

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

10th International Conference on Information Technology Convergence and Services
(ITCSE 2021), June 26~27, 2021, Sydney, Australia.



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: 978-1-925953-43-5

DOI: 10.5121/csit.2021.110901 - 10.5121/csit.2021.110918

This work is subject to copyright. All rights are reserved, whether whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the International Copyright Law and permission for use must always be obtained from Academy & Industry Research Collaboration Center. Violations are liable to prosecution under the International Copyright Law.

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

Preface

The 10th International Conference on Information Technology Convergence and Services (ITCSE 2021), June 26~27, 2021, Sydney, Australia, 10th International Conference on Digital Image Processing and Vision (ICDIPV 2021), 9th International Conference of Networks and Communications (NC 2021), 2nd International Conference on Cloud, Big Data and IoT (CBIoT 2021), 2nd International Conference on Artificial Intelligence and Machine Learning (CAIML 2021), 10th International Conference on Cryptography and Information Security (CRYPIS 2021), 10th International Conference on Advanced Computer Science and Information Technology (ICAIT 2021) and 2nd International conference on Natural Language Computing Advances (NLCA 2021) was collocated with 10th International Conference on Information Technology Convergence and Services (ITCSE 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 ITCSE 2021, ICDIPV 2021, NC 2021, CBIoT 2021, CAIML 2021, CRYPIS 2021, ICAIT 2021, NLCA 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, ITCSE 2021, ICDIPV 2021, NC 2021, CBIoT 2021, CAIML 2021, CRYPIS 2021, ICAIT 2021, NLCA 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 ITCSE 2021, ICDIPV 2021, NC 2021, CBIoT 2021, CAIML 2021, CRYPIS 2021, ICAIT 2021, NLCA 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

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Organization

Southeastern Louisiana University, USA
Wireilla Net Solutions, Australia

Program Committee Members

A. S. M. Sanwar Hosen,
Abdel-Badeeh M. Salem,
Abdellatif Bouzid-Daho,
Abhas kumar singh,
Abhay Kumar Agarwal,
Abiodun Odedoyin,
Abtoy Anouar,
Addisson Salazar,
Adrian Olaru,
Ahmed Farouk AbdelGawad,
Ajay Anil Gurjar,
AKhil Gupta,
Ali Ahmad Alawneh,
Ali Karkeh Abadi,
Alireza Valipour Baboli,
Amal azeroual,
Amando P. Singun Jr,
Amel Borgi,
Amel Ourici,
Anamika Ahrwar,
Anastasios Doulamis,
Ankur Singh Bist,
Aref Wazwaz,
Arti Jain,
Aymen Ben Said,
Azah Kamilah Muda,
Bektache Djamel,
Benyamin Ahmadnia,
Beshair Alsiddiq,
Bin Xue,
Bogdan Wiszniewski,
Boukari nassim,
Brahim Lejdel,
Cagdas Hakan Aladag,
Carla Osthoff,
Chau Kien Tsong,
Cheng Siong Chin,
Chuan-Ming Liu,
Ciprian Ion Rizescu,
Claude Tadonki,
Jeonbuk National University, South Korea
Ain Shams University, Egypt
University of Tizi ouzou, Algeria
Isroset, India
Kamla Nehru Institute of Technology, India
Obafemi Awolowo University, Nigeria
Abdelmalek Essaadi University, Morocco
Universitat Politècnica de València, Spain
University Politehnica of Bucharest, Romania
Zagazig University, Egypt
Sipna College of Engineering & Technology, India
Lovely Professional University, India
Philadelphia University, Jordan
University of Tehran, Iran
University Technical And Vocational, Iran
Mohammed V University, Morocco
Higher College of Technology, Muscat, Oman
Université de Tunis El Manar, Tunisia
University Badji Mokhtar of Annaba, Algeria
Jayoti Vidhyapeeth Women's University, India
National Technical University Of Athens, Greece
Advanced Technology, India
Dhofar University, Oman
Jaypee Institute of Information Technology (JIIT), India
University of Regina, Canada
Universiti Teknikal Malaysia Melaka, Malaysia
UBM Annaba University, Algeria
UC Davis, United States
Prince Sultan University, Saudi Arabia
National University of Defense Technology, China
Gdansk Tech, Poland
Skikda university, Algeria
University of El-Oued, Algeria
Hacettepe University, Turkey
National Laboratory for Scientific Computing, Brazil
Universiti Sains Malaysia, Malaysia
Newcastle University In Singapore, Singapore
National Taipei University of Technology, Taiwan
Politehnica University of Bucharest, Romania
Mines Paristech-Psl, France

