points, called the support vectors, is located close to the hyperplane and helps to orient the hyperplane, such that the maximum separation of the classes, i.e., larger margins, may be achieved. Support Vector Machines use mathematical functions called kernels to transform the feature space such that nonlinear boundaries are transformed to linear boundaries separable by a hyperplane. The linear kernel is the simplest and works well with linearly separable data. Some other Kernels which are commonly used [54] are RBF (Gaussian radial basis function), polynomial and sigmoid (sometimes being referred to as neural network model) kernels. Therefore, we applied the linear, RBF, polynomial and sigmoid kernels in our experiments. The parameter C is the regularization parameter controlling the trade-off between misclassifications and the width of the margin. A large value of C leads to an overfit wiggly boundary whereas a low value of C causes a smoothened boundary [54]. In our proposed model, the value of C used was 1.

In Figure 2 and 3, we show ten selected features with large positive difference and another ten features with large negative difference.



Word	BurnoutFreqNorm	NoBurnoutFreqNorm	Difference
Und	0.018621	0.010431	0.008190
Sachen	0.004325	0.000348	0.003977
mal	0.016338	0.012865	0.003473
Du	0.003604	0.000695	0.002909
Sie	0.005646	0.002782	0.002865
teilweise	0.003003	0.000348	0.002656
sagen	0.009851	0.007302	0.002549
hab	0.004085	0.001739	0.002346
Praxis	0.002643	0.000348	0.002295
Da	0.002883	0.000695	0.002188

Figure 2. Words with large positive difference. Translation: *Und: and, Sachen: things/stuff, mal: one time, Du: you, Sie: she/her/they/you(formal), teilweise: partially, sagen: to say, hab: (I) have, Praxis: (medical) practice, Da: there*



Word	BurnoutFreqNorm	NoBurnoutFreqNorm	Difference
Also	0.005887	0.009388	-0.003501
Arbeit	0.003484	0.006606	-0.003122
Patienten	0.004205	0.007302	-0.003097
Menschen	0.001081	0.004172	-0.003091
Unternehmen	0.000000	0.002782	-0.002782
Nein	0.000000	0.002782	-0.002782
irgendwas	0.000721	0.003477	-0.002756
okay	0.001321	0.003825	-0.002503
Ich	0.012134	0.014604	-0.002470
sage	0.004445	0.006606	-0.002161

Figure 3. Words with large negative difference. Translation: *Also: so/thus, Arbeit: work, Patienten: patients, Menschen: people, Unternehmen: company, Nein: no, irgendwas: something/anything, okay: okay, Ich: I, sage: (I) say*

5. Results

5.1. The Words of Burnout Patients

Figure 4 and 5 show the word clouds for the ten most used nouns for the text labeled as *burnout* and *noburnout* respectively. The size of the font reflects the quantity (number of times) of the words appearing in the dataset.

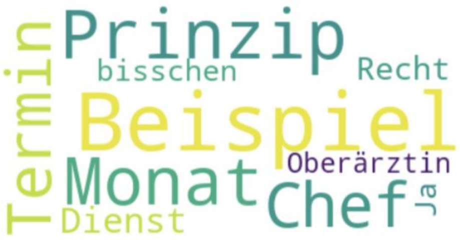


Figure 4. Top ten nouns from the class *burnout*; translation to English: *Beispiel: example, Prinzip: principle, Monat: month, Chef: boss, Termin: appointment/deadline, Dienst: shift/service, Recht: right/law, bisschen: a bit, Ja: yes, Oberärztin: senior physician (female)*

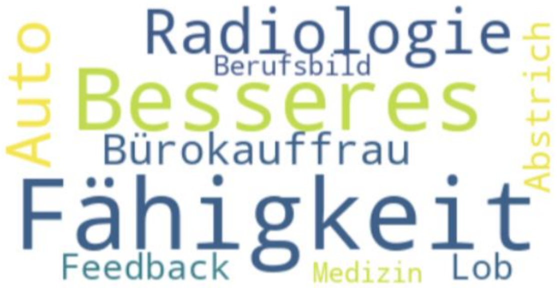


Figure 5. Top ten nouns from the class *noburnout*; translation to English: *Fähigkeit: capability/competence, Besseres: something better, Radiologie: radiology, Auto: car, Bürokauffrau: office clerk (female), Lob: praise, Feedback: feedback, Abstrich: tradeoff, Berufsbild: job profile, Medizin: medicine*

We observe in the results that for burnout patients, stress-related topics are of high importance (e.g. senior physician, shift/service, boss), as well as factors concerning the time (e.g. month, appointment/deadline). On the other side, for the texts from the class *noburnout*, positive work-related topics are more common (e.g. competence, praise, feedback).

5.2. Burnout Detection Model

Figure 6 depicts the results of our experiments, relating the number of features considered to the accuracy obtained for models using linear, RBF, polynomial and sigmoid kernel SVM. The random baseline is 73.4% (percentage of burnout cases), since the test and training set contain a majority of records of the class *burnout*.

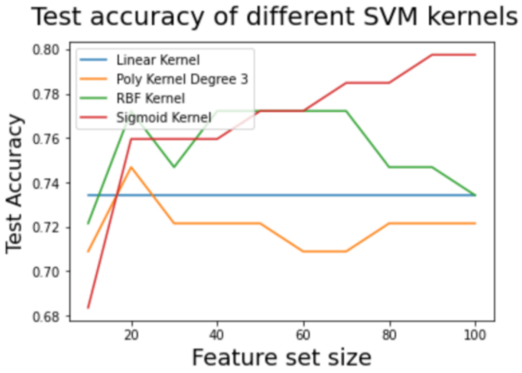


Figure 6. Test accuracy for different feature set sizes of the SVM classifiers using linear, RBF, polynomial and sigmoid kernels

The sigmoid kernel performs the best (79.7%) at around feature set size 90. It can be observed that increasing features continues to improve the accuracy in case of sigmoid kernel. We have shown the features set size up to 100, after this point increasing the feature set size caused a fall in test accuracy for the sigmoid kernel. The linear kernel likely predicts only one class and therefore lies on the random baseline. This indicates that our data is likely not linearly separable. The polynomial kernel of degree 3 goes up to an accuracy of 74.7% at around feature set size 20. The RBF kernel goes up to 77.2% at feature set size 40 and then stabilizes before falling down.

We are thus able to train a classifier to detect burnout in text snippets with an accuracy clearly above the random baseline. Figure 7 shows the confusion matrix for the SVM with sigmoid kernel for a feature set size of 90.

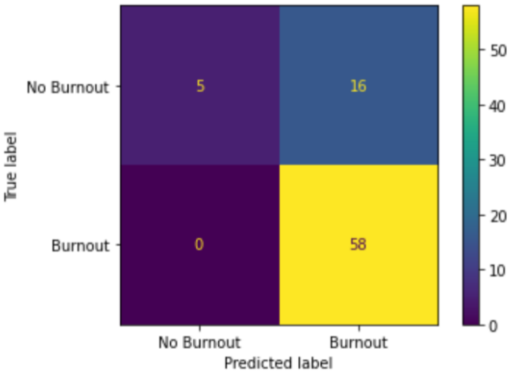


Figure 7. Confusion matrix of the SVM classifier with sigmoid kernel for a feature set size of 90.

6. DISCUSSION

In the feature selection we identified the most typical words for the classes *burnout* and *noburnout*. Interestingly, the word *Du (engl: you)* is a feature for the class *burnout*, whereas the word *Ich (engl: I)* is a feature of the class *noburnout*. This is counterintuitive since previous work has shown that the common use of first-person singular can be interpreted as a sign of depression [20] [21].

Our burnout classifier reaches an accuracy of 79.7%, which is clearly over the baseline of 73.4%, even though the dataset contains a limited number of records. We therefore see a large potential for text-based approaches to identify burnout.

The confusion matrix in Figure 7 indicates that our approach is more likely to present false positives (i.e. patients **not** having burnout being classified as having burnout) and less false negatives (i.e. patients having burnout, being classified as **not** having burnout). This can be partially explained by the fact that the dataset contains more records for the class *burnout*. In a clinical setting, this is preferable, since symptoms of burnout often overlap with other diseases or syndromes. We prefer that a doctor checks closely a patient because the tool indicated a potential burnout, and finds out he/she does have another syndrome, rather than following the prediction of the tool leading to the release of a person that might need further help.

The interaction between the tool and the clinical professional is very important. The machine can provide tools for clinical professionals, but cannot replace them. Machines are not able to take ethical decisions, and cannot carry the consequences of their decisions. Therefore, a good tool provides decision-support in a clinical context, in a similar way nowadays inventories are doing. Such inventories, as for example the Maslach Burnout Inventory, have known limitations (e.g. users faking their results [10], extreme response bias [11], defensiveness and social desirability bias [12]). Free-text questions or interview transcripts can provide interesting new possibilities in the detection of burnout. The work presented in this paper confirms that such approaches are feasible and we assume that with more data, even better results than the ones presented in this paper can be achieved.

Based on the results presented in this paper, the following challenges will need to be addressed in future work. Currently, the dataset does not differentiate between men and women, since the data is completely anonymized. However, in future work more detailed profiles of patients will need to be considered and carefully examined. It has been shown that the way men and women communicate is different [55], and that gendered wording can have an important impact. Due to the available data, we currently only include German extracts from interviews in the dataset. We expect to connect to clinical partners and extend our dataset with extracts in English and French. This will allow us to target a larger community and we plan to explore whether our findings for German are also applicable for the other languages.

7. CONCLUSION

Most work using NLP to detect depression uses data from social media, whereas the specific case of burnout is rarely addressed. In this paper, we presented BurnoutWords, a dataset based on interview extracts from burnout patients in the German language. The dataset was assembled with data from previous research and not from social media, as opposed to most of the existing research in the field. It allows first insights into the wording of burnout patients, and will be extended in the future with additional data. We also plan to investigate on the wording of burnout patients in other languages.

Since the existing work in the field of burnout detection is very limited, a comparative evaluation to similar datasets is currently not possible. Therefore, in future work, we want to address also the area of burnout detection in social media, to provide comparable validation measures of our approaches.

We showed that upon such data, a classifier using Support Vector Machines with an accuracy clearly higher than the random baseline can be trained. With the sigmoid kernel, an accuracy of almost 80% was achieved, as compared to the random baseline of 73.4% (percentage of burnout cases in the dataset). Given that this first version of the dataset is very limited, the result is very promising. This work creates the foundation for future work in the field and new methods for clinical burnout detection.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funding for this work by the Swiss National Science Foundation in the context of the Spark call.

REFERENCES

[1] American Psychological Association. 2019. Stress in america 2019.

[2] American Psychological Association. 2020. Stress in america 2020: a national mental health crisis.

[3] Elie Azoulay, Jan De Waele, Ricard Ferrer, Thomas Staudinger, Marta Borkowska, Pedro Povoa, Katerina Iliopoulou, Antonio Artigas, Stefan J Schaller, Manu Shankar Hari, et al. 2020. Symptoms of burnout in intensive care unit specialists facing the covid-19 outbreak. 10(1):1–8.

[4] Takahiro Matsuo, Daiki Kobayashi, Fumika Taki, Fumie Sakamoto, Yuki Uehara, Nobuyoshi Mori, and Tsuguya Fukui. 2020. Prevalence of health care worker burnout during the coronavirus disease 2019 (covid-19) pandemic in japan. *JAMA network open*, 3(8):e2017271–e2017271.

[5] Ferdinand Jaggi. 2019. *Burnout praxisnah*. Lehmanns Media.

[6] Wilmar B Schaufeli, Christina Maslach, and Tadeusz Marek. 2017. The future of burnout. In Professional burnout, pages 253–259. Routledge.

[7] Irvin Sam Schonfeld and Renzo Bianchi. 2016. Burnout and depression: two entities or one? *Journal of clinical psychology*, 72(1):22–37.

[8] Sybille Hautle. 2012. Das burnout-syndrom: Erhebung typischer merkmale zur herleitung diagnostischer fragen für ein selbstbeurteilungsinstrument.

[9] M Burisch. 2010. Das burnout-syndrom (4. aktual. aufl.).

[10] Christine Elizabeth Lambert. 2013. *Identifying faking on self-report personality inventories: Relative merits of traditional lie scales, new lie scales, response patterns, and response times*. Ph.D. thesis.

[11] Gaël Brulé and Ruut Veenhoven. 2017. The '10 excess' phenomenon in responses to survey questions on happiness. *Social Indicators Research*, 131(2):853–870.

[12] Margot M Williams, Richard Rogers, Allyson J Sharf, and Colin A Ross. 2019. Faking good: An investigation of social desirability and defensiveness in an inpatient sample with personality disorder traits. *Journal of personality assessment*, 101(3):253–263.

[13] Christina Maslach, Susan E Jackson, Michael P Leiter, WB Schaufeli, and RL Schwab. 1996. Maslach burnout inventory manual. menlo park. *CA: Mind Garden*, pages 191–218.

[14] C Maslach, SE Jackson, MP Leiter, WB Schaufeli, and RL Schwab. 1986. Maslach burnout inventory instruments and scoring guides forms: General, human services, & educators. *Health and Quality of life Outcomes*, 7:31.

[15] Elliot Aronson, Ayala M Pines, and Ditsa Kafry. 1983. Ausgebrannt. *Vom überdruß zur selbstentfaltung*.

[16] Viviana Abati. 2007. *Burnout: Erkennen - vorbeugen - verhindern*. SPEKTRAmedia.

[17] Cary Cherniss. 1980. *Professional burnout in human service organizations*. Praeger Publishers.

[18] Nadee Seneviratne and Carol Espy-Wilson. 2020. Deep learning based generalized models for depression classification. *arXiv preprint arXiv:2011.06739*.

[19] Andrew G Reece and Christopher M Danforth. 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6:1–12.

[20] Michelle Morales, Stefan Scherer, and Rivka Levitan. 2017. A cross-modal review of indicators for depression detection systems. In *Proceedings of the fourth workshop on computational linguistics and clinical psychology—From linguistic signal to clinical reality*, pages 1–12.

[21] Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133.

[22] Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. *arXiv preprint arXiv:1709.01848*.

[23] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49.

[24] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, 28(6):447– 455.

[25] H Andrew Schwartz, Johannes Eichstaedt, Margaret Kern, Gregory Park, Maarten Sap, David Stillwell, Michal Kosinski, and Lyle Ungar. 2014. Towards assessing changes in degree of depression through facebook. In *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 118–125.

[26] Lewis R Goldberg. 1990. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

[27] Yaakov Ophir, Christa SC Asterhan, and Baruch B Schwarz. 2019. The digital footprints of adolescent depression, social rejection and victimization of bullying on facebook. *Computers in Human Behavior*, 91:62–71.

[28] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3267–3276.

[29] Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 626–638.

[30] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.

[31] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. 2018. Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 88–97.

[32] Chenhao Lin, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. 2020. Sensemood: Depression detection on social media. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 407–411.

[33] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understand- ing. *arXiv preprint arXiv:1810.04805*.

[34] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. 2017. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844.

[35] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. 1957. *The measurement of meaning*. 47. University of Illinois press.

[36] Mika Mäntylä, Bram Adams, Giuseppe Destefanis, Daniel Graziotin, and Marco Ortu. 2016. Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity? In *Proceedings of the 13th international conference on mining software repositories*, pages 247–258.

[37] Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preotiuc-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44):11203–11208.

[38] Sharath Chandra Guntuku, H Andrew Schwartz, Adarsh Kashyap, Jessica S Gaulton, Daniel C Stokes, David A Asch, Lyle H Ungar, and Raina M Merchant. 2020. Variability in language used on social media prior to hospital visits. *Scientific reports*, 10(1):1–9.

[39] David N Milne, Glen Pink, Ben Hachey, and Rafael A Calvo. 2016. Clpsych 2016 shared task: Triaging content in online peer-support forums. In *Proceedings of the third workshop on computational linguistics and clinical psychology*, pages 118–127.

[40] Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. Clpsych 2015 shared task: Depression and ptsd on twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 31– 39.

[41] Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 51–60.

[42] Ge Ge Jackie Chen. 2014. *Visualizations for mental health topic models*. Ph.D. thesis, Massachusetts Institute of Technology.

[43] Jonathan Gratch, Ron Artstein, Gale M Lucas, Giota Stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, et al. 2014. The distress analysis interview corpus of human and computer interviews. In *LREC*, pages 3123–3128.

[44] Michel Valstar, Jonathan Gratch, Björn Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. 2016. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th international workshop on audio/visual emotion challenge*, pages 3–10.

[45] K Kroenke, RL Spitzer, and JBW Williams. 2001. The patient health questionnaire (phq-9)– overview. *J9 Gen Intern Med*, 16:606–16.

[46] Michel Valstar, Björn Schuller, Kirsty Smith, Timur Almaev, Florian Eyben, Jarek Krajewski, Roddy Cowie, and Maja Pantic. 2014. Avec 2014: 3d dimensional affect and depression recognition challenge. In *Proceedings of the 4th international workshop on audio/visual emotion challenge*, pages 3– 10.

[47] Michelle Renee Morales and Rivka Levitan. 2016. Speech vs. text: A comparative analysis of features for depression detection systems. In *2016 IEEE spoken language technology workshop (SLT)*, pages 136–143. IEEE.

[48] Sameer Pradhan, Wendy Chapman, Suresh Man, and Guergana Savova. 2014. Semeval-2014 task 7: Analysis of clinical text. In *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Citeseer.

[49] Jackson M Steinkamp, Wasif Bala, Abhinav Sharma, and Jacob J Kantrowitz. 2020. Task definition, annotated dataset, and supervised natural language processing models for symptom extraction from unstructured clinical notes. *Journal of biomedical informatics*, 102:103354.

[50] Edith Rahner. 2011. *Das Burnout-Syndrom bei Ärzten: eine qualitative Studie zur Selbstwahrnehmung von Ursachen und Lösungsansätzen*. Ph.D. thesis.

[51] Simone Albrecht. 2008. Burnout–der weg danach. *Burnout im Lichte von Theorie und Praxis. VDM, Saarbrücken*.

[52] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

[53] Corinna Cortes and Vladimir Vapnik. 1995. Support- vector networks. *Machine learning*, 20(3):273–297.

[54] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. 2001. *The elements of statistical learning*, volume 1. Springer series in statistics New York.

[55] Danielle Gaucher, Justin Friesen, and Aaron C Kay. 2011. Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of personality and social psychology*, 101(1):109.

AUTHORS

Sukanya Nath

Sukanya Nath is currently a PhD student at the Computational Linguistics group at the University of Neuchâtel. Her area of research is multiple authorship attribution, authorship verification and profiling. She is also a research assistant under Dr. Mascha Kurpicz-Briki at the Bern University of Applied Sciences for the BurnoutWords project. She has a master in Computer Science (Data Science Specialization) from the University of Bern. Contact: sukanya.nath@bfh.ch

Mascha Kurpicz-Briki

Dr. Mascha Kurpicz-Briki obtained her PhD in the area of energy-efficient cloud computing at the University of Neuchâtel. After her PhD, she worked a few years in industry, in the area of open-source engineering, cloud computing and analytics. She is now professor for data engineering at the Bern University of Applied Sciences, investigating how to apply digital methods and in particular natural language processing to social and community challenges. Contact: mascha.kurpicz@bfh.ch

FEDERATED LEARNING WITH RANDOM COMMUNICATION AND DYNAMIC AGGREGATION

Ruolin Huang, Ting Lu, Yiyang Luo, Guohua Liu and Shan Chang

College of Computer Science and Technology,
Donghua University, Shanghai, China 201620

ABSTRACT

Federated Learning (FL) is a setting that allows clients to train a joint global model collaboratively while keeping data locally. Due to FL has advantages of data confidential and distributed computing, interest in this area has increased. In this paper, we designed a new FL algorithm named FedRAD. Random communication and dynamic aggregation methods are proposed for FedRAD. Random communication method enables FL system use the combination of fixed communication interval and constrained variable intervals in a single task. Dynamic aggregation method reforms aggregation weights and makes weights update automately. Both methods aim to improve model performance. We evaluated two proposed methods respectively, and compared FedRAD with three algorithms on three hyperparameters. Results at CIFAR-10 demonstrate that each method has good performance, and FedRAD can achieve higher classification accuracy than state-of-the-art FL algorithms.

KEYWORDS

Federated Learning, Random Communication, Dynamic Aggregation, Self-learning, Distributed Computing.

1. INTRODUCTION

Local devices such as mobile phones own a lot of data. However, due to data privacy and device ability, it is impractical to conduct centralized training at central server by gathering all the data from clients[1]. To address these problems, federated learning (FL)[2] is proposed. FL allows clients to train a joint global model collaboratively while keeping data locally. In this way, FL has advantage of data confidential, distributed storing and computing.

A typical FL[3] system consists of two stages connected by communication, (1) clients train local models with their local private datasets independently, and (2) server aggregates the local models into a joint global model. Since communication and aggregation are two primary performance bottlenecks of FL, interests in these two areas have increased.

In this paper, we designed a new FL algorithm named FedRAD. Random communication and dynamic aggregation methods are proposed for FedRAD. By using the combination of fixed communication interval and constrained variable intervals, random communication method enables FL system try various intervals in a single task so that we hope it can improve model accuracy. By presenting a new form of aggregation weights and making weights update automately, dynamic aggregation method enables system utilizes mutural impact between global model and local models, so as to increase task accuracy. In addition, dynamic aggregation

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 193-202, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111816

method not only puts additional but a little computation burden on powerful server instead of resource-constrained clients, but also can apply to modern CNN. We evaluate two methods respectively, and compare FedRAD with three algorithms on three hyperparametes. Results at CIFAR-10 demonstrate that each method outperforms compared way, and FedRAD can obtain higher accuracy than state-of-the-art FL algorithms.

We organize this paper as follows. Section 2 states related works. Section 3 states two proposed methods respectively and provides an overview of FedRAD constructed by two methods, the performance of two methods and FedRAD algorithm is evaluated in section 4, section 5 summarizes the whole content of this paper.

2. Related Work

Existing algorithms mainly focus on reducing the amount of parameters in communication. For example, Suresh et al.[5] uses a constant number of model in communication, and Horvath et al.[6] ignores the mantissa of parameters in model when communicating. However, these algorithms reduced the amount of parameters at the cost of decreasing task accuracy. It is know that accuracy is important to classification task such as identification task in self-driving car[7]. In order to improve model accuracy, many works can be focused on communication scheme[3]. For example, server and clients communicate once per fixed number of interval in these algorithms[5,6], which, is called fixed communication scheme. Wang et al.[8] proved that communication interval can affect the performance of the FL system. However, it is hard to get the proper communication interval before training.

Aggregation is another performane bottleneck in FL. McMahan et al.[2] proposed the standard aggregation algorithm federated averaging (FedAvg). FedAvg provides an averaging aggregation method that aggregates parameters of local models by setting weights relevant to the sizes of local datasets. Xiao et al.[9] proved that averaging parameters may not be the optimum way. In order to improve the performance of FedAvg, Sahu et al.[10] presented FedProx keeping local updates close to the original global model by adding a proximal term to the client cost functions. Although it considers the impact of global model on local updating, it increases the amount of computation on clients. To reduce the computation on clients, Yurochkin et al.[11,12] proposed Probabilistic Federated Neural Matching (PFNM) by matching the neurons of client models before averaging. However, it only works with simple architectures, e.g. fully connected network. Obviously, all of these aggregation algorithms can not consider the impact of global model on local updating, reduce the amount of computation on clients or apply to modern architecture (e.g. convolutional neural network, CNN) at the same time.

3. Methods

In this section we introduce FedRAD. First, we state the proposed random communication method. Then, we introduce proposed dynamic aggregation method. Finally, we provide an overview of FedRAD constructed by two methods.

3.1. Random Communication

Since it is hard to use the proper communication interval before training, we first enable FL system use fixed communication scheme in the first half of training by using fixed interval. Then, system uses proposed random communication scheme in the second half of training by using constrained variable intervals. In terms of large interval leads to deterioration of the task accuracy

[8], we add a constraint on variable communication interval. Random communication method is as follows.

Let E denote total training epochs, f denote the fixed number of local training epochs, $set_t = \{t \in N^* \mid 1 \leq t \leq E\}$ denote the set of the training epoch, set_e denote the set of epochs communication happens, and set_{int} denote the set of communication intervals.

First, we partition set_t into three subsets according to f and the middle training epoch $\lfloor E/2 \rfloor$. In terms of f is not necessarily divisable by E , these three subsets are described as $set_{t1} = [1, \lfloor E/2f \rfloor * f]$, $set_{t2} = [\lfloor E/2f \rfloor * f, \lfloor E/f \rfloor * f]$ and $set_{t3} = [\lfloor E/f \rfloor * f, E]$ respectively.

Then, we construct set_{int} according to three subsets. For set_{t1}, we add f of quantity $\lfloor E/2f \rfloor$ into set_{int} . In this way, FL system train as fixed communication scheme. For set_{t2} , we sequentially take one subset of set_{t2} with the length of f without replacement. During each taking, we select a element randomly of taken subset and add it into set_e . After the final selection, we set the prior element of the first element in set_e as value 0, and sequentially calculate the difference value between each element in set_e with its prior element. The purpose of this way is to add a constraint on variable communication interval. i.e. By limitting the selected element (i.e, the selected communication epoch e) in the range of length f , we make the variable communication intervals the maximum value as $2f - 1$ and the minmum value as 1, so that prevent the FL system training with too much large interval. In this way, FL system trains as proposed dynaimc communication scheme. For set_{t3}, we do nothing. If f can be divisable by E , set_{t3} will not exist. Otherwise, for each $t \in set_{t3}$, the FL system training as fixed communication scheme does not communicate. So in order to compare proposed method with fixed method accurately in experiment, we do nothing on set_{t3}.

Finally, the construted set_{int} is applied to server. Server broadcasts global model together with one taken element in set_{int} to clients. The element should be taken in order without replacement. Clients then train the global model locally, setting the value of broadcast element as local training epochs.

Obviously, FL system in our method will follow fixed communication sheme when $t \in set_{t1}$, otherwise it will follow random scheme. In terms of combining two schemes to get good performance, random scheme has notable efficiency. Therefore we call this combination as random communication method. The algorithm is described as follows:

Algorithm 1 : Random Communication

```
Input: The total training epochs $E$, the fixed number of local training epochs $f$.
Initialize the 'middle' partition epoch $\lfloor E/2f \rfloor * f$, $set_e[1]=0$, $set_{int}$, $i=1$ and $j=1$
for $t$ in range(1, $E$) do:
    if $1 \le t \le \lfloor E/2f \rfloor * f$ and $t/f == 0$:
        $set_{int}[i++] = f$
    elif $t = \lfloor E/2f \rfloor * f + 2$:          # 'randint(a,b]' denotes selecting a integer in range of (a, b] randomly
        $set_e[++j] = randint(\lfloor E/2 \rfloor, t]$
        $set_{int}[i++] = set_e[j] - set_e[j-1] + 1$
    elif $t > \lfloor E/2f \rfloor * f + 2$ and $t/f == 0$:
        $set_e[++j] = randint(t - f, t]$
        $set_{int}[i++] = set_e[j] - set_e[j-1] + 1$
Output: $set_{int}$
```

3.2. Dynamic Aggregation

In terms of averaging is not the optimum way for aggregation [8], we proposed a new form for aggregation weights in this paper firstly. Then, by using a simple neureul network, aggregation weights can update automatically. This method is as follows.

First, we reform the aggregation weights based on fomula in FedAvg[2] as:

$$GM = \sum_{k=1}^{n} W[k] \cdot \frac{N_k}{N} \cdot M_k \qquad (1)$$

Where GM denotes global model, n denotes the amount of client, N_k denotes size of local dataset D_k, N denotes size of all D_k, M_k denotes local model and $W[k]$ denotes proposed weight of local model M_k. $W[k]$ will be described initially as:

$$W[k] = \frac{acc_k^e \cdot acc_k^e \cdot}{\sum_{k=1}^{n} acc_k^e \cdot acc_k^e \cdot} \qquad (2)$$

Where acc_k^e denotes task accuracy of M_k at current communication epoch e. In this way, we extend the impact of local models which have better performances.

Then, to make $W[k]$ update automatically, here we use a 2-layer neural network $NeuNet$ since neural network has the advantage of self-learning[13]. This work focused on how to set the optimization goals in $NeuNet$. In terms of $NeuNet$ and global model share the purpose of decreasing task loss, $NeuNet$ can use this shared purpose for self-learning. To that aim, $NeuNet$ requires a connection between its output layer and FL system (as shown in Figure 1). This connection is used to one-way deliver a loss value $SysLoss$ from server to output layer of $NeuNet$, to set $SysLoss$ as the loss for back propagation in $NeuNet$. The delivered $SysLoss$ is the average loss of global model after the last communication, that is, the average loss of each client tested on local test dataset before local training. In this way, server do not need to gather local private data from clients to test the loss of global model. Accordingly, let inputs of $NeuNet$ as weights of all M_k and loss as $SysLoss$, the updated aggregation weight $W[k]$ for M_k can be calculated as:

$$W[k] <= W[k] - \eta \cdot \frac{\partial SysLoss}{\partial W[k]} \quad (3)$$

Fomula 3 is back propagation fomula in NN[14], where η denote learning rate of $NeuNet$. Thereby, dynamic aggregation method utilizes the impact of global model on local models complementarily besides considers the influence of local models on global model according to Formula 2. The algorithm is as follows.

Algorithm 2 : Dynamic Aggregation

```
Input: The set of accuracy on local models $SET_{ACC}$, the set of local models $SET_M$, and the set of task loss $SET_{LOSS}$.
Initialize 2-layer neural network model $NeuNet$, and the average task loss $Sysloss = 0$
for $k$ in range(1, $|SET_M|$+1) do:
    $W[k] = \frac{acc_k \cdot acc_k}{\sum_{k=1}^{|SET_M|n} acc_k \cdot acc_k}$
    $SysLoss$ += $SET_M[k]$
$SysLoss = SysLoss / |SET_M|$
set $SysLoss$ as loss for back propagation in $NeuNet$, and get the updated $W[|SET_M|]$
$GM = \sum_{k=1}^{|SET_M|} W[k] \frac{N_k}{N} \cdot M_k$
Output: $GM$
```

3.3. FedRAD

Based on the random communication method designed in section 3.1 and the dynamic aggregation method in section 3.2, a new federated learning algorithm named FedRAD is proposed in this paper. Based on typical FL system[3], FedRAD consists of one server and several clients. The global model in server and the local model in each client adopt the same model, such as AlexNet. As can be seen from the server block in Figure 1, both methods proposed in this paper are applied to server, which has the advantage of placing the extra but small computation burden on server rather than the resource-constrained clients.

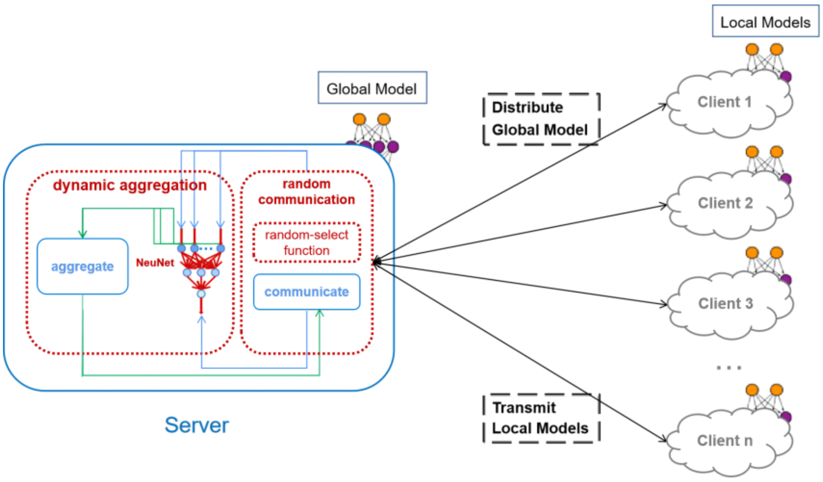


Figure 1. Structure of FedRAD

This system repeats following four steps until the training end, (1) server distributes global model and a communication interval INT generated by random communication method to clients, (2) clients first use global model to test on the local test dataset to get the loss, then train on global model using local training dataset with *INT* epochs, and test on local models to get task accuracy, (3) clients report their trained models, loss and accuracy to server, (4) server aggregates local models into a new global model according to dynamic aggregation method. The algorithm is as follows.

Algorithm 3 : FedRAD *t* is current epoch; *GM* is the global model; *Interval* is the set of communication intervals; *E* is the total of training epoch; *f* is the fixed communication interval; M_k is the local model; D_k is the local training dataset of client *k; T_k* is the local testing dataset of client *k*.

```
Server：
if $t$ == 1:
    Initialize $GM_0$
    $Interval$ <= Random Communication (E, f )
else:
    Receive $M_k$, $acc_k$, $loss_k$ clients report
    construct $SET_{ACC}$ with all $acc_k$, $SET_M$ with all $M_k$, and $SET_{LOSS}$ with all $loss_k$
    $GM$ <= Dynamic Aggregation ($SET_{ACC}$, $SET_M$, $SET_{LOSS}$)
Distributes $GM_0$ or $GM$, and $Interval[i++]$ to clients        #$i$ denotes a increment variable, which initialized as 0
Client  k：
initialize $M_k$ <= $GM$,  $e$ <= $Interval$
tests $M_k$ on $T_k$ to get $loss_k^0$
trains $M_k$ on $D_k$ with $e$ epochs locally to get  $acc_k^{+e}$, $M_k^{+e}$
communicate with server to report $loss_k^0$, $acc_k^{+e}$, $M_k^{+e}$
```

4. EXPERIMENTS

In this section, we first state experiment settings in section 4.1. Then, we evaluate two proposed methods respectively and present an evaluation of FedRAD compared with three algorithms on three hyperparametes in section 4.2.

4.1. Setup

4.1.1. Task and dataset

Our training task is image classification on CIFAR-10. We separate smaller datasets of various sizes from the training set and further use data augment method to simulate several conditions. Test images are used for a global test after each round. For different models, we record the test accuracy as the metric to compare model performance.

4.1.2. Baselines

For random communication method, we compare it with typical fixed method. For dynamic aggregation method, we compared it with FedAvg and FedProx. For FedRAD, we compare it with three algorithms FedAvg, FedProx and centralized training. In addition, in order to ensure the accuracy of comparison results, modern CNN architecture MobileNet is used as learning model among all comparison algorithms.

4.2. Results

4.2.1. Performance of two methods

Random communication Since FedAvg and FedProx both communicate as fixed scheme, here we compared proposed random method with typical fixed method. Fig. 2(a) and Table 1 show the compared results, where “fixed” denotes a FL system with fixed communication method, and

“Random” denotes system with random communication method with the same communication amount T under the setting $c = *$. We can see in Fig. 2(a) that enlarging f does not always work for all conditions, which matches the conclusion made in the previous work[2]. In addition, since it is hard to know the proper number of communication interval for certain task before training, FL system with random communcation method can use various intervals in training, so that improve model performance. It can be seen in Fig. 2(a), in case of assigning different communication interval f as 4, 5, 6, and 7, random group performs better than the fixed group, which demonstrates the availability of proposed method.

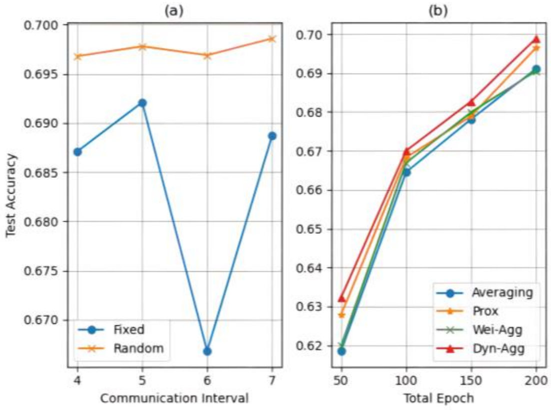


Figure 2. Comparative experiments results. (a) The performance of the model depending on the communication methods. (b) The performance of the model depending on the aggregation methods.

Dynamic aggregation Since FedAvg and FedProx have different aggregation methods, here we compared proposed dynamic method with FedAvg and FedProx. Fig. 2(b) and Table 1 show the compared results, where “Averaging” denotes a FL system with federated averaging method, “Prox” denotes system with FedProx method, “Wei-Agg” denotes system with proposed form of aggregation weights method, and “Dyn-Agg” denotes system with proposed dynamic aggregation method. It could be seen that weighted aggregation shows a little better performance than FedAvg, but it could not reach the height of FedProx. Dynamic aggregation method shows more flexible and efficient ability than others, which demonstrates the availability of proposed method.

Table 1. Trained summary on MobileNet over CIFAR-10 as shown in Figure 2.

hyperparameter	**Algorithm**	FedAvg	**Ran-Com**	FedProx	**Wei-Agg**	**Dyn-Agg**
Communication interval	4	68.71	**69.68**			
	5	69.21	**69.78**			
	6	66.68	**69.69**			
	7	68.87	**69.86**			
Iteration amount (* client amount)	5	61.85		62.79	**62.00**	**63.23**
	10	66.69		66.83	**66.69**	**67.00**
	15	67.99		67.89	**68.00**	**68.27**
	20	69.11		69.55	**69.05**	**69.89**

4.2.2. Performance of FedRAD

Dataset size It is known that model performs better when more training data is available. To simulate this scenario, we first partition the entire training CIFAR-10 dataset into n parts, where n denotes the client amount. We then augment data of original entire dataset and concatenate

n parts with data-augment parts for each client's demand. Using this strategy, we partition the training set into n sub-datasets containing 5/8/10/20 thousand (k) points each. Figure 3(a) and Table 2 show that centralized training performs better than others when dataset size is 5k, while the gap is closing as size increasing. The augment methods we used are not sufficient to fill data variety might account for the result[15]. Still it could demonstrate that FedRAD performs better than FedAvg and FedProx, and obtains comparable even slightly higher accuracy than centralized training.

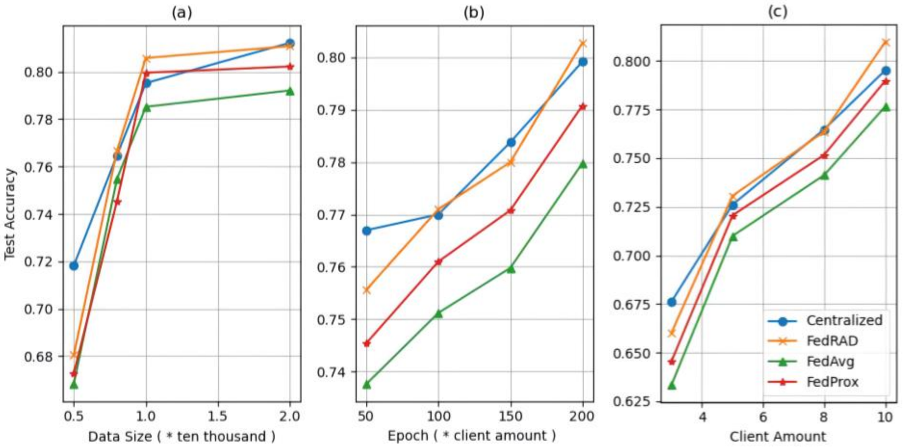


Figure 3. Comparative performances among models over three hyperparameters. (a) Influence of dataset size. (b) Influence of iteration amount. (c) Influence of client amount.

Iteration amount Training iteration also matters to model performance. Thus we further test influence of iteration amount. Results (Fig. 3(b) and Table 2) show that FedRAD obtains higher accuracy than FedAvg and FedProx, and achieves similar performance with centralized training.

Client amount We already know that model performs better as increasing the amount of training data and epoch. Challenge here is when new clients participate a FL system which already works for a while, they may not adapt to global model at short notice, which can decrease the model accuracy despite the growing data size and iteration amount. To simulate this scenario, we first let 3 clients first join the FL system, then add 2, 3, 2 clients in system respectively in each 200 epochs. Results (Fig. 3(c) and Table 2) show that FedRAD obtains higher accuracy than FedAvg and FedProx when handling new participants.

To sum up, each method we proposed outperforms typical or state-of-the-art methods. FedRAD consists of two methods obtains higher task accuracy compared with FedAvg and FedProx, and achieves similar performance with centralized training.

Table 2. Trained summary on MobileNet over CIFAR-10 as shown in Figure 3.

hyperparameter	**Algorithm**	FedAvg	FedProx	Centralized	**FedRAD**
Dataset size (* thousand)	5	66.82	67.27	68.04	**71.82**
	8	75.48	74.54	76.67	**76.48**
	10	78.52	79.96	80.58	**79.52**
	20	79.21	80.22	81.08	**81.22**
Iteration amount (* client amount)	5	73.76	74.54	75.56	**76.70**
	10	75.12	76.10	77.10	**77.00**
	15	75.98	77.08	78.00	**78.38**
	20	77.98	79.08	80.28	**79.92**

Client amount	4	63.34	64.55	66.02	**67.62**
	6	71.00	72.06	73.06	**72.61**
	8	74.13	75.17	76.37	**76.48**
	10	77.66	78.98	80.98	**79.52**

5. SUMMARY

Conclusions This paper proposed a new federated learning algorithm with random communication and dynamic aggregation (FedRAD). Random Communication method uses the combination of fixed communication interval and constrained variable intervals that FL system can try various intervals in a single task, so as to improve model performance. Dynamic aggregation method reforms aggregation weights and updates weights automately that considers mutural impact between global model and local model, aiming to increase task accuracy. Thus, FedRAD consolidates several advantages into a single framework. It considers mutural impact between global model on local model, puts additional computation burden on powerful server instead of resource-constrained clients, applies to modern architectures, and all while improves task accuracy. Though, the proposed random communication method can not address the problem of performance divergence caused by too large communication interval still as well as state-of-the-art methods, and the proposed dynamic aggregation method is a whole-wise way compared with element-wise way[11]. Thus, further works can be focused on as follows.

Future works For communication, we will try to reform communication scheme to a gradual way with introducing incremental learning (IL)[16]. Since IL has the advantage of promoting the connection of old and new tasks[17], FL system with IL can face the problem of unbalanced data distribution. And for aggregation, we will devote to design a element-wise method used for modern complex CNNs that registering neurons before aggregating[11], so as to improve model performance. Thus, we will devote to design a more effective and flexible FL algorithm than popular algorithms.

ACKNOWLEDGEMENTS

This work is supported by Shanghai Municipal Natural Science Foundation (Grant No. 21ZR1401200, Grant No.18ZR1401200), Special Fund for Innovation and Development of Shanghai Industrial Internet (Grant No. 2019-GYHLW-01004). (Corresponding author: Ting Lu)

REFERENCES

[1] Barrachina, S. , A Castelló, M Catalán, Dolz, M. F. , & Mestre, J. I. . (2021) “Pydtnn: a user-friendly and extensible framework for distributed deep learning”, The Journal of Supercomputing(4).

[2] Mcmahan, H. B. , Moore, E. , D Ramage, Hampson, S. , & Arcas, B., (2017) “Communication-efficient learning of deep networks from decentralized data”, In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp1273-1282.

[3] Zhang, C. , Xie, Y. , Bai, H. , Yu, B. , & Gao, Y. , (2021) “A survey on federated learning”, Knowledge-Based Systems, Vol. 216, No. 1, pp106775.

[4] Peter, Kairouz. , H. Brendan, McMahan. , Brendan, Avent. , Aurélien, Bellet. , & Mehdi, Bennis. , (2019) “Advances and Open Problems in Federated Learning”, arXiv preprint, arXiv:1912.04977.

[5] Suresh, A. T. , Yu, F. X. , Kumar, S. , & Mcmahan, H. B. , (2017) “Distributed mean estimation with limited communication”, InProceedings of the 34th International Conference on Machine Learning, Vol. 70, pp3329–3337.

[6] Horvath, S. , Ho, C. Y. , Horvath, L. , Sahu, A. N. , & Richtarik, P. , (2019) “Natural compression for distributed deep learning”, arXiv preprint arXiv:1905.10988.

[7] Cao, D. , Chang, S. , Lin, Z. , Liu, G. , & Sun, D. , (2019) “Understanding Distributed Poisoning Attack in Federated Learning”, IEEE, 25th International Conference on Parallel and Distributed Systems (ICPADS).

[8] Wang, H. , Yurochkin, M. , Sun, Y. , Papailiopoulos, D. , & Khazaeni, Y. , (2020) “Federated learning with matched averaging”, arXiv preprint, arXiv:2002.06440.

[9] Xiao, P. , Cheng, S. , & Stankovic, V. , (2020) “Averaging Is Probably Not the Optimum Way of Aggregating Parameters in Federated Learning”, Entropy, Vol. 22, No. 3, pp314.

[10] Li, T. , Sahu, A. K. , Zaheer, M. , Sanjabi, M. , Talwalkar, A. , & V Smith. , (2018) “Federated optimization in heterogeneous networks”, arXiv preprint, arXiv:1812.06127.

[11] Yurochkin, M. , Agarwal, M. , Ghosh, S. , Greenewald, K. , Hoang, T. N. , & Khazaeni, Y. , (2019) “Statistical model aggregation via parameter matching”, In Advances in Neural Information Processing Systems, 2019a, pp10954–10964.

[12] Yurochkin, M. , Agarwal, M. , Ghosh, S. , Greenewald, K. , Hoang, T. N. , & Khazaeni, Y. , (2019) “Bayesian nonparametric federated learning of neural networks”, In International Conference on Machine Learning, 2019b, pp7252–7261.

[13] Seo, J. W. , Jung, H. G. , & Lee, S. W. , (2021) “Self-augmentation: generalizing deep networks to unseen classes for few-shot learning”, Neural Networks, pp12.

[14] Zhang, D. , & Lou, S. , (2021) “The application research of neural network and bp algorithm in stock price pattern classification and prediction”, Future Generation Computer Systems, Vol.115, pp872-879.

[15] Perez, L. , & Wang, J. , (2017) “The effectiveness of data augmentation in image classification using deep learning”, arXiv preprint, arXiv: 1712.04621.

[16] Li, Z. , & Hoiem, D. , (2017) “Learning without forgetting”, IEEE Transactions on Pattern Analysis & Machine Intelligence.

[17] Eba, B. , Ap, A. , & Ik, B. , (2021) “A comprehensive study of class incremental learning algorithms for visual tasks”, Neural Networks, Vol.135, pp38-54.

AUTHORS

Ruolin Huang received her bachelor’s degree in 2019, from Qufu Normal University, Rizhao, China. She is currently a master student in College of Computer Science and Technology, Donghua University, Shanghai, China. Her research interests include, Federated Learning and Computer Vision.

Ting Lu is currently an associate professor in College of Computer Science and Technology, Donghua University, Shanghai, China. Her research interests include, Wireless Network and Mobile Computing et al.

Yiyang Luo is currently a master student in College of Computer Science and Technology, Donghua University, Shanghai, China.

Guohua Liu is currently a professor in College of Computer Science and Technology, Donghua University, Shanghai, China. His research interests include, Outsourcing database and Privacy Protection et al.

Shan Chang is currently a professor in College of Computer Science and Technology, Donghua University, Shanghai, China. Her research interests include, Internet of Things and Internet of Vehicles et al.

THE EVOLUTION OF VECTOR MACHINE SUPPORT IN THE FIELD OF INTRUSION DETECTION SYSTEMS

Ouafae Elaeraj and Cherkaoui Leghris

L@M, RTM Team, Faculty of Sciences and Techniques Mohammedia,
Hassan II University of Casablanca, Morocco

ABSTRACT

With the increase in Internet and local area network usage, malicious attacks and intrusions into computer systems are growing. The design and implementation of intrusion detection systems became extremely important to help maintain good network security. Support vector machines (SVM), a classic pattern recognition tool, has been widely used in intrusion detection. They make it possible to process very large data with great efficiency and are easy to use, and exhibit good prediction behavior. This paper presents a new SVM model enriched with a Gaussian kernel function based on the features of the training data for intrusion detection. The new model is tested with the CICIDS2017 dataset. The test proves better results in terms of detection efficiency and false alarm rate, which can give better coverage and make the detection more effective.