Dadmehr Rahbari,	Tallinn University of Technology, Estonia
Dan Wan,	Hunan Normal University, China
Dario Ferreira,	University Of Beira Interior, Portugal
Deyu Lin,	Nanchang University, China
Ding Wang,	Nankai University, China
Domenico Rosaci,	University "Mediterranea" of Reggio Calabria, Italy
Domenico Rotondi,	FINCONS Spa, Italy
Dong-yuan Ge,	Guangxi University of Science and Technology, China
DV Ashoka,	JSS Academy of Technical Education, INDIA
El Murabet Amina,	Abdelmalek Essaâdi University, Morocco
Elios Krinidis,	Centre for Research & Technology Hellas, Greece
Elzbieta Macioszek,	Silesian University of Technology, Poland
Emerita Lynn Margaret Batten,	Deakin University, Australia
Eng Islam Atef,	Alexandria University, Egypt
Erdal Ozdogan,	Gazi University, Turkey
Fatma Susilawati Mohamad,	Universiti Sultan Zainal Abidin, Malaysia
Felix J. Garcia Clemente,	University of Murcia, Spain
Fereshteh Mohammadi,	Shiraz University, Iran
Fernando Zacarias Flores,	Universidad Autonoma de Puebla, Mexico
Francesco Zirilli,	Sapienza Universita Roma, Italy
Francis Ibikunle,	Landmark University, Nigeria
Fulvia Pennoni,	University of Milano-Bicocca, Italy
G. Rajkumar,	N.M.S.S.Vellaichamy Nadar College, India
Gál Zoltán,	University of Debrecen, Hungary
Grigorios N. Beligiannis,	University of Patras, Greece
Grzegorz Sierpinski,	Silesian University of Technology, Poland
Guilong Liu,	Beijing Language and Culture University, China
Guo Zhiqiang,	Wuhan University of Technology, China
Gururaj H L,	VVCE, India
H. S. Ramane,	Karnatak University, India
Hamid Ali Abed AL-Asadi,	Iraq University College, Iraq
Hamid Mcheick,	Université Du Québec À Chicoutimi, Québec, Canada
Hari Krishna Garg,	Nus, Singapore
Harikumar Rajaguru,	Bannari Amman Institute of Technology, India
Hassan,	II University, Morocco
Hesham F. A. Hamed,	EL-Minia University, Egypt
Hiroshi Ban,	osaka university, Japan
Hung Tran Cong,	Posts and Telecoms Institute of Technology, VietNam
Isa Maleki,	Islamic Azad University, Iran
Israa Shaker Tawfic,	Ministry of Science and Technology, Iraq
Ivan Izonin,	Lviv Polytechnic National University, Ukraine
Jabbar,	Vardhaman College of Engineering, India
Jagadeesh HS,	APS College of Engineering (VTU), India
Janusz T. Starczewski,	Czestochowa University of Technology, Poland
Jesuk Ko,	Universidad Mayor De San Andres (Umsa), Bolivia
Jian Wang,	China University of Petroleum (East China), China
Jian Xu,	Shandong Management University, China
José Luis Abellán Miguel,	Universidad Católica De Murcia, Spain
K Abhimanyu Kumar Patro,	National Institute of Technology Raipur, India
Kadir Sercan Bayram,	Hasan Kalyoncu University, Gaziantep, Turkey
Kassem Danach,	Islamic university of Lebanon, Lebanon
Katrina I. Sundus,	University of Jordan, Jordan

Kazi Shah Nawaz Ripon,	Østfold University College, Norway
Kazim Yildiz,	Marmara University, Turkey
Kazuyuki Matsumoto,	Tokushima University, Japan
Ke-Lin Du,	Concordia University, Canada
Khalid Nazim Abdul Sattar,	Majmaah University, Saudi Arabia
Khedija Arour,	University of Jeddah, KSA
Khurram Hameed,	Edith Cowan University, Australia
Kirtikumar Patel,	IE Engineer, USA
Klenilmar L. Dias,	Federal Institute of Amapa, Brazil
L Abed Ramzi,	University of Oran 1, Algeria
Labraoui Nabila,	University of Tlemcen, Algeria
Ljubomir Lazic,	Union University Belgrade, Serbia
Loc Nguyen,	Independent Scholar, Vietnam
Luis Gomez Deniz,	University of Las Palmas de Gran Canaria, Spain
Luisa Maria Arvide Cambra,	University of Almeria, Spain
M. A. Jabbar,	Vardhaman College of Engineering, India
M. Dolores Ruiz,	University of Granada, Spain
M. Zakaria Kurdi,	University of Lynchburg, Virginia, USA
Maad M. Mijwil,	Baghdad College of economic sciences university, Iraq
Mahboubeh Shamsi,	Qom University Of Technology, Iran
Mahdi Sabri,	Islamic Azad University Urmia Branch, Iran
Malka N. Halgamuge,	University of Melbourne, Australia
Mamoun Alazab,	Charles Darwin University, Australia
Manuel Gericota,	School of Engineering/Polytechnic of Porto, Portugal
Maria Ganzha,	Polish Academy of Sciences, Poland
Mario Versaci,	DICEAM - University Mediterranea, Italy
Maumita Bhattacharya,	Charles Sturt University, Australia
Md Azher Uddin,	Ajou University, South Korea
Meera Ramadas,	University college of Bahrain, Bahrain
Mehdi Gheisari,	Iau, Iran
Messaoud Rahim,	Yahia Farès University of Medea, Algeria
Michail Kalogiannakis,	University of Crete, Greece
Mihai Horia Zaharia,	"Gheorghe Asachi" Technical University, Romania
Mohamed Aridj,	Hassiba Benbouali University Chlef, Algeria
Mohamed Skander Daas,	Frères Mentouri Constantine 1 University, Algeria
Mohamed Yacoab,	The New College, Chennai, India
Mohammad Jafarabad,	Iran University of Science & Technology, Iran
Mohammed Aref Abdul Rasheed,	Dhofar University, Oman
Mohammed Bouhorma,	Abdelmalek Essaadi University, Morocco
Mostafa Safdari Shadloo,	National Institute of Applied Science (INSA), France
Muhammad Naufal Bin Mansor,	Universiti Malaysia Perlis, Malaysia
Muhammad Sajjadur Rahim,	University of Rajshahi, Bangladesh
Mu-Song Chen,	Da-Yeh University, Taiwan
Mussa Turdalyuly,	Satbayev University, Kazakhstan
Nadine Akkari,	Lebanese University, Lebanon
Nicolas Durand,	Aix-Marseille University, France
Oliver L. Iliev,	Fon University, Republic of Macedonia
Omid Mahdi Ebadati,	Kharazmi University, Tehran
Osman Toker,	Yıldız Technical University, Turkey
Otilia Manta,	Romanian American University, Romania
P. S. Hiremath,	Kle Technological University, India
P.Mariappan,	Bishop Heber College, India

Parameshachari,	GSSS Institute of Engineering and Technology, India
Patrono Luigi,	University of Salento, Italy
Pavel Loskot,	Swansea University, United Kingdom
Peiying Zhang,	China University of Petroleum, China
Phuoc Tran-Gia,	University of Wuerzburg, Germany
Ping Zhang,	Anhui Polytechnic University, China
Piotr Kulczycki,	Polish Academy of Sciences, Poland
Pranita Mahajan,	Mumbai University, India
pranita Mahajan,	SIESGST, India
Prasan Kumar Sahoo,	Chang Gung University, Taiwan
Przemyslaw Falkowski-Gilski,	Gdansk University of Technology, Poland
Rahmat Sulaiman,	ISB Aatma Luhur, Indonesia
Rahul Saha,	Lovely Professional University, India
Rajeev Kanth,	University of Turku, Finland
Rajkumar,	N.M.S.S.Vellaichamy Nadar College, India
Rajni Jindal,	Delhi Technological University, India
Ramadan Elaies,	University of Benghazi, Libya
Ramgopal Kashyap,	Amity University Chhattisgarh, India
Rao Li,	University of South Carolina Aiken, USA
Renuka Mohanraj,	Maharishi International University, USA
Richa Purohit,	D Y Patil International University, India
Robert Hsu,	National Chung Cheng University, Taiwan
Rohit Thanki,	Prognica Labs, UAE
Ruksar Fatima,	Khaja Bandanawaz University, India
Ruofei Shen,	AI researcher - Menlo Park, USA
S Saravana kumar,	CMR University, India
Sabre Rachid,	University Dijon, France
Sabyasachi Pramanik,	Haldia Institute of Technology, India
Sahil Verma,	Chandigarh University, India
Santanu Chatterjee,	Defence Research Development Organization, India
Segun Olatinwo,	University of Pretoria, South Africa
Seifedine Kadry,	American University of the Middle East, Kuwait
Seyed Abolfazl Shahzadeh fazeli,	Yazd University, Iran
Seyyed AmirReza Abedini,	Islamic Azad University, Iran
Shadi Abudalfa,	University College of Applied Sciences, Palestine
Shahram Babaie,	Islamic Azad University, Iran
Shamneesh Sharma,	Poornima University, India
Shashikant Patil,	SVKMs NMIMS, India
Shervan Fekri-Ershad,	Azad university of Iran, Iran
Shi Dong,	Zhoukou Normal University, China
Shin-Jer Yang,	Soochow University, Taiwan
Shruti Bhargava Choubey,	Sreenidhi Institute of Science & Technology, India
Shruti Suman,	K L University, India
Siarry Patrick,	Universite Paris-Est Creteil Val de Marne, France
Sikandar Ali,	China University of Petroleum- Beijing, China
Simanta Shekhar Sarmah,	Alpha Clinical Systems Inc, USA
Simon James Fong,	University of Macau, China
SivaKumar PV,	VNR VJIET, India
Smain Femmam,	UHA University, France
Sourav Sen,	Research Scientist, Upstart Network Inc, USA
Sourav Sen,	Upstart Network Inc., USA
Sridhar Iyer,	S.G. Balekundri Institute of Technology, India