KEYWORDS

Intrusion detection System, Support vector machines, Machine Learning.

1. INTRODUCTION

Most intrusion detection systems are software defence systems. The main functions of an IDS are to monitor events that occur in a computer system or network, analyse system events, detect activities, and raise an alarm if an intrusion is detected. An IDS can be divided into three functional components [3]: an information source, an analysis engine and a decision maker.

Figure 1 shows the relationships between these three components, the Internet, and the protected systems.

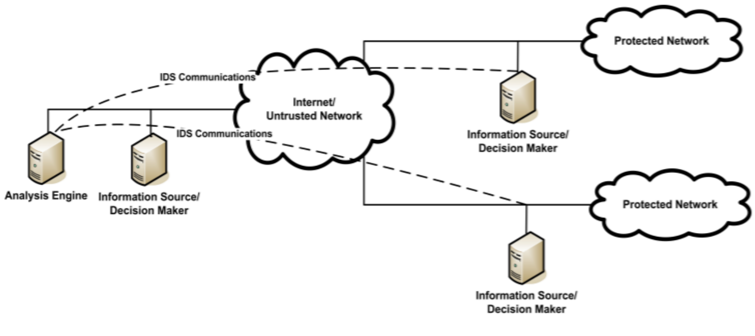


Figure 1. Relation between IDS components

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 203-210, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111817

In the first IDS component, the information source has the task of collecting and pre-filtering all relevant and necessary data to unmask intruders. During the monitoring period, the information source provides a stream of event records for analysis. This component works as an event generator. It detects and monitors different data sources and generates well-formatted event data suitable for further analysis by an analysis engine.

IDS, using anomaly detection, are able to detect unknown attacks, but with the risk of progressive distortion of the profile by repeated attacks. In this context, the creation of anomaly-accurate IDS is of major interest to be able to identify still unknown attacks. Machine learning models can be a solution to create IDS that can be deployed in real computer networks. First of all, an optimization method composed of three steps is proposed to improve the quality of the detection: 1/ data augmentation to rebalance the datasets, 2/ parameter optimization to improve model performance, and 3/ assemble learning to combine the results of the best models. This paper proposes a new SVM model to have a better coverage and make the detection more efficient. The rest of the paper is organized in the following sections: In Section 2, we briefly describe the various existing research of using SVM in IDS. In section 3, we study the proposed solution. The results are described in section 4. At the end, we summarize the paper and address some prospects.

1.1. Machine learning

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that enable computers to evolve behaviours based on empirical data, such as sensor data or databases. Among the main goals of machine learning research is to automatically exercise in recognizing complex patterns and making intelligent data-driven decisions.

ML is used in several applications including search engines, medical diagnostics, text and handwriting recognition, load forecasting, marketing and commercial diagnosis, etc. In 1994, ML was used for the first-time classification of Internet streams in the context of intrusion detection [1]. This is the start of much of the work using ML techniques in classifying Internet traffic. In this paper, the choice was made on the SVM algorithm since it is the best one, in terms of accuracy and processing time, then the KNN and the decision tree [2].

1.2. Support Vector Machines

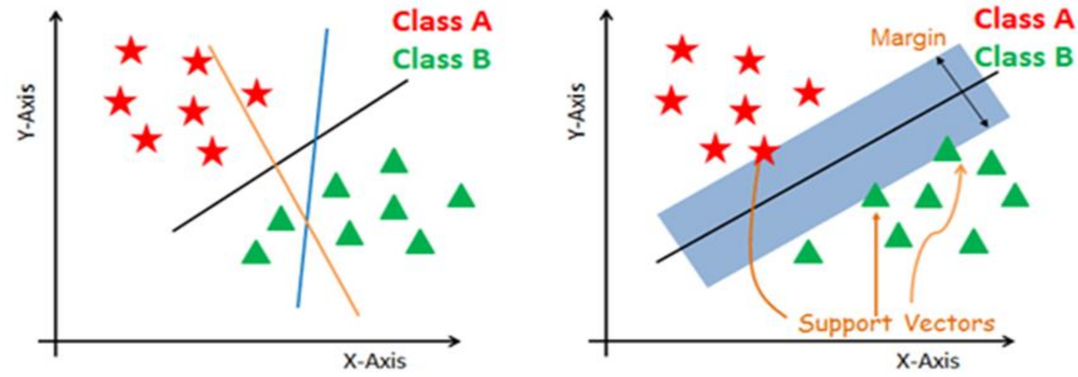
Generally, support vector machines are considered a classification approach, but they can be used in both classification and regression problems. It easily manages several continuous and categorical variables. SVM builds a hyperplane in multidimensional space to separate different classes. SVM generates an optimal hyperplane iteratively, which is used to minimize an error. The main idea of SVM is to find a maximal marginal hyperplane (MMH) that best fragments the dataset into classes.

It cannot be denied that SVM is the best learning algorithm for binary classification. This last, originally a type of classifier model based on a statistical learning technique for classification and regression with a variety of kernel functions, has been successfully applied on several pattern recognition applications. Recently, it has also been applied to information security for intrusion detection. Support Vector Machine has become one of the anomaly intrusion detection techniques because of their good generalization nature. Another advantage of SVM is that it is useful for finding an overall minimum of the actual risk using structural risk minimization, as it can be generalized well with kernel tips even in large spaces with little training conditions. some samples.

1.2.1. Operation of the SVM

The main objective is to separate the data set in the best possible way. The distance between the two closest points is called the margin. The purpose is to select a hyperplane with the maximum possible margin between the support vectors in the given dataset (Figure 2). SVM searches for the maximum marginal hyperplane in the following steps:

1. Generate hyperplanes that separate the classes. Figure 2 (a) shows three black, blue and orange hyperplanes. Here blue and orange have a higher classification error, but black correctly separates the two classes.
2. Select the right hyperplane with the maximum separation of the closest data points, as shown in the figure 2(b).



(a) **(b)**

Figure 2. Operation of the SVM

1.3. IDS using SVM

There are too many reasons why we use SVMs for intrusion detection; The first one is speed: in real time, performance is of primary importance for intrusion detection systems. Any classifier that can potentially runs "fast" is worth considering. The second reason is scalability: SVMs are relatively insensitive to the number of data points and the complexity of the classification does not depend on the dimensionality of the feature space, so they can potentially learn a larger set of models and thus be able to scale better than neural networks.

2. RELATED WORK

Intrusion detection has a long history, dating back to the work of Anderson (1980) [3]. Since long, various discrimination techniques have been proposed, ranging from support vector machines. Several complete systems have been built and operated on real computer systems. However, despite more than 25 years of research, the topic remains popular, partly because of the rapid development of information processing systems and the consequent discovery of new vulnerabilities, but also because of the fundamental difficulties in obtaining an accurate report of an intrusion.

In paper [4], the authors propose a genetic algorithm (GA) to improve the support vector machine (SVM) based intrusion detection system (IDS). As known, SVM is a relatively new classification technique and has shown superior performance to traditional learning methods in many applications. The fusion of GA and SVM has been used in this article to fortify the overall performance of SVM-based IDS. Through this fusion, SVM-based IDSs not only select the "optimal parameters" for SVMs but also an "optimal set" from the feature set.

An anomaly-based IDS using a genetic algorithm and a support vector machine (SVM) with a novel feature selection method is also proposed in [5]. The model adopts a feature selection method based on the genetic algorithm with a change in the fitness function that sums up the size of the data, increases the detection of true positives and simultaneously decreases the detection of false positives. In addition, the computational time for learning will also be reduced remarkably. The results show that the proposed method can simultaneously achieve high accuracy and low false positive rate (FPR). This study proposes a method that achieves more stable features compared to other techniques. The proposed model is experimented and tested on the KDD CUP 99 and UNSW-NB15 datasets.

Network traffic in cloud computing is characterized by a large volume of data, with high levels of redundancy. The efficient correlation-based feature selection (ECOFS) approach, proposed in [6], can handle linear and non-linear dependent data and eliminates redundant and irrelevant features. Its effectiveness has been examined using an intrusion detection system. A Libsvm-IDS has been designed to operate using the features selected by the proposed ECOFS algorithm. The results of the evaluation show that the ECOFS algorithm selects the smallest number of features, resulting in the lowest computational cost for the Libsvm-IDS, with better performance. In fact, the algorithm has achieved greater precision.

In order to optimize the training procedure of SVM-based intrusion detection systems and reduce the time consumption, a GPU-based SVM intrusion detection method is proposed in [7]. During the simulation experiments with the KDD 1999 Cup data, the GPU-based parallel computing model is adopted. The results of the simulation experiments show that the time consumption in the IDS training procedure is reduced, and the performance of the IDS is maintained as usual.

Although IDSs have been in development for many years, the large number of return alerts make system maintenance inefficient for managers. In this article [8], the use of RST (Rough Set Theory) and SVM has been blown to detect intrusions. First, RST is used to pre-process data and reduce dimensions. Then, the features selected by RST are sent to the SVM model to learn and test respectively. This method is effective in reducing the spatial density of the data. The experiments compare the results with different methods and show that RST and SVM scheme can improve the false positive rate and accuracy.

The paper [9] proposes a Factor Analysis based Support Vector Machine (FA-SVM) algorithm to develop efficient IDSs using the popular statistical technique called factor analysis (FA). To design more effective and efficient IDSs, it was essential to select the best classifiers. This work is performed on the Knowledge discovery dataset and data mining to perform tests. The performance of this approach was compared to existing approaches such as principal component analysis (PCA) using SVMs, as well as classification with SVMs itself without feature selection. The results prove that the proposed method improves the detection of intrusions and intrusions, in terms of calculating false positive rates.

Wireless fidelity (WiFi) is a widely used test area due to its mobility in the presence of the main disadvantage of securing the network. Several attempts to secure 802.11 result, in inadequate security mechanisms, that this technology is vulnerable to various attacks and intrusions. The paper [10] proposes a Normalized Gain for MAC Intrusion (NMI)-based IDS to significantly improve the performance of the IDS. The proposed NMI consists of two primary components OFSNP and DCMI. The first component is the optimal feature selection using NGand PSO (OFSNP) and the second one is the detection and categorization of 802.11 MAC intrusions

(DCMI) using the SVM classifier. Proposed NMI achieves a better trade-off between detection accuracy and learning time. The experimental results show that NMI accurately detects and classifies 802.11 specific intrusions and also reduces false positives.

3. PROPOSED SOLUTION

This research presents a comprehensive framework to select the best CICIDS2017 dataset feature sets that effectively characterize normal traffic and distinguish it from abnormal traffic using SVM.

The CICIDS2017 dataset contains the most up-to-date, benign common attacks that resemble real-world data (PCAP). It also includes the results of network traffic analysis using CICFlowMeter with labelled flows based on timestamp, source and destination IP addresses, source and destination ports, protocols and attacks (CSV files). For this dataset, we constructed the abstract behaviour of 25 users based on HTTP, HTTPS, FTP, SSH, and email protocols.

The proposed algorithm uses, to select the important features, a set from the CICIDS2017. The reduced feature CICIDS2017dataset is then used for training and design a detection model on the SVM classifier.

We partition the data into two classes: normal and attack, where the implemented attacks include Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS. The goal of our SVM implementing is to separate the normal and attack patterns. In our case, we will take the traffic of 3 days to have accurate results keeping first day as training set and the rest of the csv files as test set:

- **Monday July, 3rd 2017**
  Benign (Normal human activity)
- **Tuesday July, 4th 2017**
  Force brute
  FTP-Patator (9h20 - 10h20)
  SSH-Patator (14h00 - 15h00)
- **Wednesday July, 5th 2017**
  DoS / DDoS
  DoS slowloris (9h47 - 10h10)
  DoS Slowhttptest (10h14 - 10h35)
  DoS Hulk (10h43 - 11h)
  DoS GoldenEye (11h10 - 11h23).

The procedure of the solution note is as follows:

1- Importing libraries ;
2- Data import ;
3- Data selection and indexing ;
4- Data pre-processing :
  - Data pre-processing involves dividing the data into training and test sets ;

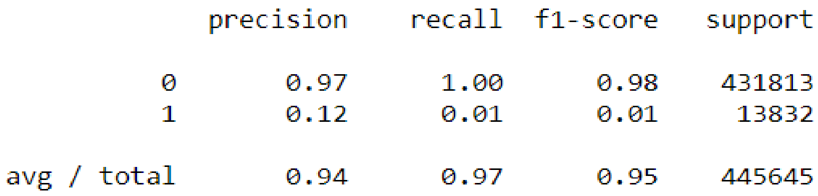
5- Training the algorithm :
  - Unsupervised outlier detection with a kernel (Specifies the type of kernel to use in the algorithm. It can be "linear", "poly", "rbf", "sigmoid", "precalculed". If none is specified, "rbf" will be used.) of parameters gamma(kernel coefficient for rbf and poly, if gamma is 0.0 then 1 / n_features will be taken) ;
  - Detection of the soft limit of the sample set.

6- Evaluation of the algorithm :

- Confusion matrix, precision, recall are the most commonly used measures for classification tasks ;

4. Testing

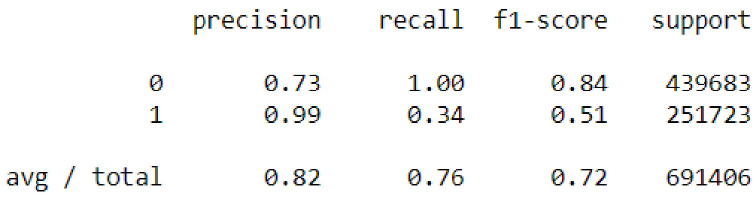
In our first experiment, we will take as training data source the Monday traffic, which is a benign traffic, and a Tuesday test set, which is composed of attacks and normal activities, and we will apply our solution with the parameter's gamma = 0.0005, kernel = 'rbf', nu=0.001. The results show that the precision reaches up to 97 % (Figure 3).



```
             precision    recall  f1-score   support

          0       0.97      1.00      0.98    431813
          1       0.12      0.01      0.01     13832

avg / total       0.94      0.97      0.95    445645
```

Figure 3. Results of the first experiment with parameters (gamma = 0.0005, kernel = 'rbf', nu=0.001)

In the second experiment, we will take as training data source the Monday traffic which is a benign traffic and a test set of Wednesday which consists of attacks plus normal activity, and we will apply our solution with the parameter's gamma = 0.0005, kernel = 'rbf', nu=0.001. The results show that the precision reaches up to 99 % (Figure 4).



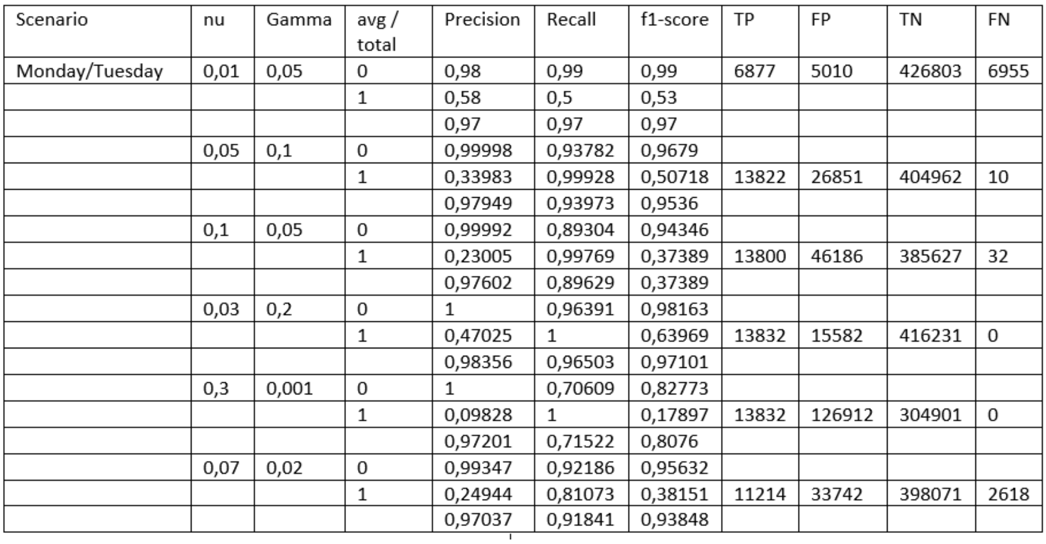
```
             precision    recall  f1-score   support

          0       0.73      1.00      0.84    439683
          1       0.99      0.34      0.51    251723

avg / total       0.82      0.76      0.72    691406
```

Figure 4. Results of the first experiment with parameters (gamma = 0.0005, kernel = 'rbf', nu=0.001)

5. Results

Table 1 shows the results of our first experiment with different values of nu and gamma.

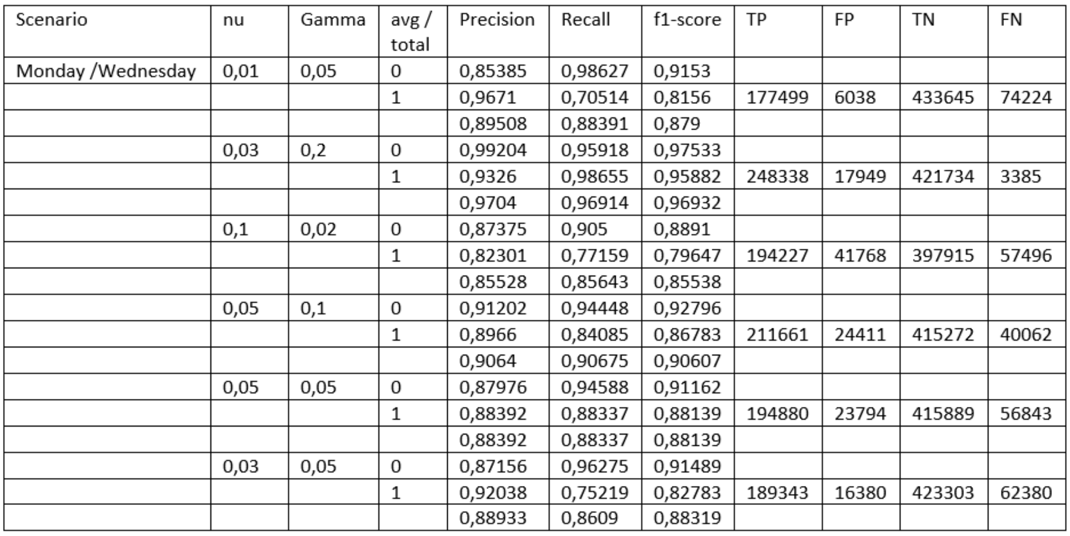
Table 1. The results of the first experiment with different values of nu and gamma



Scenario	nu	Gamma	avg / total	Precision	Recall	f1-score	TP	FP	TN	FN
Monday/Tuesday	0,01	0,05	0	0,98	0,99	0,99	6877	5010	426803	6955
			1	0,58	0,5	0,53				
				0,97	0,97	0,97				
	0,05	0,1	0	0,99998	0,93782	0,9679				
			1	0,33983	0,99928	0,50718	13822	26851	404962	10
				0,97949	0,93973	0,9536				
	0,1	0,05	0	0,99992	0,89304	0,94346				
			1	0,23005	0,99769	0,37389	13800	46186	385627	32
				0,97602	0,89629	0,37389				
	0,03	0,2	0	1	0,96391	0,98163				
			1	0,47025	1	0,63969	13832	15582	416231	0
				0,98356	0,96503	0,97101				
	0,3	0,001	0	1	0,70609	0,82773				
			1	0,09828	1	0,17897	13832	126912	304901	0
				0,97201	0,71522	0,8076				
	0,07	0,02	0	0,99347	0,92186	0,95632				
			1	0,24944	0,81073	0,38151	11214	33742	398071	2618
				0,97037	0,91841	0,93848				

Table 2 shows the results of the second experiment with different values of nu and gamma.

Table 2. The results of the second experiment with different values of nu and gamma



Scenario	nu	Gamma	avg / total	Precision	Recall	f1-score	TP	FP	TN	FN
Monday /Wednesday	0,01	0,05	0	0,85385	0,98627	0,9153				
			1	0,9671	0,70514	0,8156	177499	6038	433645	74224
				0,89508	0,88391	0,879				
	0,03	0,2	0	0,99204	0,95918	0,97533				
			1	0,9326	0,98655	0,95882	248338	17949	421734	3385
				0,9704	0,96914	0,96932				
	0,1	0,02	0	0,87375	0,905	0,8891				
			1	0,82301	0,77159	0,79647	194227	41768	397915	57496
				0,85528	0,85643	0,85538				
	0,05	0,1	0	0,91202	0,94448	0,92796				
			1	0,8966	0,84085	0,86783	211661	24411	415272	40062
				0,9064	0,90675	0,90607				
	0,05	0,05	0	0,87976	0,94588	0,91162				
			1	0,88392	0,88337	0,88139	194880	23794	415889	56843
				0,88392	0,88337	0,88139				
	0,03	0,05	0	0,87156	0,96275	0,91489				
			1	0,92038	0,75219	0,82783	189343	16380	423303	62380
				0,88933	0,8609	0,88319				

The results are significant, we get up to 98% accuracy with a zero false negative value, we do not forget that the data is voluminous Monday (11G), Tuesday (11G) and Wednesday (13G) including attacks : Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS. In addition, we notice that when we increase the value of Gamma, we have more efficient precision and a lower false negative rate.

6. CONCLUSION

In order to improve the efficiency of the carrier vector machine (SVM) based intrusion detection system (IDS), we have conducted a large number of experiments to measure the performance of carrier vector machines in the intrusion detection, using CICIDS2017 data for intrusion assessment. SVMs provide very accurate performance (99% and above).

Our future work is to apply new SVM optimization techniques in SNORT fusion to improve attack detection in computer network security.

REFERENCES

[1] J. Frank, "Machine learning and intrusion detection: Current and future directions", in Proceedings of the National 17th Computer Security Conference, Washington, October 1994 ;

[2] Elaeraj, Ouafae & Cherkaoui, Leghris & Renault, Éric, "Performance Evaluation of Some Machine Learning Algorithms for Security Intrusion Detection", in the Machine Learning for Networking Third International Conference, Paris, 2020 ;

[3] J. P. Anderson, "Computer Security Threat Monitoring and Surveillance", Technical Report, Fort Washington, 1980 ;

[4] Kim, Dong Seong, Nguyen, Ha-Nam & Park, Jong, "Genetic Algorithm to Improve SVM Based Network Intrusion Detection System", in the 19th International Conference on Advanced Information Networking and Applications, Taiwan, 2005 ;

[5] H. Gharaee and H. Hosseinvand, "A new feature selection IDS based on genetic algorithm and SVM", in the 8th International Symposium on Telecommunications (IST), United States, 2016 ;

[6] W. Wang, X. Du and N. Wang, "Building a Cloud IDS Using an Efficient Feature Selection Method and SVM", in IEEE Access, vol. 7, pp. 1345-1354, 2019 ;

[7] Xia, Yong & Shi, Zhi & Zhang, Yu & Dai, Jian, "A SVM intrusion detection method based on GPU", in the international Conference on Applied Mechanics, Mechatronics and Intelligent System, China, 2014 ;

[8] Chen, Rung-Ching, Cheng, Kai-Fan & Hsieh, Chia-Fen, "Using Rough Set and Support Vector Machine for Network Intrusion Detection", in the International Journal of Network Security & Its Applications, 2010 ;

[9] P Indira Priyadarsini, I Ramesh Babu, "Building Efficient Intrusion Detection System Using Factor Analysis and Support Vector Machines", in the international journal of engineering research & technology (IJERT) Volume 03, Issue 04, April 2014 ;

[10] Murugan, Kavitha & M, Usha, "Anomaly Based Intrusion Detection for 802.11 Networks based on Optimal Feature Selection using SVM Classifier", Springer DOI: 10.1007/s11276-016-1300-5, Wireless Networks,2016.

AUTHORS

Ouafae El AERAJ, aged 26, Research student in network security at the Faculty of Sciences and Techniques Mohammedia, Hassan II University of Casablanca, Morocco.

Cherkaoui LEGHRIS has a PhD in computer sciences from the Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Mohammed V University of Rabat, Morocco, in 2007. On 2003, He had the diploma of Higher Studies in ENSIAS and before the degree in Applied Computer Science from the Cadi Ayyad University of Marrakech. He is currently working as Professor of Higher Education at the Faculty of Science and Techniques, Hassan II University of Casablanca. He is responsible of various modules in communication networks domain.

He conducts research on networking at L@M Laboratory, RTM team. His main research focus on the IoT networks based across IPv6 protocol, multi-access network technologies and IT security. He is also an active member in the Moroccan Internet SOCiety organization, which works for openness and accessibility of the Internet for any.

REDUCING CYBER INCIDENT RESPONSE TO PROTECT CNI FROM CYBER ATTACKS USING AN N-SIEM INTEGRATION WITH AN ICTI-CNI

Igli Tafa and Kevin Shahollari

Department of Computer Engineering, Polytechnic University of Tirana, Tirana

ABSTRACT

The rapid evolution of technology has increased the role of cybersecurity and put it at the center of nationalcritical infrastructure. This role supports and guarantees the vital services of (CNI) while provides the proper functionalities for running operations between the public and private sectors. This evolution has hadthe same impact on cyberattack tools, methods, techniques used to gain unauthorized access to these computer systems that contain confidential and high-value information in the digital data sales market or asit called "darkweb".

As a result, it has become necessary to monitor all events of the National Critical Infrastructure (CNI) computer systems. This proposed system uses a centralized National SIEM (N-SIEM) specializing in the correlation of security events caused by cyber attacks, collected by CNIs systems while integrating with anInternational Cyber Threat Intelligence system (ICTI-CNI).

In addition, this conceptual model collects security breach events of CNIs systems, analyzes only cyber attacks, and correlates these security events in real-time with an intelligent automated platform while reducing the response time of security analysts.

KEYWORDS

CNI, N-SIEM, ICTI-CNI, IOC, cyber attacks security events.

1. INTRODUCTION

Today, cyber security incidents are constantly increasing in frequency and size, becoming more complex and unrestricted by state borders.[1] Consequently, when incidents occurred in these critical infrastructures, they can cause catastrophic damages for the country and its sectors by directly affecting human society.

1.1. Critical National Infrastructure (CNI)

Protecting critical national infrastructure is the primary goal of cyber security.[2] The Critical National Infrastructure (CNI) includes all the sectors of a country that work at all capacity 24 hours, seven days to ensure the existence, continuity, and vitality of the essential services used by society today. The protection and provision of this infrastructure is a responsibility shared between the two constituent sectors of the country: public and private, to ensure the availability of these services. By definition [3], Critical National Infrastructure (CNI) includes these categories: Food, Water, Chemicals, Health, Energy, Communications, Transport, Emergency Services, Civil

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 211-226, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111818

Nuclear, Finance, Government, Defence, and Space. But each country can specify its most critical infrastructure, relying on those individuals sectors that provide the most necessary and crucial services to their society. Almost all of this infrastructure (CNI) has fallen under the control of advanced computer systems.[4] Damage to the security and availability of this infrastructure (CNI) would bring chaos throughout the country. This chaos comes because these sectors do not function separately but together consequently, a cyberattack on one of the infrastructure sectors would negatively affect all other sectors. In this way, is requisite complete security between the critical national infrastructure at the local, country, regional and international levels. Attackers often target this infrastructure (CNI)for various purposes ranging from gathering financial and confidential information to terrorist purposes.

2. BIG PICTURE

The most advanced cyber attacks can paralyze the most crucial communication hubs of the country by blocking public or private services but can also have devastating effects if these attacks affect the physical damage of critical national infrastructure which can cause human loss. Cybersecurity and Infrastructure Security Agency (CISA) identified critical national infrastructure (CNI) as the primary target of SolarWinds mass hacking that happened this year and the true impact of this attack is still unknown. [5] In the future, as a result of the exposure of this information and computer systems, countries may have interruption of services provided by this infrastructure (CNI). From this study [6], CISA, National Security Agency (NSA), and Federal Bureau of Investigation (FBI) have instructed all countries and organizations to do all the recommended patches of computer systems, especially vulnerabilities that are targeted by state-sponsored hackers.

For this reason, all sectors of CNI and the services they provide that are considered critical must be monitored and protected in real time 24 hours in 7 days from cyber attacks, ensuring 99.999% availability of the necessary services in the daily life of human society. To ensure these vital services to society is required not only that monitoring of this infrastructure but more decisive is the reaction time to these cyber attacks when they occur toward CNI. According to [7], the process of sharing information between countries and organizations that have previously encountered these cyber threats to their critical national infrastructure is relevant for detecting and preventing possible future cyber attacks.

This paper will be divided as follow: in section 3 we are presenting the essential role of this infrastructure, which is vital for the maintenance of societal functions and analyze the cyber threats that can be occurred to these CNIs.In section 4 will be analyzing the advantages and disadvantages of the actual technologies where our proposed will be mostly based. The proposed approach will be then presented in section 5, a new model and architecture for monitoring this critical infrastructure from security events generating by real attacks based on National SIEM integrated with a ICNI-CTI platform. Finally, section 6 is dedicated to the cyber attackers that had attacked that infrastructure in the past and a list of their APTs, group by countries suspected of these attacks.

3. RELATED WORKS

3.1. Definition

A cyberattack is a malicious attempt by an individual or group to attack and compromise the confidentiality, integrity, and availability of the systems that store valuable information which can be stolen and used for various purposes.

3.2. Requirements of cyberattack

Cyber attacks targeting this infrastructure (CNI) require more time and experience than attacks on widely used computer systems. The impact of these factors is reduced when considering that most cyber attacks are organized by the countries and their intelligence agencies that have the proper resources to attack this type of infrastructure (CNI).

3.3. Goals of cyberattack

The principal goal of a cyber attack is to affect the actual world but the real actions happenin the virtual and artificial world where they have involved: computers, servers, databases, or other devices that may have exploitable vulnerabilities. [8].

3.4. Categories of cyber attacks to CNI

Cyber crime (CC) - The activity of using technology or exploitable vulnerabilities discovered in these computer systems [9] to gain unauthorized access to confidential data that are used for financial gain.

a. Cyberthieves are individuals or groups involve in illegal cyber attacks for monetary gain and those stealing are considered low-risk for cyberattackers and costly for victims, with some estimates placing the annual global cost as high as hundreds of billions of dollar [10]. Cyber espionage (CE) - The activity of using technology or exploitable vulnerabilities discovered in these computer systems [9] to gain unauthorized access to classified data, stealing: industry or military trade secrets for economic gain, competitive advantage, or political reasons

b. Cyberspies are individuals or groups involve in illegal cyber attacks who steal classified or proprietary information used by governments or private corporations to gain a competitive strategic advantage. These individuals often work at the behest of and take direction from, foreign government entities [10]. Cyber warfareand Cyber sabotage (CW/CS) - The activity of using technology or exploitable vulnerabilities discovered in these computer systems [9] to damage the critical national infrastructure (CNI) of the target country, an action that could bring military precedent.

c. Cyberwarriors are state-sponsored and non-state actors who develop capabilities and undertake cyber attacks in support of a country's strategic objectives [10]

4. CNI MONITORING SYSTEMS

To monitor the critical services provided by this infrastructure (CNI), a security event management system (SIEM) is required to monitor all types of security events related to cyber attacks that can occur on a CNI.

4.1. Definition of SIEM

Gartner describes this security incident management system (SIEM) as a system needed for real-time data analysis for faster detection of information leakage and cyber attacks against internal or external threats. This system also collects, stores, and reports these cyber attacks in log format to respond more rapidly against cyber incidents. Additionally can be used in the digital forensics process or for legal procedures and compliance policies. [11].

4.2. Components of SIEM

A SIEM is composed of two main components which are:

1) Security Event Management (SEM) - provides real-time analysis, correlation, normalization of data collected from computer systems, infrastructures, network devices, and applications in log data format to provide a manual or automatic response, thus simplifying the management processby cyber security analysts.

2) Security Information Management (SIM) – SIM provides real-time analysis, storage, and reporting of collected (historical) security events to provide a data collection database by providingadditional functionalities that simplify the process of investigating cyber events.

When these two microsystems are combined, a Security Information and Event Management (SIEM) is created with the focus on collecting, analyzing, storing, and presenting the data collected by technological devices and entirely monitoring theinfrastructure.

4.3. Disadvantage of SIEM

Most of today's SIEMs function by collecting, correlating, analyzing, and presenting security events in the most understandable forms by the cyber security analyst whose function is to investigate cyber incidents. In the case of critical national infrastructure, theresponse should be as fast as possible. This event affects the response time because it needs time to accurately the data received, manually reduce the false positives, and verifythe real malicious indicators.

4.4. Reaction time dependence of SIEM

This time depends on the accuracy of data collected and correlated by all their internal and external IT and ICS infrastructures. Response time increases if the cyberattack has recently been identified and there is no available information on the methods used by theattackers to compare with other organizations or the infrastructure of the countries that may have been affected by this type of cyber attack

4.5. Monitoring and Intelligence integration

These systems are integrating with a new technology called Cyber Threat Intelligence (CTI) which includes more than raw data. This technology requires more detailedinformation collected only when human analysis is involved in this process.

This detailed information includes the tactics, tools, and procedures used by an attacker to predict those techniques that may use during other cyber attacks in the future. It shouldalso include the link between compromise indicators (IP addresses, risk-related domains,or hashes associated with malicious files and intruders (motivations, goals, and information about what they are targeting). [12]

5. PROPOSED SYSTEM DESIGN

Given the increasing complexity and vulnerability of CNI, the security of this infrastructure will continue to be crucial in the future. Consequently, the response to thesecyber incidents should be as fast as possible, considering this importance.

5.1. Why a new system?

At present, there is a high number of open-source or commercial platforms of this technology used to share cyber attacks. These platforms sometimes have many Indicators of Compromise (IOCs), obtained either from cyber attacks that have occurred and verifiedor from voluntary that can manually add the (IOCs) and those are not verified.

Add to the fact that these platforms don't orientate according to the critical national infrastructure sectors where the same threat feed serves for all (CNIs) sectors. Soincreases the likelihood that some of these indicators (I0Cs) will be false positives by notifying security analysts about preventing cyber attacks that are not actually happening, thus wasting time responding to and preventing cyber attacks that are likely to occur.

5.2. The concept of new system

“Share early and share often” it is essential in times of crisis. A cyber attack on critical infrastructure can have a cascading impact on multiple sectors across multiple jurisdictions, providing little time to contain and mitigate damage and prevent any follow-on attacks. [13]

The new system should rely on threat feeds focused on exact national critical infrastructure sectors, analyzing and correlating security events occurring in the internal (CNI) with events that appeared in the (CNI) of other countries that had the same attack behavior. This system (ICTI-CNI) should work as a correlation platform with their actualcurrent N- SIEM.

5.3. Requirements of new system

1. Reciprocal cooperation is required between countries and sectors of the same category, where shared information from a country that has been attacked is usedto prevent an attack on another country within the same group.
2. The ICTI platform should create a clear channel for the classification and declassification of attack information. The exchange process must take place between specific sectors of the CNI and countries, also the same between the public and private sectors providing vital services to society.

This proposed system is supposed to provide a more effective service while offering automated information sharing and alerting each other to incidents or cyber attacks that just happened in similar infrastructure.

5.4. Architecture of the system

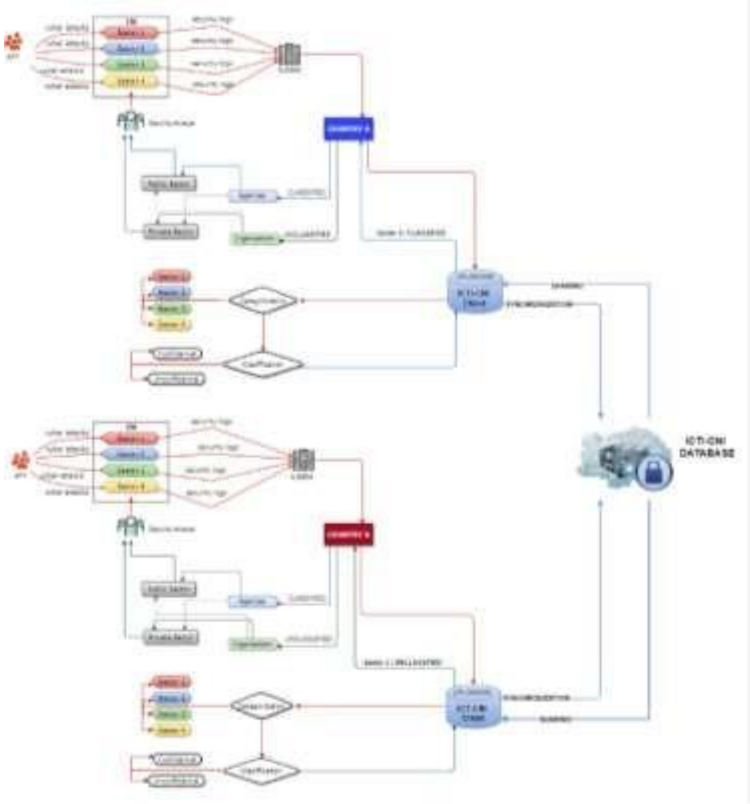


Figure 1. Architecture of the proposed system ICTI-CNI

This proposed system is composed of two components which are:

1) ICTI - CNI Client

It will be called any country that agrees to exchange information on cyber attacks occurring on their infrastructure (CNIs). Once the ICTI-Client is synchronizing with the central database, cyber attacks events that occurred in this client infrastructure (CNIs) will be an automatic correlation with other cyber incidents stored in the database.

2) ICTI - CNI Database

It will be called the central database, where all the information collected from the clients infrastructure that will be part of this exchange process will be stored and analyzed.

After the correlation of security events related to cyber attacks and according to the analysis of this information, the proper classification and categorization will be achieved. The information is then authorized to be shared with other clients within thesame group.

5.5. Integration with existing systems

All agencies that have the responsibility to protect those infrastructures can have their own commercial or open-source Threat Intelligence Feeds and continues to use even afterimplementing this proposed system. But on the ICHI, they should only share indicators of compromise

(IOCs) of attacks that have occurred in these infrastructures (CNIs) by grouping them in specific sectors.

5.6. Non-Disclosure secret agreement

All agencies that share information that corresponds to cyber attacks detected in other countries, in critical national infrastructure, must maintain the confidentiality of data andnot share it with other organizations outside the sector where it operates

6. Advanced Persistent Threats

In most cases, cyber attacks conducted on CNI are state-sponsored, and detection of thoseattack campaigns has become extremely difficult while considering the evolution of strategies, techniques, and tactics used by cyber attackers on these infrastructures.

Advanced persistent threat (APT) refers to a group state-funded or the foreign government using intelligence gathering techniques to access sensitive information targeting a specific entity. [14] Objectives of APT attacks include continuous exfiltrationof information, cyber warfare, damage to critical infrastructure, and degradation of military assets. [15]

Why is calling APT?

Advanced

Attackers use a variety of intelligence gathering techniques, including computer and conventional technologies such as telephone-interception technologies and satellite imaging. They often combine multiple targeting methods, tools, and techniques to reachand compromise their target and maintain access to it. [14]

Persistent

Attackers give priority to a specific task guided by external entities. The targeting is conducted using a “low-and-slow” approach through continuous monitoring and interaction to achieve the defined objectives. If the attackers lose access to their target, they usually will reattempt access because the purpose of this attack is to maintain alwayslong-term access to the target. [14]

Threat

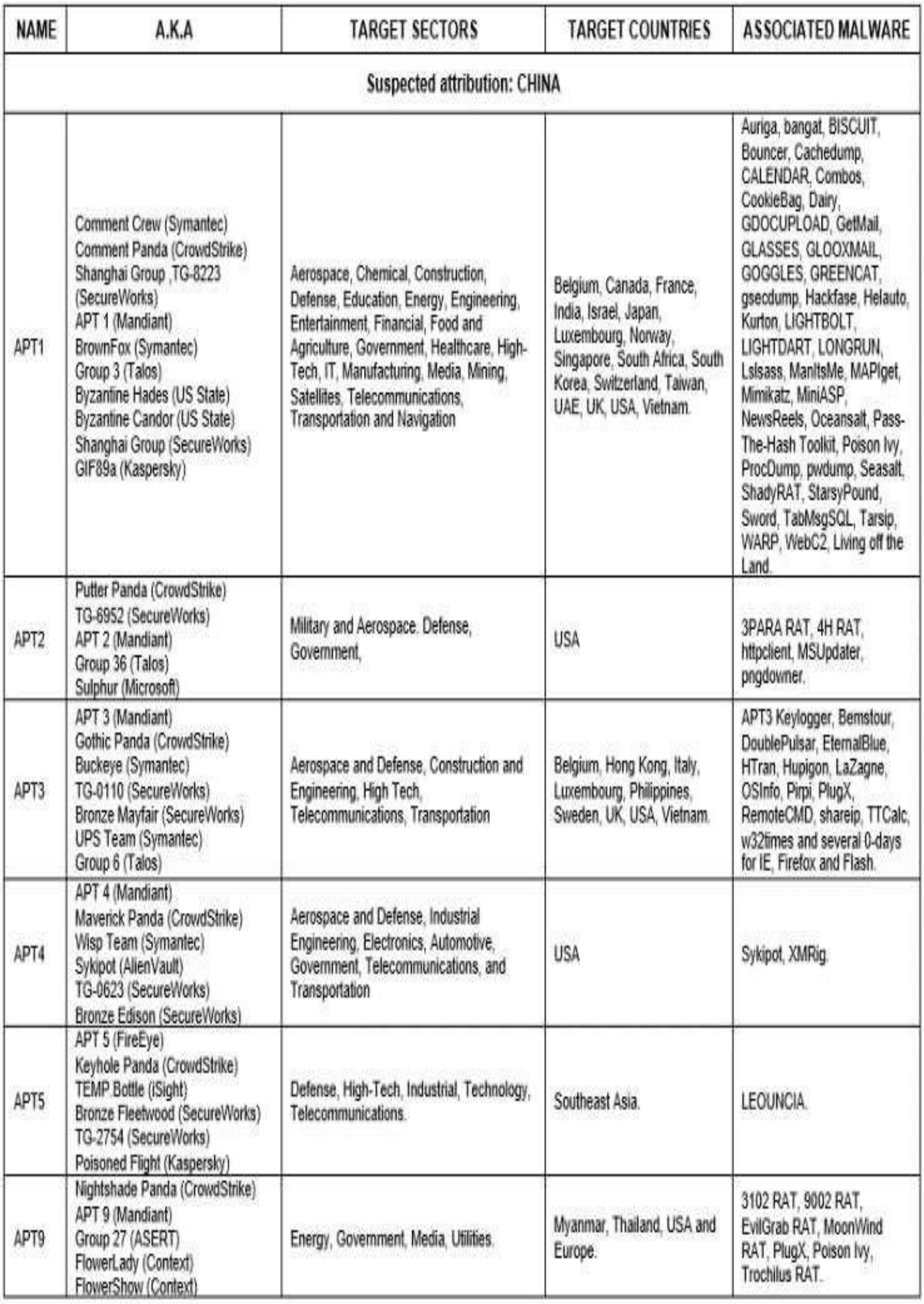
Attackers are a threat because they are skilled, motivated, organized, and well-funded also have both capability and intent to attack and damage these critical infrastructures. Those attacks are executed by coordinated human actions rather than by automated piecesof code. [14]

Below are listed the most known dangerous APTs, which are often state-sponsored, whereeach of them has attack campaigned against many sectors of CNIs, including internationaland military organizations.

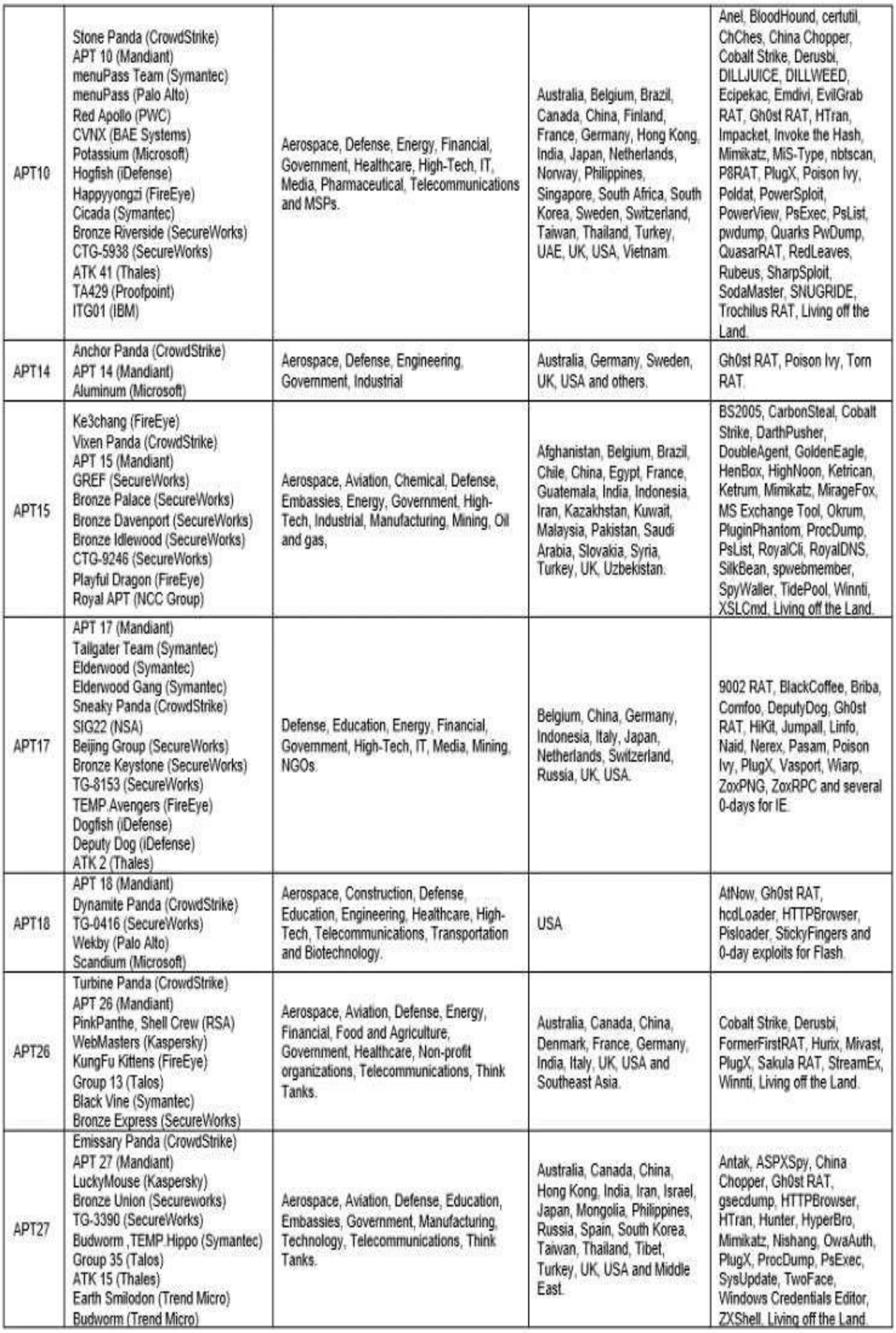
All pieces of information on APT are taken from [16] [17] [18] [19], analyzed, grouping, and presented graphically in the following tables on the most popular APTs groups that exist today.