Subhendu Kumar Pani,
Suhad Faisal Behadili,
Suphonsa Khetkeeree,
Tasher Ali Sheikh,
Thenmalar S,
Tran Cong Manh,
Tri Minh Tran,
U. Srinivasulu Reddy,
Ved Prakash Mishra,
Viranjay M. Srivastava,
Virupakshi Patil,
Wei lu,
Wenyuan Zhang,
William Simpson,
Xiaochun Cheng,
Yanrong Lu,
Yogendra Kumar Jain,
Yousfi Abdellah,
Youssef Taher,
Yu-Chen Hu,
Zakaria Laboudi,
Zamira Daw,
Zhang Ziwen,
Zhihan Lv,
Zoran Bojkovic,

Krupajal Computer Academy, India
University of Baghdad, Iraq
Mahanakorn University of Technology, Thailand
Madanapalle Institute of Technology & Science, India
SRM Institute of Science and Technology, India
Le Quy Don Technical University, Vietnam
Gunma University, Japan
National Institute of Technology, India
Amity University, Dubai
University of KwaZulu-Natal, South Africa
Sharnbasva University Kalaburagi, India
Early Warning Academy, China
Southeast University, China
Institute for Defense Analyses, USA
Middlesex University, UK
Civil Aviation University of China, China
Samrat Ashok Technological Institute, India
University Mohamed V, Rabat - Morocco
Mohammed V University, Morocco
Providence University, Taiwan
University of Oum El Bouaghi, Algeria
Raytheon Technologies Research Center, USA
Guangzhou Maritime University, China
Qingdao University, China
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)



Organized By



Academy & Industry Research Collaboration Center (AIRCC)

10th International Conference on Information Technology Convergence and Services (ITCSE 2021)

Voice Assistants with Artificial Intelligence for Improving Academic English...01-12
Vladimir Tregubov

10th International Conference on Digital Image Processing and Vision (ICDIPV 2021)

**Technique for Removing Unnecessary Superimposed Patterns
from Image using Generative Network.....**13-22
Kazutake Uehira and Hiroshi Unno

**Adaptive Filtering Remote Sensing Image Segmentation Network
based on Attention Mechanism.....**23-36
*Cong zhong Wu, Hao Dong, Xuan jie Lin, Han tong Jiang, Li quan Wang,
Xin zhi Liu and Wei kai Shi*

9th International Conference of Networks and Communications (NC 2021)

**Framework for Enterprise Local Area Network Design: an
Object-Connectivity Approach**37-45
Sunil Seneviratne and Rohan de Silva

**Global Systems Performance Analysis for Mobile Communications (GSM)
using Cellular Network CODECS.....**47-60
*Maphuthego Etu Maditsi, Thulani Phakathi, Francis Lugayizi and
Michael Esiefarienrhe*

2nd International Conference on Cloud, Big Data and IoT (CBIoT 2021)

Several Typical Paradigms of Industrial Big data Application61-68
Hu Shaolin, Zhang Qinghua, Su Naiquan and Li Xiwu

**An Intelligent System to Enhance Visually-Impaired Navigation and Disaster
Assistance using Geo-Based Positioning and Machine Learning**69-79
Wenhua Liang, Ishmael Rico and Yu Sun

**Adoption Factors of Enabling I4.0 Technologies and Benefits
in the Supply Chain**81-96
José Carlos Franceli and Silvia Novaes Zilber Turri

2nd International Conference on Artificial Intelligence and Machine Learning (CAIML 2021)

FindThatQuote: A Question-Answering Web-based System to Locate Quotes using Deep Learning and Natural-Language Processing97-104
Nathan Ji and Yu Sun

Enhancement of Consistent Depth Estimation for Monocular Videos Approach105-115
Mohamed N. Sweilam and Nikolay Tolstokulakov