6.1. Suspected attribution

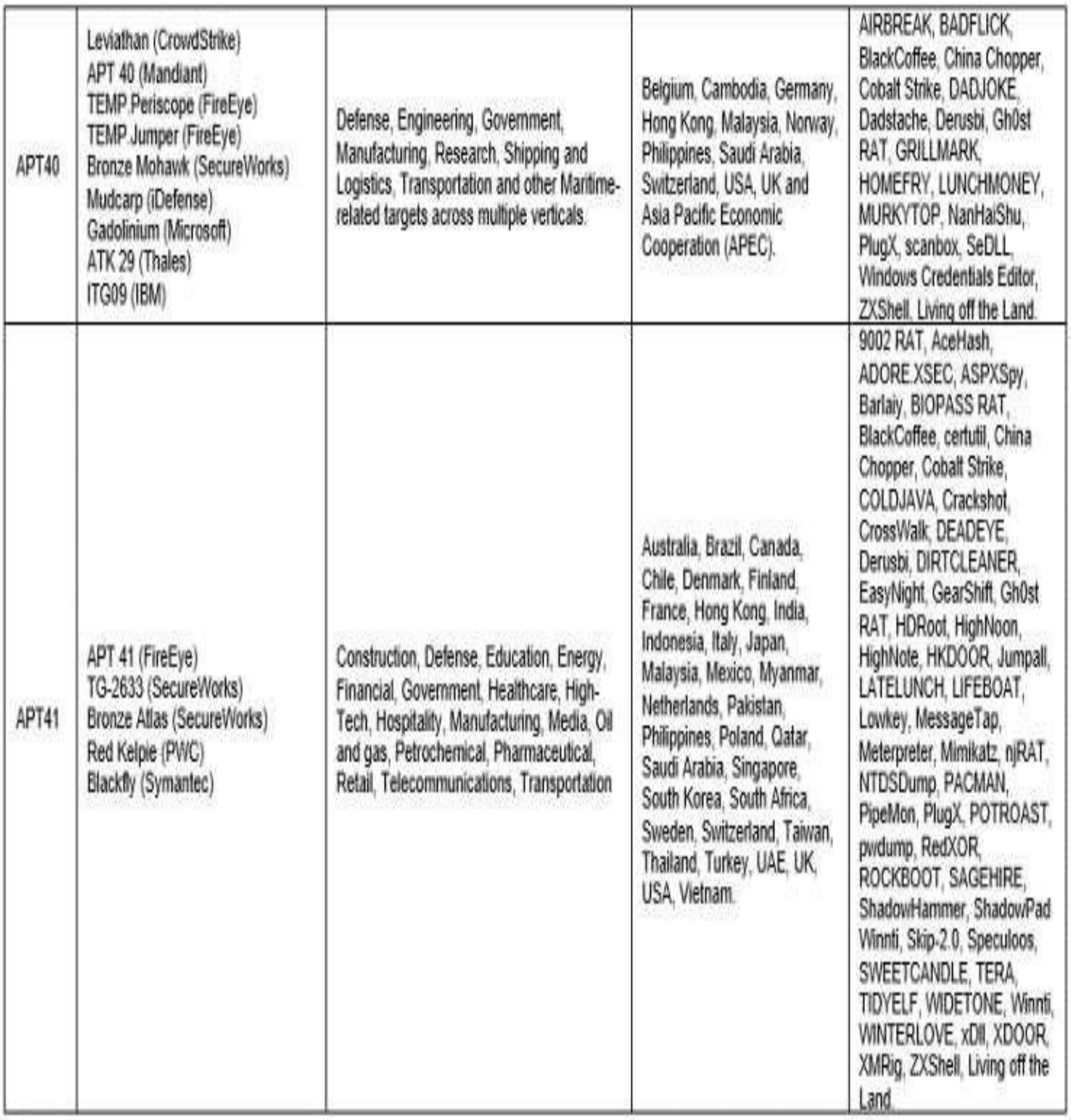
6.1.1. China



NAME	A.K.A	TARGET SECTORS	TARGET COUNTRIES	ASSOCIATED MALWARE
Suspected attribution: CHINA				
APT1	Comment Crew (Symantec) Comment Panda (CrowdStrike) Shanghai Group ,TG-8223 (SecureWorks) APT 1 (Mandiant) BrownFox (Symantec) Group 3 (Talos) Byzantine Hades (US State) Byzantine Candor (US State) Shanghai Group (SecureWorks) GIF89a (Kaspersky)	Aerospace, Chemical, Construction, Defense, Education, Energy, Engineering, Entertainment, Financial, Food and Agriculture, Government, Healthcare, High-Tech, IT, Manufacturing, Media, Mining, Satellites, Telecommunications, Transportation and Navigation	Belgium, Canada, France, India, Israel, Japan, Luxembourg, Norway, Singapore, South Africa, South Korea, Switzerland, Taiwan, UAE, UK, USA, Vietnam.	Auriga, bangat, BISCUIT, Bouncer, Cachedump, CALENDAR, Combos, CookieBag, Dairy, GDOCUPLOAD, GetMail, GLASSES, GLOOXMAIL, GOGGLES, GREENCAT, gsecdump, Hackfase, Helauto, Kurton, LIGHTBOLT, LIGHTDART, LONGRUN, Lslsass, ManltsMe, MAPIget, Mimikatz, MiniASP, NewsReels, Oceansalt, Pass-The-Hash Toolkit, Poison Ivy, ProcDump, pwdump, Seasalt, ShadyRAT, StarsyPound, Sword, TabMsgSQL, Tarsip, WARP, WebC2, Living off the Land.
APT2	Putter Panda (CrowdStrike) TG-6952 (SecureWorks) APT 2 (Mandiant) Group 36 (Talos) Sulphur (Microsoft)	Military and Aerospace. Defense, Government,	USA	3PARA RAT, 4H RAT, httpclient, MSUpdater, pngdowner.
APT3	APT 3 (Mandiant) Gothic Panda (CrowdStrike) Buckeye (Symantec) TG-0110 (SecureWorks) Bronze Mayfair (SecureWorks) UPS Team (Symantec) Group 6 (Talos)	Aerospace and Defense, Construction and Engineering, High Tech, Telecommunications, Transportation	Belgium, Hong Kong, Italy, Luxembourg, Philippines, Sweden, UK, USA, Vietnam.	APT3 Keylogger, Bemstour, DoublePulsar, EternalBlue, HTran, Hupigon, LaZagne, OSInfo, Pirpi, PlugX, RemoteCMD, shareip, TTCalc, w32times and several 0-days for IE, Firefox and Flash.
APT4	APT 4 (Mandiant) Maverick Panda (CrowdStrike) Wisp Team (Symantec) Sykipot (AlienVault) TG-0623 (SecureWorks) Bronze Edison (SecureWorks)	Aerospace and Defense, Industrial Engineering, Electronics, Automotive, Government, Telecommunications, and Transportation	USA	Sykipot, XMRig.
APT5	APT 5 (FireEye) Keyhole Panda (CrowdStrike) TEMP.Bottle (iSight) Bronze Fleetwood (SecureWorks) TG-2754 (SecureWorks) Poisoned Flight (Kaspersky)	Defense, High-Tech, Industrial, Technology, Telecommunications.	Southeast Asia.	LEOUNCIA.
APT9	Nightshade Panda (CrowdStrike) APT 9 (Mandiant) Group 27 (ASERT) FlowerLady (Context) FlowerShow (Context)	Energy, Government, Media, Utilities.	Myanmar, Thailand, USA and Europe.	3102 RAT, 9002 RAT, EvilGrab RAT, MoonWind RAT, PlugX, Poison Ivy, Trochilus RAT.

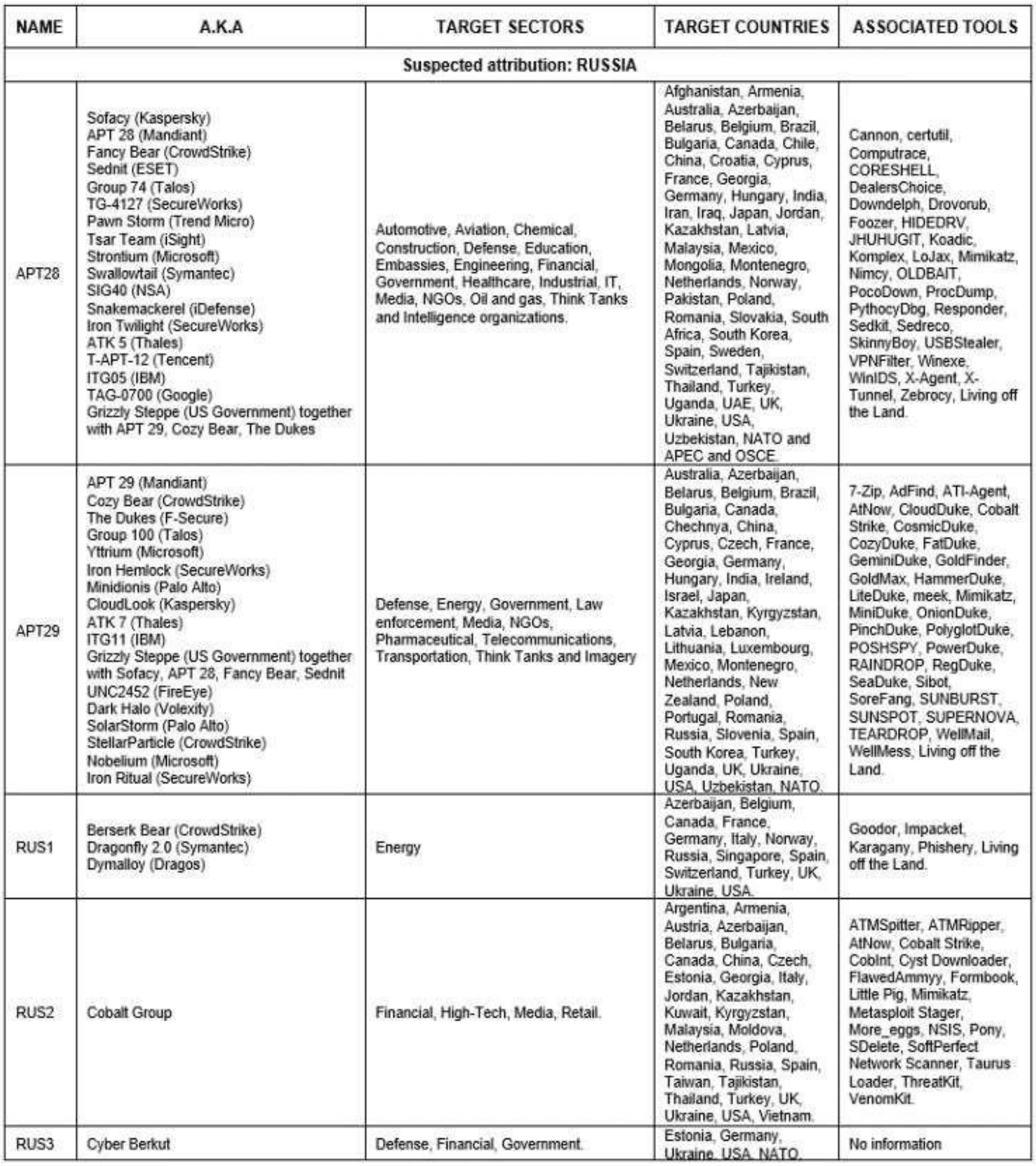


APT10	Stone Panda (CrowdStrike) APT 10 (Mandiant) menuPass Team (Symantec) menuPass (Palo Alto) Red Apollo (PWC) CVNX (BAE Systems) Potassium (Microsoft) Hogfish (iDefense) Happyyongzi (FireEye) Cicada (Symantec) Bronze Riverside (SecureWorks) CTG-5938 (SecureWorks) ATK 41 (Thales) TA429 (Proofpoint) ITG01 (IBM)	Aerospace, Defense, Energy, Financial, Government, Healthcare, High-Tech, IT, Media, Pharmaceutical, Telecommunications and MSPs.	Australia, Belgium, Brazil, Canada, China, Finland, France, Germany, Hong Kong, India, Japan, Netherlands, Norway, Philippines, Singapore, South Africa, South Korea, Sweden, Switzerland, Taiwan, Thailand, Turkey, UAE, UK, USA, Vietnam.	Anel, BloodHound, certutil, ChChes, China Chopper, Cobalt Strike, Derusbi, DILLJUICE, DILLWEED, Ecipekac, Emdivi, EvilGrab RAT, Gh0st RAT, HTran, Impacket, Invoke the Hash, Mimikatz, MiS-Type, nbtscan, P8RAT, PlugX, Poison Ivy, Poldat, PowerSploit, PowerView, PsExec, PsList, pwdump, Quarks PwDump, QuasarRAT, RedLeaves, Rubeus, SharpSploit, SodaMaster, SNUGRIDE, Trochilus RAT, Living off the Land.
APT14	Anchor Panda (CrowdStrike) APT 14 (Mandiant) Aluminum (Microsoft)	Aerospace, Defense, Engineering, Government, Industrial	Australia, Germany, Sweden, UK, USA and others.	Gh0st RAT, Poison Ivy, Torn RAT.
APT15	Ke3chang (FireEye) Vixen Panda (CrowdStrike) APT 15 (Mandiant) GREF (SecureWorks) Bronze Palace (SecureWorks) Bronze Davenport (SecureWorks) Bronze Idlewood (SecureWorks) CTG-9246 (SecureWorks) Playful Dragon (FireEye) Royal APT (NCC Group)	Aerospace, Aviation, Chemical, Defense, Embassies, Energy, Government, High-Tech, Industrial, Manufacturing, Mining, Oil and gas,	Afghanistan, Belgium, Brazil, Chile, China, Egypt, France, Guatemala, India, Indonesia, Iran, Kazakhstan, Kuwait, Malaysia, Pakistan, Saudi Arabia, Slovakia, Syria, Turkey, UK, Uzbekistan.	BS2005, CarbonSteal, Cobalt Strike, DarthPusher, DoubleAgent, GoldenEagle, HenBox, HighNoon, Ketrican, Ketrum, Mimikatz, MirageFox, MS Exchange Tool, Okrum, PluginPhantom, ProcDump, PsList, RoyalCli, RoyalDNS, SilkBean, spwebmember, SpyWaller, TidePool, Winnti, XSLCmd, Living off the Land.
APT17	APT 17 (Mandiant) Tailgater Team (Symantec) Elderwood (Symantec) Elderwood Gang (Symantec) Sneaky Panda (CrowdStrike) SIG22 (NSA) Beijing Group (SecureWorks) Bronze Keystone (SecureWorks) TG-8153 (SecureWorks) TEMP.Avengers (FireEye) Dogfish (iDefense) Deputy Dog (iDefense) ATK 2 (Thales)	Defense, Education, Energy, Financial, Government, High-Tech, IT, Media, Mining, NGOs.	Belgium, China, Germany, Indonesia, Italy, Japan, Netherlands, Switzerland, Russia, UK, USA.	9002 RAT, BlackCoffee, Briba, Comfoo, DeputyDog, Gh0st RAT, HiKit, Jumpall, Linfo, Naid, Nerex, Pasam, Poison Ivy, PlugX, Vasport, Wiarp, ZoxPNG, ZoxRPC and several 0-days for IE.
APT18	APT 18 (Mandiant) Dynamite Panda (CrowdStrike) TG-0416 (SecureWorks) Wekby (Palo Alto) Scandium (Microsoft)	Aerospace, Construction, Defense, Education, Engineering, Healthcare, High-Tech, Telecommunications, Transportation and Biotechnology.	USA	AtNow, Gh0st RAT, hcdLoader, HTTPBrowser, Pisloader, StickyFingers and 0-day exploits for Flash.
APT26	Turbine Panda (CrowdStrike) APT 26 (Mandiant) PinkPanthe, Shell Crew (RSA) WebMasters (Kaspersky) KungFu Kittens (FireEye) Group 13 (Talos) Black Vine (Symantec) Bronze Express (SecureWorks)	Aerospace, Aviation, Defense, Energy, Financial, Food and Agriculture, Government, Healthcare, Non-profit organizations, Telecommunications, Think Tanks.	Australia, Canada, China, Denmark, France, Germany, India, Italy, UK, USA and Southeast Asia.	Cobalt Strike, Derusbi, FormerFirstRAT, Hurix, Mivast, PlugX, Sakula RAT, StreamEx, Winnti, Living off the Land.
APT27	Emissary Panda (CrowdStrike) APT 27 (Mandiant) LuckyMouse (Kaspersky) Bronze Union (Secureworks) TG-3390 (SecureWorks) Budworm ,TEMP.Hippo (Symantec) Group 35 (Talos) ATK 15 (Thales) Earth Smilodon (Trend Micro) Budworm (Trend Micro)	Aerospace, Aviation, Defense, Education, Embassies, Government, Manufacturing, Technology, Telecommunications, Think Tanks.	Australia, Canada, China, Hong Kong, India, Iran, Israel, Japan, Mongolia, Philippines, Russia, Spain, South Korea, Taiwan, Thailand, Tibet, Turkey, UK, USA and Middle East.	Antak, ASPXSpy, China Chopper, Gh0st RAT, gsecdump, HTTPBrowser, HTran, Hunter, HyperBro, Mimikatz, Nishang, OwaAuth, PlugX, ProcDump, PsExec, SysUpdate, TwoFace, Windows Credentials Editor, ZXShell, Living off the Land.

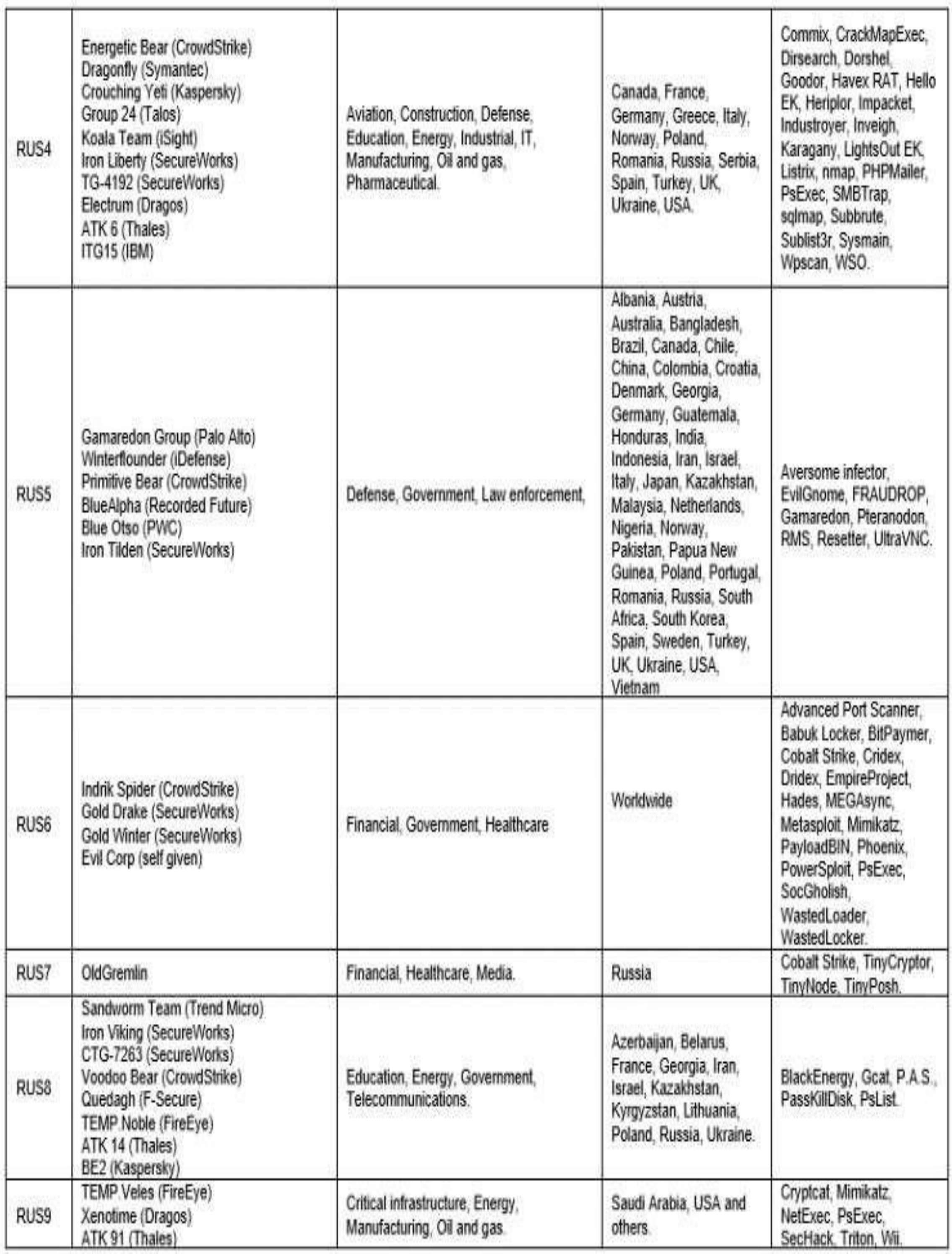


APT40	Leviathan (CrowdStrike) APT 40 (Mandiant) TEMP.Periscope (FireEye) TEMP.Jumper (FireEye) Bronze Mohawk (SecureWorks) Mudcarp (iDefense) Gadolinium (Microsoft) ATK 29 (Thales) ITG09 (IBM)	Defense, Engineering, Government, Manufacturing, Research, Shipping and Logistics, Transportation and other Maritime-related targets across multiple verticals.	Belgium, Cambodia, Germany, Hong Kong, Malaysia, Norway, Philippines, Saudi Arabia, Switzerland, USA, UK and Asia Pacific Economic Cooperation (APEC).	AIRBREAK, BADFLICK, BlackCoffee, China Chopper, Cobalt Strike, DADJOKE, Dadstache, Derusbi, Gh0st RAT, GRILLMARK, HOMEFRY, LUNCHMONEY, MURKYTOP, NanHaiShu, PlugX, scanbox, SeDLL, Windows Credentials Editor, ZXShell, Living off the Land.
APT41	APT 41 (FireEye) TG-2633 (SecureWorks) Bronze Atlas (SecureWorks) Red Kelpie (PWC) Blackfly (Symantec)	Construction, Defense, Education, Energy, Financial, Government, Healthcare, High-Tech, Hospitality, Manufacturing, Media, Oil and gas, Petrochemical, Pharmaceutical, Retail, Telecommunications, Transportation	Australia, Brazil, Canada, Chile, Denmark, Finland, France, Hong Kong, India, Indonesia, Italy, Japan, Malaysia, Mexico, Myanmar, Netherlands, Pakistan, Philippines, Poland, Qatar, Saudi Arabia, Singapore, South Korea, South Africa, Sweden, Switzerland, Taiwan, Thailand, Turkey, UAE, UK, USA, Vietnam.	9002 RAT, AceHash, ADORE.XSEC, ASPXSpy, Barlaiy, BIOPASS RAT, BlackCoffee, certutil, China Chopper, Cobalt Strike, COLDJAVA, Crackshot, CrossWalk, DEADEYE, Derusbi, DIRTCLEANER, EasyNight, GearShift, Gh0st RAT, HDRoot, HighNoon, HighNote, HKDOOR, Jumpall, LATELUNCH, LIFEBOAT, Lowkey, MessageTap, Meterpreter, Mimikatz, njRAT, NTDSDump, PACMAN, PipeMon, PlugX, POTROAST, pwdump, RedXOR, ROCKBOOT, SAGEHIRE, ShadowHammer, ShadowPad Winnti, Skip-2.0, Speculoos, SWEETCANDLE, TERA, TIDYELF, WIDETONE, Winnti, WINTERLOVE, xDll, XDOOR, XMRig, ZXShell, Living off the Land.

6.1.2. Russia

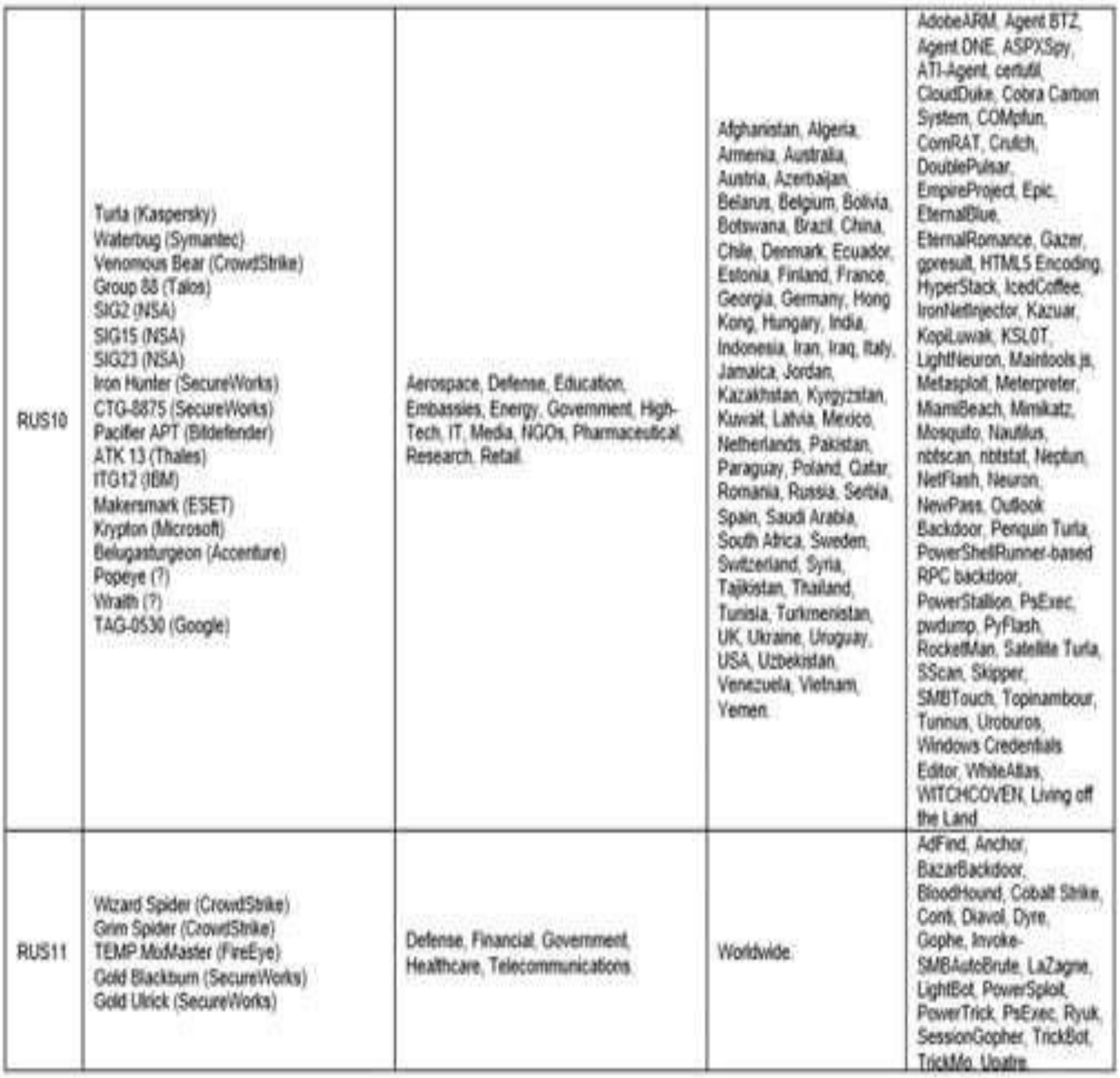


NAME	A.K.A	TARGET SECTORS	TARGET COUNTRIES	ASSOCIATED TOOLS
Suspected attribution: RUSSIA				
APT28	Sofacy (Kaspersky) APT 28 (Mandiant) Fancy Bear (CrowdStrike) Sednit (ESET) Group 74 (Talos) TG-4127 (SecureWorks) Pawn Storm (Trend Micro) Tsar Team (iSight) Strontium (Microsoft) Swallowtail (Symantec) SIG40 (NSA) Snakemackerel (iDefense) Iron Twilight (SecureWorks) ATK 5 (Thales) T-APT-12 (Tencent) ITG05 (IBM) TAG-0700 (Google) Grizzly Steppe (US Government) together with APT 29, Cozy Bear, The Dukes	Automotive, Aviation, Chemical, Construction, Defense, Education, Embassies, Engineering, Financial, Government, Healthcare, Industrial, IT, Media, NGOs, Oil and gas, Think Tanks and Intelligence organizations.	Afghanistan, Armenia, Australia, Azerbaijan, Belarus, Belgium, Brazil, Bulgaria, Canada, Chile, China, Croatia, Cyprus, France, Georgia, Germany, Hungary, India, Iran, Iraq, Japan, Jordan, Kazakhstan, Latvia, Malaysia, Mexico, Mongolia, Montenegro, Netherlands, Norway, Pakistan, Poland, Romania, Slovakia, South Africa, South Korea, Spain, Sweden, Switzerland, Tajikistan, Thailand, Turkey, Uganda, UAE, UK, Ukraine, USA, Uzbekistan, NATO and APEC and OSCE.	Cannon, certutil, Computrace, CORESHELL, DealersChoice, Downdelph, Drovorub, Foozer, HIDEDRV, JHUHUGIT, Koadic, Komplex, LoJax, Mimikatz, Nimcy, OLDBAIT, PocoDown, ProcDump, PythocyDbg, Responder, Sedkit, Sedreco, SkinnyBoy, USBStealer, VPNFilter, Winexe, WinIDS, X-Agent, X-Tunnel, Zebrocy, Living off the Land.
APT29	APT 29 (Mandiant) Cozy Bear (CrowdStrike) The Dukes (F-Secure) Group 100 (Talos) Yttrium (Microsoft) Iron Hemlock (SecureWorks) Minidionis (Palo Alto) CloudLook (Kaspersky) ATK 7 (Thales) ITG11 (IBM) Grizzly Steppe (US Government) together with Sofacy, APT 28, Fancy Bear, Sednit UNC2452 (FireEye) Dark Halo (Volexity) SolarStorm (Palo Alto) StellarParticle (CrowdStrike) Nobelium (Microsoft) Iron Ritual (SecureWorks)	Defense, Energy, Government, Law enforcement, Media, NGOs, Pharmaceutical, Telecommunications, Transportation, Think Tanks and Imagery	Australia, Azerbaijan, Belarus, Belgium, Brazil, Bulgaria, Canada, Chechnya, China, Cyprus, Czech, France, Georgia, Germany, Hungary, India, Ireland, Israel, Japan, Kazakhstan, Kyrgyzstan, Latvia, Lebanon, Lithuania, Luxembourg, Mexico, Montenegro, Netherlands, New Zealand, Poland, Portugal, Romania, Russia, Slovenia, Spain, South Korea, Turkey, Uganda, UK, Ukraine, USA, Uzbekistan, NATO.	7-Zip, AdFind, ATI-Agent, AtNow, CloudDuke, Cobalt Strike, CosmicDuke, CozyDuke, FatDuke, GeminiDuke, GoldFinder, GoldMax, HammerDuke, LiteDuke, meek, Mimikatz, MiniDuke, OnionDuke, PinchDuke, PolyglotDuke, POSHSPY, PowerDuke, RAINDROP, RegDuke, SeaDuke, Sibot, SoreFang, SUNBURST, SUNSPOT, SUPERNOVA, TEARDROP, WellMail, WellMess, Living off the Land.
RUS1	Berserk Bear (CrowdStrike) Dragonfly 2.0 (Symantec) Dymalloy (Dragos)	Energy	Azerbaijan, Belgium, Canada, France, Germany, Italy, Norway, Russia, Singapore, Spain, Switzerland, Turkey, UK, Ukraine, USA.	Goodor, Impacket, Karagany, Phishery, Living off the Land.
RUS2	Cobalt Group	Financial, High-Tech, Media, Retail.	Argentina, Armenia, Austria, Azerbaijan, Belarus, Bulgaria, Canada, China, Czech, Estonia, Georgia, Italy, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Malaysia, Moldova, Netherlands, Poland, Romania, Russia, Spain, Taiwan, Tajikistan, Thailand, Turkey, UK, Ukraine, USA, Vietnam.	ATMSpitter, ATMRipper, AtNow, Cobalt Strike, CobInt, Cyst Downloader, FlawedAmmyy, Formbook, Little Pig, Mimikatz, Metasploit Stager, More_eggs, NSIS, Pony, SDelete, SoftPerfect Network Scanner, Taurus Loader, ThreatKit, VenomKit.
RUS3	Cyber Berkut	Defense, Financial, Government.	Estonia, Germany, Ukraine, USA, NATO.	No information



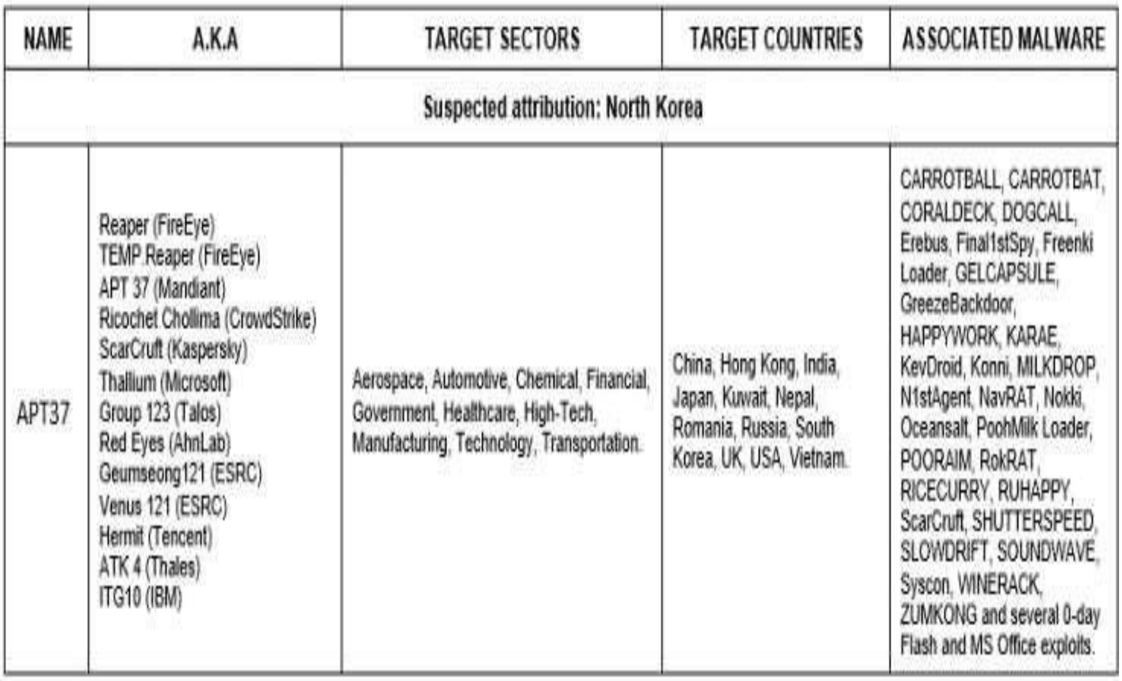
RUS4	Energetic Bear (CrowdStrike) Dragonfly (Symantec) Crouching Yeti (Kaspersky) Group 24 (Talos) Koala Team (iSight) Iron Liberty (SecureWorks) TG-4192 (SecureWorks) Electrum (Dragos) ATK 6 (Thales) ITG15 (IBM)	Aviation, Construction, Defense, Education, Energy, Industrial, IT, Manufacturing, Oil and gas, Pharmaceutical.	Canada, France, Germany, Greece, Italy, Norway, Poland, Romania, Russia, Serbia, Spain, Turkey, UK, Ukraine, USA.	Commix, CrackMapExec, Dirsearch, Dorshel, Goodor, Havex RAT, Hello EK, Heriplor, Impacket, Industroyer, Inveigh, Karagany, LightsOut EK, Listrix, nmap, PHPMailer, PsExec, SMBTrap, sqlmap, Subbrute, Sublist3r, Sysmain, Wpscan, WSO.
RUS5	Gamaredon Group (Palo Alto) Winterflounder (iDefense) Primitive Bear (CrowdStrike) BlueAlpha (Recorded Future) Blue Otso (PWC) Iron Tilden (SecureWorks)	Defense, Government, Law enforcement,	Albania, Austria, Australia, Bangladesh, Brazil, Canada, Chile, China, Colombia, Croatia, Denmark, Georgia, Germany, Guatemala, Honduras, India, Indonesia, Iran, Israel, Italy, Japan, Kazakhstan, Malaysia, Netherlands, Nigeria, Norway, Pakistan, Papua New Guinea, Poland, Portugal, Romania, Russia, South Africa, South Korea, Spain, Sweden, Turkey, UK, Ukraine, USA, Vietnam	Aversome infector, EvilGnome, FRAUDROP, Gamaredon, Pteranodon, RMS, Resetter, UltraVNC.
RUS6	Indrik Spider (CrowdStrike) Gold Drake (SecureWorks) Gold Winter (SecureWorks) Evil Corp (self given)	Financial, Government, Healthcare	Worldwide	Advanced Port Scanner, Babuk Locker, BitPaymer, Cobalt Strike, Cridex, Dridex, EmpireProject, Hades, MEGAsync, Metasploit, Mimikatz, PayloadBIN, Phoenix, PowerSploit, PsExec, SocGholish, WastedLoader, WastedLocker.
RUS7	OldGremlin	Financial, Healthcare, Media.	Russia	Cobalt Strike, TinyCryptor, TinyNode, TinyPosh.
RUS8	Sandworm Team (Trend Micro) Iron Viking (SecureWorks) CTG-7263 (SecureWorks) Voodoo Bear (CrowdStrike) Quedagh (F-Secure) TEMP.Noble (FireEye) ATK 14 (Thales) BE2 (Kaspersky)	Education, Energy, Government, Telecommunications.	Azerbaijan, Belarus, France, Georgia, Iran, Israel, Kazakhstan, Kyrgyzstan, Lithuania, Poland, Russia, Ukraine.	BlackEnergy, Gcat, P.A.S., PassKillDisk, PsList.
RUS9	TEMP.Veles (FireEye) Xenotime (Dragos) ATK 91 (Thales)	Critical infrastructure, Energy, Manufacturing, Oil and gas.	Saudi Arabia, USA and others.	Cryptcat, Mimikatz, NetExec, PsExec, SecHack, Triton, Wii.

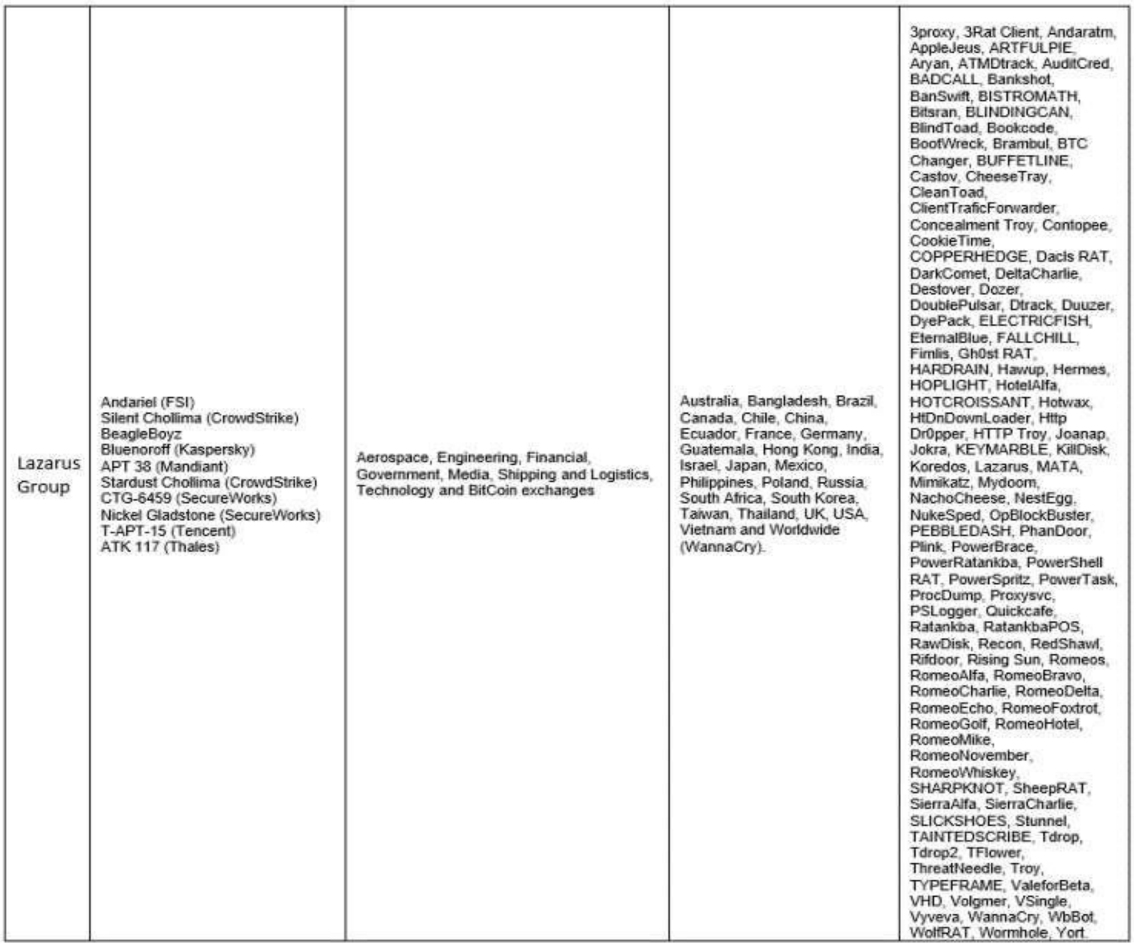
6.1.3. Iran



RUS10	Turla (Kaspersky) Waterbug (Symantec) Venomous Bear (CrowdStrike) Group 88 (Talos) SIG2 (NSA) SIG15 (NSA) SIG23 (NSA) Iron Hunter (SecureWorks) CTG-8875 (SecureWorks) Pacifier APT (Bitdefender) ATK 13 (Thales) ITG12 (IBM) Makersmark (ESET) Krypton (Microsoft) Belugasturgeon (Accenture) Popeye (?) Wraith (?) TAG-0530 (Google)	Aerospace, Defense, Education, Embassies, Energy, Government, High-Tech, IT, Media, NGOs, Pharmaceutical Research, Retail	Afghanistan, Algeria, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bolivia, Botswana, Brazil, China, Chile, Denmark, Ecuador, Estonia, Finland, France, Georgia, Germany, Hong Kong, Hungary, India, Indonesia, Iran, Iraq, Italy, Jamaica, Jordan, Kazakhstan, Kyrgyzstan, Kuwait, Latvia, Mexico, Netherlands, Pakistan, Paraguay, Poland, Qatar, Romania, Russia, Serbia, Spain, Saudi Arabia, South Africa, Sweden, Switzerland, Syria, Tajikistan, Thailand, Tunisia, Turkmenistan, UK, Ukraine, Uruguay, USA, Uzbekistan, Venezuela, Vietnam, Yemen.	AdobeARM, Agent BTZ, Agent.DNE, ASPXSpy, ATI-Agent, certutil, CloudDuke, Cobra Carbon System, COMpfun, ComRAT, Crutch, DoublePulsar, EmpireProject, Epic, EternalBlue, EternalRomance, Gazer, gpresult, HTML5 Encoding, HyperStack, IcedCoffee, IronNetInjector, Kazuar, KopiLuwak, KSL0T, LightNeuron, Maintools.js, Metasploit, Meterpreter, MiamiBeach, Mimikatz, Mosquito, Nautilus, nbtscan, nbtstat, Neptun, NetFlash, Neuron, NewPass, Outlook Backdoor, Penquin Turla, PowerShellRunner-based RPC backdoor, PowerStallion, PsExec, pwdump, PyFlash, RocketMan, Satellite Turla, SScan, Skipper, SMBTouch, Topinambour, Tunnus, Uroburos, Windows Credentials Editor, WhiteAtlas, WITCHCOVEN, Living off the Land
RUS11	Wizard Spider (CrowdStrike) Grim Spider (CrowdStrike) TEMP.MixMaster (FireEye) Gold Blackburn (SecureWorks) Gold Ulrick (SecureWorks)	Defense, Financial, Government, Healthcare, Telecommunications.	Worldwide.	AdFind, Anchor, BazarBackdoor, BloodHound, Cobalt Strike, Conti, Diavol, Dyre, Gophe, Invoke-SMBAutoBrute, LaZagne, LightBot, PowerSploit, PowerTrick, PsExec, Ryuk, SessionGopher, TrickBot, TrickMo, Upatre

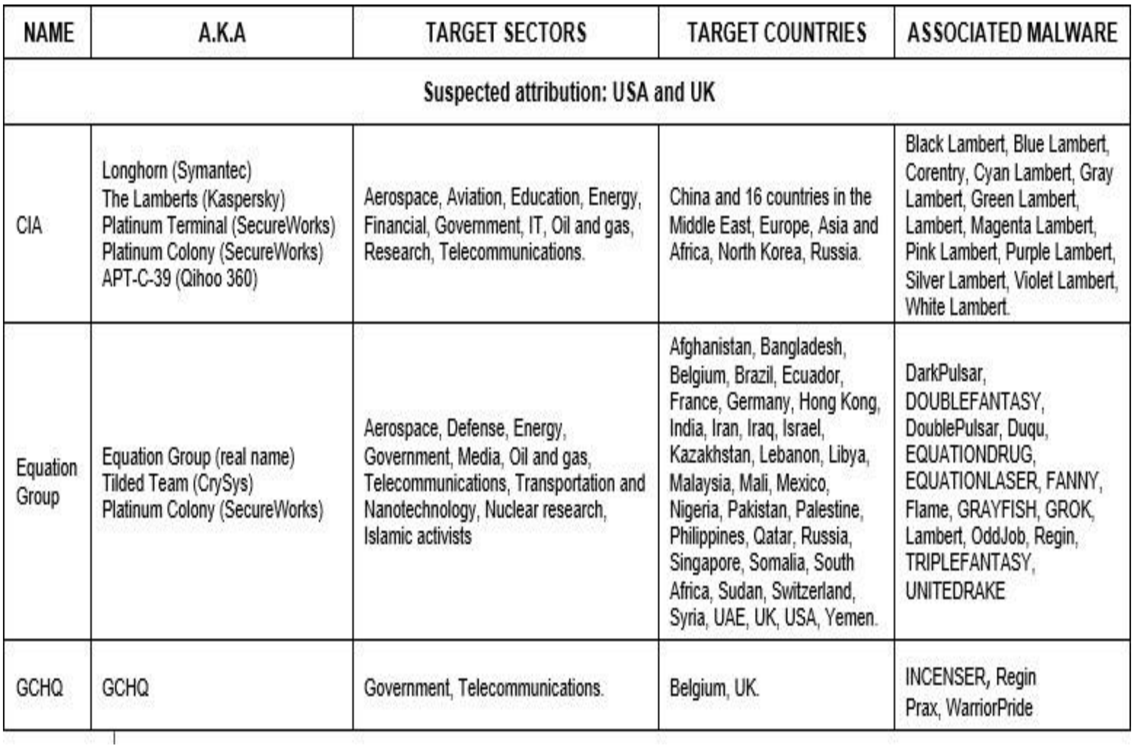
6.1.4. North Korea





NAME	A.K.A	TARGET SECTORS	TARGET COUNTRIES	ASSOCIATED MALWARE
Suspected attribution: North Korea				
APT37	Reaper (FireEye) TEMP.Reaper (FireEye) APT 37 (Mandiant) Ricochet Chollima (CrowdStrike) ScarCruft (Kaspersky) Thallium (Microsoft) Group 123 (Talos) Red Eyes (AhnLab) Geumseong121 (ESRC) Venus 121 (ESRC) Hermit (Tencent) ATK 4 (Thales) ITG10 (IBM)	Aerospace, Automotive, Chemical, Financial, Government, Healthcare, High-Tech, Manufacturing, Technology, Transportation.	China, Hong Kong, India, Japan, Kuwait, Nepal, Romania, Russia, South Korea, UK, USA, Vietnam.	CARROTBALL, CARROTBAT, CORALDECK, DOGCALL, Erebus, Final1stSpy, Freenki Loader, GELCAPSULE, GreezeBackdoor, HAPPYWORK, KARAE, KevDroid, Konni, MILKDROP, N1stAgent, NavRAT, Nokki, Oceansalt, PoohMilk Loader, POORAIM, RokRAT, RICECURRY, RUHAPPY, ScarCruft, SHUTTERSPEED, SLOWDRIFT, SOUNDWAVE, Syscon, WINERACK, ZUMKONG and several 0-day Flash and MS Office exploits.
Lazarus Group	Andariel (FSI) Silent Chollima (CrowdStrike) BeagleBoyz Bluenoroff (Kaspersky) APT 38 (Mandiant) Stardust Chollima (CrowdStrike) CTG-6459 (SecureWorks) Nickel Gladstone (SecureWorks) T-APT-15 (Tencent) ATK 117 (Thales)	Aerospace, Engineering, Financial, Government, Media, Shipping and Logistics, Technology and BitCoin exchanges	Australia, Bangladesh, Brazil, Canada, Chile, China, Ecuador, France, Germany, Guatemala, Hong Kong, India, Israel, Japan, Mexico, Philippines, Poland, Russia, South Africa, South Korea, Taiwan, Thailand, UK, USA, Vietnam and Worldwide (WannaCry).	3proxy, 3Rat Client, Andaratm, AppleJeus, ARTFULPIE, Aryan, ATMDtrack, AuditCred, BADCALL, Bankshot, BanSwift, BISTROMATH, Bitsran, BLINDINGCAN, BlindToad, Bookcode, BootWreck, Brambul, BTC Changer, BUFFETLINE, Castov, CheeseTray, CleanToad, ClientTraficForwarder, Concealment Troy, Contopee, CookieTime, COPPERHEDGE, Dacls RAT, DarkComet, DeltaCharlie, Destover, Dozer, DoublePulsar, Dtrack, Duuzer, DyePack, ELECTRICFISH, EternalBlue, FALLCHILL, Fimlis, Gh0st RAT, HARDRAIN, Hawup, Hermes, HOPLIGHT, HotelAlfa, HOTCROISSANT, Hotwax, HtDnDownLoader, Http Dr0pper, HTTP Troy, Joanap, Jokra, KEYMARBLE, KillDisk, Koredos, Lazarus, MATA, Mimikatz, Mydoom, NachoCheese, NestEgg, NukeSped, OpBlockBuster, PEBBLEDASH, PhanDoor, Plink, PowerBrace, PowerRatankba, PowerShell RAT, PowerSpritz, PowerTask, ProcDump, Proxysvc, PSLogger, Quickcafe, Ratankba, RatankbaPOS, RawDisk, Recon, RedShawl, Rifdoor, Rising Sun, Romeos, RomeoAlfa, RomeoBravo, RomeoCharlie, RomeoDelta, RomeoEcho, RomeoFoxtrot, RomeoGolf, RomeoHotel, RomeoMike, RomeoNovember, RomeoWhiskey, SHARPKNOT, SheepRAT, SierraAlfa, SierraCharlie, SLICKSHOES, Stunnel, TAINTEDSCRIBE, Tdrop, Tdrop2, TFlower, ThreatNeedle, Troy, TYPEFRAME, ValeforBeta, VHD, Volgmer, VSingle, Vyveva, WannaCry, WbBot, WolfRAT, Wormhole, Yort.