An Intelligent Mobile App to Detect Drowsy Driving with Artificial Intelligence117-129
Thomas Xiao and Yu Sun

10th International Conference on Cryptography and Information Security (CRYPIS 2021)

Applications of SKREM-Like Symmetric Key Ciphers.....131-147
Mircea-Adrian Digulescu

Secure Cloud Key Management based on Robust Secret Sharing149-161
Ahmed Bentajer, Mustapha Hedabou, Sara Ennaama and Abderrahim Tahiri

Hiding Data in Plain Sight: Towards Provably Unbreakable Encryption with Short Secret Keys and One-Way Functions163-183
Mircea-Adrian Digulescu

Electromagnetic Analysis of an Ultra-Lightweight Cipher: PRESENT185-205
Nilupulee A. Gunathilake, Ahmed Al-Dubai, William J. Buchanan, Owen Lo

10th International Conference on Advanced Computer Science and Information Technology (ICAIT 2021)

Web Scraper Utilizes Google Street view Images to Power a University Tour207-219
Peiyuan Sun and Yu Sun

A Video Note Taking System to Make Online Video Learning Easier221-228
Haochen Han and Yu Sun

2nd International conference on Natural Language Computing Advances (NLCA 2021)

Rational Mobile Application to Detect Language and Compose Annotations: Notespeak App229-237
Yingzhi Ma and Yu Sun

VOICE ASSISTANTS WITH ARTIFICIAL INTELLIGENCE FOR IMPROVING ACADEMIC ENGLISH

Vladimir Tregubov

Department of Business and Logistics, Yuri Gagarin State
Technical University of Saratov, Saratov, Russia

ABSTRACT

The article describes applications of using voice recognition technology based on artificial intelligence to the educational process. The author presents a comparative analysis of existing examples artificial intelligence in the educational process. Artificial intelligence uses in specialized software it makes educational process more convenient for both the students and the teachers. There is a description of an application "Academic phrase bank" developed by author. The application consists of two specialising actions for Google assistant. The application allows to increase academic vocabulary, train of creating grammatically correct academic expressions, and memorize templates of academic phrases. In active mode, this application helps to create correct phrases of academic English and improve the abilities of understanding English speech.

KEYWORDS

Academic English, Google home, personal assistant, academic publications, applications for learning English.

1. INTRODUCTION

Active development of artificial intelligence (AI) leads to the fact that AI is widely spread in various fields, including education [1]. We use the term "artificial intelligence" in the narrow meaning as a specialized software system for implementation a natural user interface. AI allows learners to gain some advantages over traditional learning techniques []. An educational platform with AI can be adjusted to meet the needs of students and consider their personal features (initial level of knowledge, speed of learning, current interests) for identifying the most effective way of studying. The interaction students with a smart platform reveal the weaknesses of students and motivate them for additional studying, so it makes possible to create a system of adaptive learning [2].

There are some AI-based technologies popular in the educational process [3]: adaptive knowledge assessment, spaced repetition, virtual assistants, adaptive feedback, formal verification of creative works, etc. Let us briefly consider the features of these technologies and the mechanisms of their implementation with AI.

Computer-aided assessment with AI [4] implies using of specialized applications that allow to test, and automatic check various types of creative work, such as essays, etc. System with computer-aided assessment analyse answers of students, providing feedback, finding factual or

semantic errors. The data gathered from student' answers are using for preparation an individual learning path adapted to the current knowledge of students.

The spaced repetition technique based on the theory that a proper time interval of information representation allows maximizing the rate of learning [5]. AI technology in spaced repetition software applications allows recording and tracking progress, when the information was presented, how effective it was for a particular learner, and then to adapt the time intervals of the information representation when it is most appropriate for the learner.

Traditional educational system uses different types of feedback, such as checking and evaluation of student's work, interaction tests, different teacher's activities. Artificial intelligence makes it possible to create specialized chatbots, which can interact with the student through a dialogue interface, analyse their responses, and provide quality feedback to make easier the teacher's work. In the future, these chatbots will be transformed into virtual assistants and be able to respond to students' requests more accurately and quickly in the natural language [6].

The technology of virtual assistants (VAS) is a perspective variant of the AI implementation [7]. One of the first in the mid-nineties, Microsoft has released the office assistant "Clippie", a specialized virtual assistant, which should help users to effectively use Microsoft Office. This feature got many negative responses from users, so in later versions of Microsoft Office, this assistant was removed because it showed low efficiency and negative attitude of users.

The surge in popularity of virtual assistants has been observed over the last few years, and they are especially widely used in modern smartphones. Many software companies like Amazon, Google, Microsoft have released their own virtual assistants [8]. The growing interest to this topic related to dramatically improvement of natural speech recognition, which allows to implement an effective voice interface [9]. A significant increasing of the usefulness of virtual assistants also correlated with the development of the Internet of things (IoT).

Virtual assistants allow to manage IoT devices, organize and analyse information obtained from both virtual and physical sources. Modern personal assistants use AI extensively for automation tasks of everyday activities, such as creation of reminders, prompting required information in the current context, constructing routes, grammar checking, translating, etc.

There is a particular interest in using AI for adaptive learning [10]. Adaptive learning is an educational method of using modern computer algorithms to teach students in accordance with their individual characteristics. Adaptive learning is considered: the emotional state of a student, his ability to perceive different types of information, the current level of learning skills, gender, etc. Adaptation is that the curriculum customized to the student, try to build an individual learning trajectory.

In many countries, the adaptive learning based on AI is being actively implemented with supporting of the government. There are a number of countries where it is enshrined as a legal framework. In Brazil, for example, the Geekie system. This system allows pupils to prepare for the final examination at school. At the beginning of the training, a student takes some tests and selects the own learning goal, then the system with AI analyses the student's preferences, prepares individual learning trajectories for them, highlights the necessary and suitable content.

In learning, a specialized program captures every student's action and transmits information to the system. Each student can have a personal pace of learning and the system accumulates information about students and adapts the learning process to their abilities. The developer points out that the using of this system has increased results of the final exams for all students [11].

In Russia, some private companies are also trying to create an adaptive platform with AI elements. Stepik created a learning platform which generates educational materials based on the level of user knowledge, which selects the content suitable for a particular stage of learning. At present, this educational platform is positioned as a template for open online courses with adaptive recommendations. The platform allows users to create interactive lessons with feedback and verification of the teacher's tasks.

Thus, there are three main reasons why AI is being intensively embedded into the educational process:

- AI helps to make the educational process more effective and convenient for both students and teachers. In Russia, some large online schools create and implement training courses with AI.
- AI increases the involvement of students in the educational process through gamification. Most training simulators use artificial intelligence to create the effective game mechanics. For example, a language school SkyEng [12] integrated AI into the educational process, it provides adaptive and personalized learning, and performs task verification in real time. AI analyses each lesson, assesses student progress, and evaluates teacher performance. Such learning system allows to change a path of student teaching.
- implementation of AI into educational process provides automation of it. Modern educational resources can effectively function without a human: chatbots answer questions and conduct lessons. This trend is increasing every year.

2. VOICE USER INTERFACE

Natural User Interface (NUI) describes interaction between a person and a computer with intuitive actions that do not require special training of users [8]. The purpose of using a user-friendly interface is to hide the complexity of a system as much as possible. Even if the user is inexperienced or requires a rather complicated interaction, it will be able to interact with the system by NUI. Examples of interfaces with NUI include interfaces based on touch, gestures, body movements (Kinect), voice, etc. Neurocomputer interfaces are expected to be created soon.

In recent years, virtual assistants with voice control have become widespread in various fields. Virtual assistants have become commercial products thanks to Apple, which introduced the Siri assistant for iPhone in 2011. Virtual assistants use a Voice User Interface (VUI), which is a human-computer interface with voice input to control computers and devices. The VUI interface has a long history since the 1950s, when the first prototypes of devices with VUI appeared [13]. However, initially these interfaces were not highly effective, a significant breakthrough has been achieved in the last decade, thanks to the developing natural language recognition with AI. VUI have become more convenient and useful, and these technologies have gained wide popularity due to the spread of smart speakers (Amazon Echo, Google Home, Yandex.Station) and other devices with voice recognition.

There are a few significant advantages of voice input in comparison with using keyboard [13]:

- high speed of information input (faster than typing on keyboard).
- voice input can be used simultaneously with other activities (you can react without switching from the current context).
- intuitive interface.
- high level of empathy (voice messages have more empathy than text messages, interaction through voice can be more pleasant and convenient for a person).

VUI is becoming widely used as a communication for many various intelligent devices. For example, all modern smart speakers do not have a graphical display and keyboard, so you can communicate with them only by VUI. A virtual chatbot is a program that imitates the real interaction with a person. Interfaces of modern chatbots based on VUI. These types of interfaces are implemented like text chat in Facebook Messenger or Skype Instant Messaging for dialog with an intelligent device.

There are standard development tools for creation VUI. For example, a script describing the interaction between users and a script of interaction between a user and a computer have a remarkably similar structure. The only exception is that the chat-bot interface can display images and hyperlinks, while the VUI can only play sounds.