6.1.5. USA and UK



NAME	A.K.A	TARGET SECTORS	TARGET COUNTRIES	ASSOCIATED MALWARE
Suspected attribution: USA and UK				
CIA	Longhorn (Symantec) The Lamberts (Kaspersky) Platinum Terminal (SecureWorks) Platinum Colony (SecureWorks) APT-C-39 (Qihoo 360)	Aerospace, Aviation, Education, Energy, Financial, Government, IT, Oil and gas, Research, Telecommunications.	China and 16 countries in the Middle East, Europe, Asia and Africa, North Korea, Russia.	Black Lambert, Blue Lambert, Corentry, Cyan Lambert, Gray Lambert, Green Lambert, Lambert, Magenta Lambert, Pink Lambert, Purple Lambert, Silver Lambert, Violet Lambert, White Lambert.
Equation Group	Equation Group (real name) Tilded Team (CrySys) Platinum Colony (SecureWorks)	Aerospace, Defense, Energy, Government, Media, Oil and gas, Telecommunications, Transportation and Nanotechnology, Nuclear research, Islamic activists	Afghanistan, Bangladesh, Belgium, Brazil, Ecuador, France, Germany, Hong Kong, India, Iran, Iraq, Israel, Kazakhstan, Lebanon, Libya, Malaysia, Mali, Mexico, Nigeria, Pakistan, Palestine, Philippines, Qatar, Russia, Singapore, Somalia, South Africa, Sudan, Switzerland, Syria, UAE, UK, USA, Yemen.	DarkPulsar, DOUBLEFANTASY, DoublePulsar, Duqu, EQUATIONDRUG, EQUATIONLASER, FANNY, Flame, GRAYFISH, GROK, Lambert, OddJob, Regin, TRIPLEFANTASY, UNITEDRAKE
GCHQ	GCHQ	Government, Telecommunications.	Belgium, UK.	INCENSER, Regin Prax, WarriorPride

7. CONCLUSIONS

Most of these CNI are connected to the Internet because it is easier to remotely manage these complex sectors while using technology devices than to physically send a technician engineer to manually checking the functionality of every individual of them.

As the use of the Internet increases, so does the probability of these infrastructures being cyber-attacked by APTs groups to steal unauthorized confidential information for various purposes such as intelligence gathering, financial gain, or sabotaging vital services of a country.

In this paper, we proposed a new system to monitor and protect these critical infrastructures (CNI) based on National SIEM integrated with a CNI-CTI platform from cyber-attacks that occurred in different sectors of CNI.

The way to achieve this advanced cyber intelligence is by improving the exchange information process on cyber attacks across all countries' infrastructure (CNI) sectors.

The efficiency and velocity of this system improve security analyst's response time to take accelerated countermeasures by applying specific policies to their CNIs.

REFERENCES

[1] ENISA (European Union Agency for Network and Information Security), "Cyber Security Information Sharing: An Overview of Regulatory and Non-regulatory Approaches", 2015, pp. 06. online:https://www.enisa.europa.eu/publications/cybersecurity-information- sharing.

[2] Presidency of the Council of Ministers, "NATIONAL STRATEGIC FRAMEWORK FOR CYBERSPACE SECURITY", 2013, pp. 12-15, online:https://www.sicurezzanazionale.gov.it/sisr.nsf/wp-content/uploads/2014/02/italian- national-strategic-framework-for-cyberspace-security.pdf

[3] Cabinet Office UK, "Public Summary of Sector Security and Resilience Plans UK" 2018, pp. 5, online

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786206/20190215_PublicSummaryOfSectorSecurityAndResiliencePlans2018.pdf

[4] Misha Glenny, “DarkMarket. How Hackers Became the New Mafia” 2012, pp. 11

[5] CISA (Cybersecurity and Infrastructure Security Agency), “Advanced Persistent Threat Compromise of Government Agencies, Critical Infrastructure, and Private Sector Organizations Alert (AA20-352A)” April 15 2021, online https://us-cert.cisa.gov/ncas/alerts/aa20-352a

[6] NSA (National Security Agency), “APT29 Targets U.S. and Allied Networks”, U/OO/132340-21, April 2021 Ver. 1.0, pp. 21, 0292, online:https://media.defense.gov/2021/Apr/15/2002621240/- 1/-1/0/CSA_SVR_TARGETS_US_ALLIES_UOO13234021.PDF

[7] Executive Order No 13636 President of USA, “Improving Critical Infrastructure Cybersecurity”, The White House, February 12, 2013. Sec. 4-c, online: https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/executive-order-improving- critical-infrastructure-cybersecurity

[8] Kenneth Geers, “The Cyber Threat to National Critical Infrastructures: Beyond Theory”, July 7, 2009, pp. 2, online https://ccdcoe.org/uploads/2018/10/Geers2009_The-Cyber-Threat-to-National-Critical-Infrastructures.pdf

[9] The Federal Council, “National strategy for the protection of Switzerland against cyber risks (NCS) 2018-2022” 2018, pp, 4 online: https://www.ncsc.admin.ch/ncsc/en/home/strategie/strategie-ncss-2018-2022.html

[10] Eric A. Fischer, Edward C. Liu, John W. Rollins, Catherine A. Theohary, Congressional Research Service, “The 2013 Cybersecurity Executive Order: Overview and Considerations for Congress” pp. 3, online: https://sgp.fas.org/crs/misc/R42984.pdf

[11] Gartner, “Security Information And Event Management”, online:https://www.gartner.com/en/information-technology/glossary/security-information-event-management

[12] FireEye, “Cyber Threat Intelligence 101” online: https://www.fireeye.com/mandiant/threat-intelligence/what-is-cyber-threat-intelligence.html

[13] INSA (Intelligence and National Security Alliance), “Managing A Cyber Attack On Critical Infrastructure: Challenges Of Federal, State, Local, And Private Sector Collaboration” August 2018, Pp. 13, online:https://www.insaonline.org/wp-content/uploads/2018/08/INSA-Managing-Cyber-Attack-Critical-Infrastructure.pdf

[14] ENISA (European Union Agency for Network and Information Security), “Advanced persistent threat incident handling” September 2014, pp.2 online https://www.enisa.europa.eu/topics/trainings-for-cybersecurity-specialists/online-training-material/documents/advanced_persistent_threat_incident_handling_toolset

[15] N. Pissanidis, H. Rõigas, M. Veenendaal, “Countering Advanced Persistent Threats through Security Intelligence and Big Data Analytics” pg.1, NATO CCD COE Publications, Tallinn, online: https://www.ccdcoe.org/uploads/2018/10/Art-15-Countering-Advanced-Persistent-Threats-through-Security-Intelligence-and-Big-Data-Analytics.pdf

[16] FireEye, “Advanced Persistent Threat Groups”, online: https://www.fireeye.com/current- threats/apt-groups.html

[17] MITRE ATT&CK, “Advanced Persistent Threat Groups”, online: https://attack.mitre.org/groups/G0082/

[18] ThaiCERT (Thailand Computer Emergency Response Team) “Advanced Persistent Threat Groups”, online: https://apt.thaicert.or.th/cgi-bin/listgroups.cgi

[19] Malpedia, “Advanced Persistent Threat Groups”, online: https://malpedia.caad.fkie.fraunhofer.de/

IMPROVING THE REQUIREMENTS ENGINEERING PROCESS THROUGH AUTOMATED SUPPORT: AN INDUSTRIAL CASE STUDY

Fabio Alexandre M.H. Silva, Bruno A. Bonifacio, Fabio Oliveira Ferreira,
Fabio Coelho Ramos, Marcos Aurelio Dias and Andre Ferreira Neto

Sidia Institute of Science & Technology, Manaus, Amazonas, Brazil

ABSTRACT

Although Distributed Software Development (DSD) has been a growing trend in the software industry, performing requirements management in such conditions implies overcoming new limitations resulting from geographic separation. SIDIA is a Research and Development (R&D) Institute, located in Brazil, responsible for producing improvements on the Android Platform for Samsung Products in all Latin America. As we work in collaboration stakeholders provided by Mobile Network Operators (MNO) from Latin countries, it is common that software requirements be provided by external stakeholders. As such, it is difficult to manage these requirements due to the coordination of many different stakeholders in a distributed setting. In order to minimize the risks, we developed a tool to assist our requirements management and development process. This experience paper explores the experience in designing and deploying a software approach that facilitates (I) Distributed Software Development, (II) minimizes requirements error rate, (III) teams and task allocations and (IV) requirements managements. We also report three lessons learned from adopting automated support in the DDS environment.

KEYWORDS

Industrial case study, requirement management, DSD, distributed software development, RM, automation, industrial experience.

1. INTRODUCTION

Distributed Software Development (DSD) has been a growing trend as the software industry is experiencing increasing commercial globalization [1]. In this scenario, many companies have been adopting DSD in their software products to accelerate the time to market for new products, better customer satisfaction, and higher product quality [2].

On the other hand, working with distributed teams also face new challenges, particularly in requirements management: communication, software documentation and project coordination [3]. As a means to overcome these challenges, the software industry has sought to automate their process and tasks. In the context of SIDIA, there was the need for tools that are essential for collaboration among team members, enabling the facilitation, automation, and control of the entire requirements management process [4]. However, the existing tools are rarely tailored to the needs of a collaborating group of engineers [5]. Therefore, SIDIA had to develop its own tools that meet the company's needs.

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 227-236, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111819

SIDIA is a R&D Institute and a Samsung Company strategic partner, located in Manaus-Brazil, that develops innovative software solutions in various areas, such as machine learning, games, data mining and others related to mobile products. SIDIA is responsible for the development of embedded software and improvements on that Android Platform for Samsung Products in all Latin America. The institute collaborates with Samsung Mobile division, located in Korea, and external stakeholders provided by Mobile Network Operators (MNOs) from other Latin American countries (e.g., Brazil, Mexico, Chile, Peru). For this reason, to meet the demands of MNOs, SIDIA works on a DSD environment. MNOs are the main Samsung clients as relates to the acquisition of Samsung's mobile products. Thus, these stakeholders act as middlemen between MNOs and Samsung, who constantly provide software requirements that need to be implemented into Samsung's mobile products. There are several external stakeholders that present a given MNO in a particular country. For instance, there is a stakeholder in Ecuador who represents all of Claro's requirements in that country. Given that there are many countries in Latin America and each country with several MNOs, the management of all the requirements becomes a difficult process and this could lead to error-prone software products.

In this context, the requirements management process becomes difficult due to the coordination of many different stakeholders in a distributed setting, due to geographic dispersion, language and time zone differences. This has led to a challenge of implementing and validating requirements (e.g., requirement consistency, requirement integration problems and wrongly implemented requirements), which leads to long delays and risks during the software development process.

In order to minimize these challenges, we developed a tool to assist in our requirements management process. In this paper we report the experience in designing and deploying this tool, referred to as Checklist Tool, whose main objective is to facilitate the requirements management process. The Checklist Tool improves requirements testing and validation through integration between systems in the context of DSD. Our results show important improvements in team productivity (e.g., minimizing the time to execute tasks), minimizing error rates (with a reduction in 30% error rates) and task allocation (e.g., one developer can simultaneously do more than one task). We also report the lessons learned from adopting automated support in a DSD environment.

This paper is structured as follows: Section II provides some related works. Section III describes the SIDIA process and the support tool added. In Section IV we present the results achieved by using the proposed tool. In Section V we present the conclusion and propose some future directions.

2. RELATED WORKS

In relation to software engineering, one of the areas mostly affected by a DSD environment is Requirements Engineering (RE). To overcome this difficulty software industry moves to automate the requirements management process [7]. According to [6] DSD requires software tools (management tools, development tools, etc.) to minimize problems such as: geographic dispersion, control and coordination breakdown, communication, team engagement and socio-cultural differences. Moreover, It is important to propose and analyze tools in real scenarios [4][12]. We describe some existing requirements management tools in the following paragraphs.

Sinha et al. [8] proposed a distributed requirement management tool called EGRET (eclipse-based global requirements tool), after interacting for more than one year with approximately 30 IBM employees, involved in distributed development. The EGRET prototype was tested in three projects at the requirements definition stage. Users reported a good experience: "*found the tool very useful for capturing requirements, having discussions, and tracking requirement changes*". Goda Software presents a solution for requirements management in the form of Analyst Pro [9]. It facilitates requirements specification, tracking and visual traceability analysis. It is a scalable solution, which can provide a collaborative environment that allows sharing of common pool of project information among stakeholders. The requirements can be tracked through design and testing.

Vitech Corporation developed a tool for requirements management, CORE 5.1 [10]. Main features of CORE 5.1 are: reducing schedule risk, improving communication, enabling collaboration, defining and verifying requirements. It also ensures completeness and consistency as well as provides facility to plan tests at an early stage. It also ensures up-to-date documentation and improves planning, visibility and control.

Projectricity developed a requirements management tool called Projectricity [11]. It is a web-based project management platform that enables project team members to efficiently communicate and work collaboratively no matter where they are located. It manages the requirements at all levels of project. It has ability to manage project and task information. Also it is able to manage requirements, test plans, change requests, traceability, and problems in requirements, risks and documentation.

As we can observe, the most common feature in the above-mentioned requirement tools is the issue tracker, some of them include attachment of requirement documents. It is no surprise that issue traceability is very important for requirements management since it is an important component for functionality testing and software validation. In the context of our proposed tool, we combine some features from the related works, including: requirements tracking, requirements specification, risk reduction, communication improvement, collaboration and requirements validation. We describe this in more details in the next section.

3. THE SIDIA MOBILE PRODUCT REQUIREMENT PROCESS

SIDIA develops and updates embedded software for Samsung products commercialized in Latin America. The process of developing and updating this software is divided into three main categories: new models, Operating System (O.S.) upgrade, or maintenance release or MR (for products already in the market). During the software development process, the main objective is the generation of releases. Releases are software versions containing bug fixes, security updates and requirements provided by MNOs. Therefore, once a software version is released, be it a new model, O.S. upgrade or MR, this version goes through a series of tests including validation of MNOs. Once the release is approved, it is then propagated to the respective mobile devices and the end user can download and install. Figure 1 below presents the release process with respect to the requirements management process.

The Android Platform development process starts with requirements definition, by external stakeholders representing MNOs (represented by step (1) in Figure 1 above). The MNOs define, refine and add the requirements on the external system containing information such as O.S. version, device information, mobile applications (apps), wallpapers, and other features (represented as System Requirements in Figure 1).

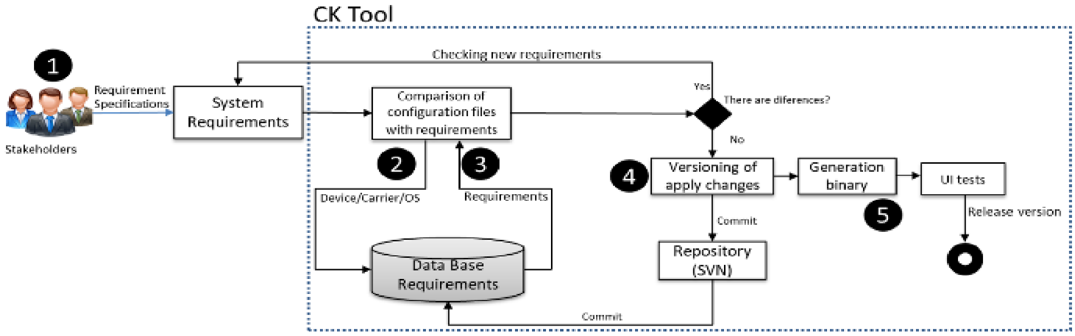


Figure 1. The Release Process from the Requirements Management Perspective.

After that, SIDIA's development team collects requirements and implement according requested by each MNO to each model. This goes through a series of verifications, comparing the requirements from MNOs with that stored in a local data base (steps (2) and (3) of Figure 1). This verification goes on until there are no more differences between the requirements in the database and requirements from MNOs. Upon completion of this phase, the new requirement changes are embedded to the software (step (4) of Figure 1) and stored in a repository. In parallel, the tool is integrated with an SVN (subversion) server and a model compile feature.

system, that uses continuous integration to control file versions and built binaries and tests. Once embedded, a binary is generated (step (5) of Figure 1) and this goes through a series of UI tests. Once the tests are successful, the binary is released.

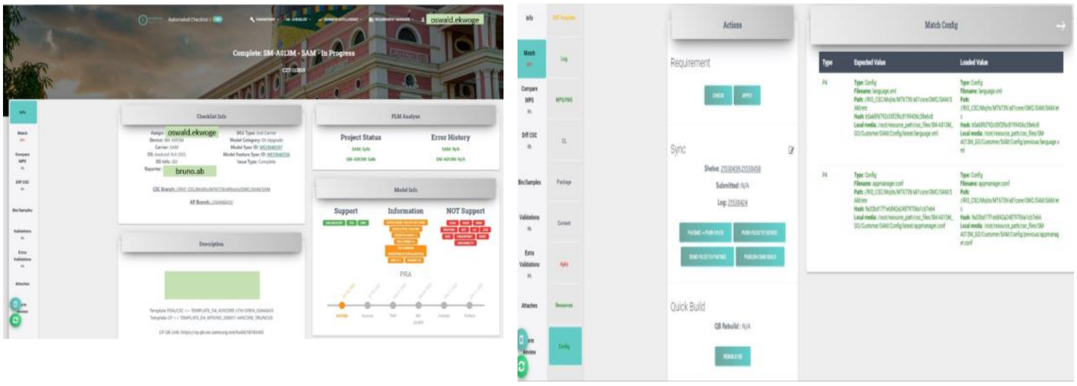
The main verification step of this process (steps (2) and (3)) has been manually done in the past and this has led to human errors, missing or wrong requirements, applications with wrong versions, which led to long release delays, and rework. Some of these errors have led to serious consequences like delays in market delivery and consequently monetary loss. This led to the need for an automated tool which verifies and applies the requirements with very little human intervention.

For this reason, the Checklist Tool was developed. This solution aimed to automate the requirements validation and testing, minimizing errors occurrence and rework related to missing or wrong requirements. This tool is described in more details in the next section.

4. CHECKLIST TOOL

The tool is divide into three modules: (1) Requirements, (2) Business Intelligence, and (3) Requirement Manager. Initially we developed the Business Intelligence Module to capture activity logs from developers. This feature was important to collect the team's data and create a dataset containing information for each developer. Such information includes: time taken to execute tasks, previously executed or applied MNO requirements, devices to which MNO requirements were applied and average errors committed while executing or applying the requirements. Based on this information, this module is able to recommend tasks to developers.

It is worth noting that even though the tool can recommend tasks to developers, the developers can also manually choose the tasks or tasks can be assigned to them. To use this module, the developers must first authenticate with their ID. Once authenticated, the Checklist Tool can then assign the developer with tasks to apply a software requirement. This is shown in below.



Complete: SM-A013M - SAM - In Progress

Checklist Info

Assign: oswald.ekwoge

Reporter: bruno.ab

PLM Analyse

Project Status

Error History

Model Info

Support

Information

NOT Support

PRA

Description

Actions

Requirement

Sync

Shelve: 21530459,21530458

Submitted: N/A

Log: 21530424

Quick Build

QB Rebuild : N/A

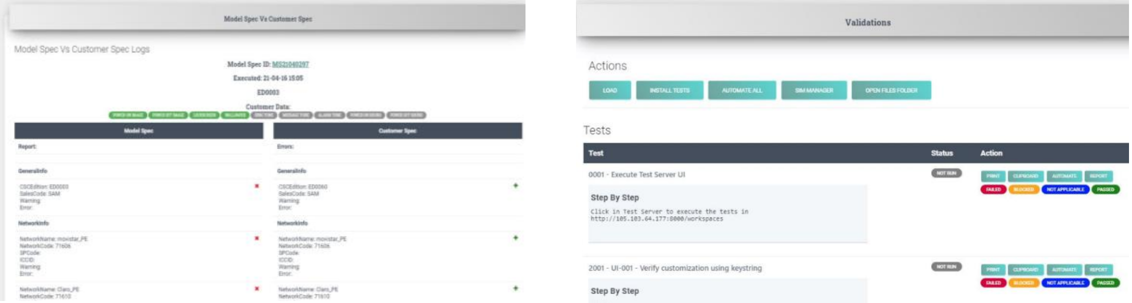
Match Config

Type	Expected Value	Loaded Value
P4	Type: Config Filename: language.xml	Type: Config Filename: language.xml
P4	Type: Config Filename: appmanager.conf	Type: Config Filename: appmanager.conf

Figure 2. Checklist Tool Dashboard and Requirements Verification and Execution.

As shown in Figure 2.1, the task generated by the Checklist Tool is composed of: developer to whom task is assigned, the device under test (shown as example was the Samsung Galaxy A01 Core (identified by code name SM-A013M)), the MNO, in this case Movistar (we identify the MNO by ID, in this example we used SAM as code for Movistar for Peru), the operating system used (GO - Android GO) and its version (11.0), the model category, in this case, O.S Upgrade. In addition, the Checklist Tool shows the requirements that have to be applied using tags for each requirements comprised of the model specification and feature specification. On the left hand corner of the dashboard, there are certain actions that the developer can choose, such as Match, Compare model specification, Diff on User Interface (UI), Validations, Extra Validations, Attaches and Form Review. When the developer clicks on the Match button, this takes him/her to a new screen where he/she can check if there are new requirements, and in case there are, he/she can apply them (Figure 2.2). On the right hand side of this screen is the Match Config, which compares the expected value with newly loaded value. This is shown when the Check button is clicked. Green text implies expected and loaded values are the same; red means the loaded value is different from the expected value. In this case, the apply button can be clicked to apply the new requirements. After application, the Check button can then be clicked again to do another verification.

In addition, the developer can verify the Device model specifications. This verifies features like power on image, power off image, lock screen image, wallpapers, ring tone, message tone, alarm tone, power on sound and power off sound for both Samsung's and the customer's (MNO's) specifications. This is shown in Figure 3.1 below.



Model Spec Vs Customer Spec

Model Spec Vs Customer Spec Logs

Model Spec ID: MS21040297

Executed: 21-04-16 15:05

ED0003

Customer Data:

Model Spec	Customer Spec
Report:	Errors:
GeneralInfo	GeneralInfo
CSCEdition: ED0003 SalesCode: SAM Warning: Error:	CSCEdition: ED0060 SalesCode: SAM Warning: Error:
NetworkInfo	NetworkInfo
NetworkName: movistar_PE NetworkCode: 71606 SPCode: ICCID: Warning: Error:	NetworkName: movistar_PE NetworkCode: 71606 SPCode: ICCID: Warning: Error:
NetworkName: Claro_PE NetworkCode: 71610	NetworkName: Claro_PE NetworkCode: 71610

Validations

Actions

Tests

Test	Status	Action
0001 - Execute Test Server UI Step By Step Click in Test Server to execute the tests in http://105.103.64.177:8000/workspaces	NOT RUN	PRINT CLIPBOARD AUTOMATE REPORT FAILED BLOCKED NOT APPLICABLE PASSED
2001 - UI-001 - Verify customization using keystring Step By Step	NOT RUN	PRINT CLIPBOARD AUTOMATE REPORT FAILED BLOCKED NOT APPLICABLE PASSED

Figure 3. Requirements Validation and Device Model Specification Verification.

As shown in Figure 3.2, the left column shows the Model Spec while the right column shows the customer spec. Once each feature is verified, the color is changed to green; the features still to be verified are left in grey. In addition to the model spec verification, features are also verified. These features can include application permissions, device permissions, network features and application features. This verification is usually done in an external system and the tool just compares the results with the information in the repository, in this case, Perforce (P4). The tool just compares the differences and this is done using the tool's Diff UI.

Another important feature of the tool is validation. The Validations feature will validate the newly applied requirements. Upon clicking the Load button, all the tests for that particular model and MNO are loaded. The tests are run when the developer clicks the Install Tests button. In addition to displaying the tests, the tool also displays the step by step so that the developer can understand exactly what test is being executed and how it is executed. The tool then displays the test status, which can be: PASSED: all tests were passed; NOT APPLICABLE: when test was not applicable; FAILED – when a requirements was wrongly applied; BLOCKED: when an external situation blocked the test from being executed, for instance, samples or binary available; and ALREADY EXECUTED: when test was performed and approved by previous version. One of the tests involves taking screenshots as evidence. The screenshots are stored under "Attaches" in the tool. After applying all requirements, the tool then stores all modifications and test results in a repository for future use and this cycle continues.

Before this tool was developed, the entire process was done manually. Today, most of the tasks have been automated. There are many other features that have been added to the tool, but have been left out of this work, and many more features are being implemented. We hope to publish this in future works. In the next section, we present some important results obtained by using this automated tool. Furthermore, we also present the results of a survey done with developers about their experience using the tool, as well as some lessons learned.

5. CASE STUDY

After the team started using the tool, two main metrics were evaluated: average time to execute tasks and average errors. The time to execute tasks was calculated based on different task phases which are: time to collect and validate requirement, time to apply requirement, time taken for versioning (embedding software and build generation), time to execute tests and total time taken to execute all these steps. This is summarized in Table 1 below. 3

Table 1. Time taken to Execute Tasks Results.

Type	*Requirement Collection and Validation*	*Requirement application*	*Versioning* *(embedding software + build generation)*	*Test execution*	*Total Time Taken*
Manual	~30 min	~30 min	~2 hours	~5 hours	~8 hours
Automated	~3 min	~1 min	~1 hour	~2:30 hours	~3:30 hours

As can be observed from Table 1, without the tool, developers used approximately 30 minutes to collect and validate requirements; with the tool, it took just approximately 3 minutes. This implies a 90% time gain by using the tool. In terms of application requirement, it took approximately 30 minutes to do this manually, while the tool performed this activity in about 1 minute, implying a 96% time gain. As relates to versioning, it took about 2 hours to perform this activity when manually done as opposed to just about 1 hour when executed using the tool, implying about a 50% time gain. Finally, when the tests were manually executed, it took approximately 5 hours to perform this activity while the tool reduced this time to almost half the time (two and a half hours). It is worth noting that by the time this version of the tool was developed, just over 80% of the tests were automated. The team is currently working to automate all tests. In total, it took almost a day's work to apply a requirement when done manually, as opposed to just about three and a half hours when performed using the tool. Therefore, by using the tool to execute tasks, we gained about 37.5% of time.

In terms of average errors, we collected data (logs) from 4,500 tasks half of which were manually executed and half were executed using the tool. The task selection and division was random in order to avoid any bias in our results. With respect to manual execution, it was observed that 12% of the tasks presented an issue. On the other hand, with automated tests, about 2% of the tasks presented some issue. This presents about 83% error reduction. These results are very good even though there are several improvements being done on the tool with the aim of achieving near 100% error reduction, especially if all tests can be automated.

In terms of experience of use, we conducted a survey with developers in order to understand their experience using the tool. In total, 36 developers participated (denoted P1 – P36). The survey was a questionnaire composed of just two questions: (i) In relation to the automated tool, how do you classify your experience using the tool, given that 1 is "very bad", 2 is "bad", 3 is "neither good nor bad", 4 is "good" and 5 is "very good"? (ii) Could you describe your experience with the tool in a few words and if possible, suggestions for improvements? Both questions were mandatory even though some participants responded to question (ii) with "I have nothing to say." The results are summarized in Figure 4 below. As can be observed, 15 participants (42%) had a very good experience with the tool, 15 (42%) had a good experience, 4 participants (about 11%) neither had a good nor bad experience, while 2 participants (5%) had a bad experience. Those two participants (P3 and P9) who had a bad experience with the tool respectively explained their reasons and provided suggestions for improvement as follows: "*The tool's buttons and processes are confusing. I will suggest that the UX and usability be improved*.", "*The tool is complete and helps a lot in performing our activities. However, the tool suffers from constant updates which implies the constant execution of a local server by the developer*.".

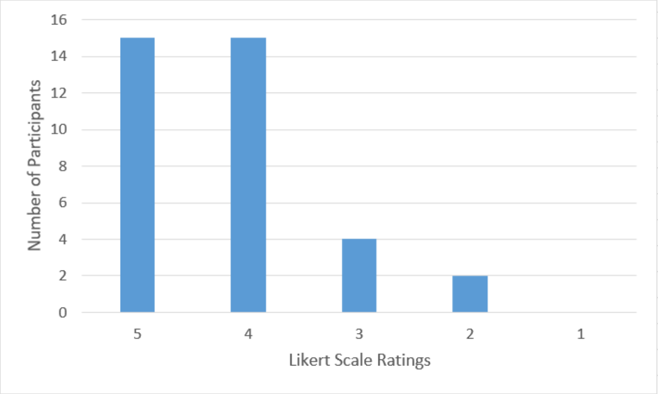


Figure 4. Experience of use results.

Participants (P11, P18, P22 and P34) who neither had a good nor bad experience shared their experiences with focus on UX, transparency and issues faced while working from home during the pandemic: "*The tool has led to a lot of improvements in the process and everything can now be done on a single tool. However, process automation has led to less transparency which is a disadvantage for newly integrated developers*.", "*executing certain tasks when working from home has been an issue, and this has led to some tasks taking longer to execute.", "The UX needs to be improved. Certain buttons must have their positions changed as it can be confusing at times*." "*Improve the execution time for those working at home due to delays caused by the VPN*." (NB: As relates to VPN, during the pandemic, all teams were forced to work from Home Office and as such, in order to guarantee protection on Samsung's network, it was necessary for every employee to install Samsung's VPN. As such, employees reported delays in connectivity with several Samsung applications. This also affected the automated tool which had to access several of these third-party applications and hence a delay in executing certain tasks.).

As for participants who had a good or very good experience, some of them reported certain issues with the tool, the main which are: UX, VPN, bug reports. For instance, participant P5 reported that "*I have a very good experience using the tool. However, the tool has presented several bugs. I have the impression that new features are tested in production. If this is the case, I would suggest that a test environment be created in order to avoid bad user experience*.".

In terms of positive aspects about the tool, we classified participants' experiences into categories: execution time, error rate, robustness, standardization and continuous improvement. This is summarized in Table 2 below.

Table 2. Positive aspects of the tool.

Feature	***Description***	***Participants***	***Example***
Execution time	Time taken to execute tasks	P8, P10, P17, P18, P20, P21	P17: "*The tools had greatly reduced the execution time of certain tasks.*"
Error rate	Rate at which errors occur while executing tasks	P24, P27, P29	P24: "*The tools has facilitated the whole process and also greatly reduced the error margin.*"
Robustness	Reliability of results	P2, P10, P27	P2: "*The tool has ensured reliability in the analysis and application of requirements.*"
Standardization	Task execution follows the same standard	P10, P33	P10: "*...I can easily say the process has become more standardized, with quality and rapidity.*"
Continuous improvement	Constant tool updates	P28, P37	P27: "*The tool is always undergoing continuous improvements, this is excellent. Congrats to all involved!*"

Based on these results, there were some lessons learned

2.3. Lessons Learned

Lesson Learned #1: The release process can be considered an "external body of knowledge" built by a set of external stakeholders. In this context, the tool helped to specify everything that is known considering MNO and Samsung requirements.

With the tool, it was possible to provide greater quality in the specification and application of requirements, leading to lower error rates and quicker time-to-market.

Lesson Learned #2: The tool must consider device model characteristics and specific features from MNOs. This information has to be linked aiming to maintain requirement consistency. The historical data can be used to guarantee that requirements do not contain internal contradictions.

The tool has been able to provide standardization between requirements and device model characteristics. This has led to an alignment between the team, MNO and Samsung.

Lesson Learned #3: The tool has become an infrastructure that supports verification, validation, and testing of releases on the device set. It can be supported by using device farms. We started a simple infrastructure that needs to be refined to consider: more devices connected at the same time and access (with levels of control) by external stakeholders to help validate the application of requirements.

Continuous improvement is important to correct problems related to bugs, time to execute tasks and results reliability. In addition, the tool should be able to simultaneously execute several activities which will lead to even more time gain and even faster time-to-market.

In the next section, we present our conclusion and future works.

6. CONCLUSIONS

As stated by Portillo-Rodriguez [4] and Anwer [12] there is a need for tools that support software engineering processes in the context of DSD and more specially requirements management. Most of the existing tools considers issue tracker as the main feature. Other strategies to manage requirements in DSD scenario have yet to be explored [1]. Our proposed tool, named Checklist Tool, considers three modules: (1) Requirement; (2) Business Intelligence, and; (3) Requirement Management.

Checklist Tool is part of a software tool-based approach that facilitates: (I) Distributed Software Development, (II) minimizing the requirements errors rate, and (III) teams and task allocations. The Checklist Tool integrates external requirement system, and processing to apply requirements and versioning changes, maintain changes history to help developer’s team on management applied requirements.

We have important contribution to productivity team. The automated support assists to minimize time to execution tasks, wrong requirements and overload team. However, we realize there is a need to build a supporting infrastructure that allows validation of applied requirements and release testing in a device family. It is important to remember that releases consider features of the Mobile Network Operators (some even cultural), the manufacturer (in this case, Samsung) and the operating system (Android). Another interesting aspect to be investigated is how to minimize the impact of developer misunderstanding of the requirement. In our case we applied the business intelligence module to recommend a set of requirements for certain developer profiles. However, it is an aspect that still needs further investigation.

ACKNOWLEDGEMENTS

The authors would like to thank everyone, just everyone!

REFERENCES

[1] M. El Bajta et al. “Software Project Management Approaches for Global Software Development: A Systematic Mapping Study” Tsinghua Science and Technology, 2018, 3(6):690–714

[2] M. Lormans, H. Van Dik, A. Van Dersen, E. Nocker, A. de Zeeuw, “Managing Evolving Requirements in an Outsourcing Context: An Industrial Experience Report” In: Proceedings. 7th International Workshop on Principles of Software Evolution, 2004, pp. 148 – 158.

[3] G. Kanakis, Fischer, S., Khelladi, D.E. and Egyed, A., 2019, May. Supporting a flexible grouping mechanism for collaborating engineering teams. In Proceedings of the 14th International Conference on Global Software Engineering (pp. 119-128). IEEE Press.

[4] J. Portillo-Rodriguez, A. Vizcaino, M. Piattini, S. Beecham, Tools used in global software engineering: a systematic mapping review, Information and Software Technology (2012).

[5] Akbar, M. A., Sang, J., Khan, A. A., Mahmood, S., Qadri, S. F., Hu, H., & Xiang, H. (2019). Success factors influencing requirements change management process in global software development. Journal of Computer Languages, 51, 112-130.

[6] Akbar, M. A., Shafiq, M., Kamal, T., & Hamza, M. (2019). Towards the Successful Requirements Change Management in the Domain of Offshore Software Development Outsourcing: Preliminary Results. International Journal of Computing and Digital Systems, 8(03), 205-215.

[7] M. Mukhtar, Z. H. Chuhan, Z. Ahmad, Tools for Requirements Management in GSD: A Survey, International Journal of Scientific & Engineering Research, Volume 6, Issue 4, 2010.

[8] V. Sinha, B. Sengupta and S. Chandra "EGRET: A Collaborative Tool for distributed requirements management", TR, report RI06001, IBM Research 2005.

[9] Goda Software - http://www.analysttool.com - Accessed in: 2020.01.06

[10] Vitech Corporation - http://www.vitechcorp.com/solutions - Accessed in: 2020.01.06

[11] Projectricity Project - http://www.projectricity.com/ - Accessed in: 2020.01.06

[12] S. Anwer, L. Wen, Z. Wang, S. Mahmood, Comparative Analysis of Requirement Change Management Challenges Between In-House and Global Software Development: Findings of Literature and Industry Survey. In: IEEE Access (Volume: 7), pp. 116585 – 116611, 2019 Lee, S.hyun. & Kim Mi Na, (2008) "This is my paper", ABC *Transactions on ECE*, Vol. 10, No. 5, pp120-122.

A NOVEL PRIVACY-PRESERVING SCHEME IN IOT-BASED SOCIAL DISTANCING TECHNOLOGIES

Arwa Alrawais[1], Fatemah Alharbi[2], Moteeb Almoteri[3],
Sara A Aljwair[4] and Sara SAljwair[5]

[1,4,5] College of Computer Engineering and Sciences,
Prince Sattam Bin Abdulaziz University, Alkharj, 16278, Saudi Arabia
[2]College of Computer Science and Engineering, Taibah University,
Yanbu 46522, Saudi Arabia
[3]College of Business Administration, King Saud University,
Riyadh, 11451, Saudi Arabia

ABSTRACT

The COVID-19 pandemic has swapped the world, causing enormous cases, which led to high mortality rates across the globe. Internet of Things (IoT) based social distancing techniques and many current and emerging technologies have contributed to the fight against the spread of pandemics and reduce the number of positive cases. These technologies generate massive data, which will pose a significant threat to data owners' privacy by revealing their lifestyle and personal information since that data is stored and managed by a third party like a cloud. This paper provides a new privacy-preserving scheme based on anonymization using an improved slicing technique and implying distributed fog computing. Our implementation shows that the proposed approach ensures data privacy against a third party intending to violate it for any purpose. Furthermore, our results illustrate our scheme's efficiency and effectiveness.

KEYWORDS

Anonymization, Fog computing, IoT, Privacy, Social distancing technologies.

1. INTRODUCTION

Last year, the world fell prey to the COVID-19 pandemic and made itself prone to the negative impact. The pandemic has devastated the world with a huge number of positive cases, where the total confirmed cases reported over 124 million, including more than two million deaths [1], during one year. One of the most effective techniques to constrain this danger is social distancing, which means reducing physical contact among people. Furthermore, studies have pointed out that social distancing techniques assist in reducing the pandemic, and even single-day negligence can reciprocate the graph. This light up the opportunity in existing technologies in maintaining the social distancing between individuals. The role of these technologies is effective as it facilitates and even maintains social distancing by measuring the distances between individuals and alerts them in case, they cross the regulated distance or if a nearby individual is infected. Mobile technologies, Bluetooth, IoT devices, and other emerging technologies are used in place of social distancing.

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 237-247, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111820

In this scenario, social distancing technologies mainly depend on sharing and collecting data by IoT devices to perform their task accurately and correctly. For instance: • Sensors measure an individual’s health status to detect symptoms of COVID-19. • Wearable devices track an individual’s movements then alert him to keep the imposed distance. Consequently, this enormous data collected by IoT devices will be stored and analyzed through the cloud. By its nature, this personal data is not confidential (e.g., location, identity, and medical information) [2]. However, when data is stored and analyzed, this will create a serious challenge, which can significantly exploit and affect people’s privacy. The researchers in [2] defined privacy as a right of an individual to control his data by figuring out who is entitled to access his data and how they use it. When this data is used in an undisclosed manner by the cloud, such as publishing it or selling it to parties, it is considered a privacy violation. In [3], they mentioned that the impact of privacy violation, especially in publishing Location-based data, could negatively affect people in several manners, such as in employment opportunities and insurance policies. The purpose of enabling (Location-based Services) LBS is to locate places and facilitate access. When using this service in an undisclosed manner (such as monitoring movements), a whole idea of an individual will form. Cloud may sell this information to other companies, and when you apply for a job, you might get rejected because of this disclosure data. The statistics in [4] stated that 54% of adults reject downloading tracking applications for social distancing, were rejection reason of 30% of them to preserve their privacy. In this paper, we propose a novel scheme to address users’ concerns about storing their data in the cloud for social distancing purposes. Applying a new approach based on anonymization using a slicing-fog privacy-preserving technique on the collected data before transferring them to the cloud. We highlight our contributions as follows: We are introducing a novel scheme called slicing-fog privacy-preserving technique by improving the slicing technique used in anonymization to protect users’ privacy in social distancing technology from the cloud itself. Taking advantage of the distributed fog computing architecture to implement the proposed scheme. Demonstrate the efficient performance of implementing the proposed scheme in terms of computational time. Demonstrate how efficiently the proposed scheme preserves privacy by using two metrics: entropy and estimation error. Logically demonstrates the ability of the proposed scheme to deter three privacy threats: the identity disclosure threat, the attribute disclosure threat, and the correlation analysis attack. The paper is organized as follows. Section 2 illustrates the recent related work. In Section 3, a description of the proposed system is provided. We follow it by explaining the proposed scheme methodology, discussing the implementation, and demonstrating the results in Section 4. We draw our conclusion in Section 5.

2. RELATED WORK

In this section, we mainly summarize the recent and significant privacy-preserving approaches in the research communities. In [3], the authors defined the security and privacy preserving of data collected by the social distancing technologies as a challenging problem. Several proposed mechanisms are presented to preserve the location, identity, and health information in other works. Authors in [5], [6] proposed a scheme in privacy-preserving of LBS by using lightweight cryptography. In another works [2], [7], the authors investigated the privacy-preserving approaches in IoT and discussed the pros and cons of each technique besides mentioning the future issues and open problems. In [8], the authors addressed the role of IoT technology in the COVID-19 pandemic setting aside its contributions to social distancing while clarifying their concerns in terms of security and privacy issues and mentioned that one of the biggest challenges facing the IoT technology is the data collected requires big storage centers. Similarly, researchers in [9] have proposed an approach that demonstrated the effectiveness of IoT in monitoring patients of COVID-19 remotely and studying their cases. In [10], the researchers conducted the first privacy study on 41 official contact tracing applications. They discovered the privacy and security concerns in some applications, where it is possible to access the application’s fingerprint and track some users. As another study in [11] also analyzed 48 COVID-19 applications to assess their privacy elements, focusing on three criteria: data retention, right to opt-out, and compliance and polling many participants. In another work in [4], the authors described three scenarios that contact tracing applications and highlight the actors that threaten privacy and cause sensitive data to leak in these applications. These applications include application developers who may perform some malicious activities such as selling data to a third party and suggest several privacy guidelines, which could apply to any contact tracing application. The researchers in [12] proposed an application that measures distance and gives a real-time alert if social distancing violates. To preserve the user's privacy, the researchers suggested that it does not request any personal information at the time of registration, except for email and the account identification assigned to each user residing on the server side. The work in [13] presented a new initiative by designing an open-source application that combines two tasks based on alerting users if they breach the social distancing as well as tracking the contacts via Bluetooth Low Energy (BLE). The application warns the users if an infection appears, considering the security and privacy concerns by storing the timestamp and deleting the data after 28 days of collecting it. While researchers in [14] pointed out that Ultra-wideband (UWB) technology is capable of effectively collecting the location data and measuring the distance more accurately than BLE and Wi-Fi, with consuming significantly lower power. Relatively fewer efforts in social distancing techniques focus on privacy preservation and not considering all the privacy challenging issues. Even if they conceal the identity, the accumulated storage of data in the cloud poses a danger when analyzing this data, especially location tracking data. It may lead to disclosing the identity or lifestyle. In this paper, we attempt to fill this gap.

3. PROPOSED SYSTEM STRUCTURE

The proposed system structure consists of three layers: IoT layer, fog computing layer, and cloud computing layer, as Fig. 1. It begins with the IoT devices (representing the lowest layer). IoT devices are connected via the Internet to collect data such as location or measure a user's health. The collected information intends to provide services that benefit the end-user, such as alert the user in case of violating the imposed distance or in case there are infected people around him. Due to limited storage capacity in IoT devices, the data is stored and processed by the cloud (representing the top layer). The IoT layer and cloud layer connected through fog computing (representing the middle layer), which is distributed computing that brings storage and computing closer to end-users (the IoT layer). Fog computing aims to improve response time and performance by reducing the overload on the cloud, where it intermediates the layer between the IoT layer and cloud. Our proposed scheme could be placed in the fog computing layer for the purpose of protecting the privacy of end-users from the cloud. This is due to several reasons:

- Fog computing has features that address cloud limitations, as we mentioned earlier.
- It is less subject able to violations [15]; therefore, we can consider it as a secure environment.
- Unlike the cloud, it is not provided from an external source. The data is stored in a distributed manner, not centralized as in the cloud [16], which reduces the surface of privacy violation.
- Finally, the most significant incentive in fog computing ability is to apply policies on data before it is sent to the cloud [17].

4. PROPOSED PRIVACY PRESERVATION APPROACH

The building block of the proposed scheme is data generated by IoT devices, so it is imperative to understand its nature. Firstly, the data formed in a table of rows (records) and columns (attributes). A single record is a set of attributes that describe a single user. Secondly, in terms of privacy, these attributes are classified into three categories [18]:

- Identifiers (ID): they are unique attributes that identify the individual. For instance, name and national identity.
- Quasi-Identifiers (QI): they are attributes that identify an individual when combining two or more attributes. For instance, gender, age, and address.
- Sensitive Attribute (SA): these are the sensitive attributes of the individual, which should not be revealed. For instance, medical data, historical location, and current location.