Using chat bots for learning a foreign language has the following advantages [14]:

- Gamification. Chatbots use gamification, they have built-in game mechanics to motivate students to continue learning. For example, learners receive points or other achievements for regular training or achieving a certain goal. Gamification turns learning into a hobby, it motivates users to study even during periods of rest.
- Individualization. Personal lessons with a teacher who will correct your grammatical or phonetic errors are quite expensive, and when the classes are in a group, you must wait your turn to work with the teacher. Lessons with a chat-bot removes these restrictions, you can get individual attention at any time and freely.
- Availability. Lessons with a chat-bot can carry out at any time. The chatbot instantly answers any your questions, and you can train for several hours. This significantly increases the intensity of class and involvement in the education process.

No evaluations from the teacher and no fear of errors. For many students, the language barrier is due to an internal fear of making a mistake in dialogues. Chatbot allows to overcome this fear, remove the language barrier, help a learner to start fluently communicating.

Below, there is a description of several chatbots which designed specifically for teaching English. We compared two main categories of chatbots that can be used for self-study of English. The first category is the chat bots, which are specially designed for learning English. The second category is the chat bots for practical tasks, with them we can train English in real situations, which is an effective way of learning a foreign language.

3. CHATBOTS FOR LANGUAGE LEARNING

Modern applications use AI for creating a language environment, assess the current language level, etc. Learning language with chatbots is faster than in traditional ways. With several test dialogues at beginning, the chat-bot identifies the strengths and weaknesses of student, and then configures the scripts of the dialog according to the student's language capabilities. Chatbots are adaptable, then more the learner performs the exercises and the more information collected by a chatbot, then the better it will form tasks for students. The AI allows to create individual tasks for students, make them own educational path, and then teach at a pace that is convenient for them.

Mondly [15] is a specialized program for English learning, which has two versions of the interface: an internet site and a mobile application for smartphones. The user can communicate with the Mondly by keyboard or by VOI through a microphone. The application uses a dialogue mode, when a user answering a question, he should choose the best option to answer or pronounce it aloud (the user also can type his answer in a special dialog box). If the chat-bot does not understand what the user told, it will ask him to repeat the answer. If there is a pause in the

conversation, the chatbot offers several different answers for continuous of the conversation. The user can also choose how to listen of the answers in a male or female voice. Developers carry out some experiments by using a virtual reality to immerse students into the language environment with chatbots. They created Mondly VR, a special application for learning English in virtual reality glasses.

Andy [15] was designed to improve conversation practice in English. This chat-bot is positioned as a virtual teacher and you can talk to it on various topics, play language games, build vocabulary, and learn grammar rules. The user can choose a topic of discussion and get a short explanation for the corresponding grammar rule. The application supports different levels of language skills (from Beginner to Advanced). If the users have questions during the training, they can receive additional information and feedback. At the end of each topic, the user has to test himself. If there were mistakes in the test, the chatbot explains how to amend them and show the correct variant, and for unfamiliar words shows their definitions and examples of using in context. The user can talk to the chatbot on a free topic. Communication can be both written and oral by user's choice, and the chatbot corrects errors.

Virtual assistants help to practice language skills in a real-life situation. Online shops and services (travel industry, catering, etc.) use chatbots for helping customers to shop goods or book a place.

Hipmunk was designed by a company that helps in a travel company. Hipmunk help to find information for travellers about hotels, flights, car rental, etc. The internet site offers a virtual assistant, and the user can get help from him through conversation on natural language. Conversation with the chatbot starts by saying "Hello Hipmunk", then the user can ask a question, for example, "What is the cheapest flight from Moscow to San Francisco in the first week of March?" and get a reasonable answer. The chatbot allows user to practice English questions. If Hipmunk understands your questions and can answer you, it means a sufficient level of your pronunciation.

Mona is a chatbot for helping users to make profitable purchases. Mona uses databases of large online stores and finds products that maximize the customer needs. User can ask the chatbot in a free form about what kind of good he is trying to find, what range of prices, and so on, then the chatbot will pick for user the best option. User can interact with this chat bot through a dialog. In this mode, Mona asks leading questions and provides additional information in reply messages in a form of a menu to select an option. This is a good way to improve your ability quickly reading and answering English questions. This type of communication allows for user to memorize words about household items, products in stores, terms related to fashion and clothing, home appliances, etc.

Android smartphones have Google Assistant as a virtual assistant that can use for daily English conversation practice. Google assistant has a natural user interface and can help to perform different tasks, for example, it can set an alarm clock at a given time by voice request. The user should be able to build different types of English sentences. If the user's pronunciation is wrong, the assistant will not execute the command. This is an effective way to improve pronunciation and train skills of making sentences without context or nonverbal clues.