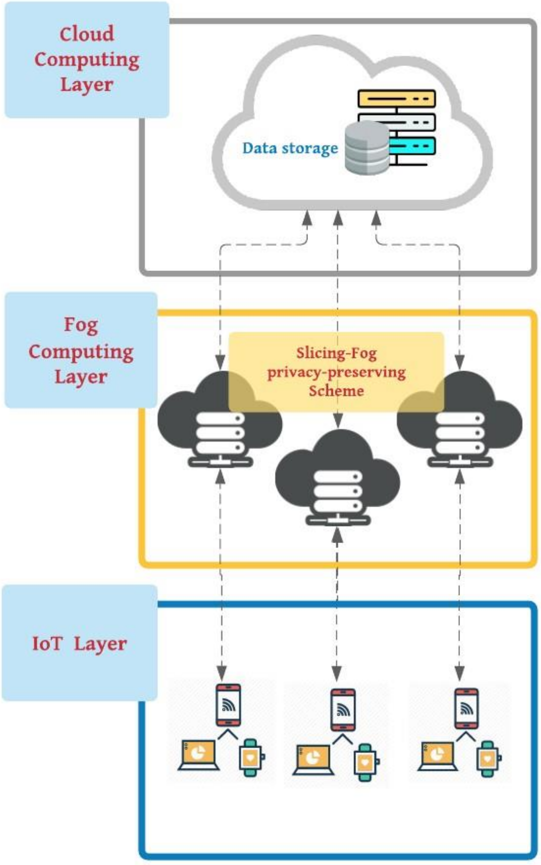


Figure 1. Proposed System Structure

According to [3], the general principle of privacy preserving is defined as the sensitive data available for public access, which must be kept private, such as data stored in the cloud. Where it may lose its privacy, if it is not well preserved or published by the cloud. To achieve this principle, there are many mechanisms to protect the privacy. Based on our observation and study in the field, we find that the slicing technique is an effective and efficient approach to maintaining user data collected for the social distancing purpose. The essential of the slicing technique is to remove the correlation between the attributes of a single user record, which is obtained in several steps.

To illustrate the details of the steps of the proposed privacy-preserving scheme, we assumed data resulting from social distancing processes then performed the steps of the proposed scheme on it by the example shown in Tab. 1.

Step 1: Apply Vertical Partition: the table is partitioned into columns and each column contains a set of attributes as shown in Tab. 2. This partition is according to the correlation of (QI) attributes where the correlation is intense between (QI) attributes in each column to have better utilization. As for the Sensitive Attribute (SA) must be in one column to remove their correlation with the other attributes. Before doing the vertical partition, we have added a primary step in our scheme. Wherein the fog layer will replace the identifier attribute with fog-ID.

Step 2: Apply Horizontal Partition: records are divided into buckets based on a certain duration as shown in Tab. 3, (where) and each bucket contains a set of records.

Step 3: Apply Permutation: last step and most important one in the slicing technique is the random permutation between the values of the columns in each bucket shown in Tab. 4. This step aims to break the correlation between these columns, especially the sensitive attribute columns, as we mentioned earlier in the first step.

The proposed scheme structure is shown in Fig. 2. In this paper, we rely on IoT-based social distancing techniques. We assumed some attributes of the data collected from IoT devices when implementing social distancing techniques. We considered location, health status, and time as sensitive attributes, and the rest are quasi-identifiers attributes. Essentially, the data is sent to fog for temporary storage to perform real-time social distancing operations. Then, the fog implements the proposed scheme on the collected data (i.e., collected by the end of the day) before sending it to the cloud for permanent storage. Considering that the data will return to its natural form when it retrieves from cloud to fog, as needed. Finally, we achieved the purpose of the proposed scheme so that the sliced data stored in the cloud does not reveal the identity of users to the cloud, as shown in Tab. 5. Additionally, if the cloud event is exposed to an attack or the cloud is intended to publish the database.

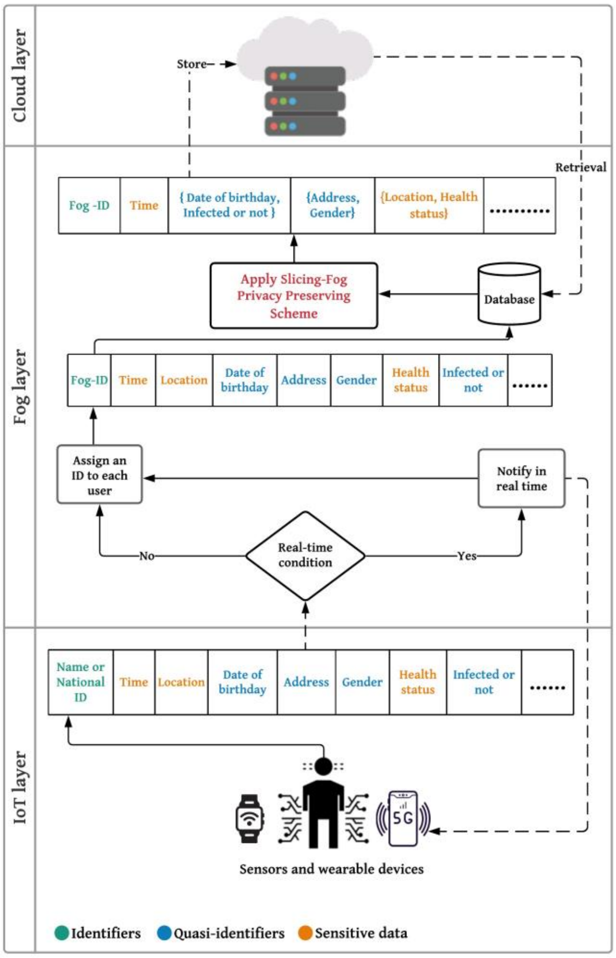


Figure 2. Proposed Privacy-Preserving Scheme.

5. Implementation And Result

In this section, we describe our implementation setup and discuss our scheme evaluation to validate its effectiveness

5.1. Experimental Setup

For performance evaluation purposes, a PC fulfilled our experimental setup requirements. The PC is equipped with Intel® Core™ i7-7700 CPU @ 3.60GHz 3.60 GHz and 16 GB RAM running 64-bit Windows 10 20H2 operating system. We implement our scheme in C# Window Forms language utilizing .NET Framework 4.8. In our implementation, we utilize a set of data collected from 5000 users.

Table 1. Original table of collected data from IoT devices

National ID	Time	Location	Date of birth	Address	Gender	Health status	Infected or not
1087658432	15:02	24.1635° N, 47.3339° E	30-2-1990	Riyadh-11564	female	Healthy	not
1117239099	15:05	26.1415° N, 43.7321° E	13-6-1991	Riyadh-11564	male	Healthy	Infected
1000326722	15:08	24.1532° N, 47.2718° E	5-2-2001	Kharj-16244	male	asthma	Infected
1219876278	15:10	26.3592° N, 43.9818° E	30-7-1990	Qassem-51431	male	Healthy	Infected
1768987657	15:13	26.3072° N, 50.1783° E	19-2-1999	Khobar-34424	female	asthma	not
1032457890	15:17	26.3223° N, 50.2168° E	26-6-1995	Khobar-34424	male	Blood pressure	not
1167975421	15:17	24.7311° N, 46.6701° E	7-2-1988	Riyadh-11564	male	diabetes	Infected
112568900	15:18	24.6951° N, 46.6806° E	9-11-1997	Riyadh-11564	female	asthma	not

Table 2. Step 1 replace all identifier attribute with fog-id and apply vertical partition

Fog-ID	Time	{Date of birth, Infected or not}	{Address, Gender}	{Location, Health status}
333	15:02	{30-2-1990, Not}	{Riyadh-11564, Female}	{24.1635° N, 47.3339° E, Healthy}
549	15:05	{13-6-1991, Infected}	{Riyadh-11564, Male}	{26.1415° N, 43.7321° E, Healthy}
278	15:08	{5-2-2001, Infected}	{Kharj-16244, Male}	{24.1532° N, 47.2718° E, Asthma}
999	15:10	{30-7-1990, Infected}	{Qassem-51431, Male}	{26.3592° N, 43.9818° E, Healthy}
123	15:13	{19-2-1999, Not}	{Khobar-34424, Female}	{26.3072° N, 50.1783° E, Asthma}
345	15:17	{26-6-1995, Not}	{Khobar-34424, Male}	{26.3223° N, 50.2168° E, Blood pressure}
789	15:17	{7-2-1988, Infected}	{Riyadh-11564, Male}	{24.7311° N, 46.6701° E, Diabetes}
267	15:18	{9-11-1997, Not}	{Riyadh-11564, Female}	{24.6951° N, 46.6806° E, Asthma}

Table 3. Step 2 applying horizontal partition at specific span of time, in this scenario period = 10 minutes

Fog-ID	Time	{Date of birth, Infected or not}	{Address, Gender}	{Location, Health status}
333	15:02	{30-2-1990, Not}	{Riyadh-11564, Female}	{24.1635° N, 47.3339° E, Healthy}
549	15:05	{13-6-1991, Infected}	{Riyadh-11564, Male}	{26.1415° N, 43.7321° E, Healthy}
278	15:08	{5-2-2001, Infected}	{Kharj-16244, Male}	{24.1532° N, 47.2718° E, Asthma}
999	15:10	{30-7-1990, Infected}	{Qassem-51431, Male}	{26.3592° N, 43.9818° E, Healthy}
123	15:13	{19-2-1999, Not}	{Khobar-34424, Female}	{26.3072° N, 50.1783° E, Asthma}
345	15:17	{26-6-1995, Not}	{Khobar-34424, Male}	{26.3223° N, 50.2168° E, Blood pressure}
789	15:17	{7-2-1988, Infected}	{Riyadh-11564, Male}	{24.7311° N, 46.6701° E, Diabetes}
267	15:18	{9-11-1997, Not}	{Riyadh-11564, Female}	{24.6951° N, 46.6806° E, Asthma}

Table 4. Step 3 applying random permutation in each bucket to ensure removing the correlation between the attributes of a single record

Fog-ID	Time	{Date of birth, Infected or not}	{Address, Gender}	{Location, Health status}
333	15:02	{30-7-1990, Infected}	{Kharj-16244, Male}	{26.1415° N, 43.7321° E, Healthy}
549	15:05	{5-2-2001, Infected}	{Qassem-51431, Male}	{24.1635° N, 47.3339° E, Healthy}
278	15:08	{30-2-1990, Not}	{Riyadh-11564, Female}	{26.3592° N, 43.9818° E, Healthy}
999	15:10	{13-6-1991, Infected}	{Riyadh-11564, Male}	{24.1532° N, 47.2718° E, Asthma}
123	15:13	{26-6-1995, Not}	{Riyadh-11564, Male}	{24.7311° N, 46.6701° E, Diabetes}
345	15:17	{9-11-1997, Not}	{Khobar-34424, Female}	{24.6951° N, 46.6806° E, Asthma}
789	15:17	{19-2-1999, Not}	{Riyadh-11564, Female}	{26.3072° N, 50.1783° E, Asthma}
267	15:18	{7-2-1988, Infected}	{Riyadh-11564, Male}	{26.3223° N, 50.2168° E, Blood pressure}

Table 5. Shows the efficiency of the scheme, as all users' data do not refer to a specific induvial comparing with original table

Fog-ID	Time	{Date of birth, Infected or not}	{Address, Gender}	{Location, Health status}
333	15:02	{30-7-1990, Infected}	{Kharj-16244, Male}	{26.1415° N, 43.7321° E, Healthy}
549	15:05	{5-2-2001, Infected}	{Qassem-51431, Male}	{24.1635° N, 47.3339° E, Healthy}
278	15:08	{30-2-1990, Not}	{Riyadh-11564, Female}	{26.3592° N, 43.9818° E, Healthy}
999	15:10	{13-6-1991, Infected}	{Riyadh-11564, Male}	{24.1532° N, 47.2718° E, Asthma}
123	15:13	{26-6-1995, Not}	{Riyadh-11564, Male}	{24.7311° N, 46.6701° E, Diabetes}
345	15:17	{9-11-1997, Not}	{Khobar-34424, Female}	{24.6951° N, 46.6806° E, Asthma}
789	15:17	{19-2-1999, Not}	{Riyadh-11564, Female}	{26.3072° N, 50.1783° E, Asthma}
267	15:18	{7-2-1988, Infected}	{Riyadh-11564, Male}	{26.3223° N, 50.2168° E, Blood pressure}

5.2. Result

In our evaluation, we focus on three factors: the performance factor, the privacy factor, and the logical factor.

5.2.1. Performance Factor

The proposed scheme performance is evaluated by measuring the computing time of implementing the scheme on users' data. The computing time greatly affects real-time analysis. Our evaluation result is illustrated in Fig. 3. The computation time of implementing the proposed scheme is measured in milliseconds (*ms*) on various inputs of data. The results show the efficient performance of the scheme, as it consumes a reasonable period when the number of data is increased. This shows that the scheme is lightweight on the device. Therefore, our scheme does not consume resources, consequently power.

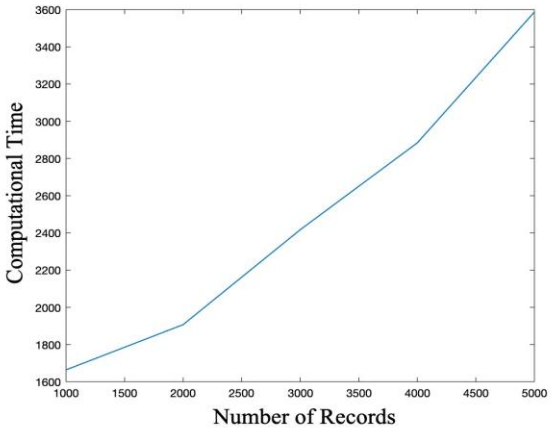


Figure 3. Computational Time

5.2.2. Privacy Factor

It is worth mentioning that our privacy-preserving scheme is based on the principle of anonymization by slicing technique but in a novel way. To evaluate the privacy efficiency of our proposed scheme, we have used entropy and estimation error as a privacy metric. The entropy metric represents the amount of accurate and correct information about a specific user that cloud provider deduces from the collected data. The equation below defined the entropy [19]:

$$\mathrm{E} = -\sum_{\mathrm{i}=0}^{\mathrm{n}} \mathrm{P_i} * \mathrm{Log_2(P_i)} \quad (1)$$

Where n is the number of data sent and Pi is the probability of that data i belongs to a specific user. Therefore, in our scheme the data is collected in the cloud is sliced and does not identify a specific user. Consequently, the entropy is at the maximum, which is the one making the highest level of privacy-preserving.

Estimation error means the rate of the attacker or the violator falling in fault. The attacker must not be able to detect the correct user data, by increasing the value of the estimation error [19]. To achieve this goal in the proposed scheme, when permutations are made between a larger set of data in each bucket, especially the data of columns, which contain sensitive attributes, greater than the estimation error rate of the attacker. Its value and relationship to the entropy metric is illustrated in the following equation:

$$EE = (E) * 100\% \quad (2)$$

Where E is the Entropy value. As we mentioned that the maximum entropy value in the proposed scheme is 1, and when that value is offset in the error estimation equation, the result becomes 100%. This indicates the maximum rate of the attacker and even the cloud provider could be wrong in detecting the correct user data.

5.2.3. Logical Factor

After measuring the effectiveness of the approach in preserving privacy, we can logically demonstrate the validity of the approach in deterring privacy threats.

- The identity disclosure threat: giving a fog-id to each user instead of his name and identification, is a deterrent against the threat of identity disclosure.
- The attribute disclosure threat: this threat discloses the user identity and sensitive data, through combining data from more than one attribute. In the proposed scheme, the process of data permutation in each bucket is to remove the correlation between the sensitive attributes. Thus, the information about any user can not be identified when combining two or more attributes.
- Correlation Analysis Attack: the attacker collects and tracks the user’s history of location data, health status, or other serial data.

Then it analyzes it to predict the new data [19]. To deter this type of attack, the main principle of the slicing technique is to remove the correlation between each user’s attributes by permutations within the bucket. Furthermore, the approach provides proper preservation of the utility. For instance, when conducting statistics that usually executed in social distancing such as the number of infected people of ages between 30 and 35 years, it will give correct results. Due to sliced data, which is divided vertically, according to the most associated attribute as mentioned previously. Finally, the implementation of this proposed approach is not limited to social distancing techniques rather, it can be implemented on various IoT applications.

6. CONCLUSION

Social distancing has greatly contributed for limiting the spread of pandemics in recent years. While the benefits associated with implementing social distancing technology are numerous, privacy issues have been raised. Consequently, we have proposed a new scheme to preserve the privacy based on improving the slicing technique and imply fog computing. This approach has enhanced the level of privacy in social distancing techniques. In our future work, we plan to apply our scheme in more complex environment where privacy preserving is demanded.

ACKNOWLEDGMENT

This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project No.2019/01/10411.

REFERENCES

[1] “World Health Organization: Who coronavirus (covid-19) dashboard.” https://covid19.who.int (accessed Mar. 28, 2020).

[2] A. A. A. Sen, F. A. Eassa, K. Jambi and M. Yamin, “Preserving Privacy in Internet of Things: A Survey,” International Journal of Information Technology, vol. 10, no. 2, pp. 189-200, 2018.

[3] C. T. Nguyen, Y.M. Saputra. N. V. Huynh, N.T. Nguyen and T. V. Khoa et al., “A comprehensive survey of enabling and emerging technologies for social distancing—part ii: Emerging technologies and open issues,” IEEE Access, vol. 8, pp. 154209–154236, 2020.

[4] M. Shukla, R. MA, S. Lodha, G. Shroff and R. Raskar, “Privacy guidelines for contact tracing applications,” arXiv preprint arXiv: 2004.13328, 2020.

[5] H. Shen, M. Zhang, H. Wang, F. Guo and W. Susilo, “A Lightweight Privacy-Preserving Fair Meeting Location Determination Scheme,” IEEE Internet of Things Journal, vol. 7, no. 4, pp. 3083-3093, 2020.

[6] Y. Pu, J. Luo, Y. Wang, C. Hu and Y. Huo et al., “Privacy preserving scheme for location based services using cryptographic approach,” in Proceedings of PAC, Washington, DC, USA, pp. 125–126, 2018.

[7] A. Alrawais, A. Alhothaily, C. Hu and X. Cheng, “Fog computing for the internet of things: Security and privacy issues,” IEEE Internet Computing, vol. 21, no. 2, pp. 34–42, 2017.

[8] M. Kamal, A. Aljohani and E. Alanazi, “IoT meets covid-19: Status, challenges, and opportunities,” arXiv preprint arXiv: 2007.12268, 2020.

[9] S. Jaafari, A. Alhasani, E. Alghosn, R. Alfahhad and S. M. Almutairi, “Certain investigations on IoT system for Covid-19,” in Proceedings of ICCIT-1441, Tabuk, Saudi Arabia, pp. 1–4. IEEE, 2020.

[10] H. Wen, Q. Zhao, Z. Lin, D. Xuan and N. Shroff, “A study of the privacy of covid-19 contact tracing apps,” in Proceedings of SecureComm:2020, Washington, DC, USA, pp. 297–317. Springer, 2020.

[11] T. Sharma, T. Wang and M. Bashir, “Advocating for users’ privacy protections: A case study of Covid-19 apps,” in Proceedings of MobileHCI'20, New York, NY, USA, pp. 1–4, 2020.

[12] A. Ksentini and B. Brik, “An edge-based social distancing detection service to mitigate Covid-19 propagation,” IEEE Internet of Things Magazine, vol. 3, no. 3, pp. 35–39, 2020.

[13] Y. C. Ho, Y. H. Chen, S. H. Hung, C. H. Huang and P. Po et al., “Social distancing 2.0 with privacy-preserving contact tracing to avoid a second wave of Covid-19,” arXiv preprint arXiv: 2006.16611, 2020.

[14] F. S. CA and S. A. Kabeer, “Smart access card system to mitigate the covid-19 outbreak,” in Proceedings of ICCAKM, Dubai, UAE, pp. 168–173. IEEE, 2021.

[15] V. Puri, S. Sachdeva and P. Kaur, “Data anonymization for privacy protection in fog-enhanced smart homes,” in Proceedings of ICSC, Noida, India, pp. 201– 205. IEEE, 2020.

[16] P. Bellavista, J. Berrocal, A. Corradi, S. K. Das and L. Foschini et al., “A survey on fog computing for the internet of things,” Pervasive and mobile computing, vol. 52, pp. 71–99, 2019.

[17] A. A. A. Sen and M. Yamin, “Advantages of using fog in IoT applications,” International Journal of Information Technology, vol. 13, no. 3, pp. 829–837, 2021.

[18] A. Kumar and M. Gyanchandani, "A comparative survey on privacy preservation and privacy measuring techniques in data publishing," in Proceedings of ICICCS, Madurai, India, pp. 1902–1906. IEEE, 2018.

[19] S. S. Albouq, A. A. A. Sen, A. Namoun, N. M. Bahbouh, and A. B. Alkhodre et al., "A double obfuscation approach for protecting the privacy of IoT location based applications," IEEE Access, vol. 8, pp. 129415–129431, 2020.

AUTHORS

Arwa Alrawais received the M.S. degree in computer science and the Ph.D. degree from the Department of Computer Science, The George Washington University, Washington, DC, USA, in 2011 and 2017, respectively. She holds a patent in system and method for remote authentication with dynamic usernames. Her current research interests include network security, wireless and mobile security, and algorithm design and analysis. She serves as a Professional Reviewer for several conferences and journals of the IEEE and ACM.

Fatemah Alharbi is an assistant professor in the Computer Science department at Taibah University, Yanbu, Saudi Arabia. She received her Ph.D. from University of California, Riverside, in 2020. Her research interests are on system and network security, including vulnerability discovery, Internet of Thing (IoT), applied cryptography, applied program analysis, system building, and measurement of real-world security problems.

MoteebAlmoteri received the B.Sc. degree in computer science (information assurance emphasis) from the University of Findlay, OH, USA, in 2010, the M.Sc. degree in information security and assurance from Robert Morris University, PA, USA, in 2012, and the Ph.D. degree in computer science from the Florida Institute of Technology, FL, USA, in 2017. He has been an Assistant Professor with the Department of MIS, Business Administration College, King Saud University, since February 2018, and the Chairman of the MIS Department, King Saud University, since January 2019. His research interests include cloud security, security shared responsibility, information security strategies, information security management, cyber security GRC, and computer vision.

Sara AAljwair received the B.S. degree in Information System from the Department of Information System in 2019 and the M.S. degree in Engineering of Cybersecurity from the Department of Computer Engineering in 2021, The Prince Sattam bin Abdulaziz University, Al-kharj, KSA. Her current research interests include the security and privacy of the Internet of things, algorithm design and analysis.

Sara SAljwair received the B.S. degree in Information System from the Department of Information System in 2017 and The M.S. degree in engineering of cybersecurity from the Department of Computer Engineering, The Prince Sattam bin Abdulaziz University, Al-kharj, KSA, in 2021. Her current research interests include security and privacy in the Internet of Things applications and algorithm design and analysis.

METHOD FOR ORTHOGONAL EDGE ROUTING OF DIRECTED LAYERED GRAPHS WITH EDGE CROSSINGS REDUCTION

Jordan Raykov

JDElite Consulting, Boulder, Colorado, USA

ABSTRACT

This paper presents a method for automated orthogonal edge routing of directed layered graphs using the described edge crossings reduction heuristic algorithm. The method assumes the nodes are pre-arranged on a rectangular grid composed of layers across the flow direction and lanes along the flow direction. Both layers and lanes are separated by rectangular areas defined as pipes. Each pipe has associated segment tracks. The edges are represented as orthogonal polylines consisting of line segments and routed along the shortest paths. Each segment is assigned to a pipe and to a segment track in it. The edge crossings reduction uses an iterative algorithm to resolve crossings between segments. Conflicting segments are reassigned to adjacent segment tracks, either by swapping with adjacent segments, or by inserting new tracks and calculating the shortest paths of edges. The algorithm proved to be efficient and was implemented in an interactive graph design tool.

KEYWORDS

Directed Graphs, Orthogonal Edge Routing, Crossings Reduction Algorithms.

1. INTRODUCTION

Graph drawings are commonly used to model or document data flows, complex data structures or processes, system behaviour and management, knowledge representation, and many others. Most often the orthogonal edge drawing is being used as the common graph presentation mode. The main goal of the tools for automated generation of graph drawings is to produce diagrams with uncluttered edge paths around the nodes and with minimal number of crossings between the edges. There are already plenty of techniques for orthogonal drawing of edges. Some known solutions use sophisticated path-finding algorithms [1][3]. Other solutions, based on the layered node layout, introduce abstract virtual nodes on each intermediate layer in search of optimal visibility routes [6][7]. Some other solutions reduce the task to the consideration of acyclic graphs only [2][4]. Most existing solutions are narrowly oriented toward specific classes of diagrams [8]. Those solutions do not cover all cases, especially for large and complex graphs or when it comes to implementations in interactive diagramming tools.

As we know, the reduction of edge crossings is a complex mathematical problem in the computational complexity theory, considered to be an NP-complete decision problem [5]. Most often the NP-complete problems are handled by using heuristic methods. As the graph size grows, the reduction of edge crossings becomes increasingly difficult. When the graph nodes do not have strict positioning constraints, rearranging some nodes may eliminate part of the crossings between edges connected to these nodes. However, in most cases the clarity of the

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 249-259, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111821

drawing has the highest priority, which restricts the positions of most nodes, often within certain areas of the diagram. Currently existing solutions use heuristic techniques that provide partial crossings minimization. With more complex graphs these proposed techniques usually do not achieve satisfactory results, introducing unpredictable clutter or aesthetically unappealing and difficult to read diagrams.

We begin with some definitions and terms as well as with a brief description of the general approach to reduce edge crossings. We borrow some concepts and terminology from [5] related to the orthogonal drawing conventions for directed graphs, as well as some considerations for edge crossings reduction. Additionally, we present more definitions reflecting the practical aspects of the design.

2. DESCRIPTION

The common definition of a directed graph (digraph) is $G = (V, E)$, consisting of a set V of vertices (nodes) and a set E of edges (connections), where each connection is represented (in our case) as ordered pair (u, v) of vertices (nodes). We will use the term *node,* rather than *vertex*, since it is more intuitive to assign to it certain properties that are necessary at some of the steps described below. The nodes can have different flowchart shapes and have assigned connection points $(p_1, \ldots, p_n)$ where connections are attached. We will use the term *connection* for the pair (u, v) and the term *edge* for the polyline in the final drawing.

2.1. Definitions

In the presented method the flow direction F of the graph G can be either vertical (top to bottom), or horizontal (left to right), creating congruent drawings. Switching between flow directions is supported by a rotational transformation of the coordinate system. All graph drawing elements have coordinates in the coordinate system corresponding to the flow direction. When the flow direction is switched, all coordinates are recalculated using the transformation matrix. The positions of the nodes are mapped to a grid consisting of two kinds of rectangular areas: layers $I_L = (L_0, L_1, \ldots, L_p)$ across the flow direction, and lanes $I_M = (M_0, M_1, \ldots, M_q)$ along the flow direction. The nodes are positioned in the cells $C = (c_0, c_1 \ldots, c_{p+q})$ at the intersections of the layers and the lanes. Each cell position is defined by a pair (l, m) where l € L_i and m € M_j. This means that each node can be identified by the pair (l, m) of the cell in which it resides. This pair defines the *position* of the node, while its *XY* coordinates define its *location*.

The layers are separated by rectangular areas defined as layer pipes $P_L = (l_0, l_1, \ldots, l_{p+1})$, and the lanes are separated by rectangular areas defined as lane pipes $P_M = (m_0, m_1, \ldots, m_{q+1})$. Each pipe contains zero or more segment tracks $T = (t_0, t_1, \ldots, t_k)$ running lengthwise within the pipe and presumably positioned at a predetermined distance from each other. The edges are represented as polylines of segments. Each edge corresponds to a connection (u, v) between two nodes (source and target) and consists of a polygonal chain of line segments $(s_0, \ldots, s_r)$.

In the first phase, the shortest paths for routing the edges are determined by applying Dijkstra's algorithm following orthogonal graph visibility as explained later. For each edge, the segments created using the set of weighted graph vertices are assigned to layer pipes or to lane pipes correspondingly. The segments within each pipe are positioned on different segment tracks that define the segment locations within the pipe. The segment tracks are created as needed whenever a segment is assigned to a pipe. Each segment track can contain one or more non-overlapping segments. In certain cases the edges shortest paths are allowed to cross through empty cells. This approach excludes the possibility for the edges to cross graph nodes and guarantees the spacing between adjacent segments and between segments and nodes, and is the precondition for an efficient reduction of edge crossings.

The subsequent phase implements steps based on the presented heuristic algorithm to eliminate or to reduce to a reasonable minimum the number of crossings between edges. These steps include iterations for: (a) resolving the crossings between segments around the source ends of the edges, (b) resolving the crossings between segments around the target ends of the edges, and (c) resolving the crossings between middle segments in pipes and segments of adjacent edges. To eliminate these crossings, the conflicting segments are moved to adjacent segment tracks or to new parallel segment tracks created in their pipes. Since the segments are chained, moving the segments does not break the chains of segments; rather, the segments are effectively 'rubber banded' while still calculating the shortest paths of the edges. Each connection connects two nodes. Each resulting edge consists of a polygonal chain of line segments. The segments are chained in such a way that consecutive segments share their start and end points. The node at the originating end of an edge is a *source* node, and the node at the terminating end is a *target* node. Each segment chain of a routed edge starts at a connection point on the source node, referred to as an *output port*, and ends at a connection point on the target node, referred to as an *input port*. Multiple connections do not share ports on any node.

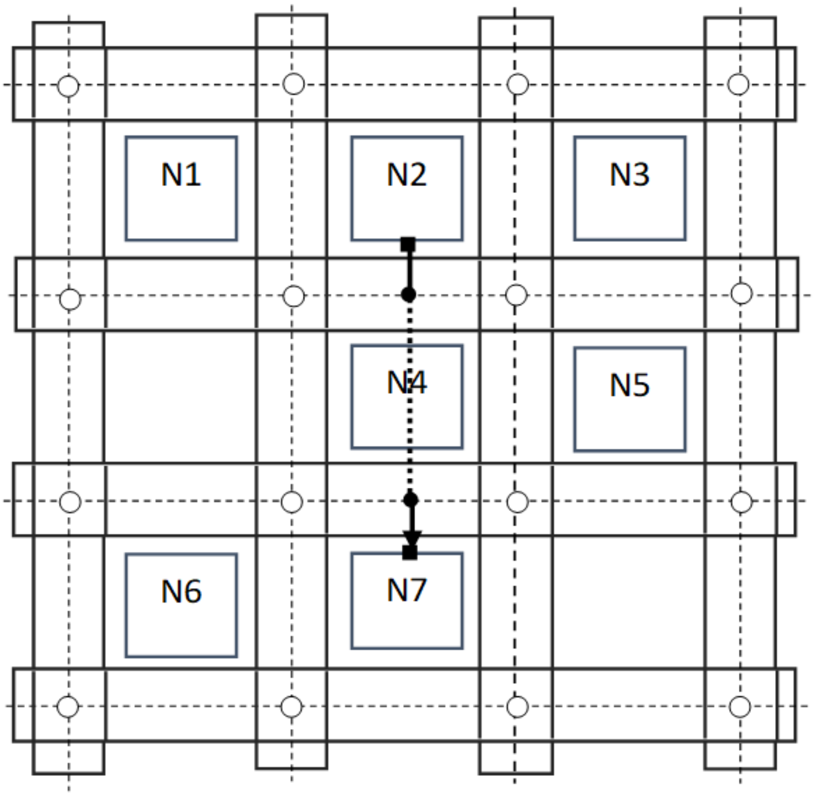


Figure 1. Rectangular grid

As illustrated in Figure 1, the first segment of an edge is referred to as a *source segment handle*, and the last segment of an edge is referred to as a *target segment handle*. The source segment handle starts at an output port on a source node and extends orthogonally to a point referred to as a *source anchor point*. The target segment handle starts at a point referred to as a *target anchor point* and extends orthogonally to an input port of a target node. The two anchor points delineate the remaining part of the segment chain of the edge. The rectangular areas around the layers are the layer pipes, and the rectangular areas around the lanes are the lane pipes. Most of the segments of the edges are laid out along the layer pipes and the lane pipes. The median line of each pipe that is running lengthwise the pipe is referred to as a *main median line*. The crossings between the main median lines of the layer pipes and the main median lines of the lane pipes are *dummy vertices* $D = (d_{00}, d_{0p}, \ldots, d_{i0}, \ldots, d_{ij}, \ldots, d_{p+1,q+1})$, marked with circles. They are used to calculate the end points of the line segments.

The dimensions of the nodes may depend on the number of the corresponding ports, on the shape of the nodes, and possibly on some additional graphical content. The dimensions of the layers and the lanes are calculated based on the contained nodes. The dimensions of the layer pipes and the lane pipes are determined by the number of tracks. The overall size of the drawing grid is determined eventually by the dimensions of the contained drawing elements.

2.2. Routing the edges

During the edge routing stage, the set of connections E is traversed in sequence. The routing of each edge connection follows a sequence of procedural steps. The first step is to determine a new output port at the source node and a new input port at the target node. The next step is to determine a source anchor point and a target anchor point and to determine a source segment handle as the first segment of the edge and a target segment handle as its last segment. For each of the two nodes the initial locations of the anchor points are determined in two steps identical for both the source anchor point and the target anchor point: (a) finding the orthogonal projection of the respective port onto the main median line of the adjacent pipe (Figure 1), and (b) inspecting the neighbouring cells in the layer and the lane where the node is positioned. In most cases the anchor point is located on the main median line of the corresponding adjacent pipe. However, if some empty neighbouring cells are detected, the anchor point is pushed further away.

The locations of the anchor points are used to calculate the shortest path for each edge. After the edge is routed, the locations of the anchor points are adjusted and moved to the respective segment tracks where the segments are assigned. After the source segment handle and the target segment handle are determined by the locations of the anchor points, the next step is to determine the shortest orthogonal path between the two anchor points. This is achieved by applying Dijkstra's algorithm. The set of weighted graph vertices, supplied to the algorithm, includes the source anchor point as a source vertex, the target anchor point as a destination vertex, and all the dummy vertices across the grid. Running Dijkstra's algorithm produces a resulting set of vertices $W = (w_0, w_1, \ldots, w_n)$, defining the shortest orthogonal path between the source vertex w_0 and the destination vertex w_n. Here w_0 is the source anchor point, and w_n is the target anchor point, and $Y = (w_1, \ldots, w_{n-1}) \subset D$. The path trajectory delineates the vertices where bends are located. Each collinear group of vertices from the resulting set Y defines a new segment (w_i, w_j) along the main median line of the pipe. Each new segment is assigned to that pipe and positioned respectively on a segment track (Figure 2).

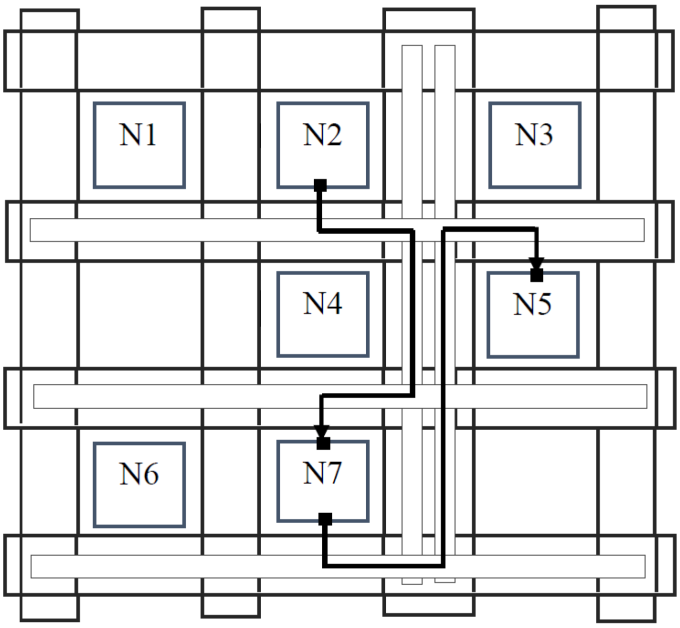


Figure 2. Routing edge connections

There are cases when there is a direct visibility between the two anchor points. In such simple cases Dijkstra's algorithm is applied to a reduced set of vertices, determined by the rectangle enclosing the source and target node.

The final step of the procedure for routing an edge is to position the created segments on the segment tracks in the pipes they are assigned to, along the calculated shortest path. At this point the locations of some anchor points may change. Initially the anchor points are positioned on the main median lines of the corresponding pipes. After the shortest path has been calculated, the locations of the anchor points are adjusted with respect to the segment tracks on which the segments are positioned. In case there is a single segment track in a pipe, the segment track coincides with the main median line of the pipe and the location of the anchor point does not change. Otherwise, the anchor point is reallocated on the respective segment track. The flowchart in Figure 3 illustrates the steps to route an edge.

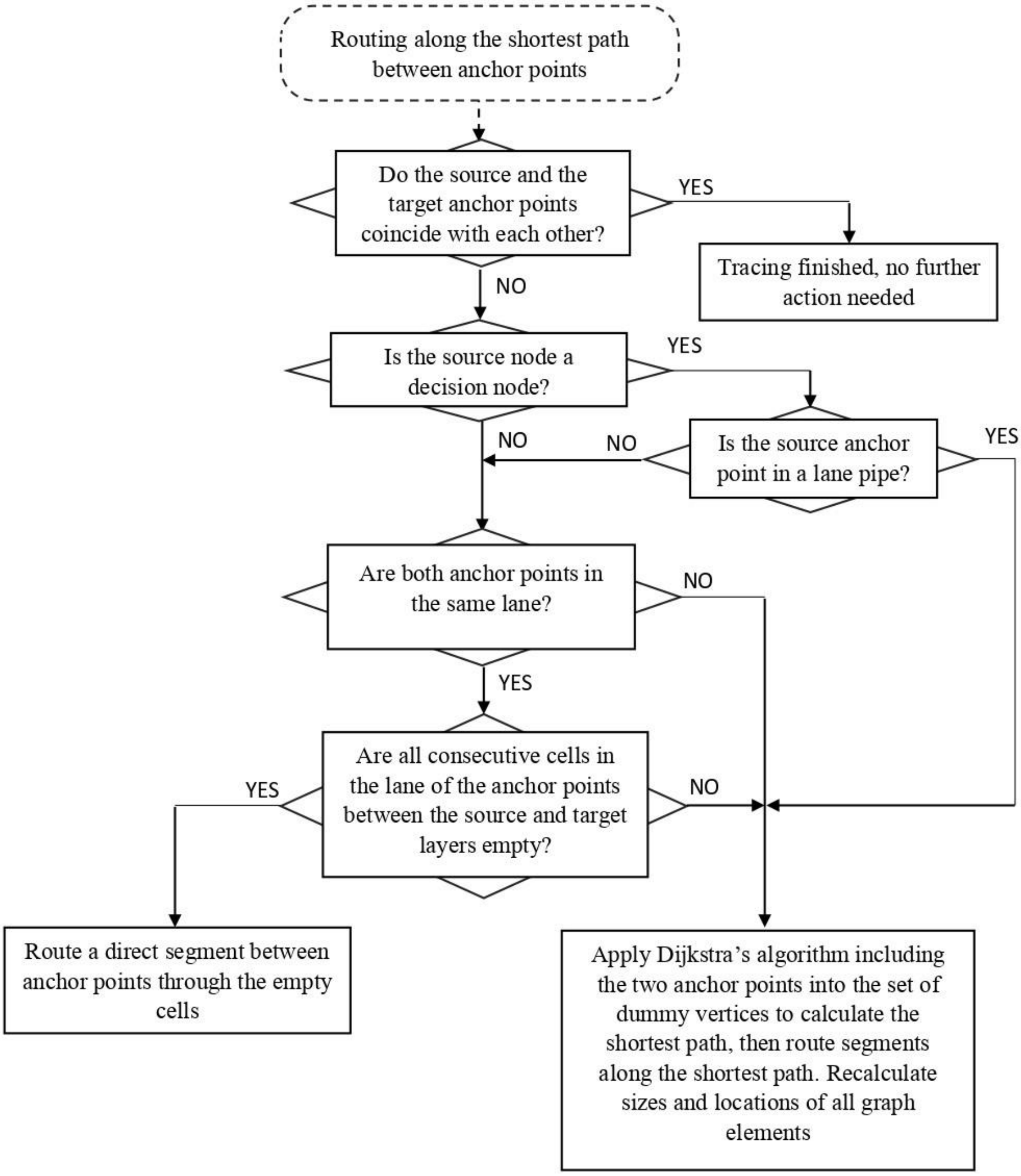

Routing along the shortest path between anchor points
Do the source and the target anchor points coincide with each other?
YES
Tracing finished, no further action needed
NO
Is the source node a decision node?
YES
Is the source anchor point in a lane pipe?
NO
NO
YES
Are both anchor points in the same lane?
NO
YES
Are all consecutive cells in the lane of the anchor points between the source and target layers empty?
YES
NO
Route a direct segment between anchor points through the empty cells
Apply Dijkstra's algorithm including the two anchor points into the set of dummy vertices to calculate the shortest path, then route segments along the shortest path. Recalculate sizes and locations of all graph elements

Figure 3. Routing edge steps

2.3. Crossing reduction

The crossings reduction starts with examination of the created set of edges *E* for crossings. The analysis is performed iteratively by steps where each step includes immediate actions to fix the detected crossings. Most of the intersections are eliminated either by swapping segment handles, or by moving some of the segments to different parallel segment tracks. The crossings reduction is performed by the following steps in this order: (a) iterative detection and fixing of the crossings between source segment handles and segments adjacent to their sibling source segment handles, (b) iterative detection and fixing of the crossings between target segment handles and segments adjacent to their sibling target segment handles, and (c) iterative detection and fixing of the crossings between some segments in pipes and segments adjacent to other segments in the same pipes. These procedures are repeated in this order until a satisfactory condition is reached.

The flowchart in Figure 4 illustrates an example of the procedure (c) for the detection and fixing of the crossings between segments in some pipes and segments adjacent to other segments in the same pipes. The procedures (a) and (b) have similar steps. The first step here is to initialize a pipe crossings counter to calculate the number of crossings. The next step is the iteration traversal of the set of edges *E* to identify a group consisting of zero or more ordered pairs of segments $S_X=((s_i, s_j), \ldots,(s_m, s_n))$ so that the segments of a pair selected from this pipe crossings group are assigned to one and the same pipe but are positioned on different segment tracks from this pipe, and at least one segment adjacent to a segment selected from the pair intersects with another segment selected from the same pair. Any adjacent segment in this case is expected to be located on a perpendicular pipe. The next step is to check the number of pairs in the pipe crossings group. If the number of pairs is equal to zero, the procedure ends by reporting zero value. If the number of pairs is greater than zero, the next step is to select the first pair from the pipe crossings group, to remove the first segment from the first segment track, to remove the second segment from the second segment track; and if the first segment does not have an overlapping conflict with any of the remaining segments positioned on the second segment track and if the second segment does not have an overlapping conflict with any of the remaining segments positioned on the first segment track, to position the first segment on the second segment track and to position the second segment on the first segment track, otherwise to add an additional segment track to the pipe that the segments from the first pair are assigned to, to position the first segment on the additional segment track, to reposition the second segment on the second segment track, to remove the first pair from the pipe crossings group and to recalculate the locations of all affected graphical elements. The next step is to check the current number of pairs in the pipe crossings group. If the current number of pairs is equal to zero, the procedure ends by reporting zero value. If the current number of pairs is greater than zero, the next step is to check if this is the first iteration of this procedure. If the result is negative (not the first iteration), the next step is to check if the number of pairs in the pipe crossings group is less than the value in the pipe crossings counter. If the result from this step is negative, the next step is to check if the number of pairs in the pipe crossings group is equal to the value in the pipe crossings counter. If the result from this step is positive, it means that the minimal number of crossings between segments in pipes and segments adjacent to other segments in the same pipes is reached, and the iterations end by reporting the number of pairs in the pipe crossings counter. This number is the actual number of unresolved crossings. If the result from the step is negative, it means that the number of pairs in the pipe crossings group is greater than the value in the pipe crossings counter and a predefined minimal number of crossings was detected in the previous iteration. The next step is to roll back the current iteration to the previous one, and to end the iterations by reporting the number of pairs in the pipe crossings counter. If the result from this step is positive (in the case this is the first iteration of the procedure), or if the result from this step is positive (meaning that the number of pairs in the pipe crossings group is less than the value in the pipe crossings counter), the next step is to set the value of the pipe crossings counter to be equal to the number of pairs in the pipe crossings group, and to return to the previous step for the next iteration.

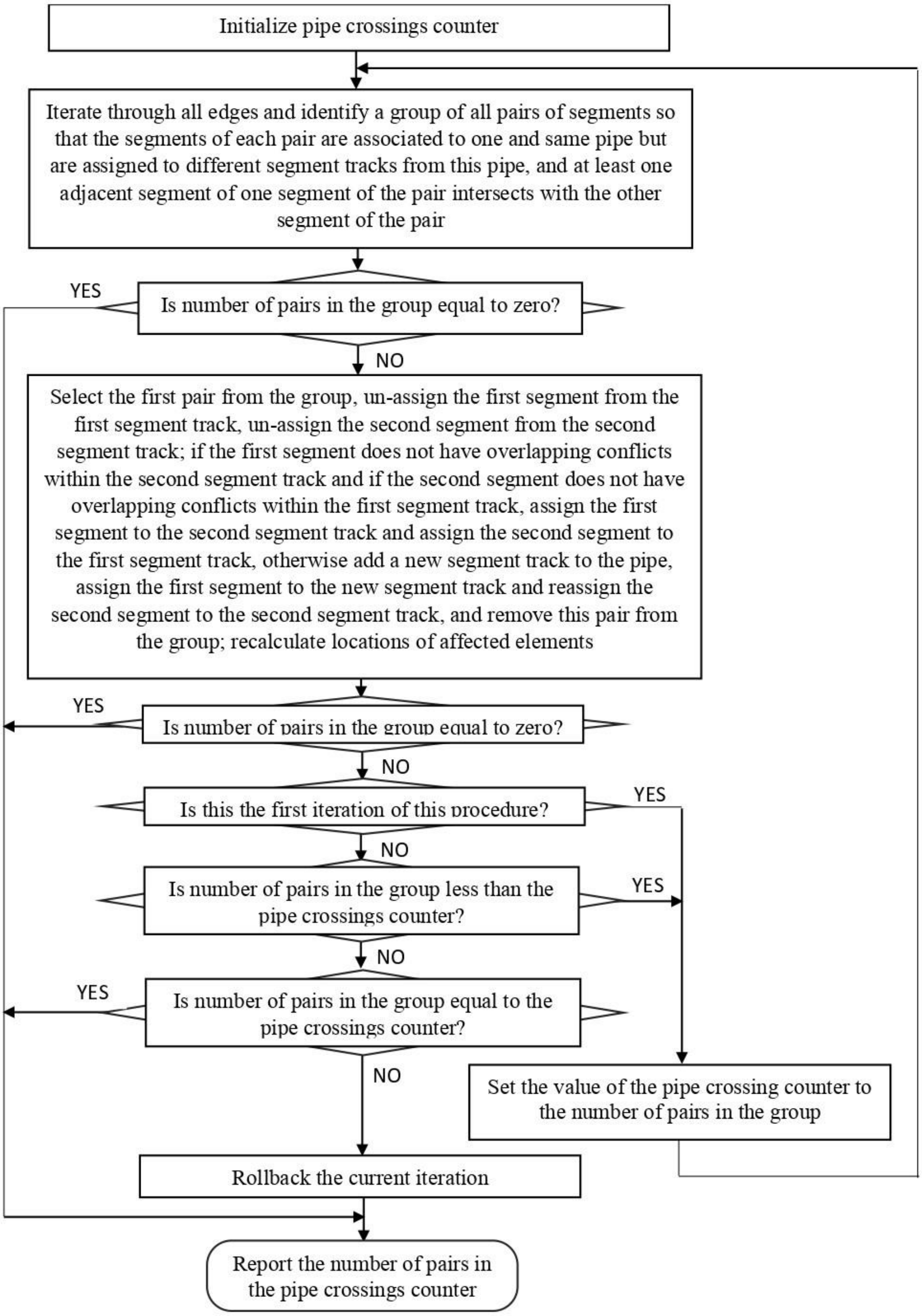

Initialize pipe crossings counter
Iterate through all edges and identify a group of all pairs of segments so that the segments of each pair are associated to one and same pipe but are assigned to different segment tracks from this pipe, and at least one adjacent segment of one segment of the pair intersects with the other segment of the pair
YES
Is number of pairs in the group equal to zero?
NO
Select the first pair from the group, un-assign the first segment from the first segment track, un-assign the second segment from the second segment track; if the first segment does not have overlapping conflicts within the second segment track and if the second segment does not have overlapping conflicts within the first segment track, assign the first segment to the second segment track and assign the second segment to the first segment track, otherwise add a new segment track to the pipe, assign the first segment to the new segment track and reassign the second segment to the second segment track, and remove this pair from the group; recalculate locations of affected elements
YES
Is number of pairs in the group equal to zero?
NO
Is this the first iteration of this procedure?
YES
NO
Is number of pairs in the group less than the pipe crossings counter?
YES
NO
YES
Is number of pairs in the group equal to the pipe crossings counter?
NO
Set the value of the pipe crossing counter to the number of pairs in the group
Rollback the current iteration
Report the number of pairs in the pipe crossings counter

Figure 4. Fixing crossings in adjacent pipes

The flowchart in Figure 5 illustrates an implementation of crossings reduction algorithm at global level with an example of an iterative sequence of all steps. The first step initializes a total crossings counter to monitor the number of unresolved edge crossings after each iteration, as well as a current crossings counter to accumulate the number of unresolved edge crossings at each step. The next step is to activate a procedure to fix crossings between source segment handles and adjacent segments. This procedure iterates through all edges to detect and fix crossings between source segment handles and segments adjacent to their sibling source segment handles. At its completion, this procedure reports the number of unresolved crossings, and this number is used to increment the value of the current crossings counter. The next step is to activate a procedure to fix crossings between target segment handles and adjacent segments. This procedure iterates through all edges in order to detect and fix crossings between target segment handles and segments adjacent to their sibling target segment handles. At its completion, this procedure reports the number of unresolved crossings, and this number is used to increment the value of the current crossings counter. The next step is to activate a procedure to fix crossings in adjacent pipes. This procedure iterates through all pipes in order to detect and fix crossings between segments in pipes and segments adjacent to other segments in the same pipes. At its completion, this procedure reports the number of unresolved crossings, and this number is used to increment the value of the current crossings counter. The next step is to check the value of the current crossings counter. If this value is equal to zero, the sequence ends by reporting the zero value. If this value is not equal to zero, the next step is to check if this is the first iteration of this sequence. If the result is negative (not the first iteration), the next step is to check if the value of the total crossings counter is less than the value of the current crossings counter. If this value is negative, meaning that the minimal total number of unresolved crossings has been reached, the iterations of the sequence end with the reporting the number of unresolved crossings. If the result from this step is positive (this is the first iteration of the sequence), or if the result from the previous step is positive (the value of the total crossings counter is less than the value of the current crossings counter), the next step is to set the value of the total crossings counter to the value of the current crossings counter, to reset the current crossings counter, and to return to the initial step for the next iteration of the sequence.

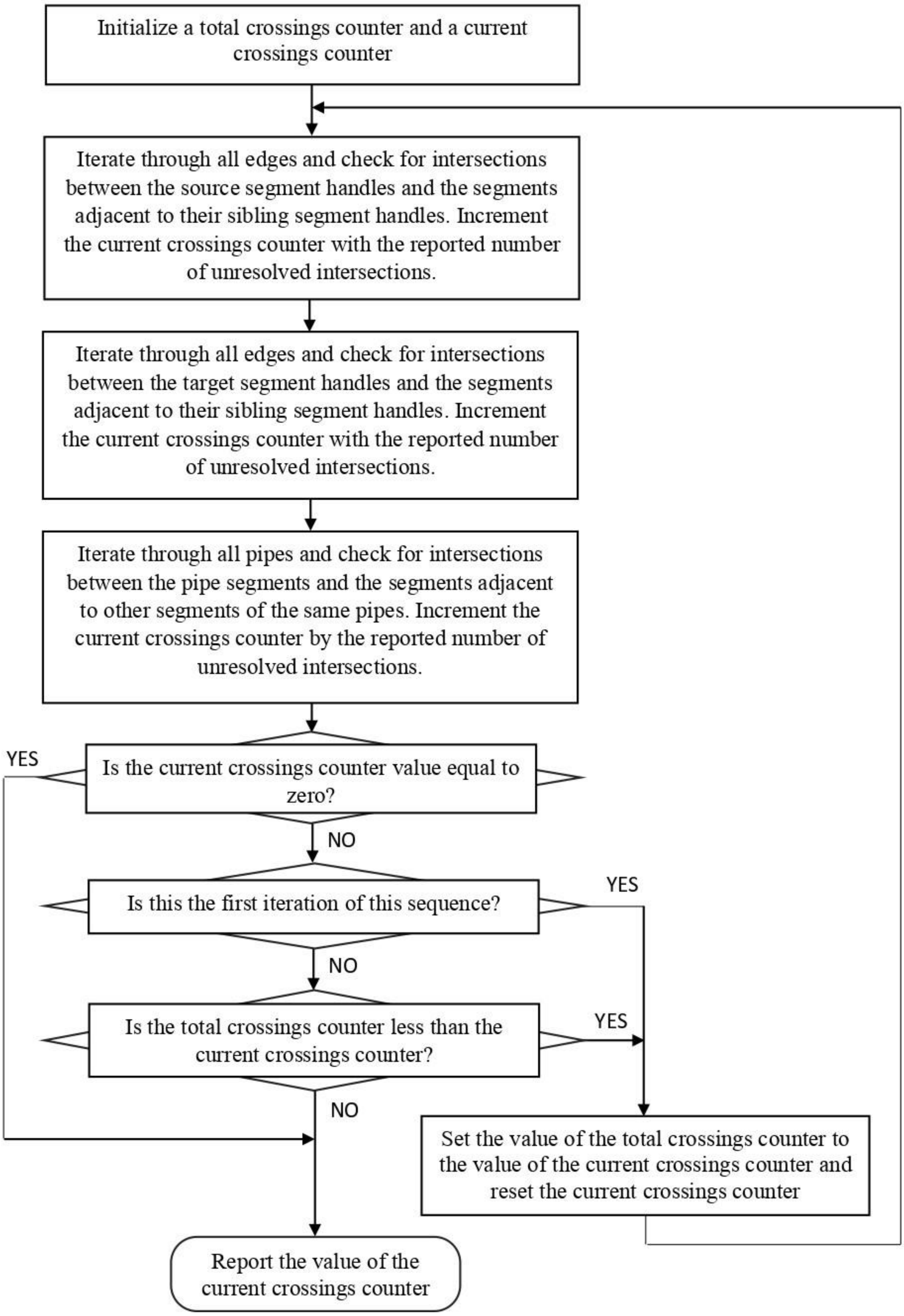

Initialize a total crossings counter and a current crossings counter
Iterate through all edges and check for intersections between the source segment handles and the segments adjacent to their sibling segment handles. Increment the current crossings counter with the reported number of unresolved intersections.
Iterate through all edges and check for intersections between the target segment handles and the segments adjacent to their sibling segment handles. Increment the current crossings counter with the reported number of unresolved intersections.
Iterate through all pipes and check for intersections between the pipe segments and the segments adjacent to other segments of the same pipes. Increment the current crossings counter by the reported number of unresolved intersections.
Is the current crossings counter value equal to zero?
YES
NO
Is this the first iteration of this sequence?
YES
NO
Is the total crossings counter less than the current crossings counter?
YES
NO
Set the value of the total crossings counter to the value of the current crossings counter and reset the current crossings counter
Report the value of the current crossings counter

Figure 5. Fixing crossings over the entire diagram

3. CONCLUSION

The described method and algorithm were implemented in JDElite Diagram Builder, a Web graph interactive flowcharting editor (www.jdelite.com). In order to build a diagram, the user only needs to drag-and-drop nodes from a palette and to enter the connections between them by mouse which is followed by automatic routing of the edges and refreshing the view. The editor is especially efficient to create medium sized complex diagrams, showing satisfactory refresh time. It shows longer refresh time for very large flowcharts.

REFERENCES

[1] C. A. Duncan and M. T. Goodrich, “Planar Orthogonal and Polyline Drawing Algorithms,” in Handbook of Graph Drawing and Visualization, R. Tamassia, Ed.; Boca Raton, FL, USA: Chapman and Hall/CRC Press, 2013, ch. 7, pp. 223-246. Available: http://cs.brown.edu/people/rtamassi/gdhandbook/chapters/orthogonal.pdf.

[2] C. D. Schulze et al., “Drawing Layered Graphs with Port Constraints,” J. Vis. Lang. Comput., vol. 25, no. 2, Apr. 2014. Available: https://www.researchgate.net/publication/258433383_Drawing_Layered_Graphs_with_Port_Constraints.

[3] D. P. Dobkin et al., “Implementing a General-Purpose Edge Router,” in Graph Drawing: 5th International Symposium, GD ‘97, Rome, Italy, September, 18-20, 1997, Proceedings, G. DiBattista, Ed.; Berlin, Heidelberg, Germany: Springer-Verlag, 1997, pp. 262-271. Available: http://dpd.cs.princeton.edu/Papers/DGKN97.pdf

[4] E. R. Gansner et al., “A Technique for Drawing Directed Graphs,” IEEE Trans. Softw. Eng., vol. 19, no. 3, pp. 214-230, Mar. 1993, doi: 10.1109/32.221135. Available: https://www.graphviz.org/Documentation/TSE93.pdf

[5] G. DiBatista et al., Graph Drawing: Algorithms for the Visualization of Graphs. Englewood Cliffs, NJ, USA: Prentice Hall, 1999. ISBN: 0-13-301615-3

[6] M. Forster, “Applying Crossing Reduction Strategies to Layered Compound Graphs,” in Graph Drawing: 10th International Symposium, GD 2002, Irvine, CA, USA, August 26-28, 2002, Revised Papers, M. T. Goodrich and S. G. Kobourov, Eds.; Berlin, Heidelberg, Germany: Springer-Verlag, 2002, pp. 276-284. Available: https://pdfs.semanticscholar.org/5d6f/de3d38417a5a92380ab09f3d84cffb8d0054.pdf

[7] M. Wybrow et al., “Orthogonal Connector Routing,” in Graph Drawing: 17th International Symposium, GD 2009, Chicago, IL, USA, September 22-25, 2009, Revised Papers, D. Eppstein and E. R. Gansner, Eds.; Berlin, Heidelberg, Germany: Springer-Verlag, 2010, pp. 219-231. Available: http://citeseerx.ist.psu.edu/viewdock/download;jsessionid=579B148C3160452F0EE0A3F0115E7A36?doi=10.1.1.159.2326&rep=rep1&type=pdf

[8] M. Cermak et al., "Edge Routing and Bundling for Graphs with Fixed Node Positions," in Proc. 15th International Conference on Information Visualisation, Jul. 2011, pp. 475-481, doi: 10.1109/IV.2011.47. Available: https://www.computer.org/csdl/proceedings/iv/2011/0868/00/06004087.pdf

AUTHOR

Jordan Raykov

Short Biography: Many years of experience in software design and engineering. Currently a software consultant at JDElite Consulting, Boulder, CO, USA

A COMPREHENSIVE SURVEY OF ENERGY-EFFICIENCY APPROACHES IN WIRED NETWORKS

Rahil Gandotra[1] and Levi Perigo[2]

[1]Interdisciplinary Telecom Program, University of Colorado Boulder, USA
[2]Department of Computer Science, University of Colorado Boulder, USA

ABSTRACT

Energy consumption by the network infrastructure is growing expeditiously with the rise of the Internet. Critical research efforts have been pursued by academia, industry and governments to make networks, such as the Internet, operate more energy efficiently and reduce their power consumption. This work presents an in-depth survey of the approaches to reduce energy consumption in wired networks by first categorizing existing research into broad categories and then presenting the specific techniques, research challenges, and important conclusions. At abroad level, we present five categories of approaches for energy efficiency in wired networks – (i) sleeping of network elements, (ii) link rate adaptation, (iii) proxying, (iv) store and forward, and (v) network traffic aggregation. Additionally, this survey reviews work in energy modeling and measurement, energy-related standards and metrics, and enumerates discussion points for future work and motivations.

KEYWORDS

Energy efficiency, energy proportionality, energy-aware protocols, wired networks.

1. INTRODUCTION

The Internet has proven to be one of the most important technological innovations. It acts as the primary catalyst in the global digital revolution, and is considered a public utility, along with running water and electricity. Fig. 1 compares access to running water to the Internet in the USA. Although this is not a like-for-like comparison, it is still important to note the pace of Internet proliferation when compared to other public utilities. Recent statistics estimate the global Internet adoption rate to be around 65.6%, or 5.1 billion users [1], and these numbers are increasing rapidly.

In addition to the escalation of Internet adoption, the digitization of services including over the top (OTT) video streaming, ecommerce, voice over IP (VoIP), and the Internet of things (IoT), has established pressure on service and network providers to expand their network hardware infrastructure. This network hardware expansion results in an upward trend of the energy consumed by the Internet globally. While Koomey's law states that post-2000 the energy efficiency of computing hardware has doubled every 2.6 years, it is still slower than the increase in data traffic, which follows Moore's law, doubling every 18 months [2]. Although the energy consumption by the wired-networking infrastructure is a small fraction of the total consumption by the information and communications technologies, the absolute numbers indicate that efforts to reduce energy consumption in computer networks are warranted [3].

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA - 2021
pp. 261-282, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111822

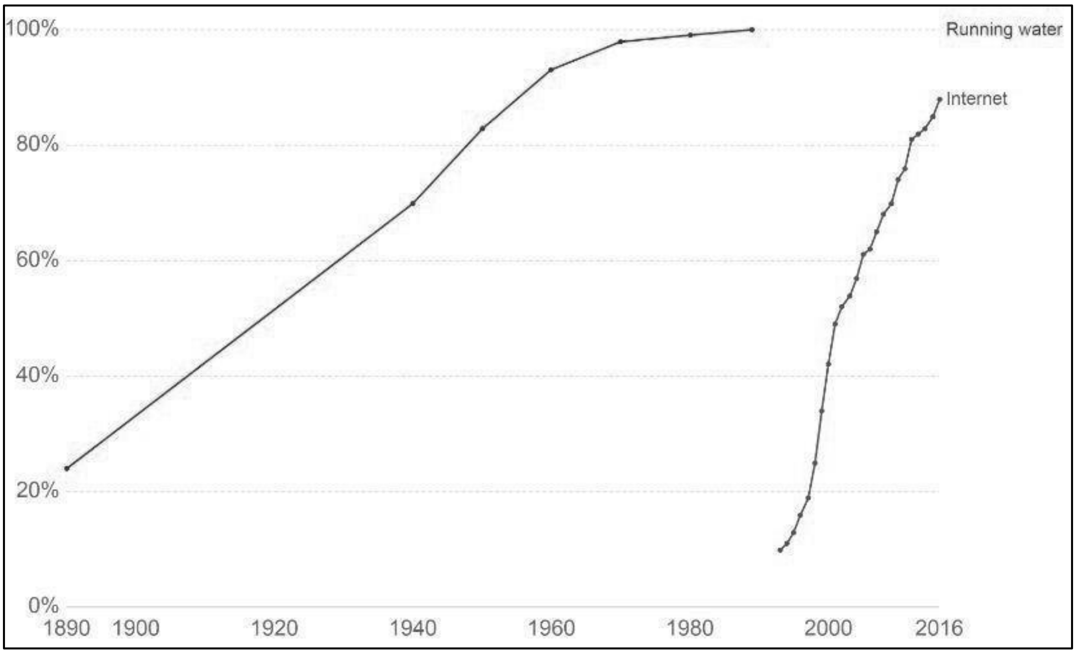


Figure 1. Access to running water vs the Internet in the USA

In the past two decades, there has been a surge of research dedicated to make the network infrastructure more energy efficient. The three major areas of focus have been, (i) system-oriented energy management, (ii) energy-aware network design, and (iii) energy-aware protocol design. System-oriented approaches strive to make the underlying hardware more energy-efficient using techniques like frequency scaling, dynamic voltage scaling (DVS), providing additional sleep states (like C-states in Intel) and performance states (like P-states in Intel). Energy-aware network design approaches seek to find the optimal network topologies that are most suited for energy savings while maintaining performance and reliability standards. Energy-aware protocol design approaches attempt to rethink and remodel existing protocols, and propose new protocols, to incorporate energy-awareness with the goal to provide opportunities to reduce network energy consumption. The system-oriented approaches are comparatively more mature than the other two approaches. This survey paper, therefore, excludes system-oriented approaches and focuses on network design and protocol-oriented approaches for wired networks. We study the work done in the last twenty years, while analyzing and discussing the research challenges, with the aim to present the state-of-the-art and realize important conclusions that can benefit future work in this field.

The remainder of the paper is organized as follows: Section II discusses the broad categories of the existing work, Section III presents a detailed survey of the specific approaches in each category, Section IV presents work that is closely related, but are not included in Section III, and Section V presents conclusions and discussion points for future research.

2. CATEGORIZATION OF EXISTING WORK

Computer networks are a critical part of the Internet infrastructure acting as highways connecting users to online services. They are conventionally designed to handle peak-traffic loads with sufficient redundancy and Quality-of-Service (QoS). Owing to the best-effort nature of the Internet Protocol (IP), network architectures and applications are developed to provide reliable transmission in case of failures. While such techniques have promoted the growth of the Internet, they are not typically energy-proportional, i.e. their energy consumption is not proportional to the traffic load. The practice of over-provisioning and observations of under-utilization and non-energy-proportional behavior during the majority of times prompted researchers to find methods to reduce the energy-consumption of wired networks. Since individual network hardware devices and their components have become increasingly reliable, and network protocols have become more robust, new techniques can be leveraged to determine better energy savings.

Different parts of the wired-network infrastructure – network topology, hardware devices, electrical/optical links, protocols, and network management techniques – exhibit different runtime behaviors and offer different opportunities to lower the energy consumption. We categorize the existing work into five distinct categories: **sleeping of network elements**, **link rate adaptation**, **proxying**, **store and forward**, and **network traffic aggregation**, and provide an overview of them in this section. Each category targets a specific area of wired-networking with its own research challenges, implementation-specific techniques, and scope for future work. Fig. 2 shows the taxonomy of existing research in different categories.The detailed survey of specific approaches in each category is presented in section III.

The concept of **sleeping of network elements** is based on the intuition that because computer networks are provisioned for peak load and are generally under-utilized, parts of the network or individual devices can be put to sleep during off-peak hours. This category of work was one of the earliest approaches (2003) to save network energy consumption and has seen a multitude of proposals. The work involved in this approach is two-fold: identifying/predicting low-utilization periods suitable for energy savings, and reconfiguring the network to put network elements to sleep. The target network element to be put to sleep could either be the physical device(s), or individual components of one device. Although networking devices traditionally provide limited support for sleeping, approaches in this category propose techniques which either make assumptions about the underlying hardware support, or find ways to work around this limitation. Waking up a sleeping element and understanding the impact of sleeping on network protocols were some of the research challenges in this category.

Link rate adaption gained popularity as a potential approach after research proved that high link rates (1/10 Gbps) consume more energy as compared to low link rates (10/100 Mbps), and that a desirable level of performance could be achieved by running links at lower speeds. The work involved in this approach is two-fold: identifying/predicting low-utilization periods wherein the links could be operated at slower rates, and reconfiguring the network to configure specific links at lower speeds. Both reactive approaches, based on matching the current utilization with link speeds, and proactive approaches, predicting current/future traffic patterns from historical utilization data, were proposed. Research challenges in this category include understanding the tradeoffs between frequent link-rate change and performance, and the impact of these changes on protocols, especially cost-based routing and switching protocols.

Proxying is based on the hypothesis that the periodic network protocol traffic does not allow end-hosts to sleep effectively. Although this approach does not directly target energy savings from the network, it involves the network elements to act as proxies to enable infrastructure-level savings. The techniques proposed in this category attempted to counter the requirements necessitating networked end-hosts to be connected and responsive even while sleeping, i.e. always-on should not mean always powered on. A survey of office and commercial equipment indicated that most of the computers were not put into sleep modes due to the above requirement of always maintaining network presence [27], and this prompted researchers to consider network proxying and find ways to offload some of the end-host responsibilities to a proxy. The placement of the proxy in the network and the tradeoff between energy savings by end-hosts sleeping and the additional energy consumption by the proxy were some of the research challenges in this area of work.

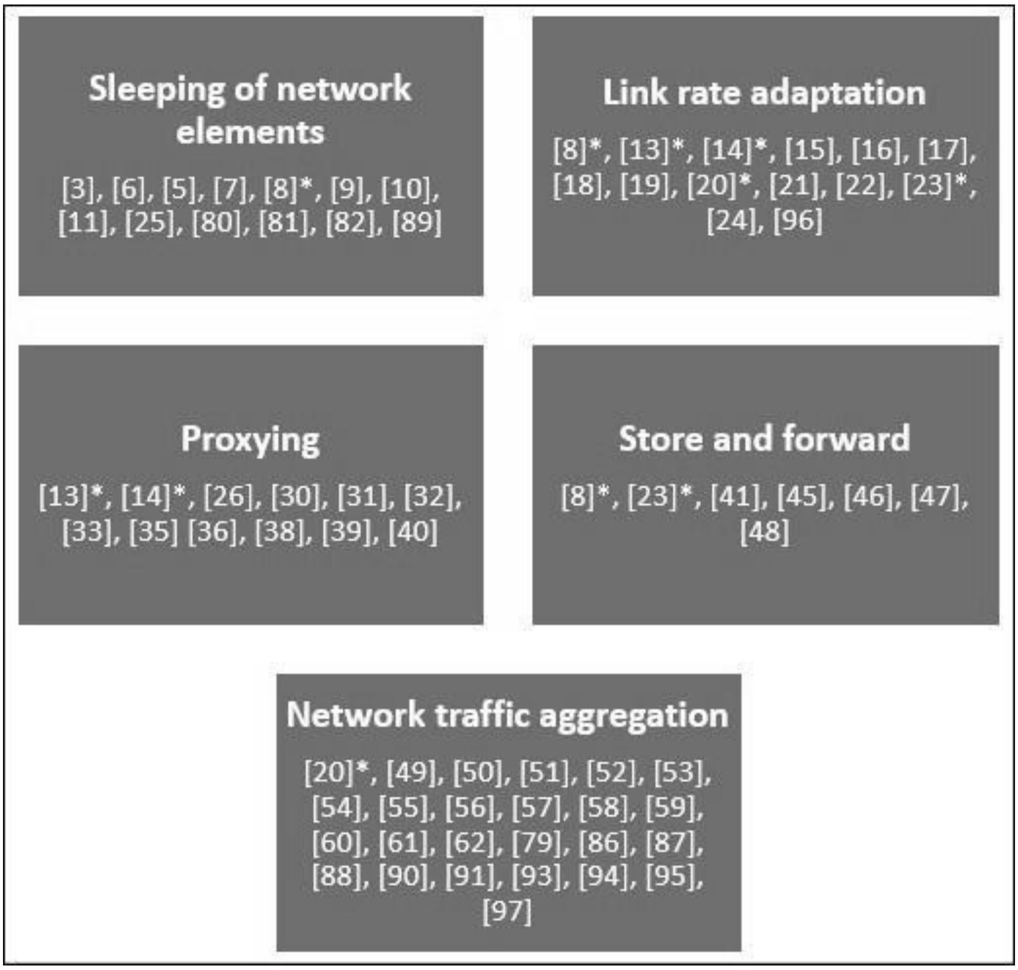


Figure 2. Categorization of existing work (* indicates work appears in multiple categories)

The observation that many web applications exhibit bursty traffic with high inter-arrival traffic times led to the development of **store and forward** techniques. Although this is not a mature area and does not have many research proposals, it is still an important area of work to study with insights for future works. While the traditional practice in network engineering is to avoid bursts using congestion avoidance and QoS queues, the insight behind the techniques in this area is to shape traffic in bursts in order to allow network elements to be put in low power/sleep mode in between these bursts. The difference between sleeping of network elements and store and forward is that techniques in this area use explicit proactive mechanisms to shape and engineer traffic with the intention to create sufficient bunching that would allow efficient sleeping, while techniques in the former aim to reactively put network elements to sleep based on utilizations. Store and forward techniques could be considered as a subset of the sleeping of network elements category, but we believe there are sufficient design differences in the two areas to warrant separate categories. Developing practical ways to implement this approach, synchronizing the bursts across the network to enable efficient sleeping, and understanding the tradeoffs between performance and energy savings were some of the research challenges in this area.

Network traffic aggregation techniques are based on the assumption that networks are over-provisioned and under-utilized, and that a subset of devices and links would suffice during off-peak hours. While we do not include server load aggregation techniques – clustering compute resources onto minimal physical servers to save energy – in this area, specific techniques that involve manipulating network traffic to enable server-load aggregation are included. The difference between network aggregation techniques and sleeping of network elements is that approaches in this area attempt to proactively find minimum-power subsets of the infrastructure that can provide the desired level of performance while enabling energy savings, whereas, techniques on sleeping of network elements attempt to put network elements to sleep reactively while considering the entire topology. These techniques could be included in the first category, but we believe that the insight and the way forward is different enough to justify separate categories. Research challenges in this area focus on traffic engineering problems such as NP-complete [37] and finding methods to reduce the computing time that allows for on-the-fly solutions to efficiently save energy.

3. DETAILED SURVEY

In this section we present a chronological survey of specific approaches in each category.

3.1. Sleeping of network elements

Gupta and Singh [3] were among the earliest researchers to examine the energy consumption of networking devices. A previous study which analyzed the energy consumption of office and telecommunications equipment in commercial buildings revealed that Internet devices consumed 6.05 TW-h of energy in the USA [4]. This realization prompted the researchers to find ways and suggest directions to save energy going forward. They proposed sleeping of devices/interfaces as a solution and explored coordinated and uncoordinated sleeping. In coordinated sleeping, the network devices collectively made decisions on which elements to put to sleep, while in uncoordinated sleeping, devices made such decisions in isolation, independent of others. Analyzing sample traces from their autonomous system (AS), they showed that sleeping was a reasonable solution, however, modifications to existing protocols and the Internet architecture would be needed to maximize the amount of energy conserved.

Soteriou and Peh [6] earmarked links interconnecting routers as a major source of energy consumption and proposed a dynamic power management policy employing on/off links. The work involved deriving a power-performance connectivity graph to identify candidate on/off links, developing a deadlock-free routing algorithm based on the information from the graph, and implementing an on/off decision mechanism. Input buffer utilization of a router was used to determine if a link could be turned off. Their results indicated a 37.5% reduction in energy consumption for an 8-ary 2-mesh topology with a moderate latency increase.

Gupta et al. extended their prior work in [5] to investigate the feasibility of sleeping in LAN devices. LAN switches are targeted as they comprise the bulk of LAN devices and therefore consume the largest amount of energy. Three different sleep models were proposed – (i) Simple sleep – wherein timers are employed to wake up a sleeping device/interface and all incoming packets during sleep time are lost, (ii) Hardware assisted sleep (HAS) – wherein the incoming packet is lost but it wakes up a sleeping device/interface, and (iii) Hardware assisted buffered sleep (HABS) – wherein the incoming packet is buffered and it wakes up a sleeping device/interface. Both HAS and HABS assumed hardware support which was not available at the time. A need to modify existing protocols (like STP and TCP) to allow for efficient sleeping was reiterated.

While the work in [5] was a preliminary study to evaluate the viability of sleeping as a solution to save energy, in [7] Gupta et al. presented practical approaches to achieve it, using technology available in the hardware of the time. The target area was Ethernet interfaces and their proposal leveraged smart timer-based Ethernet transceivers that automatically turned off when no power was detected on the other end. The algorithms to decide when to turn off a link included – (i) On/Off-1 – wherein the upstream interface of a link determines if the link can be put to sleep by estimating the number of packet arrivals in a time period t, and finding the maximum value of t for which the probability of packet arrivals being less than a buffer threshold is less than 10%, and (ii) On/Off-2 – which is a modification of the previous algorithm, wherein instead of both the upstream and downstream interfaces waking up after time period t and reevaluating if they can go back to sleep, only the upstream interface wakes up and runs the algorithm, to allow for higher energy savings. The simulation results on real-world traces indicated that 37% of the interfaces could be put to sleep anywhere from 40-98% of the time. Another conclusion from their work was that sleeping interfaces allowed portions of the internal switching fabric to be put to sleep as well.

Nedevschi et al. [8] studied opportunistic sleeping in which link interfaces sleep when no packet arrivals are observed for a period of time. This approach assumed hardware support of wake-on-arrival – the circuitry that sensed traffic on the interface and has the capability to wake up the device. The simulation results on Abilene showed that opportunistic sleeping was suitable for LANs with high idle times, however, it was not suitable for high-speed links with low idle times. The percentages of time links could sleep with either constant bit rate traffic or bursty traffic were minimal, proving that this approach was suitable only in the specific LAN scenarios.

Following similar intuitions as before, Ananthanarayanan and Katz [9] proposed two energy-saving schemes for switches and studied the tradeoffs between performance and energy consumption. Their schemes fall under the umbrella of uncoordinated sleeping in the way that they are run on each switch independently. The concept of shadow port is introduced wherein a shadow port is associated with a cluster of normal ports allowing packets to be buffered at the shadow port in case the associated normal ports are sleeping. The Time Window Prediction (TWP) scheme is similar to the On/Off-1 algorithm in [7], wherein predictions are made about the number of packets traversing a port in a time window, and if this number is below a set threshold then the port is powered off. However, the difference in this scheme was that the sleep window is made adaptive by setting a bound on the increase in latency – if the latency of the buffered packets increases above a set bound, the sleep window is reactively reduced. The Power Save Mode (PSM) scheme was a modification to TWP to make it more aggressive for energy-savings by not considering traffic flows while making sleep decisions. This enabled higher energy savings at the cost of increased latency. The simulation results employing an enterprise network's traffic patterns demonstrate a 18-21% potential for energy savings using the TWP scheme. Their tradeoff study indicated that the wake-on-arrival capability and large buffer sizes could allow for a significant increase in energy-savings while ensuring minimal performance degradation.

Chiaraviglio et al. [10] proposed an algorithm that creates an ordered list of nodes by decreasing power consumption, and for each element in the list, it first powers down the node and then checks if network-wide connectivity is maintained. If the network is disconnected, the node is powered back on and the algorithm proceeds to the next list element. Similar procedure is employed to power down links ensuring connectivity and maximum link load constraints are met. The simulation results on topologies similar to those of national ISPs confirmed the possibility to save more than 23% of energy.

The IEEE 802.3az – Energy Efficient Ethernet (EEE) standard was approved in September 2010 with the aim to reduce the energy consumption of computer networks [11]. While there was more than one scheme initially proposed for this standard, the approach of low-power idle, developed at Intel, was finally adopted [12]. Ethernet interfaces traditionally transmitted an auxiliary signal called IDLE when no data packets are transmitted to keep the transmitters and receivers synchronized. In 802.3az, in the case of no traffic, the link enters a low-power state and sends a Refresh signal over short durations allowing the link to consume less energy over large durations. Estimated savings by using EEE were projected to be $410 million/year in the USA, and over $1 billion/year globally.

Herrería et al. enhanced the techniques of opportunistic sleeping in [25] by proposing four interface states – active, idle, transition, and sleep. The interface transitioned to sleep as soon as its buffer was empty, to allow for larger sleep durations, and the wake up back to active state was triggered either by a timer or the buffer length crossing a certain threshold. The simulation results showed a potential to save 75% energy without noticeable impact to performance.

Distributed algorithms to make sleeping decisions are presented in [80-82]. In [80], the approach involved selecting a router as the power saving router (PSR) by random election, and the PSR checking if network connectivity would be maintained if itself is disconnected from the network for a certain time period. If connectivity could be maintained, the PSR recomputed the routing table which would then be broadcast through the network. The numerical results from this approach indicated that up to 18% of the power could be saved for the whole network. [81] utilized the periodic link state information (LSA) to decide which links to switch off in a distributed way. The proposed algorithm did not require the knowledge of the actual and past/future traffic matrices, and the results from realistic case studies indicated energy savings of up to 50%. In [82], Patota et al. proposed DAFNES – a distributed algorithm for network energy saving based on stress-centrality. The stress centrality of a network node n refers to the number of shortest paths between any two endpoints which pass through node n. The proposed approach falls under the category of coordinated sleeping wherein switches, one at a time, made the decision to power off linecards independently, but the selection of the switch required synchronization by the exchange of control messages. The algorithm involved the computing and exchange of the stress centrality values between all switches, and the switch with the lowest stress centrality switched off its linecards in each iteration. The testing results showed the potential to save up to 50% of network energy.

[89] aimed to reduce the power consumption in data center networks by optimizing the number of SDN controllers active to serve the OpenFlow switches. The proposed heuristic first sorted all switches by decreasing order of the number of aggregated flow arrival rates, and then assigned each switch to a partially filled controller with the smallest but sufficient capacity. If no active controller was found with sufficient capacity, then a new controller node would be turned on.

3.2. Link rate adaptation

The concept of link rate adaptation was first proposed by Christensen et al. [13] and was based on the realization that higher link speeds equated to higher energy consumption. The aim was to match link data rates with the traffic levels to make networks more energy proportional. In [14], Gunaratne et al. preliminarily explored rate adaptation for links between PCs and access-layer switches. The proposed algorithm made decisions to increase or decrease link rates based on the queue lengths on personal computer (PC)NICs and switch interfaces, and the simulation experiments on a university campus network confirmed that it was feasible to operate at lower data rates with no significant increase in delay.

The work to develop a buffer threshold policy was extended in [15], [16] and [17]. In [15], a Markov model is developed for a state-dependent service rate, single-server queue while making the assumptions of standard Poisson arrival rates and exponential service rates. The proposed model is augmented by ensuring that service rate transitions occur only at service completions to replicate traditional Ethernet behavior. Single threshold policy, wherein service rates are switched to higher or lower rates if the output buffer utilization crosses a set threshold, and dual threshold policy, wherein upper and lower thresholds are defined and the service rate is switched to a higher rate when the output buffer utilization cross the upper threshold, and switched to a lower rate when the output buffer utilization falls below the lower threshold, are presented. The simulation experiments performed on traces from two university campuses indicated that links could be run at lower rates 99% of the time with a 4 ms average increase in delay. To prevent frequent link rate changes adversely impacting performance in case of smooth and bursty traffic, current link utilization information is incorporated into the buffer threshold policy in [16]. An important insight in this work was that links should be operated at lower data rates for at most 50% utilization, which equates to 5% utilization at the higher rate, and thus, if the current link utilization exceeds 5% at the higher rate, the link is not switched to the lower rate in order to balance performance with energy savings. An additional policy – time-out threshold policy –was proposed in [17] with the aim to reduce complexity of a NIC having link rata adaptation capability. The notion behind this scheme was to hard-set the time a link would run at the higher data rate and once this timer expired, and if the buffer utilization is below the set threshold, the link would switch to the lower data rate, and if not, it would continue to operate at the high rate. Transition from low to high data rates were immediate, triggered by the buffer utilization crossing the set threshold. Callegari et al. extended this work in [24] by incorporating transition times into their Markov model for a dual threshold policy, but their simulation experiments concluded that link rate adaptation was unsuitable for the NIC buffer sizes of the time.

A mechanism to implement link rate adaptation was proposed in [18]. A MAC handshake protocol was presented wherein a MAC frame containing the desired line rate was exchanged between the two nodes comprising a link. A switch to a lower data rate initiated by one node was permitted only if the other node responded with an ACK; if the other node cannot switch to the lower data rate, it responded with a NACK, and the first node continued operation at its original rate. A switch to a higher data rate initiated by one node would always be responded with an ACK by the other node to ensure no performance degradation. This mechanism was formalized as Rapid PHY Selection (RPS) in [19] and was proposed as one of the candidate approaches for IEEE 802.3az. The emulation results indicated little to no impact, i.e. minimal increase in delay and no packet loss, on TCP and UDP file transfers using this scheme.

Nedevschi et al. studied rate adaptation in [8] with the motivation being that operating links at slower speeds have twofold benefits – operating at slower frequencies consumes less energy, and operating at slower frequencies allows the use of dynamic voltage scaling (DVS) – making energy consumption scale quadratically with operating frequency. A practical algorithm is proposed which uses exponentially weighted moving average (EWMA) to make estimations about packet arrival times and uses current link utilization and operating rate information to ensure a rate change does not violate delay constraints. Two important conclusions from the simulation experiments performed on traces from Abilene were - (i) the granularity and distribution of operating rates played an important factor in the number of transitions and therefore the amount of energy saved; uniformly distributed link rates performed significantly better than exponentially distributed rates available in most networking hardware, and (ii) the time to transition between data rates also impacts the amount of energy savings, with higher transition times leading to reduced savings and higher delay.

Mahadevan et al. [20] studied link state adaptation (LSA) and performed experiments by simulating a Web 2.0 workload in a production data center topology. The proposed strawman scheme reactively adapted link rates based on current utilizations, while the service level (SL) aware scheme added constraints to guarantee a minimum level of performance was maintained – ensuring the link utilization was below 70% at any given time. The schemes assumed the existence of an Oracle that had perfect knowledge about upcoming traffic. The results showed that 16% energy savings were possible employing the LSA scheme, while SL-aware LSA consumed slightly more energy but provided significantly better performance. The deployment considerations described two approaches – reactively making changes by collecting link utilization information using protocols such as SNMP, and proactively making changes by predicting future traffic patterns using simple models such as AR(1). However, the prediction-based approach led to over and under-estimations in their simulations.

Abts et al. [21] leveraged the capability of modern plesiochronous links to operate in a dynamic range to make data centers more energy proportional. The intuition is that high-speed communication channels generally comprise of multiple links operating plesiochronously independent of each other and of the core router logic rate, and this independence is exploited to match operating rates with estimated bandwidth requirements. The proposed mechanism involved the switch tracking the amount of traffic traversing through it in a given time period, and if it exceeded a threshold defined for each link, the operating rate was doubled, and if it was less than the threshold, the operating rate was halved. The simulation experiments performed on synthetic and real-world traces (Google's data center) and assuming perfectly energy-proportional channels indicated a potential to save 15-36% of energy depending on traffic patterns.

Ginis [22] described the various link rates and power states for ADSL2 and ADSL2+ with the aim to achieve energy savings when averaged over a long period of time. Staessens et al. [23] proposed leveraging the capabilities offered by software-defined networks (SDN) to perform link rate adaptation. Since an SDN controller has a global view of the network, collecting and analyzing traffic patterns is less complicated, and this allows for easier deployment of the developed link rate policies [62].[96] combined the strategies of sleeping of network elements and link rate adaptation into one scheme for SDN-based data centers. The proposed algorithm proactively checked for inactive links and switches to power them off, and also reduced the link speeds of ports with low utilization.

3.3. Proxying

The concept of proxying in Ethernet interfaces was introduced in [13]. The authors suggested adding some intelligence to the PC NICs allowing them to respond to non-critical packets and waking up the PC from sleep in case of packets requiring a response. [26] presented the initial design and development behind ideas for a proxying Ethernet adapter. A study conducted by the Lawrence Berkeley National Laboratory in 2004 revealed that existing energy management features such as sleep states were disabled in around 95% of all PCs [27]. The major reasons behind this included certain applications not allowing the PC to sleep due to open sessions by exchanging keepalive packets, or network administrators manually disabling the features to ensure seamless firmware upgrades. Industry standards for networked energy management such as Wake-on-LAN (WOL), which used a Magic Packet (MAC address of the receiving NIC 16 times) to wake up a sleeping PC [28], and Advanced Configuration and Power Interface (ACPI), which specified standard interfaces for communication between applications and the BIOS, and extended WOL by providing direct wake-up using IP or ARP packets [29], were proposed, but rarely used for the purpose of energy savings. The authors in [26] suggested a control logic, wherein incoming packets either were ignored, handled at the proxy, or triggered the PC wake-up. The energy savings using this approach were estimated to be around 1 TW-h/year, implying an $80 million savings assuming 8 cents/kWh.

Gunaratne et al. [14] proposed two approaches in this area – protocol proxying and split TCP connections. Protocol proxying involved categorizing all traffic as – (i) no response required, such as broadcast, bridging, and routing packets, (ii) minimal response required, such as ARP and ICMP packets, and (iii) wake-up required, such as TCP SYN and SNMP Get packets. Traces from a campus PC indicated that protocol proxying could discard or minimally respond to 91% of all incoming packets. Split TCP connections involved adding a shim layer between the application and transport socket interface, which allowed the application to be presented with a persistent connection while allowing the underlying client to sleep. The initial experiments confirmed the feasibility of this approach. To avoid hardware changes to PC NICs, placing the proxy functionality on the LAN switches was discussed in [30].

Allman et al. [31] discussed the various architectural components required to support selective connectivity of end systems – (i) assistant – the mechanism which filtered traffic and took the corresponding action, (ii) exposing selective connectivity –making the protocol stack and the neighbors aware about the end system's energy state, (iii) evolving soft state – maintaining a proxyable state or a limbo state which enabled the distinction between a sleeping host and a non-existing host, (iv) host-based control – which ensured that the end system had the capability to dictate its selective connectivity policy, which allowed different policies to coexist in one network, (v) application primitives – making applications proxy-aware by using less number of general primitives instead of more specifics, and (vi) security – understanding the various vulnerabilities of offloading end system functionality to an external proxy. Purushothaman et al. [38] analyzed proxying for the specific Peer-to-Peer application Gnutella. The experiments confirmed that end hosts could spend large amounts of time sleeping both while downloading or uploading files. A prototype of such a proxy was developed in [39].

Nedevschi et al. [32] performed an in-depth evaluation of the potential of energy savings and the effectiveness of using proxying as a solution. Network data collected from 250 enterprise end systems indicated that incoming traffic comprised of significant portions of unicast, multicast, and broadcast traffic, therefore suggesting a need to tackle all three traffic classes by the proxy. The authors identified both broadcast protocols, such as ARP, DHCP and NBNS, and multicast protocols, such as HSRP, PIM and IGMP, which were suitable candidates to be either discarded by the proxy or requiring a simple response without having to wake up the end system. A prototype proxy is implemented using the Click Modular Router and the experiments performed showed that additional delays incurred due to proxying were minimal and lower than TCP SYN timeouts. Sabhanatarajan and Gordon-Ross [40] presented a partitioned TCAM-based proxying technique for Smart-NICs (SNIC). The simulation results indicated 62% lower energy-delay as compared to the existing non-partitioned router approaches.

Agarwal et al. [36] developed a proxy prototype, Somniloquy, employing USB-based network interfaces with the capability to support BitTorrent, instant messaging and web download applications. [34] and [35] studied and presented the generic architecture of a network connectivity proxy and its responsibilities. The major research challenges identified for this approach include high memory and processing requirements of the proxy, determining its optimal location in the network, awareness about end systems' energy states, application independency and support for mobile hosts or over different subnets. The concept of thin clients is presented in [33] in which low-power client machines replace end user desktops with the applications being hosted on another dedicated machine serving multiple thin clients. Experimental results indicated energy savings of upto 66% as compared to that of a traditional desktop environment.

3.4. Store and forward

The techniques in this category aspire to predict, control, and make the most use of idle times [44]. To enhance the typical sleeping of network elements techniques, Nedevschi et al. [8] proposed shaping traffic into small bursts at the edge of the network allowing the edge devices to transmit packets in bursts throughout the network in order to increase the sleep times possible for all devices. The concept of a buffer interval is introduced wherein an ingress router buffers all incoming traffic for some time and periodically transmits all buffered traffic in a burst to its neighbors, allowing for alternating periods of sleep and transmissions. A practical algorithm is presented in which routers aggregate packets targeted for one destination together and after every buffer interval, the bursts are transmitted serially. This allowed for an ingress router to receive packets from multiple upstream routers as a single burst. Comparing with the typical wake-on-arrival sleeping techniques, the uncoordinated buffer and burst technique performed significantly better for constant bit rate traffic. The simulation results indicated lower transition times and hierarchical topologies would benefit this approach.

Intuitions in this area originated from delay-tolerant networking [43] which was a store and forward architecture for the interplanetary Internet to allow data to propagate opportunistically as connectivity subsequently allowed. Baldi and Ofek [41] proposed leveraging pipeline forwarding of IP packets in order to save network energy. The concept of pipeline forwarding was introduced in [42] wherein IP switches synchronized transmissions by either using a central time authority, like GPS, or a distributed network time protocol, in order to achieve deterministic QoS. The authors in [41] described a parallel network based on WDM-based fiber infrastructure, implementing time-based scheduling, in conjunction with the Internet. The objective was to shunt a large portion of the traffic from the Internet to this ‘super-highway’ to achieve dictated performance and energy savings.

The authors in [45] analyzed the opportunistic sleeping algorithm proposed in [7] and concluded that queuing packets for some time before transmitting allowed for fewer transitions and therefore reduced the energy costs related with frequent state changes. In [46], experiments were performed to compare frame and burst transmissions for Energy Efficient Ethernet (EEE), and the results indicated energy savings ranging from 5 to 70% for end users and of about 50% for data centers. [47] presented an analytical model for energy saving using burst transmissions in EEE, adding the elements of maximum allowed queue size and the maximum added queuing time to the 802.3az model. In [48], the authors provided an analytical comparison of the frame and burst modes to quantify the efficiency of each method. The comparison results indicated significantly higher energy savings using burst mode with a bounded increase in delay.

Staessens et al. [23] proposed employing burst mode operation in OpenFlow-based networks wherein packets are buffered at a node and then transmitted at the maximum rate. The authors mentioned that the approach worked on small scales allowing the number of elements to be switched off between bursts to be limited and emphasized the need for large packet buffers for efficient operation. An additional OpenFlow message, OFPC_BURST_MODE, was proposed which allowed switches to advertise their burst mode capability to the SDN controller.

3.5. Network traffic aggregation

Chabarek et al. were among the earliest researchers to benchmark and study the energy consumption of network devices [49]. The work involved creating a generic energy consumption model, optimizing the multicommodity network-flow problem by formulating it as a mixed-integer problem and solving it to find minimum-power system configurations. The concept was to leverage the relationship between energy consumption, network configuration and provisioning, and experimental results indicated that substantial amounts of energy could be saved by incorporating energy awareness into existing routing protocols. Mahadevan et al. [20] proposed Network Traffic Consolidation (NTC) wherein traffic is engineered to flow over fewer links allowing non-utilized links and switches to be turned off. A service level-aware NTC approach is also proposed which incorporates path-availability constraints into the scheme to sacrifice some energy efficiency at the cost of additional network redundancy. A web 2.0 workload is simulated in a tiered data center topology and the results showed that using the NTC scheme allows for a 58% reduction in the energy consumption, while the service level-aware scheme allows for a 16% reduction with minimal latency increase.

Cianfrani et al. [50] leveraged the link-state exchange behavior of routing protocols, such as OSPF, to design a network-wide strategy to save energy. The proposed Energy-Aware Routing (EAR) algorithm is three-phased – (i) electing some routers as exporters, typically the ones having the maximum number of neighbors, which are used to calculate minimum-power trees (MPT), (ii) the remaining routers, called importers, use the exporters as reference to run a modified version of Dijkstra's algorithm to detect powered down links, and (iii) computing the shortest-path trees (SPT) based on the modified topology. The objective of the algorithm was to consider a subset of routers' SPTs to select routing paths allowing some links to be powered off, and the experiment results assuming a real IP network topology showed that more than 50% of the links could be switched off. Vasic and Kostic [51] presented Energy-Aware Traffic engineering (EATe), a distributed online technique that leverages sleeping of links and routers to save energy by spreading the traffic load over multiple links. The technique employed is to shift all traffic from the links and routers with the minimum utilization to the remaining resources, allowing the network to become energy proportional to the traffic load.

Heller et al. [52] proposed ElasticTree which aimed to reduce the energy consumption of data center networks by turning off switches and links not required during times of low-utilization. The work involved collecting traffic statistics such as the topology, current traffic matrix, and the desired fault tolerance levels, modelling and optimizing the problem while satisfying all constraints, and re-provisioning the devices using the OpenFlow protocol to maintain a minimum power subset of the network. Three different models are presented – a formal model based on a standard multicommodity flow problem, a greedy-bin packing model, and a topology-aware heuristic model requiring only the port counters as the input. The experiments conducted on tree-based topologies like Fat-Tree indicated a potential to save up to 50% of network energy with the ability to handle traffic fluctuations. Zhang et al. [53] proposed a centralized energy-aware traffic engineering scheme, GreenTE, combining device, component and link-level solutions. The general TE problem is formulated as a multicommodity flow with the specific variables of link and line-card energy states modeled using mixed-integer programming (MIP) problem. Since MIP problems are generally NP-hard, practical heuristics such as constraints on the maximum link utilization and the use of candidate paths instead of searching in all paths are used to reduce the computation time. The simulation results using production topologies from Abilene and GEANT showed a reduction in the energy consumption of line-cards' by 27-42% while maintaining link utilization levels below 50%. The energy-aware routing algorithm was also formulated as an integer linear programming (ILP) problem in [54], and solved analytically for links in [55], and for both links and nodes in [56]. Puype et al. [57] proposed to leverage multilayer traffic engineering (MLTE) to traffic engineer around and shut down energy inefficient portions of the IP-over-optical networks. In [58], the authors presented the Responsive Energy-Proportional Networks (REsPoNse) framework that proactively identified energy-critical paths by analyzing the traffic matrices, installed the corresponding routing entries into three different route tables – always-on, on-demand, and failover – and reactively modified the state of network elements according to the demand. The objective was to overcome the high computation times of computing energy-aware routing tables by analyzing the tradeoff between optimality and scalability. Mohammadpour and Bakhshi [61] employed a realistic power model of network devices from [63] to develop a routing algorithm, RLA-ENAR, formulated as an ILP problem and the experiments using Abilene network showed 40% more energy savings as compared to OSPF-TE.

Wang et al. proposed CARPO [59], a correlation-aware power optimization algorithm that consolidated network flows onto a subset of switches and links in a data center network allowing for the remaining network elements to be switched off. The work involved understanding the correlations between flows for the purpose of traffic aggregation, developing a heuristic algorithm and using OpenFlow to modify flow entries accordingly. The experiments performed on Wikipedia traces showed the potential to save 46% of network energy in a data center. In [79], Chiaraviglio et al. formulated the problem of reducing the power consumption of backbone networks as an ILP formulation. To reduce the computation time of the proposed algorithm, additional constraints, reduced notations and simple heuristics are employed. The algorithms aimed to find the minimal set of routers and links to satisfy a given traffic demand under connectivity and quality-of-service constraints, while assuming that the traffic matrix at a given time and the power consumption of each link and router is known. The test results from both real and synthetic topologies indicated that up to 35% of power could be saved, especially during off-peak times, when traffic is low. While virtual-machine (VM) consolidation to save server energy in data centers had received much research focus, the authors in [60] leveraged this VM migration approach to power off unused switches and attempt to increase the number of inactive switches. The proposed approach involved adding bypass links between the physical machine, called a honey machine, and upper-tiered switches, aggregating all VM’s on one rack connecting to the bypassed switch to the honey machine, and powering off the unused switch to enable energy savings. The simulation results showed that the energy savings using this approach were up to 7.8% in a fat tree network as compared to the conventional VM-migration schemes. Son et al. [90] developed a VM-consolidation scheme based on the historical monitoring data of the host and network utilization. The algorithm first groups VMs based on their connectivity, and then sorts VM groups according to their resource requirements. Connected VM groups are consolidated on a single host in order to minimize the number of active transit switches. Also, VMs are migrated from over-utilized hosts to partially utilized hosts in order to maintain SLA requirements.

In [86], Zhu et al. discussed the energy efficiencies of common routing and scheduling algorithms by analyzing their traffic aggregation capabilities. The simulation results using Fat Tree and 3-level BCube data center topologies indicated that using priority based shortest routing – selecting the highest priority, i.e. the least congested, path amongst all shortest paths – with exclusive flow scheduling – transferring flows one by one and allowing each flow exclusive use of the total link bandwidth – was the most energy efficient strategy. Wei et al. [87] studied the energy-efficient traffic engineering problem in hybrid SDN/IP networks. The proposed fast heuristic algorithm aimed to reduce the number of active links in the network and power them off by optimizing OSPF link weights and traffic-splitting ratios in SDN devices. OSPF link weights were optimized for energy efficiency by increasing the costs of congested or sleeping links, in order to concentrate traffic on the least amount of links. SDN traffic was optimized for energy efficiency by moving traffic flows from low-utilization links to high-utilization links. The simulation tests performed in NS2 on real topologies indicated an energy savings improvement of 13.2% on exiting energy aware OSPF algorithms. In [95], the authors considered both the data plane and in-band control plane traffic to minimize the number of links required to satisfy a given traffic deman. [88] presented an energy monitoring and management application (EMMA) for SDN-based 5G backhaul networks. The scheme aimed to limit the number of active links and nodes by trying to fit any new flow into the current active network while meeting the flow requirements. If no suitable path was found, additional links and/or nodes were turned on, while also checking for possible best paths for existing flows. The algorithm was implemented using Mininet and the ONOS SDN controller, and experimental results showed that EMMA performed very close to the optimum solution.

Maleki et al. [93] presented a method to reduce the number of active links in the network by leveraging SDN features considering the GEANT network. The proposed scheme – Shared Path First (SPF) – calculated the shortest path between a source and destination for the first traffic flow from a particular source. For the subsequent flows from that source, existing active parts were first checked to determine if the new capacity requirements can be catered to, if not, new shortest paths are considered to accommodate the new flow. The experiment results indicated that 41% of links could be saved as compared to the shortest path first approach. [91] presented an approximate algorithm to save energy in cloud-based content distribution systems (CDNs). The strategy involved processing historical traffic data to determine the maximum value of traffic for each hour, which in turn was used to ascertain the number of vCDN functions required to be deployed. Next, the autoregressive integrated moving average (ARIMA) static forecasting model is employed to predict the future traffic load. Based on that and the number of redundant functions required, the minimum number of vCDN functions was computed. In [94], the authors leveraged the IEEE 802.3az functionality in bundles of links to reduce the operation costs in data centers and wired access networks. They presented several algorithms to select output links in bundles in order to increase the effective time they could be out in low-power modes while satisfying the QoS requirements of different applications. The greedy algorithm (GA) filled links to their maximum capacities before allocating new links, the bounded greedy algorithm avoided filling links to their full capacities in order the bound the packet delay and losses in GA, the conservative algorithm evenly spread the load amongst the minimum number of links in a bundle determined in a time interval. Additional modifications are also discussed such as the spare port algorithm wherein all best-effort traffic is concentrated on the least number of links, while the low-latency flows are mapped on to the remaining unused links in the bundle, and the two queues algorithm, wherein two queues are created for the best-effort traffic (low-priority queue) and the low-latency traffic (high-priority traffic) for each physical link. The experiments were performed on the ONOS SDN controller using real network traffic obtained from CAIDA, and the results indicated that the proposed algorithms consumed at least 18% less link energy than the baseline uniform-distribution algorithm. [97] proposed a bi-level optimization problem for ISP networks where the upper level represented the energy management function, and the lower level represented a multi-path routing protocol. Then, it was reformulated as a one-level MILP replacing the second level problem by different sets of optimality conditions. Numerical evaluations on Abilene, Geant and Polska networks indicated that the iterative cutting plane and branch-and-cut algorithms were close in terms of CPU time.

4. COMPLEMENTARY WORK

This section presents three additional categories of work – modeling and measurement, standards work, and energy efficiency metrics – which are closely related to the work presented in Section III.

4.1. Modeling and measurement

Research efforts have been put into modeling the energy consumption of a network device and its components. The work in [64] involved estimating the energy consumption of packet switching fabrics using statistical modeling. The discrete-time batch Markovian arrival process is used to create a stochastic traffic model and the results emphasized the importance of moving the energy optimization process from the circuitry to the system level. [65] and [66] employed analytical models to estimate the energy consumption of on-chip switching interconnects. The approach in

[67] was an architectural-level estimation aimed to incorporate the realtime nature of dynamic contention between packets into the model. The simulation results led to important conclusions – storing a packet in the buffer consumed far more power than transmitting it on the interconnect, energy consumed by the buffers was a significant part of the total energy consumption, and buffer energy would increase as the packet flow throughput increases. The model developed in [63] incorporated per-packet processing and per-byte store and forward handling estimations.
Chabarek et al. [49] measured the power consumption of two widely-used routers to come up with a generic power consumption model. The authors simulated a range of different configurations and operating conditions to ensure the comprehensiveness of the model. The experiments revealed that the power consumption of a router was dependent on the underlying chassis, the installed line cards, and the configuration and traffic utilization of the device. Mahadevan et al. [68] presented a power measurement study of a variety of network devices such as hubs, edge switches, core switches, routers, and wireless access points. The experiment observations noted that the power consumed by a network device is determined by factors such as the number of active ports, the line speed configured for each port, and the firmware version on the device, while traffic utilization and packet size had minimal-to-no impact. In [78] they present a large power profile study conducted in an enterprise network, comprising of 90 live switches from various vendors. The work in [69] showed that the energy consumption depends on the volume of computed traffic and device reconfigurations, while queue management policies and BGP updates had no impact. In [83], Orgerie et al. propose an end-to-end cost model and simulator for evaluating power consumption in large-scale networks. Their model computes the energy consumption per equipment depending on the amount of bandwidth traversing, the length of a transfer, and the type of equipment. The simulator module is integrated with NS2 and takes user inputs such as a network topology, network traffic, and the energy consumption values for the network equipment used computed from their model. The simulator also supports advanced functionalities such as dynamic on/off links and adaptive link rate. The authors extend their work in [84] to redesign the simulator to be integrated with NS3 which allows running of native Linux code to provide more accurate consumption values.

Chabarek et al. extended their prior work in [70] to develop a network-wide power consumption framework. The work involved employing an application programming interface to interface with existing network management tools, comparing the device configurations with community database benchmark measurements, and auditing to infer any missing data to enable power consumption estimations. Hossain et al. [92] measured and modeled the energy consumption of Ethernet switches considering parameters such as link speeds, traffic, the number of connections. Additionally, full factorial and linear regression analysis is performed to identify the most influential parameters. Their results showed that link capacities and the number of connections have an impact on the power consumption while changing traffic had little impact. [85] modeled energy consumption of the software stacks of Ethernet and Infiniband NICs related to VM migration. The experiment results indicated that transferring the same quantity of data over Infiniband in connected mode was more energy efficient as compared to Infiniband in datagram mode or Gigabit Ethernet. However, for message centric traffic, wherein the volume of effective data is low and the number of transmitted packets is high, Gigabit Ethernet was more efficient, closely followed by Infiniband in datagram mode.In [86], Zhu et al. developed a network energy monitoring prototype using OpenNaaS to obtain energy usage information from OpenFlow switches. They used SNMP to fetch and control the power state of switches and ports by creating power meter drivers for SNMP access to different meter vendors.

4.2. Standards work

The European Telecommunications Standards Institute (ETSI) published a standard, ES-203 237 – the Green Abstraction Layer, that specified the Green Standard Interface (GSI) for a uniform

way for interactions between the energy-aware hardware and the control framework [71]. The proposed GSI intended to provide the functionalities of discovery – control plane retrieving information about the different energy states supported by the data plane, provisioning – control plane configuring different energy states, and monitoring – exchanging relevant device parameters. The goal of this standard was to represent an abstraction of the energy-aware capabilities of networking devices to higher-layer protocols.
The Internet Engineering Task Force (IETF) published the Energy Management Framework (eMAN) [72], which presented a physical reference model and an information model for devices and device components within, or connected to, networks. The framework modeled relationships and capabilities between energy objects such as power, power state, energy, demand, power attributes and battery. This was the first attempt to standardize monitoring and control for power and energy of networked devices using a Management Information Base (MIB) [73].

The Society of Cable Telecommunications Engineers' (SCTE) Energy Management Subcommittee (EMS) published a standard ANSI/SCTE 216 2015 – Adaptive Power System Interface Specification (APSIS) [74]. They proposed interfacing APSIS applications with the network elements over standard management protocols such as SNMP, NETCONF, IPDR, and HTTP. Other standards such as SCTE 184 – providing guidelines for balancing energy-efficient operations with essential business requirements, SCTE 211 – defining energy metrics, and ANSI/SCTE 212 – defining a framework to establish energy baselines, were also published.

4.3. Energy efficiency metrics

Metrics have been proposed in this field to measure and compare the impact of energy saving schemes; [75] surveyed and presented them. Equipment-level metrics included ECR – energy consumption in Watt/Gbps, EER – energy efficiency in Gbps/Watt, EPI – energy proportionality in percentage, and FLOPS per watt – peak power in terms of operations per second. Facility-level metrics included PUE – Power Usage Effectiveness is the ratio of total data center power to the power drawn by IT equipment, DCiE – Data Center infrastructure Efficiency in percentage, and DCP – Data Center Productivity measured the amount of useful work done by the data center. Country-level metrics included EPI – Environmental Performance Index, ESI – Environmental Sustainability Index, and EVI – Environmental Vulnerability Index.

5. CONCLUSIONS

This paper presented a detailed survey of energy-efficiency approaches in wired networks focusing on energy-aware protocols and network design. We categorized the existing work into different broad categories – sleeping of network elements, link rate adaptation, proxying, store and forward, and network traffic aggregation - and described the specific research efforts in each category. Additionally, work in modeling and measurement, standards, and metrics is also discussed. The research challenges, test results, and important conclusions are indicated to make this survey holistic. Below we examine discussion points for future work in this field.

5.1. Energy proportionality does not imply energy efficiency

While research efforts are being invested to make devices more energy proportional and energy efficient, it is to be noted that one does not imply the other. As pointed out in [68], while some devices may consume power more proportionally to their load, higher values of absolute consumption still make them inefficient. Contrarily, some devices may consume less power while transmitting the same amount of traffic as others, negligible variations with respect to traffic load make them non-energy proportional. The efforts concentrated on future networking hardware must take this into consideration.

5.2. Energy efficiency does not imply decreased energy consumption

The fact that per unit computational power and unit energy efficiencies are improving does not mean that the energy consumption of the Internet is decreasing. Absolute numbers suggested an upward trend in the energy consumed and it is projected to continue increasing [76]. Improving energy efficiency reduces the implicit cost to use the resource making it more affordable and thus leading to increased use of the resource – a phenomenon known as the rebound effect in energy economics [77]. Therefore, future efforts should be focused on decreasing the absolute energy consumption of the Internet to ensure its long-term sustenance.

5.3. Incorporate energy efficiency into network management

Network management has traditionally comprised of five major areas – fault, configuration, accounting/administration, performance, and security. However, the criticality of energy management warrants its inclusion in the traditional definition of network management to ensure that both academia and industry do not consider energy management an afterthought.

5.4. User awareness

While the supply side of the information and communications technology industry has started to understand the importance of energy efficiency, the demand side still lags. Since end users are aware and responsible for remunerating a small portion of the total energy costs, the urgency has not trickled down. We recommend educational and governmental efforts to enable everyone to make energy-conscious decisions.

REFERENCES

[1] Internetworldstats.com. (2019). World Internet Users Statistics and 2019 World Population Stats. [online] Available at: https://www.internetworldstats.com/stats.htm [Accessed 11 October 2021].

[2] J. Koomey and S. Naffziger, Efficiency's brief reprieve: Moore's Law slowdown hits performance more than energy efficiency, in IEEE Spectrum, pp. [http://spectrum. ieee. org/computing/hardware/moores-law-might-be-slowing-down-but-not-energyefficiency], Apr. 2015.

[3] M. Gupta and S. Singh, Greening of the Internet, in Proceedings of ACM SIGCOMM ’03, Karlsruhe, Germany, Aug. 2003.

[4] K. W. Roth, F. Goldstein, and J. Kleinman, Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings Volume I: Energy Consumption Baseline, Tech. Rep. Vol I, National Technical Information Service (NTIS), US Department of Commerce, Jan. 2002.

[5] M. Gupta, S. Grover, and S. Singh, A Feasibility Study for Power Management in LAN Switches, in Proceedings of 12th IEEE International Conference on Network Protocols (ICNP), Berlin, Germany, pp. 361-371, Oct. 2004.

[6] V. Soteriou and L.-S. Peh, Dynamic power management for power optimization of interconnection networks using on/off links, in Proceedings of the 11th Symposium on High Performance Interconnects, pp. 15-20, 2003.

[7] M. Gupta and S. Singh, Using Low-Power Modes for Energy Conservation in Ethernet LANs, in Proceedings of the 26th Annual IEEE Conference on Computer Communications (INFOCOM 2007), (Anchorage, Alaska), pp. 2451 – 2455, May 2007.

[8] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, Reducing Network Energy Consumption via Sleeping and RateAdaptation, in Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation (NDSI2008), (San Francisco, California, USA), Apr. 2008.

[9] G. Ananthanarayanan and R. H. Katz, Greening the Switch, in Proceedings of the USENIX Workshop on Power Aware Computing and Systems (HotPower), held at the Symposium on Operating Systems Design and Implementation (OSDI 2008), (San Diego, California, USA), Dec. 2008.

[10] L. Chiaraviglio, M. Mellia, and F. Neri, Energy-aware Backbone Networks: a Case Study, in Proceedings of the 1st International Workshop on Green Communications (GreenComm) in conjunction with the IEEE International Conference on Communications, (Dresden, Germany), June 2009.

[11] K. Christensen, P. Reviriego, B. Nordman, M. Mostowfi, and J. A. Maestro, IEEE 802.3az: the road to energy efficient ethernet, IEEE Communications. Magazine, vol. 48, no. 11, pp. 50–56, Nov. 2010.

[12] H. Sedarat, O. Barkan, and W. Woodruff, Low-power idle mode for network transceiver, United States Patent 8156359, Sep.2009.

[13] K. Christensen, B. Nordman, and R. Brown, Power management in networked devices, IEEE Computer, vol. 37, pp. 91–93, Aug.2004.

[14] C. Gunaratne, K. Christensen, and B. Nordman, Managing energy consumption costs in desktop PCs and LAN switches with proxying, split TCP connections and scaling of link speed, International Journal of Network Management, vol. 15, pp. 297–310, Sep. 2005.

[15] C. Gunaratne, K. Christensen, and S. W. Suen, Ethernet Adaptive Link Rate (ALR): Analysis of a buffer threshold policy, in Proceedings of the IEEE Global Communications Conference (GLOBECOM 2006), (San Francisco, California, USA), Nov. 2006.

[16] C. Gunaratne and K. Christensen, Ethernet Adaptive Link Rate: System Design and Performance Evaluation,in Proceedings of the IEEE Conference on Local Computer Networks, pp. 28-35, Nov. 2006.

[17] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen, Reducing the Energy Consumption of Ethernet with Adaptive Link Rate (ALR), IEEE Transactions on Computers, vol. 57, pp. 448–461, Apr. 2008.

[18] H. Anand, C. Reardon, R. Subramaniyan, and A. George, Ethernet Adaptive Link Rate (ALR): Analysis of a MAC handshake protocol, in Proceedings of the 31st IEEE Conference on Local Computer Networks, pp. 533–534, Nov. 2006.

[19] F. Blanquicet and K. Christensen, An Initial Performance Evaluation of Rapid PHY Selection (RPS) for Energy Efficient Ethernet, in Proceedings of the 32nd IEEE Conference on Local Computer Networks, pp. 223-225, Oct. 2007.

[20] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, Energy Aware Network Operations, in Proceedings of the IEEE Global Internet Symposium, (Rio de Janeiro, Brazil), Apr. 2009.

[21] D. Abts, M. R. Marty, P. M. Wells, P. Klausler, and H. Liu, Energy proportional datacenter networks, SIGARCH Computer Architecture News, vol. 38, no. 3, pp. 338–347, Jun. 2010.

[22] G. Ginis, Low-Power Modes for ADSL2 and ADSL2+, SPAA021, Broadband Communications Group, Texas Instruments, Jan. 2005.

[23] D. Staessens, S. Sharma, D. Colle, M. Pickavet, and P. Demeester, Software defined networking: Meeting carrier grade requirements, in Proceedings of the 18th IEEE Workshop LANMAN, pp. 1–6, 2011.

[24] C. Callegari, R. G. Garroppo, S. Giordano, and G. Nencioni, A new Markov model for evaluating the ALR Dual-Threshold Policy, 2nd IEEE Workshop on Green Communications, Hawaii, USA, Nov. 30 – Dec. 4, 2009.

[25] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, Opportunistic power saving algorithms for Ethernet devices, Computer Networks, vol. 55, no. 9, pp. 2051–2064, Jun. 2011.

[26] K. J. Christensen, C. Gunaratne, B. Nordman, and A. D. George,The next frontier for communications networks: Power management, Computer Communications, vol. 27, pp. 1758–1770, 2004.

[27] J.A. Robertson et al., After-hours Power Status of Office Equipment and Energy Use of Miscellaneous Plug-Load Equipment, Lawrence Berkley Laboratory, May 2004.

[28] AMD, Magic Packet Technology, Publication# 20213, Rev. A, Nov. 1995.

[29] Advanced Configuration and Power Interface Specification, Rev. 5.0a, Nov. 2013.

[30] K. Christensen and F. Gulledge, Enabling Power Management for Network-Attached Computers, International Journal of Network Management, vol. 8, no. 2, pp. 120–30, Apr. 1998.

[31] M. Allman, K. Christensen, B. Nordman, and V. Paxson, Enabling and Energy-Efficient Future Internet Through Selectively Connected End Systems, Proc. ACM SIGCOMM HotNets Workshop (HotNets 07), Atlanta, GA, Nov. 2007.

[32] S. Nedevschi et al., Skilled in the Art of Being Idle: Reducing Energy Waste in Networked Systems, in Proceedings of the. 6th ACM/USENIX Symposium on Networked Systems Design and Implementation (NSDI 2009), Boston, MA, USA, Apr. 2009.

[33] W. Vereecken et al., Power Efficiency of Thin Clients, European Transactions on Telecommunications, vol.21, no. 6, pp. 479-490, 2010.

[34] R. Khan, R. Bolla, M. Repetto, R. Bruschi, and M. Giribaldi, Smart Proxying for Reducing Network Energy Consumption, in IEEE International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), July 2012.

[35] R. Bolla, M. Giribaldi, R. Khan, and M. Repetto, Network Connectivity Proxy: An Optimal Strategy for Reducing Energy Waste in Network Edge Devices, in The 24th Tyrrhenian International Workshop on Digital Communications, 2013.

[36] Y. Agarwal et al., Somniloquy: augmenting network interfaces to reduce pc energy usage, in Proc. of the USENIX symposium on Networked systems design and implementation (NSDI'09), 2009.

[37] M. K. Girish, B. Zhou, and J. Q. Hu, Formulation of the traffic engineering problems in MPLS based IP networks, in Proceedings of the 5th IEEE Symposium on Computers and Communications (ISCC 2000), pp. 214-219, 2000.

[38] P. Purushothaman, M. Navada, R. Subramaniyan, C. Reardon, and A. D. George, Power-Proxying on the NIC: a Case Study with the Gnutella File-Sharing Protocol, in Proceedings of the 31st IEEE Conference on Local Computer Networks (LCN 2006), Nov. 2006.

[39] M. Jimeno and K. Christensen, A Prototype Power Management Proxy for Gnutella Peer-to-Peer File Sharing, in Proceedings of the 32nd IEEE Conference on Local Computer Networks (LCN 2007),Oct. 2007.

[40] K. Sabhanatarajan and A. Gordon-Ross, A Resource Efficient Content Inspection System for Next Generation Smart NICs, in Proceedings of the IEEE International Conference on Computer Design (ICCD 2008), pp. 156–163, Oct. 2008.

[41] M. Baldi and Y. Ofek, Time for a Greener Internet, in Proceedings of the 1st International Workshop on Green Communications (GreenComm) in conjunction with the IEEE International Conference on Communications,June 2009.

[42] C.-S. Li, Y. Ofek, and M. Yung, Time-driven priority flow control for real-time heterogeneous Internetworking, IEEE INFOCOM'96, Mar. 1996.

[43] V. Cerf et al., Delay-Tolerant Networking Architecture, IETF Request for Comments #4838, April 2007.

[44] A. Akella, R. Balan and S. Seshan.. Protocols for Low Power, SIGCOMM CCR, Jan. 2002.

[45] M. Rodríguez-Pérez, S. Herrería-Alonso, M. Fernández-Veiga, and C. López-García, Improved opportunistic sleeping algorithms for LAN switches, in Proceedings of the IEEE Globecom, Honolulu, HI, USA, Dec. 2009.

[46] P. Reviriego, J. A. Maestro, J. A. Hernandez, and D. Larrabeiti, Burst transmission for Energy-Efficient Ethernet, IEEE Internet Comput., vol. 14, no. 4, pp. 50–57, Jul. 2010.

[47] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, A power saving model for burst transmission in energy-efficient Ethernet, IEEE Commun. Lett., vol. 15, no. 5, pp. 584– 586, May 2011.

[48] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, How efficient is energy-efficient ethernet?, in 3rd International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 1-7. 2011.

[49] J. Chabarek et al., Power Awareness in Network Design and Routing, in Proc. IEEE 27th IEEE Conf. on Computer Communications (INFOCOM 2008), Phoenix, AZ,pp. 457- 465, Apr. 2008.

[50] Cianfrani, V. Eramo, M. Listanti, M. Marazza, and E. Vittorini, An Energy Saving Routing Algorithm for a Green OSPF Protocol, in INFOCOM IEEE Conference on Computer Communications Workshops, pp. 1-5, Mar. 2010.

[51] N. Vasic and D. Kostic, Energy-Aware Traffic Engineering, in Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, Apr. 2010.

[52] B. Heller et al., ElasticTree: Saving Energy in Data Center Networks, in Proc. of USENIX Symposium on Networked Systems Design and Implementation (NSDI), Apr. 2010.

[53] M. Zhang, C. Yi, B. Liu and B. Zhang, GreenTE: Power-Aware Traffic Engineering, in Proceedings of the 18th IEEE International Conference on Network Protocols, pp. 21-30, Oct. 2010.

[54] L. Chiaraviglio, M. Mellia, and F. Neri, Reducing Power Consumption in Backbone Networks, in Proceedings of the IEEE International Conference on Communications (ICC 2009), (Dresden, Germany), June 2009.

[55] W. Fisher, M. Suchara, and J. Rexford, Greening Backbone Networks: Reducing Energy Consumption by Shutting Off Cables in Bundled Links, in Proceedings of 1st ACM SIGCOMM workshop on green networking, (New Delhi, India), Aug. 2010.

[56] A. Bianzino, C. Chaudet, D. Rossi, and J.-L. Rougier, Energy-Awareness in Network Dimensioning: a Fixed Charge Network Flow Formulation, in 1st International Conference on Energy-Efficient Computing and Networking (e-Energy 2010), Extended Abstract, (Passau, Germany), Apr. 2010.

[57] A. Puype, W. Vereecken, D. Cole, M. Pickavet, and P. Demeester, Multilayer traffic engineering for energy efficiency, Photonic Network Communications, vol. 21, no. 2, pp. 127-140, Apr. 2011.

[58] N. Vasic et al., Identifying and Using Energy-Critical Paths, in ACM CoNEXT, Dec. 2011.

[59] X. Wang and X. Wang, CARPO: Correlation-Aware Power Optimization in Data Center Networks, in Proc. of IEEE INFOCOM, pp. 1125-1133, Mar. 2012.

[60] H. Shirayanagi, H. Yamada, and K. Kono, Honeyguide: A VM Migration-Aware Network Topology for Saving Energy Consumption in Data Center Networks, IEICE TRANSACTIONS on Information and Systems, pp. 2055-2064, Sept. 2013.

[61] E. Mohammadpour and B. Bakhshi, RLA-ENAR: A Realistic Near-Optimal Energy-Aware Routing, in 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), pp. 1-5, Sep. 2016.

[62] R. Gandotra and L. Perigo, SDNMA: A Software-Defined, Dynamic Network Manipulation Application to Enhance BGP Functionality, IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 1007-1014, 2018.

[63] A. Vishwanat, K. Hinton, R. W. A. Ayre, and R. S. Tucker, Modeling Energy Consumption in High-Capacity Routers and Switches, IEEE Journal on Selected Areas in Communications, vol. 32, no. 8, pp. 1524-1532, Aug. 2014.

[64] G. Wassal and M. A. Hasan, Low-Power System-Level Design of VLSI Packet Switching Fabrics, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 20, no. 6, pp. 723-728, 2001.

[65] S. Patel, S. M. Chai, S. Yalamanchili, and D. E. Schimmel, Power Constrained Design of Multiprocessor Interconnection Networks, in Proceedings of the International Conference on Computer Design VLSI in Computers and Processors, pp. 408-416, 1997.

[66] A. Langen, A. Brinkmann, and U. Ruckert, High level estimation of the area and power consumption of on-chip interconnects, in Proceedings of the 13th Annual IEEE International ASIC/SOC Conference, pp. 297-301, 2000.

[67] T. T. Ye, L. Benini, and G. D. Micheli, Analysis of Power Consumption on Switch Fabrics in Network Routers, in Proceedings of the Design Automation Conference, pp. 524-529, 2002.

[68] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, A Power Benchmarking Framework for Network Devices, in Proceedings of IFIP Networking 2009, May 2009.

[69] Adelin, P. Owezarski, and T. Gayraud, On the Impact of Monitoring Router Energy Consumption for Greening the Internet, in Proceedings of the 11th IEEE/ACM International Conference on Grid Computing, pp. 298-304, 2010.

[70] J. Chabarek and P. Barford, Energy Audit: Monitoring Power Consumption in Diverse Network Environments, in Proceedings of the International Green Computing Conference, pp. 1-10, Jun. 2013.

[71] European Telecommunications Standards Institute, Green Abstraction Layer (GAL), ES 203 237 (ETSI Standard), 2014. [Online]. Available at: https://www.etsi.org/deliver/etsi_es/203200_203299/203237/01.01.01_60/es_203237v010101p.pdf.

[72] J. Parello, B. Claise, B. Schoening, and J. Quittek, Energy Management Framework, RFC 7326 (Informational), Internet Engineering Task Force, September 2014. [Online]. Available at: https://tools.ietf.org/html/rfc7326.

[73] M. Chandramouli, B. Claise, B. Schoening, J. Quittek, and T. Dietz, Monitoring and Control MIB for Power and Energy, RFC 7460 (Standards Track), March 2015. [Online]. Available at: https://tools.ietf.org/html/rfc7460.

[74] Society of Cable Telecommunications Engineers (SCTE), Adaptive Power System Interface Specification (APSIS), American National Standard, 2015. [Online]. Available at: https://www.scte.org/SCTEDocs/Standards/ANSI_SCTE%20216%202015.pdf.

[75] A. P. Bianzino, A. K. Raju, and D. Rossi, Apple-to-Apple: A Framework Analysis for Energy-Efficiency in Networks, Proc. of SIGMETRICS, 2nd GreenMetrics workshop, 2010.

[76] A. Raghavan and J. Ma, The Energy and Emergy of the Internet, in Proceedings of the 10th ACM Workshop on Hot Topics in Networks, Nov. 2011.

[77] J. D. Khazzoom, Economic implications of mandated efficiency in standards for household appliances, Energy Journal, vol. 1, no. 4, pp. 21–40, 1980.

[78] P. Mahadevan, S. Banerjee, and P. Sharma, Energy proportionality of an enterprise network, in Proceedings of the first ACM SIGCOMM workshop on Green networking, pp. 53-60, Aug. 2010.

[79] L. Chiaraviglio, M. Mellia, and F.Neri, Minimizing ISP network energy cost: Formulation and solutions, IEEE/ACM Transactions on Networking, vol. 20, no. 2, pp. 463-476, Apr. 2012.

[80] K.-H. Ho and C.-C. Cheung, Green distributed routing protocol for sleep coordination in wired core networks, in Proceedings of 6th International Conference on Networked Computing, pp. 1-6, May 2010.

[81] A. P. Bianzino, L. Chiaraviglio, M. Mellia, and J. L. Rougier, GRiDA: Green distributed algorithm for energy-efficient IP backbone networks, Computer Networks, vol. 56, no. 14, pp. 3219-3232, Sep. 2012.

[82] F. Patota et al., DAFNES: A distributed algorithm for network energy saving based on stress-centrality, Computer Networks, vol. 94, pp. 263-284, Jan. 2016.

[83] A.-C. Orgerie, L. Lefevre, I. Guerin-Lassous, and D. M. L. Pacheco, ECOFEN: An end-to-end energy cost model and simulator for evaluatingpower consumption in large-scale networks, in Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, pp. 1-6, Jun. 2011.

[84] A.-C. Orgerie et al., Simulation toolbox for studying energy consumption in wired networks, in Proceedings of 13th International Conference on Network and Service Management (CNSM), pp. 1-5, Nov. 2017.

[85] V. D. Maio, R. Prodan, S. Benedict, and G. Kecskemeti, Modelling energy consumption of network transfers and virtual machine migration, Future Generation Computer Systems, vol. 56, pp. 388-406, Mar. 2016.

[86] H. Zhu, X. Liao, C. d. Laat, and P. Grosso, Joint flow routing-scheduling for energy efficient software defined data center networks, Journal of Network and Computer Applications, vol. 63, pp. 110-124, Mar. 2016.

[87] Y. Wei, X. Zhang, L. Xie, and S. Leng, Energy-aware traffic engineering in hybrid SDN/IP backbone networks, Journal of Communications and Networks, vol. 18, no. 4, pp. 559-566, Sep. 2016.

[88] S. S. Tadesse, C. Casetti, and C. F. Chiasserini, Energy-efficient traffic allocation in SDN-based backhaul networks: Theory and implementation, in Proceedings of 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 209-215, Jan. 2017.

[89] K. Xie et al., E3MC: Improving energy efficiency via elastic multi-controller SDN in data center networks, IEEE Access, vol. 4, pp. 6780-6791, Oct. 2016.

[90] J. Son, A. V. Dastjerdi, R. N. Calheiros, and R. Buyya, SLA-aware and energy-efficient dynamic overbooking in SDN-based cloud data centers, IEEE Transactions on Sustainable Computing, vol. 2, no. 2, May 2017.

[91] Liao, G. Sun, G. Yang, and V. Chang, Energy-efficient virtual content distribution network provisioning in cloud-based data centers, Future Generation Computer Systems, vol. 83, pp. 347-357, Jun. 2018.

[92] M.M. Hossain, E. Rondeau, J.-P. Georges, and T. Bastogne, Modeling the power consumption of Ethernet switch, inProceedings of International SEEDS Conference: Sustainable Ecological Engineering Design for Society, pp. 469-480, Sep. 2015.

[93] A. Maleki, M. Hossain, J.-P. Georges, E. Rondeau, and T. Divoux, An SDN perspective to mitigate the energy consumption of core networks –GÉANT2, in Proceedings of International Sustainable Ecological Engineering Design for Society (SEEDS) Conference, pp. 233-244, Sep. 2017.

[94] P. Fondo-Ferreiro, M. Rodríguez-Pérez, M. Fernández-Veiga, and S. Herrería-Alonso, Matching SDN and legacy networking hardware for energy efficiency and bounded delay, Sensors, vol. 18, no. 11, Nov. 2018.

[95] A. Fernandez-Fernandez, C. Cervello-Pastor, and L. Ochoa-Aday, Achieving energy efficiency: An energy-aware approach in SDN, in Proceedings of IEEE Global Communications Conference (GLOBECOM), pp. 1-7, Dec. 2016.

[96] M. d. S. Conterato, T. C. Ferreto, F. Rossi, W. d. S. Marques, and P. S. S. d. Souza, Reducing energy consumption in SDN-based data center networks through flow consolidation strategies, in Proceedings of 34th ACM/SIGAPP Symposium on Applied Computing, pp. 1384-1391, Apr. 2019.

[97] Bouras, R. Figueiredo, M. Poss, and F, Zhou, Minimizing energy and link utilization in ISP backbone networks with multi-path routing: A bi-level approach, Optimization Letters, pp. 1-19, Nov. 2019.

SMARTPHONE MODEL FINGERPRINTING USING WIFI RADIATION PATTERNS

Thomas Burton and Kasper Rasmussen

Department of Computer Science, University of Oxford, Oxford, UK

ABSTRACT

This paper aims to demonstrate the feasibility of our proposed method for fingerprinting different classes of wireless devices. Our method relies on the observation that different device types, or indeed different models of the same type, have different wireless radiation patterns. We show in detail how a small set of stationary receivers can measure the radiation pattern of a transmitting device in a completely passive manner. As the observed device moves, our method can gather enough data to characterize the shape of the radiation pattern, which can be used to determine the type of the transmitting device from a database of patterns. We demonstrate that the patterns produced by different models of smartphones are easily different enough to be identified. Our measurements are repeatably measurable using RSS with commercial-off-the-shelf hardware. We then use simulations to show the success of our method as a classifier.

KEYWORDS

Wireless Radiation Patterns, Device Fingerprinting, Identification.

1. INTRODUCTION

Antennas do not radiate power equally in all directions, and the resulting pattern of transmission energy is called a radiation pattern (or antenna pattern). The ability to remotely measure the radiation pattern of a device has a range of potential applications, from checking compliance with emission regulations and standards, optimising transmission power, smart routing and beamforming, to fingerprinting and identifying devices based on unique pattern shapes.

To demonstrate the power of the idea, we focus on fingerprinting smartphone models, which by itself has a number of interesting applications in intrusion detection; commercial analysis of the phone models in a group of users (e.g., for app development); and network analysis to understand what device types are using a network, for example, ensuring compliance with bring your own device to work policies. While fingerprinting provides a convenient application with which to demonstrate our radiation pattern measurement technique, the resulting fingerprinting scheme has a lot of benefits, making it a good competitor to existing approaches.

Existing approaches often focus on how the chipset or firmware behaves under certain circumstances. Fingerprinting in such schemes is conducted by either interrogating the device with differently formed packets or observing the device's particular behavioural characteristics. This can yield good results but will often require additional specialist hardware to inject packets or measure properties that cannot be achieved with commercial-off-the-shelf (COTS) networking infrastructure hardware. Additionally, when actively sending out probing packets, you alert the device to the use of fingerprinting.

David C. Wyld et al. (Eds): MLDS, NECO, SEMIT, IBCOM, SPPR, SCAI, CSIA, ICCSEA, WiMoA - 2021
pp. 283-303, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111823

Our method is entirely passive and relies on received signal strength (RSS), which is widely available across almost all networking hardware as it plays a vital role in network management and troubleshooting. This makes our fingerprinting method an ideal choice for situations where an existing wireless infrastructure already exists, e.g., a company WiFi installation or a LoRa city-wide network.

A receiver in a fixed location will measure the RSS of a transmission from another device differently depending on the transmitting device's radiation pattern, location, and orientation. When compared to measurements by other collaborating devices that form a measurement infrastructure, this will allow the calculation of several points on the radiation pattern of the transmitting device. In this paper, we use multiple receivers to measure several points on the pattern simultaneously for a single packet and attempt to find parts of known patterns that match the shape. We use this to create a fingerprinting method for smartphone models, exploiting the fact that each manufacturer has a slightly different antenna design and a different internal structure of the phone. As a by-product, the method also produces an estimate of the location and orientation of the device.

We summarize our main contributions as follows:

- We identify the challenges of using RSS for fingerprinting when not using location as a proxy for identity.
- We propose a methodology for passively measuring and reconstructing the radiation pattern of nearby wireless devices.
- We measure and analyze patterns from a range of smartphones to show that different models have significantly different patterns.
- We analyze how attacks on this type of system can be performed to break the model fingerprinting mechanism, and we discuss the required setup to mitigate those attacks.

The paper is organised as follows: in Section 2 we cover the necessary background knowledge to understand our proposal, followed by a discussion of related work in Section 3. In Section 4 we define the system model and adversary model. Section 5 presents the challenges to using radiation patterns for fingerprinting. In Section 6 we propose our method of fingerprinting. In Section 7and 8 we evaluate our solution with simulations and we compare patterns measured from a selection of smartphones. In Section 9 we perform an analysis to show the requirements to be secure against attacks. Finally, we conclude in Section 10.

2. TECHNICAL BACKGROUND

We now discuss the background knowledge related to radiation patterns necessary to understand our proposed method.

Directivity is a measure of how directional an antenna's pattern is, and it is one way to represent a radiation pattern. Directivity is defined as ``The ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions'' [14]. From the definition, the directivity for a given elevation and azimuth, $D(\theta, \phi)$, is calculated by:

$$D(\theta, \phi) = \frac{U(\theta, \phi)}{P_{rad}/(4\pi)} \tag{1}$$

Where P_{rad} is the total radiated power output and U is the radiation intensity at the angle θ, ϕ. P_{rad} may be found through several methods, including: using the chipset specification; through calculation using input power and efficiency; an estimate can be retrieved from the US Federal Communications Commission (FCC) certification test report; or by calculation of $\mathrm{U}(\theta, \phi)$ integrated over the spherical surface if the pattern is represented by an equation [1].

$$P_{rad} = \int_0^{2\pi} \int_0^{\pi} U \sin\theta \; d\theta \, d \tag{2}$$

Directivity may be expressed as a ratio in dimensionless units or in decibels relative to another antenna. Most commonly, this is the theoretically perfect isotropic radiator, with a directivity of 1 expressed as a ratio or 0 dBi in all directions. The equations to convert between the two are as follows:

$$D_{dBi} = 10 \cdot log_{10} D \tag{3}$$

$$D = 10^{\left(\frac{D_{dBi}}{10}\right)} \tag{4}$$

3. Related Work

There are two types of fingerprinting of interest: 1) unique device fingerprinting, where every unique device is distinguishable from every other; and 2) device type fingerprinting, where devices are categorised into groups based on some element of common hardware, software, or behaviour.

Unique device fingerprinting gives a very narrow and specific identity, whereas device type fingerprinting gives a very broad identity which may have further sub-categories.

3.1. Unique Device Fingerprinting

Existing fingerprinting or identification of wireless devices is done though a range of methods [29] that have different benefits and drawbacks when it comes to simplicity, computation, reliability, and required hardware. These methods range from simple addresses and cryptographic techniques to the analysis of signal properties.

The simplest method is for a sender to attach an identifier to their message. This is common across many types of networks and at different network layers. For example, TCP/IP uses IP addresses, and IEEE 802 uses MAC addresses. The downside is that an attached identifier can easily be spoofed or modified to achieve various goals. Cryptography can be used to supplement this approach by providing authentication when a device claims to have a particular identity. However, using cryptography leads to some issues, including initial key sharing and key revocation problems and cryptography being computationally expensive for low powered devices.

With access to a network, it has been shown that profiling the network traffic can be used to uniquely identify devices [9, 22]. While this may allow us to determine the software and operating system running on the device, it does not necessarily narrow the device down to one particular model. Also, an adversary can easily spoof this as network traffic is controlled by software that can be easily modified.

Unique device fingerprinting can fall into two categories at the PHY layer: 1) location-independent and 2) location-depended identity.

Location-independent techniques use radiometric properties of a signal for unique device fingerprinting. Radiometric fingerprinting methods can be broadly placed into three categories [11]: transient-based [12, 26], modulation-based [5, 23], and spectral-based techniques [3, 20]. These techniques have a very high accuracy rate, ranging from 70% to 99%, but the drawback of these methods is that they are difficult to deploy and need specialist hardware.

Location-dependent identity methods use a property where the location of a device produces some measurable change of that property. Existing work using received signal strength (RSS) [8, 6] and channel state information (CSI) [16, 17] has focused on using location as a proxy for identity. Using these methods, spatially distanced entities that claim to be the same entity can be distinguished. The downside of this method is that devices (or at least the genuine device in the case of an intruder detection system) must be static.

In order to use location-dependent properties (i.e. RSS or CSI) more generally on mobile devices, the location as a proxy for identity element must be removed. One such method proposed by Hua et al. [13] uses CSI data to infer the carrier frequency offsets (CFO), which arise due to fundamental physical properties of the device, which remains fairly consistent over time but differs significantly depending on the device. However, like many of these fingerprinting methods, the property is not inherent to a particular device type. To later determine the type, an enrolment step must first be performed.

Higher-level fingerprinting techniques have the advantage that they are easier to deploy as the data required is more accessible at the application layer of devices. Fingerprinting particular sensors onboard a device, e.g., the microphone or camera, is possible because applications running on a device can directly access the data feed from the sensors, and some sensors display unique characteristics even across devices of the same type [2]. The sensors can also be used to measure the properties of the device themselves. Perez et al. [24] demonstrated that the electromagnetic emissions of a smartphone can be measured onboard using the device's magnetometer or remotely and then be used to fingerprint the device providing 98.9% accuracy. However, with direct access to the application layer of smartphones, it is already possible to directly retrieve some identifiable information already, such as the manufacturer and model.

3.2. Device Type Fingerprinting

Device type fingerprinting aims to classify devices based on their hardware or software. This can be achieved by manipulating the preamble of packets to see how a receiver handles non-standard or malformed packets, for example, whether it drops the packet. Bratus et al. [4] demonstrated the feasibility for access point (AP) fingerprinting, and Ramsey et al. [25] showed a 99% accuracy when using this method to classify eight transceiver types. A different method by Gao et al. [15] classifies APs based on the time shift of a `packet train' when it is received and sent out by the AP. Although these methods have the advantage of being invariant with respect to the environment, they limit the number of possible unique fingerprints to the number of vendor implementations. For example, if Apple decided to use the same chips over their whole product range, the model would be indistinguishable if these methods were applied to smartphones.

We focus on using RSS for device fingerprinting because it has the large advantage that it is widely accessible across a range of hardware, making it easy to use and process without complex modifications to hardware or software. The focus is on WiFi antennas versus other communication technologies (e.g., Bluetooth and Zigbee) due to the ubiquity across the whole range of consumer electronic devices (including all smartphone models and generations), frequency of data transmission by background applications, and high probability of being in an active state.

3.3. Radiation Patterns

The work on using radiation patterns [7, 21] for improved localization supports the use of radiation patterns for device fingerprinting and model fingerprinting. Coca and Valentin [7] explored the impact of radiation patterns on range-based localization schemes. They found that devices exhibited their own subtle unique patterns, even among the same make and model devices. While Coca proposed additional onboard compasses and a calibration certificate for Wireless Sensor Network nodes, we think this can be leveraged for device fingerprinting. Mwila et al. [21] presented a Gauss-Newton approach to optimize localization. This approach also relies on knowing the orientation of the various entities and therefore relies on cooperation between the infrastructure and device to localize. We do not use this technique because applying it directly to our work is more challenging as there are more unknowns, making the optimization problem significantly harder.

For a device, such as a smartphone, the uneven distribution of energy may be caused by several factors, including the antenna pattern of the antenna itself [1] and the structure of the components packaged within the device causing reflection, diffraction, and attenuation. From now on, we use the term radiation pattern as we refer to the pattern produced by the device as a whole, not just the antenna.

4. SYSTEM AND ATTACKER MODEL

We consider a scenario based on a hybrid `bring your own device' to work and company-issued device environment. Some devices like desktops are company-issued, but most employees also bring in their personal smartphones. These devices are allowed to connect to the network, so they have Internet access away from their desks. However, the company has a policy that requires employees to keep software up-to-date if they want to use their personal devices. To enforce this, the company bans smartphones that cannot get the latest iOS and Android security updates, effectively banning some old models and some manufactures entirely. If a device not on the allowed list $\mathcal{L}$ is detected on the network, then it can be investigated further by other means.

4.1. System Model

The fingerprinting system is deployed in the area of interest $\mathcal{A}$. A set of receivers are in fixed positions around the environment. These sniff all network traffic on the network channel and they record the sender MAC address, received signal strength, and a packet identifier (e.g. a hash of the encrypted packet). The records are then sent to a central *calculation server* (CS) through wired connections. The packets are linked to a device using the MAC address, which we assume does not change while in use. Any attempt to modify the MAC address would cause communication problems at the network layer rendering an attack on identity pointless in this case. The multiple RSS measurements at the different receivers are linked together by the encrypted packet hash. CS knows the position of each receiver, has an RSS map of the environment, and has access to a public database of patterns, so enrolment of every device model is not required. However, an allowed model list $\mathcal{L}$ is maintained and only devices that are allowed on the network are listed. The CS performs the fingerprinting process we later discuss in Section 6, and the model with the closest match to a set of reference patterns will be found, and if a device fails to match a device from$\mathcal{L}$, it will be flagged for further investigation.

For the purpose of later descriptions, every transmitter and receiver has an x, y, and z position and each entity also has an orientation. Rotation is possible around the 3 axis azimuth ϕ (z-axis), pitch θ (x-axis), and roll ψ (y-axis).

The system has the following guarantees: 1) after a coverage metric c has been reached, a device will be flagged if it is classified as a device not in $\mathcal{L}$ for a set amount of time t (c and t are tuneable parameters); 2) if there is insufficient data to determine the model of the device, i.e. to many packets have been dropped (discussed in Section 6.1) or there is insufficient pattern coverage (discussed in Section 6.2), it will also be flagged.

4.2. Attacker Model

The attacker's goal is to access the network with a smartphone not in $\mathcal{L}$ and continue to use the network without being flagged. To achieve this, an adversary must trick the system into believing that the device not on the allowed list is one of the device models on the allowed list.

We make the following assumptions about the attacker's capabilities: the adversary is allowed A1) to choose any device on the not allowed list; A2) use the device in any location within$\mathcal{A}$; A3) full knowledge of how the system works; A4) access to the database of receiver positions and public database of patterns; A5) to make correct estimates of the RSS map for the environment; A6) to make external modifications to their device to modify the pattern using blocking materials placed externally onto the device (this includes a conventional off-the-shelf phone cover or custom made blocking materials).

The adversary may not N1) modify or prevent communications between the receivers and CS; N2) modify the hardware or software on the receivers or CS; N3) modify data in any of the databases used by CS; N4) interfere with the site survey phase; N5) modify the internals of their device or the operating system; N6) move objects in the environment to cause slow-fading in select directions at a distance; N7) perform beamforming to attack; N8) use multiple smartphones; N9) change their pattern once beginning their attack, as this would require them to re-measure their pattern after each change; N10) we must also make the assumption that if two models have the same pattern but are forced to run different software (i.e. the devices have the same hardware but for whatever reason one model cannot use the latest security updates), then they are both placed on the banned list.

5. CHALLENGES

Several challenges make the process of fingerprinting using radiation patterns difficult to perform in practice:

Measuring directivity at a distance. The core of this approach is being able to measure the directivity of a transmitter at a distance. This is not a value that can be measured on its own. Instead, it must be calculated by comparing the RSS measurements to samples measured for a *reference transmitter* with a known directivity. This is discussed in more detail in Section 6.

Resolving low resolution data. Ideally, RSS measurements would be taken using a huge number of receivers (i.e., hundreds) from different directions. This would give extensive coverage of the radiation pattern, which would give many points on the pattern to match and potentially, a machine learning approach could be taken to classify the devices. In reality, there would be a much smaller number of receivers, and with a small number of receivers, a close match for several reference patterns may be found for some orientation. Therefore, measurements need to be made for many packets over time as the device moves through the environment. This should then expose different parts of the radiation pattern to the receivers and give larger pattern coverage.

Temporary slow-fading and fast-fading. Objects moving around the environment may have a temporary effect on the received power from particular locations. By taking measurements from many packets, any temporary objects causing slow-fading will only temporarily affect a small number of measurements. Using many packets will also reduce the impact of fast-fading.

6. Model Fingerprinting Method

We will initially discuss the process at a high level to give an intuition of the process before going deeper into the details of each step.

Suppose the directivity is known for some directions from a device. In that case, this is matched to a known pattern by finding the pattern and orientation that has the smallest error to determine the model of the device, as shown in Figure 1. In this example, with the 4 data points, there is only one possible pattern it matches. However, calculating the directivity values from RSS measurements is the more complex part of the process. A set of receivers are used to measure RSS samples from a series of packets to achieve this. The RSS value measured by each receiver is a combination of many factors, including the device's directivity in the direction towards the receiver. It is impossible to calculate the directivity from a single measured RSS value as the device could be at any range, orientation, transmission power, and have any pattern shape. However, by combining multiple RSS measurements from multiple receivers, different candidate positions, orientations, and pattern shapes can be tested to minimise the RSS error from the expected values to estimate the most likely values for position and orientation for each candidate pattern. With an estimate for these values, there is enough information to estimate the directivity in the direction from the transmitter to each receiver by comparing the measurements to measurements previously collection from a reference transmitter with known directivity in the same environment. This is discussed further in Section 6.1.

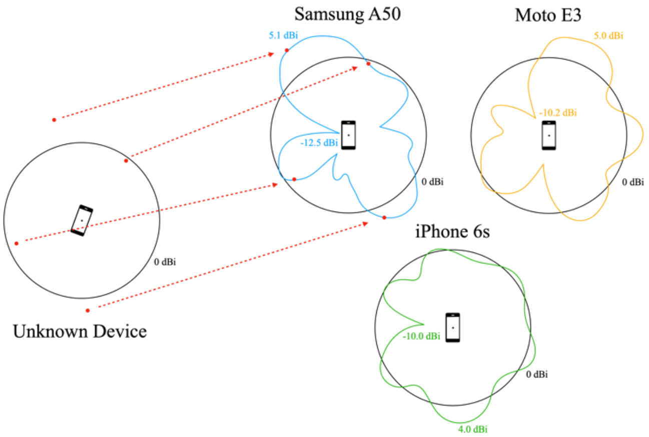


Figure 1. Diagram showing directivity measurements from an unknown device being matched to a known pattern. The patterns are shown in 2D relative to a black circle of 0dBi, representing a theoretical isotropic radiator. The largest and smallest directivity are also labelled.

To rebuild the shape of a pattern, RSS samples need to be collected for multiple packets as the device moves to increase the pattern coverage. The device is classified as the pattern with the lowest overall error. A series of packets and substantial pattern coverage is required to increase accuracy by reducing the impact of fast-fading and temporary slow-fading, ensuring important pattern artefacts are not missing from the inferred pattern, and increasing the difficulty of attacks. Therefore, classification will become more accurate with larger numbers of packets. The confidence of the classification result can be quantified by measuring the pattern coverage of all the possible patterns, and this coverage can be represented as a coverage metric. Classification is discussed further in Section 6.2.

Before identification can be performed, there are some preliminary tasks that must be conducted: Firstly, the pattern of all the devices one may wish to identify must be enrolled (This first part of this step can be skipped if the operator has access to a public database of patterns). These patterns are called the *reference patterns* and are measured by rotating the devices around two axes at set increments at a fixed distance from a receiver to collect samples at all directions from the device. This measures the signal strength at different azimuths and elevations. The receiver must be in the far-field region of the transmitter to ensure the pattern has fully formed. During this phase, the *reference transmitter* pattern must also be enrolled. We will explain the purpose of the *reference transmitter* shortly. Although it has an omnidirectional pattern measuring the reference transmitter's pattern is an important calibration step because the directivity is only constant in one plane. The directivity is calculated using the measured RSS samples to create the pattern. Equation (1) is used to calculate the directivity. To calculate $\mathrm{P_{rad}}$, a polynomial equation is fitted to the collected RSS data, and then the integration step is performed on this equation. We found that for 2D patterns, the number of coefficients required for a good fit varied between 20 and 45 depending on the pattern.

Secondly, a site survey must be performed using the *reference transmitter* that has a known pattern and orientation. RSS samples are collected from many points around the environment to form a power map in a similar way to many existing fingerprint-based localization schemes [18, 28]. The RSS map is built by placing the reference transmitter at different positions throughout the environment (the transmitter is placed at a fixed orientation α) and measuring the RSS at each receiver. For each location there is now a set of RSS values with 1 value for each receiver that is in range $\{\mathrm{P_{0_{ref}}}, \dots, \mathrm{P_{n_{ref}}}\}$.

Now that the preliminary tasks have been completed, we come to the actual identification. A device enters the area of operation $\mathcal{A}$ of the fingerprinting system. The aim is to classify it as one of the *reference patterns*. To match a device to a *reference pattern*, the directivity must be measured remotely and then the best matching *reference pattern* must be determined.

The number of receivers is variable and there is a trade-off of several factors including cost of hardware deployment, computational cost, number of packets required for accurate classification. At least n receivers are required to detect a packet for the calculations to proceed with that packet, with n being a configurable value. Enforcing a minimum n ensures that the number of receivers required to have a high classification accuracy is maintained. Depending on the receiver deployment, n may be smaller than the total number of receivers deployed. This is necessary if the area of operation is large and some areas would be out of range of some receivers. Throughout the remainder of the explanation, we use 4 receivers as an example and consider the 2-dimensional case for simplicity. For the 3D case, an extra component is added to any pattern access, and there are more possible orientations to consider.

The method is broken down into two parts: firstly, for every packet that is received, the RSS data measured by the receivers is processed to calculate the directivity, this is explained in Section 6.1; and secondly, at any time classification may be performed to find the best match along with a coverage metric c, explained in Section 6.2.

6.1. Remotely Measure Directivity

RSS values are measured for a packet recorded by each receiver to give$\mathrm{M} = \{\mathrm{P}_0, \ldots, \mathrm{P}_\mathrm{n}\}$. If less than n receivers receive the packet, the packet is dropped, but the packet still counts towards the number of packets used to calculate c.

For each possible *reference pattern*–which we refer to as the candidate *reference patterns* as we do not yet know which reference pattern is correct–the location and orientation with the best match to the measured data is found using the method we now describe.

We model RSS as a combination of transmission power (T_px), the directivity of the transmitter (D_t), the directivity of the receiver, slow-fading caused by the environment, path loss, and the effects of fast-fading [10, 27]. The impact of fast-fading is reduced by repeating this process for many packets, so we do not include this in our model equations. We assume that the slow-fading effects of the environment, path loss, and directivity of the receiver are constant from the same transmission location. Therefore, we simplify this by combining these constant values into the single constant E.

$$RSS = T_{px} + D_t + E \tag{5}$$

(5) is rearranged to separate the constant E.

$$E = RSS - T_{px} - D_t \tag{6}$$

Using (6), we can now compare different transmitters at a single location as the E from (6) will be the equal for both the correct reference pattern and the reference transmitter if the location is correct.

The values for the *reference transmitter* and each candidate *reference pattern* are substituted in to (6) so there is a system of n equations, one equation for each receiver, in the form:

$$L_{n_{ref}} - T_{px_{ref}} - D_{ref} = M_n - T_{px} - D_t \tag{7}$$

On the left-hand side of (7) there are the values for the reference transmitter, and on the right-hand side, there are the values for the candidate reference pattern. The sides are both equal because if they are at the same location, the environmental effects are also the same. Looking at each variable of (7) individually: $\mathrm{L}_{\mathrm{n}_\mathrm{ref}}$ is the power received by the receiver n from the reference transmitter at location L, the reference transmitter transmission power $\mathrm{T}_{\mathrm{px}_\mathrm{ref}}$, the directivity of the reference transmitter in the direction of the receiver D_ref, M_n is the actual RSS measured from the device to be identified by the receiver n,T_px the transmission power of the device corresponding to the candidate reference pattern, and D_t the directivity of the candidate reference pattern in the direction of receiver n.

To find the values D_ref and D_t we must fetch $\mathrm{pattern}_\mathrm{ref}[\alpha']$ and $\mathrm{pattern}_\mathrm{candidate}[\phi']$ respectively which are the directivity values from the patterns measured in phase 1. To calculate relative angles α' and ϕ' from the position of the various entities and the rotation values α and ϕ of the transmitters, the dot product of the transmitters rotation matrix and vector defining the relative position of the receiver and transmitter is calculated. The relative angle is then calculated between the two entities, converting from Cartesian coordinates to polar coordinates. In 2D space, these several steps are shown in the following equation for ϕ':

$$\phi' = atan2\Big(\sin\phi \cdot (R_x - T_x) + \cos\phi \cdot \big(R_y - T_y\big), \cos\phi \cdot (R_x - T_x) - \sin\phi \cdot \big(R_y - T_y\big)\Big) \tag{8}$$

(8) is the equation also used to calculate α'. Calculating the relative angles is more complex for 3D environments and radiation patterns as we must consider the rotation of the transmitter around 3 axes. However, this can be achieved relatively easily using the same method by adding an z component to the position vector and two additional rotation matrices for rotation around all 3 axes.

Using equation (7), the minimum squared difference between the two sides based on the RSS samples must be found to identify the closest location and orientation for each reference pattern. From the set of possible locations and orientations, the minimum value is found to determine location L and orientation ϕ:

$$\underset{L\in sampleLocations,\phi\in[-\pi,\pi]}{\operatorname{argmin}} \left(\sum_{n=0}^{R}\left(\left(L_{n_{ref}} - T_{px_{ref}} - D_{ref}\right) - \left(M_n - T_{px} - D_t\right)\right)^2\right) \tag{9}$$

The values within the sum correspond to the values from (7). $\mathrm{M_n}$ is the actual measured RSS value for receiver n; $\mathrm{T_{px}}$ is the transmission power of the device corresponding to the reference pattern; $\mathrm{D_t}$ is the directivity from the reference pattern or more precisely $\mathrm{pattern_{candidate}}[\phi']$, where ϕ' is the relative angle to the receiver from the candidate location with a candidate rotation of ϕ; $\mathrm{L_{n_{ref}}}$ is the RSS value measured from the reference transmitter during the site survey by receiver n; $\mathrm{T_{px_{ref}}}$ is the transmission power of the reference transmitter; and $\mathrm{D_{ref}}$ is the directivity of the reference transmitter or more precisely $\mathrm{pattern_{ref}}[\alpha']$, where α' is the relative angle to the receiver when the site survey was performed.

Using the minimum value from (9), the location L and orientation ϕ of the best match has now been estimated for each reference pattern. For each reference pattern, the process is now reversed to estimate the *measured Directivity* in the direction of each receiver.

$$measuredDirectivity = M_n + T_{px_{ref}} + D_{ref} - T_{px} - L_{n_{ref}} \tag{10}$$

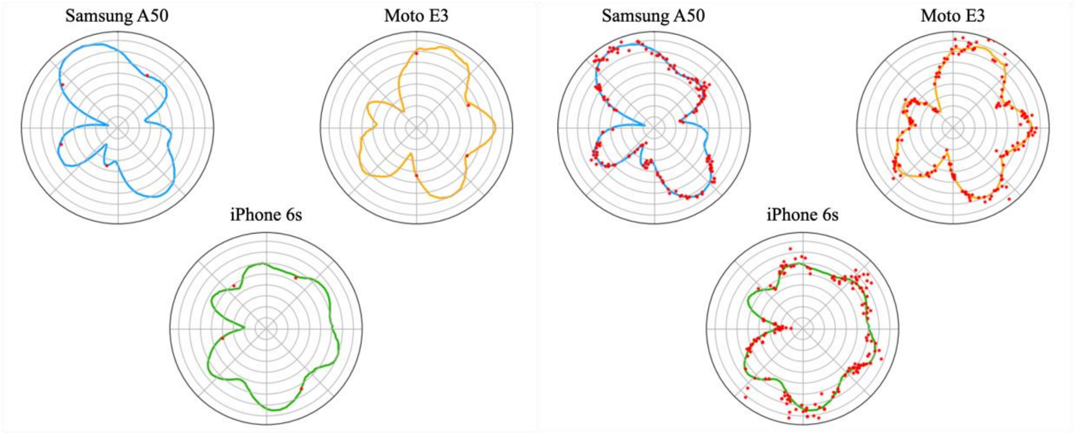
The (*measured Directivity*, *azimuth*) pair for each data point is calculated and stored for each reference pattern, where the azimuth is the direction to the receiver in the device's reference frame. If overlaid on the reference pattern database in 2D it may look something like Figure 2(a) with the red dots marking the *measuredDirectivity* values. In addition, the location and orientation can also be saved so the estimated movement can later be viewed for each possible pattern if desired. To be clear, the best matching location and orientation are found individually for each reference pattern, and therefore, the *measuredDirectivity* is calculated individually for each reference pattern. This can be seen in the Figure 2(a) example as the angles between the red dots on each pattern are different.

As previously stated, this process is performed on a series of packets to find the best reference pattern match later, so this process needs to be repeated for every received packet. As more packet samples are collected, the result will look more like Figure 2(b).

6.2. Classification

The identity is then resolved by finding the reference pattern with the closest match to the calculated *measuredDirectivity* data. This is performed by calculating the squared sum of the difference between the reference pattern directivity and the *measured Directivity* values. The reference pattern with the lowest average difference between the measured values and the actual pattern is the estimated device model.

To account for fast-fading and temporary slow-fading effects, this should only be performed when a sufficient number of measured points are taken. Ideally, the device should also have some movement during the collection time to expose more pattern segments to the receivers. As the location and orientation output correctness is not guaranteed, it is not possible to use these values to enforce movement, which is crucial for increasing the accuracy and, as discussed later on, preventing attacks on the system. However, if we look at each of the reference patterns and see that they all have significant coverage of the entire pattern, we can be confident that the transmitter has exposed a significant proportion of its pattern to various receivers. Therefore, a coverage metric c is used to give some indication of the confidence of the classification.



(a) (b)

Figure 2. Examples of reference patterns database overlaid with (a) first set of *measuredDirectivity* data calculated after 1 packet and (b) after 40 packets with movement of the transmitter. The coloured lines show the whole pattern, and the red dots mark the *measuredDirectivity* values. Measurements are shown on a dBi scale.

To calculate the coverage metric c for any instance of classification, each reference pattern is split up into equal wedges with an angle of Ω, as shown in Figure 3, and the smallest value ofΩ is found that ensures that there are at least $w\%$ of data points in every wedge for all the reference patterns–we refer to $w\%$ as the wedge requirement and it is a tuneable parameter. For ease of comparison, Ω is normalised to between 0 and 1 and inverted so that a higher value indicates higher coverage, with 1 showing there is complete coverage with the wedges as small as the resolution of the pattern and 0 indicating that the coverage requirement is only met if Ω is $360°$. It is important to note that if the operator is not careful, the coverage metric begins to break down past a certain point. For example, if the wedge requirement w is 1% then it is impossible to meet the wedge requirement if patterns are split into more than 100 wedges. The relationship between this coverage metric and classification accuracy is discussed further and evaluated in Section 7.3.

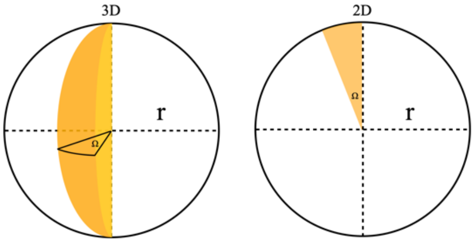


Figure 3. Radiation pattern sphere or circle sliced into wedges of Ω

The operator can use c in a variety of ways. For example, suppose a device cannot attain a threshold c after a set amount of time. In that case, the classification can be set to fail. Alternatively, the system can be configured only to accept classification results above the threshold.

6.3. Optimisations

This technique relies on an exhaustive search of the possible patterns, locations, and orientations, so optimisations are essential.

Firstly, the values for ϕ' and α' for each combination of location and orientation can be pre-calculated, so no matrix or trigonometric functions need to be computed at run time.

Secondly, a pattern can be stored in the form of a polynomial equation or as an array. Although storing a pattern as an equation saves some storage space, it comes at a high computational cost. Fetching a single directivity value requires calculating many exponents for the equation versus fetching a single array value. Storing patterns as arrays has a far lower computational cost, and the directivity value can be fetched using the azimuth and elevation as indexes.

Thirdly, during classification, the number of records to be stored for each device to be identified is approximately $nReceivers \times nPackets \times nReferencePatterns$.

With a huge number of reference patterns and/or devices to be identified, the data to be stored may grow to an unmanageable amount. To resolve this, once a storage size limit set by the operator is reached, the data from the bottom half of the matching patterns are deleted, and those reference patterns are no longer considered. Not only does this reduce the storage requirements, but it also reduces the computational cost of checking each reference pattern.

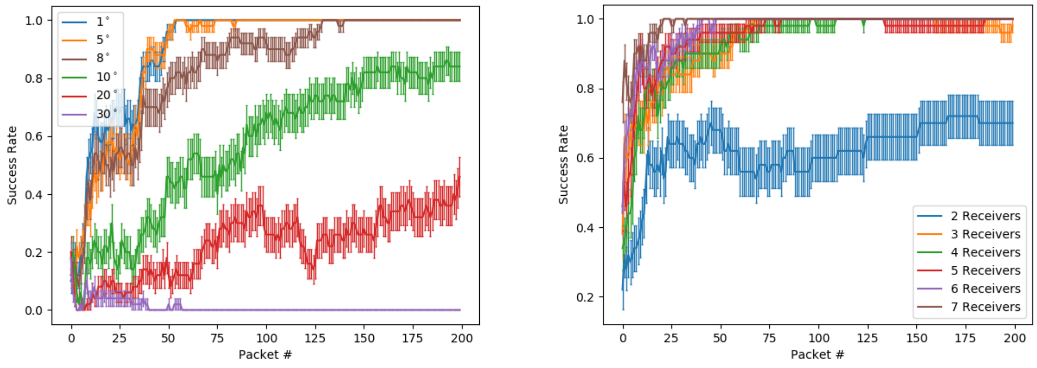
Fourthly, further processing time optimisations can be performed to reduce the number of locations to check, which may be expensive for large numbers of locations from the site survey and reference patterns. The computational requirements can be reduced in two ways 1) for the first i samples, the rough area can be calculated using a non-linear least squares optimiser for which it is assumed all patterns are isotropic and only the region with a high probability of being the locations with this less accurate method is searched thoroughly; and 2) after the first i samples a movement speed constraint can be added to prevent unnecessary checking of some locations that would be impossible to reach (e.g., locations more than 10m/s distance away from the previous packet can be discounted in the case of limited human movement).

7. EVALUATION

We evaluate our proposed method using radiation patterns collected from actual smartphones–we discuss the pattern collection further in Section 8–and simulations to explore the relationship between classification accuracy and several factors, including the number of packets, pattern fidelity, number of receivers, and our coverage metric.

For each of our simulations, we simulated 20m by 20m environments with receivers placed around the perimeter. We generated the simulated site survey data with 1m spacings between each survey point. A device to identify was then moved through the environment, and RSS samples were calculated, and random Gaussian noise with a standard deviation of 1.5dBm was added. The classification was then performed, and the results were outputted. 7 reference patterns collected from real devices as is discussed in Section 8.4 were used, this included the 2 modified patterns from the iPhone 6S, and only 1 pattern was used when there were multiple devices of the same make and model.

For the initial simulations, a device to simulate is selected from the list of reference patterns, and the success rate of classification is measured. Each simulation consisted of 50 runs, and the output plots we present show the mean result for the 50 runs with binomial proportion 95% confidence intervals marked for the success rate data. For each simulation, 200 sets of simulated RSS samples were used. Given that this is used for identity verification, we also tested it as a binary classifier and plotted the true positive rate (TPR) and false positive rate (FPR) on a receiver operating characteristic (ROC) curve for various coverage metrics c. Again, this used simulations with 200 sets of RSS data, but 420 runs of the simulation were performed.



(a) (b)

Figure 4. Success rate of classification using different (a) numbers of receivers and (b) pattern resolutions (i.e., the angle step between measurements) over 50 simulation runs with 95% confidence intervals marked.

7.1. Pattern Fidelity

The resolution of the pattern measurements impacts classification accuracy, as with more measurement points of the pattern, the pattern is of higher fidelity. With a lower pattern resolution, essential portions of patterns may be missed. This has the impact of significantly reducing the accuracy of classification. To demonstrate this, we performed simulations using various resolutions. We removed data between resolution steps and rounded the azimuth calculations to the nearest step to create the different fidelity patterns. The results in figure 4(a) show a significant increase in the number of packets required for accurate classification followed by degradation in classification accuracy as the resolution was reduced.

7.2. Number of Receivers

To demonstrate the relationship between the number of packets and the success rate of classification for different numbers of receivers, we varied the number of receivers in our simulations. The results are shown in figure 4(b) and–as expected–with 2 receivers, the success rate is significantly lower than with a larger number of receivers. For 3 receivers and above, the graph also shows that more receivers require fewer packets to achieve a higher success rate.

7.3. Coverage Metric

As already stated, exposing a greater proportion of the pattern is essential for increasing accuracy. However, there is no guarantee of movement or rotation, which is how greater pattern coverage is achieved. The coverage metric c is calculated for this reason, which is an indication of how much of the transmitter's pattern has been covered by the directivity measurements. Ideally, a high c would imply a high classification accuracy. That way, after any classification instance, one would have an estimated device identity and an indication of the likelihood of this being correct.

To demonstrate this increase in classification accuracy, simulations were performed. In each of the simulations, after every packet, classification was performed. This allowed the TPR and FPR to be recorded along with the coverage metric of that classification instance. The results of this are shown in Figure 5. The results demonstrate that as the coverage metric increases, the performance of the classifier improves. From this, we can conclude that with a higher coverage metric, we can be more confident in the accuracy of the classification.

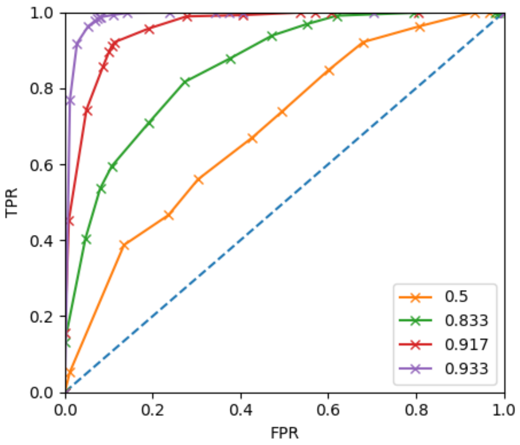


Figure 5. ROC curve of the classifier using a wedge requirement of 1%. Each line shows the curve for different output coverage metrics c.

8. Smartphone Pattern Uniqueness

To use WiFi radiation patterns for device type fingerprinting the patterns of each type must have the properties of 1) uniqueness and 2) repeatability. To explore these two properties in smartphone patterns we measured the pattern around 1 axis for 6 different smartphones.

8.1. Data Collection Framework

We used Raspberry Pi Model 4 boards running Ubuntu 19.10 (GNU/Linux 5.3.0-1022-raspi2 aarch64) as the platform for data collection. A USB external WiFi adapter with the Mediatek 7601u chipset placed into monitor mode was used for collecting wireless packet information, including sender MAC address and RSS. A separate Raspberry Pi controls a stepper motor that was used to rotate the phone 1° per second at a distance of 4m from the receiver. The transmitting device being measured is configured to transmit a packet approximately every 100ms. The Raspberry Pi clocks were synchronised to within 500ms and each packet received is matched to a rotation value by closest time. Therefore, the difference between the actual rotation value when the packet is transmitted and the measured rotation value should be ±1° providing a pattern resolution of 1°.To prevent interference from other devices a wireless router was used to setup a network using a channel that did not overlap with any other surrounding networks. The MAC address of the device being measured was used to filter out packets from devices related to the data collection setup and rogue devices (e.g. probe requests).

The pattern in the horizontal plane with the device flat on its back was measured for 6 devices: 1) a Samsung Galaxy A50, 2) an iPhone X, 3) an iPhone 6S, 4) an iPhone 5S, 5) a Motorola Moto E3, and 6) another Motorola Moto E3.

8.2. Pattern Correlation

Firstly, we found that repeat pattern measurements collected on different days from a single device visually match and have a high linear correlation, as shown in Figure 6. Secondly, we found that devices of different models have a lower correlation. Each pattern was overlaid with one another and one was rotated in 1 degree increments while the normalised cross-correlation was calculated for each degree of rotation and the rotation that yielded the highest cross-correlation was taken as the best match. Figure 7 shows the maximum normalised cross-correlation between different pattern measurements of different devices. When multiple measurements of the pattern are made, the average maximum normalised cross-correlation of matching devices was 0.91. The average maximum for non-matching devices was 0.54.

This demonstrates that there is pattern uniqueness between the devices we measured and repeatability across devices of the same make and model.

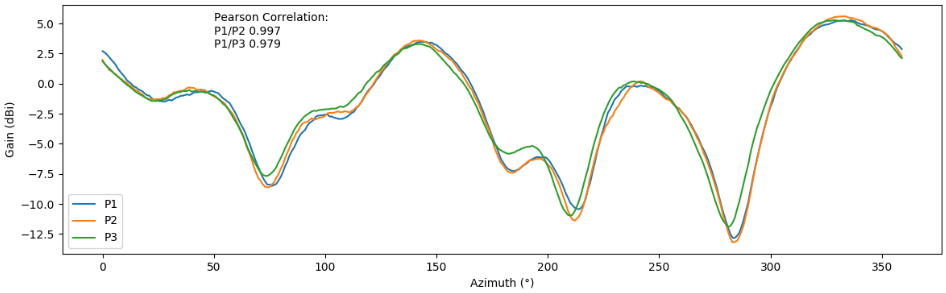


Figure 6. The diagram shows repeated measurements of the radiation pattern from the same Samsung Galaxy A50. This demonstrates that repeated pattern measurements result in the same pattern.

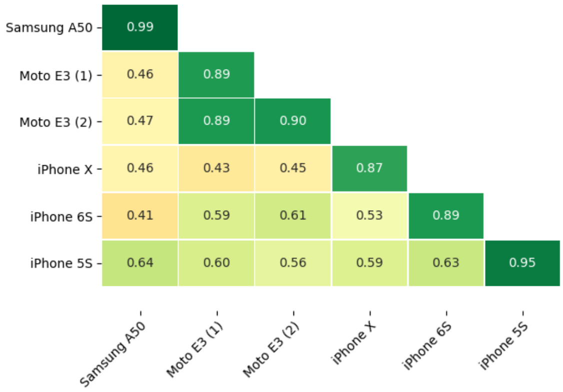


Figure 7. Normalised cross-correlation of smartphone radiation patterns. Each box gives an average of the cross-correlation when comparing the corresponding device on the y-axis with the device on the x-axis. Each device had its pattern measured at different times and was compared to every other device. The boxes on the diagonal consist of 3 averaged values as we do not compare a measurement run to itself. All the other boxes are the average of 9 values (3x3 runs).

8.3. Patterns from the Same Model

Coca [7] discussed the potential for devices to have their own unique patterns when compared to devices of the same model. However, in our experiments, conducted in an office environment, the patterns for the two Moto E3 devices did not show a significant difference. The level of difference was very similar to that of the repeat measurements for the iPhone X and 6S, which is likely caused by the effects of fast fading.

While our experiments for comparing two devices of the same model were limited, in that we only used two devices, the result shows that it is not always possible to uniquely distinguish devices of the same model based on the radiation pattern. However, it is possible that some models will exhibit larger differences across devices depending on other factors, such as manufacturing techniques and device construction. It is reasonable to expect devices that use an antenna that is part of an integrated circuit board or clamped to the chassis or other internal components would have very little manufacturing variations across devices. But in cases where a flexible wire that is not held into position is used, it is possible that patterns may vary across devices. The Samsung Galaxy A50, for example, uses a wire antenna for WiFi which is somewhat flexible as it is not clamped to the frame or circuitry.

8.4. Pattern Modifications

It is possible to modify a pattern using beamforming or by adjusting the device somehow, for example, adding additional material that will affect the energy distribution from the device. Only a tiny subset of current generation smartphones perform beamforming, so we do not consider this further.

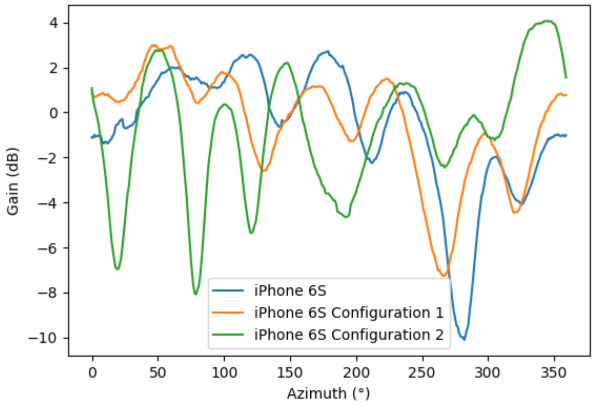


Figure 8. Pattern plots of iPhone 6S. The first without any blockers. The second and third have different configurations of blockers made from foil.

To explore how patterns can be modified we placed multiple layers of aluminium foil in two different configurations around the bottom end of an iPhone 6S. Figure 8 shows the pattern of the phone with different foil blocker configurations versus the unmodified pattern. Due to the complexity of modelling multipath fading effects, we would not expect the result to be as simple as reducing the gain in the direction of the blocker and an increasing gain proportionally in other directions. While noting we do not have the full picture, as the setup only measures power in one axis of rotation, as expected the blockers had some unpredictable effects on the pattern. Although at 180º–which was in the direction of the blocker–the gain was reduced, in other directions the increase is less clear. From 270º to 360º the results show an increase in gain with the blockers, but not by the same ratio and the alignment of peaks and troughs has changed. The other parts of the pattern have changed in an unpredictable way.

8.5. The Environment and Limitations of this Method

As with all schemes that utilise RSS, the environment plays a significant role and although we assume that the environment remains static using a constant E value, this will be affected by moving objects, such as people, doors, and furniture. As already mentioned, temporary changes are dealt with as the process is performed for many packets so the environment should return to its original state for some packets. In the case of long duration changes the site survey can be conducted with multiple states (e.g., with a door open), but the survey cannot be conducted with every possible state due to the state space explosion problem. However, multiple state surveys should be conducted when appropriate. Importantly, the movement of objects will have a limited impact to the RSS measurements in comparison to the pattern of the device which can cause differences of 14+ dBm.The limitation of this work is that in some settings, this assumption that the environment will return to the original site survey state may not apply.

9. Security Analysis

The adversary aims to break the fingerprinting mechanism, such that the system believes a smartphone that is not in $\mathcal{L}$ is a device in $\mathcal{L}$.Although the scheme produces quite a good location and orientation estimate as a by-product, we make no guarantees about those, so this is not considered in this analysis. For an attack to succeed, the attacker must find or produce a pattern that matches n points on a pattern from an allowed model where n is equal to the number of receivers. The angles between the matches in the pattern must correspond to a location in the environment where the angles to the receivers can be replicated. This must be done at multiple orientations and/or locations so the calculated points meet the pattern coverage requirement for each reference pattern.

9.1. Misclassification Attack

In the simplest form of attack, an attacker would try to orient himself such that his transmissions would be mistaken for transmissions from a valid device. To do this, the attacker must find patterns with similarities and exploit those similarities. Although devices have significantly different patterns, as shown in Figure 7, there are overlaps in the patterns. These overlaps can be found easily, as shown in Figure 9, to make a pattern indistinguishable from another. The number of points of overlap between two patterns is the upper bound for the number of receivers that an attacker can trick the system into thinking one pattern is the other with a single packet. For example, the Samsung A50 and iPhone 6S have up to 12 points of overlap in 2D; in 3D, there would be lines of overlap with the full pattern. The attacker may need to consider the difference in transmission power of the devices, but this would have a limited change on the number of overlaps. Using the total number of points of overlap is the best-case scenario for the attacker and requires the reasonably strong assumption that the attacker has complete control over the angles from the transmitter to the receivers. In practice, an attacker does not control the position of the receivers as these are fixed and thus has limited control over the required angle between matching points on the pattern.

A more realistic approach to consider the attacker's capabilities is to present the attacker with a range of environments with different layouts of receivers and determine if the attacker can create a match with a different pattern at multiple locations that satisfies the required angles to the receivers, with some flexibility (e.g. $\pm$1dBm). As the number of receivers increases, the more difficult it is for an attacker to find patterns that match at the required angles. As expected, for a simulated $\mathcal{A}$ layout, we found that for 2 receivers, 539 combinations of positions and orientations matched for the Samsung A50 and Moto E3, 21 matches for 3 receivers, and 4 & 5 receivers resulted in 0 matches. The number of matches will vary depending on the layout and patterns used. However, critically, this type of attack can be effectively mitigated, as the overlaps can be pre-calculated before system deployment. The receivers can be positioned to prevent possible overlaps of patterns that would otherwise make the scheme susceptible to attacks. First, the angles between matches for allowed and banned patterns need to be calculated. Then a layout of receivers needs to be found that prevents any location within the environment from achieving those angles to the receivers. As more devices are added to the pattern database, the calculations may need to be rerun and receivers potentially moved.

Additionally, although the attacker can match an allowed device, the weakness with this attack is that the system will still recognise the banned device as a strong match. If the system returned the incorrect but allowed device on the first attack attempt, on subsequent checks, it may not. In addition, by forcing the attacker to achieve a high pattern coverage, they must find multiple combinations of points of overlap at the required angles, and the probability of remaining indistinguishable from another pattern reduces as each wedge is covered. The situation is even more hopeless for the attacker if the location of some, or all, of the receivers is unknown.

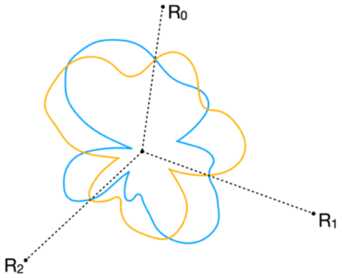


Figure 9. Example of an attacker finding matches that line up with 3 receivers. For a fixed transmission power, the overlap of directivity values will produce the same RSS values at the receivers, making the patterns indistinguishable from the perspective of the receivers.

9.2. Pattern Modification Attack

As described above, carefully crafting a situation where a device is misclassified is very difficult. Even if the attacker happens to succeed once, the attacker's actual pattern is still a strong match. To improve his chances, the attacker could modify his pattern to match an allowed pattern closely. However, as shown in Section 8.4, modifying a pattern by adding external blockers in a controlled way is difficult. This makes modifying a banned pattern to closely match an allowed pattern improbable.

An alternative way for the attacker to improve his chances is, somewhat counterintuitively, to modify the pattern, so that is it very different from any other pattern present in the database. Creating a new unique pattern with some overlap with existing patterns is not difficult. Once the attacker has created a new pattern, he can carry out attacks using the same methods discussed in the first attack. The greater the number of significant peaks and troughs, the easier it is for the attacker to maximise overlaps.

Noting what has already been discussed regarding matching patterns, the probability of successful attacks is reduced if the attacker is forced to find multiple matches at different points on the pattern–the coverage requirement introduced by the coverage metric forces this. Without very similar patterns, the attacker will not find many parts of the pattern that match at the required angles as dictated by the receivers’ positions. The number of receivers n and the threshold coverage metric c the operator requires can be adjusted to prevent this attack because the attacker is subject to the following constraints: First, the pattern generated by the attacker must match an allowed device on at least n points simultaneously; Second, the attacker is forced to move to multiple locations or use multiple orientations due to the coverage requirement, i.e., they must find multiple matches. The larger the required c, the more matches they must find.

10.Conclusion

In this paper, we have proposed a new method for fingerprinting different models of smartphones using differences in their radiation patterns. We showed how a small set of stationary receivers could measure the radiation pattern of a transmitting device in a completely passive manner. Our novel measurement and pattern recreation method can obtain the radiation pattern of nearby devices and compare them to a database of known device fingerprints. Based on this, the presence of rogue devices can be detected without requiring any special-purpose hardware or collaboration from the devices themselves. Finally, we discussed how the proposed scheme can be configured by adjusting the number of receivers, coverage requirements, and the number of packets between identification checks to mitigate different forms of attacks on the scheme. There is further scope for future work to explore how different environments affect the performance of the system.

REFERENCES

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, UNITED STATES: John Wiley & Sons, Incorporated, 2016. [Online]. Available: http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=4205879

[2] G. Baldini and G. Steri, "A Survey of Techniques for the Identification of Mobile Phones Using the Physical Fingerprints of the Built-In Components," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1761–1789, 2017, conference Name: IEEE Communications Surveys Tutorials.

[3] C. Bertoncini, K. Rudd, B. Nousain, and M. Hinders, "Wavelet Fingerprinting of Radio-Frequency Identification (RFID) Tags," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 12, pp. 4843–4850, Dec. 2012, conference Name: IEEE Transactions on Industrial Electronics.

[4] S. Bratus, C. Cornelius, D. Kotz, and D. Peebles, "Active behavioral fingerprinting of wireless devices," in *Proceedings of the first ACM conference on Wireless network security*, ser. WiSec '08. New York, NY, USA: Association for Computing Machinery, Mar. 2008, pp. 56–61. [Online]. Available: https://doi.org/10.1145/1352533.1352543

[5] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless Device Identification with Radiometric Signatures," in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, ser. MobiCom '08. New York, NY, USA: ACM, 2008, pp. 116–127, event-place: San Francisco, California, USA. [Online]. Available: http://doi.acm.org/10.1145/1409944.1409959

[6] Y. Chen and J. Yang, "Chapter 8 - Defending Against Identity-Based Attacks in Wireless Networks," in *Handbook on Securing Cyber-Physical Critical Infrastructure*, S. K. Das, K. Kant, and N. Zhang, Eds. Boston: Morgan Kaufmann, Jan. 2012, pp. 191–222. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978012415815300008X

[7] E. Coca and P. Valentin, "Antenna Radiation Pattern Influence on the Localization Accuracy in Wireless Sensor Networks," *Advances in Electrical and Computer Engineering*, vol. 13, pp. 43–46, May 2013.

[8] D. B. Faria and D. R. Cheriton, "Detecting identity-based attacks in wireless networks using signalprints," in *Proceedings of the 5th ACM workshop on wireless security*, ser. WiSe '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 43–52, event-place: Los Angeles, California tex.numpages: 10. [Online]. Available: https://doi.org/10.1145/1161289.1161298

[9] J. François, H. Abdelnur, R. State, and O. Festor, "PTF: Passive Temporal Fingerprinting," in *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*, May 2011, pp. 289–296, iSSN: 1573-0077.

[10] H. T. Friis, "A Note on a Simple Transmission Formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[11] X. Guo, Z. Zhang, and J. Chang, "Survey of Mobile Device Authentication Methods Based on RF Fingerprint," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 1–6.

[12] J. Hall, M. Barbeau, and E. Kranakis, "Detection of Transient in Radio Frequency Finger printing using Signal Phase," *Wireless and Optical Communications*, p. 6, 2003.

[13] J. Hua, H. Sun, Z. Shen, Z.Qian, and S. Zhong, "Accurate and Efficient Wireless Device Fingerprinting Using Channel State Information," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Apr. 2018, pp. 1700–1708.

[14] IEEE, "IEEE Standard for Definitions of Terms for Antennas," *IEEE Std 145-2013 (Revision of IEEE Std 145-1993)*, pp. 1–50, Mar. 2014, conference Name: IEEE Std 145-2013 (Revision of IEEE Std 145-1993).

[15] Ke Gao, C. Corbett, and R. Beyah, "A passive approach to wireless device fingerprinting," in *2010 IEEE/IFIP International Conference on Dependable Systems Networks (DSN)*, Jun. 2010, pp. 383–392, iSSN: 2158-3927.

[16] F. J. Liu, X. Wang, and S. L. Primak, "A two dimensional quantization algorithm for CIR-based physical layer authentication," in *2013 IEEE International Conference on Communications (ICC)*, Jun. 2013, pp. 4724–4728, iSSN: 1938-1883.

[17] F.J. Liu, Xianbin Wang, and H. Tang, “Robust physical layer authentication using inherent properties of channel impulse response,” in *2011 - MILCOM 2011 Military Communications Conference*, Nov. 2011, pp. 538–542, iSSN: 2155-7586.

[18] H. Liu, H. Darabi, P. Banerjee, and J. Liu, “Survey of Wireless Indoor Positioning Techniques and Systems,” *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[20] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, “Deep Learning for RF Device Fingerprinting in Cognitive Communication Networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 160–167, Feb. 2018, conference Name: IEEE Journal of Selected Topics in Signal Processing.

[21] M. K. Mwila, K. Djouani, and A. Kurien, “The use of antenna radiation pattern in node localisation algorithms for wireless sensor networks,” in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug. 2014, pp. 856–862, iSSN: 2376-6506.

[22] C. Neumann, O. Heen, and S. Onno, “An Empirical Study of Passive 802.11 Device Fingerprinting,” in *2012 32nd International Conference on Distributed Computing Systems Workshops*, Jun. 2012, pp. 593–602, iSSN: 2332-5666.

[23] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu, and Y. Yan, “Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme,” *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 349–360, Feb. 2019, conference Name: IEEE Internet of Things Journal.

[24] B. Perez, M. Musolesi, and G. Stringhini, “Fatal Attraction: Identifying Mobile Devices Through Electromagnetic Emissions,” in *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WiSec ’19. New York, NY, USA: ACM, 2019, pp. 163–173, event-place: Miami, Florida. [Online]. Available: http://doi.acm.org/10.1145/3317549.3319726

[25] B. W. Ramsey, B. E. Mullins, M. A. Temple, and M. R. Grimaila, “Wireless Intrusion Detection and Device Fingerprinting through Preamble Manipulation,” *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 585–596, Sep. 2015.

[26] K. B. Rasmussen and S. Capkun, “Implications of radio fingerprinting on the security of sensor networks,” in *2007 Third International Conference on Security and Privacy in Communications Networks and the Workshops - SecureComm 2007*, Sep. 2007, pp. 331–340.

[27] S. Y. Seidel and T. S. Rappaport, “914 MHz path loss prediction models for indoor wireless communications in multifloored buildings,” *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 2, pp. 207–217, 1992.

[28] P. Xiang, P. Ji, and D. Zhang, “Enhance RSS-Based Indoor Localization Accuracy by Leveraging Environmental Physical Features,” Jul. 2018, iSSN: 1530-8669 Library Catalog: www.hindawi.com Pages: e8956757 Publisher: Hindawi Volume: 2018. [Online]. Available: https://www.hindawi.com/journals/wcmc/2018/8956757/

[29] Q. Xu, R. Zheng, W. Saad, and Z. Han, “Device Fingerprinting in Wireless Networks: Challenges and Opportunities,” *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 94–104, 2016, conference Name: IEEE Communications Surveys Tutorials.

AUTHOR INDEX