Google Nest is a line of smart speakers created by Google. It has Google Assistant and allows for users to interact with VUI. By Google Nest users can receive answers from the Google search engine, launch applications, create reminders, etc. The devices enable users to implement of the Internet of things at home, the gadget can turn on and off devices, control lighting and TV, etc. Use Google Assistant for IoT, we can without Google Home device, for this you need to have a

smartphone with parameters (Android 5.0+ / iOS 10.0+). In Table 1, we summarize some applications for Google Assistant and describe their functionality.

Table 1. Applications for Google assistant

Name	Developer	Description
Stories	Google	This app is also known as "Tell the story". The user can listen to several classic and modern stories in English. There are many stories for children. The application improves listening skills, and it is an effective way to get access to a lot of low complexity English texts.
Creative Coach	Google	The application develops a creative potential of users. It can be used in various situations, for example, for generation ideas for photos and drawings. If you are going to improve written English, you can use the command "Ask written prompt". In response, the application will generate a theme and a few lines for a new story. This application is quite universal and suitable both for those who are already fluent in English and want to improve their creative skills, and for those who have just recently started learning the English language.
Dictionary	Google	A dictionary application for finding the meaning of a word. The app lets you ask Google Home "define [word]." or "what does [word] mean?" by voice. The answer will be a description or a translation of the word.
Fun Facts	Google	The application trains listening skills, allows user to listen to texts in English and memorize new words in context. The app tells some fun facts about different things. The fact descriptions are short and use simple language, so it is a good way for training English.
Mr Vocab	Vocab Assistant	The Vocab application is aimed at people who are preparing for the GRE. The application offers a few dictionaries games and allows checking your current vocabulary.
News	Google	Listening to news in English allows you to refill your vocabulary with everyday words, as well as to train your listening skills. This application offers various options, news sources (BBC, Financial Times, etc.).
Spell Check	Google	Sometimes you may find that you know how to say an English word... ..but you do not know how to spell it. The spell checker provided by Google lets you check the spelling of any word: just ask, "How do you spell [word]?" or say Google Home "Spell [word]". While this is no substitute for precise word learning, spell check is a great tool for helping you confidently learn new words.
Vocal Notes	Maildover LLC	User can use voice memos if he wants to dictate rather than write in English. The application can be useful for sending messages in English, or for improving your confidence in using both spoken and written English.

4. ALGORITHM OF APPLICATION CREATING

Academic English proficiency is an essential skill of a modern scientist and requires constant practice. Researchers need continuous language training for improvement their language competencies. To expand opportunities for self-development of academic English skills, we

developed application for Google assistant "Academic phrase bank trivia" and "Academic trivia". The first application helps to train the vocabulary of academic English. The second helps to develop skills of building grammatically correct sentences in academic English, memorize phrases of academic vocabulary, and train the correct pronunciation.

For the developing our application, we use the Google action console, which allows to create an application for Google Assistant. In the console we step by step create an application and make some configurations using a specialising online editor. After completing all steps, we will get a fully working application that can be tested on a computer simulator or on a smartphone and publishing in Google Assistant directory. Figure 1 demonstrates the algorithm of creating the application.

The first step of the algorithm is to collect the exercises that will be used to create our application. The collected information is checking for incorrectness, if some errors are detected, the initial data are correcting, and the first step is repeated. After the final correction, the data is entered into a special Goggle Sheet template. The next step is to configure the voice interface to run the application and build an invocation model. Then a specialized online environment is using for building of an action. This environment is also using for building of logic, responses and intent training phrases. At next step of development, the built application can be tested in a simulator. At the next step of development, the application can be tested on a smartphone. The last step of the algorithm involves create a realise and publishing the application to Google actions directory.

When we were creating the application, we used some rules of development. Questions for the application should be brief and clear. It should be considered that the question is displayed on the smartphone screen, and there is interaction only by voice, so it is difficult for a player to remember a long question, especially in a foreign language. Therefore, questions and answers should be as short as possible. In the guide, Google recommends that the answers contain only one or two words [16].

If there are many correct answers to a question, you can specify them using the "|" separator. Any of these synonyms will be considered a valid answer. The Google guide recommends varying types of questions, which ensures dynamic learning and engagement, and questions should be as compact and clear as possible. It is recommended that the quiz be culturally and gender sensitive so that it is interesting to as many users as possible.

An action creates by the Trivia template, go to the Action console, and create a New Project. Then select the Template command in the Advanced Parameters section at the bottom of the page and specify Trivia Template in the appearing window. Then open the file from templates in Google Sheet editor. Then copy and paste the URL of the sheet in your browser and open the file in the online spreadsheet editor.

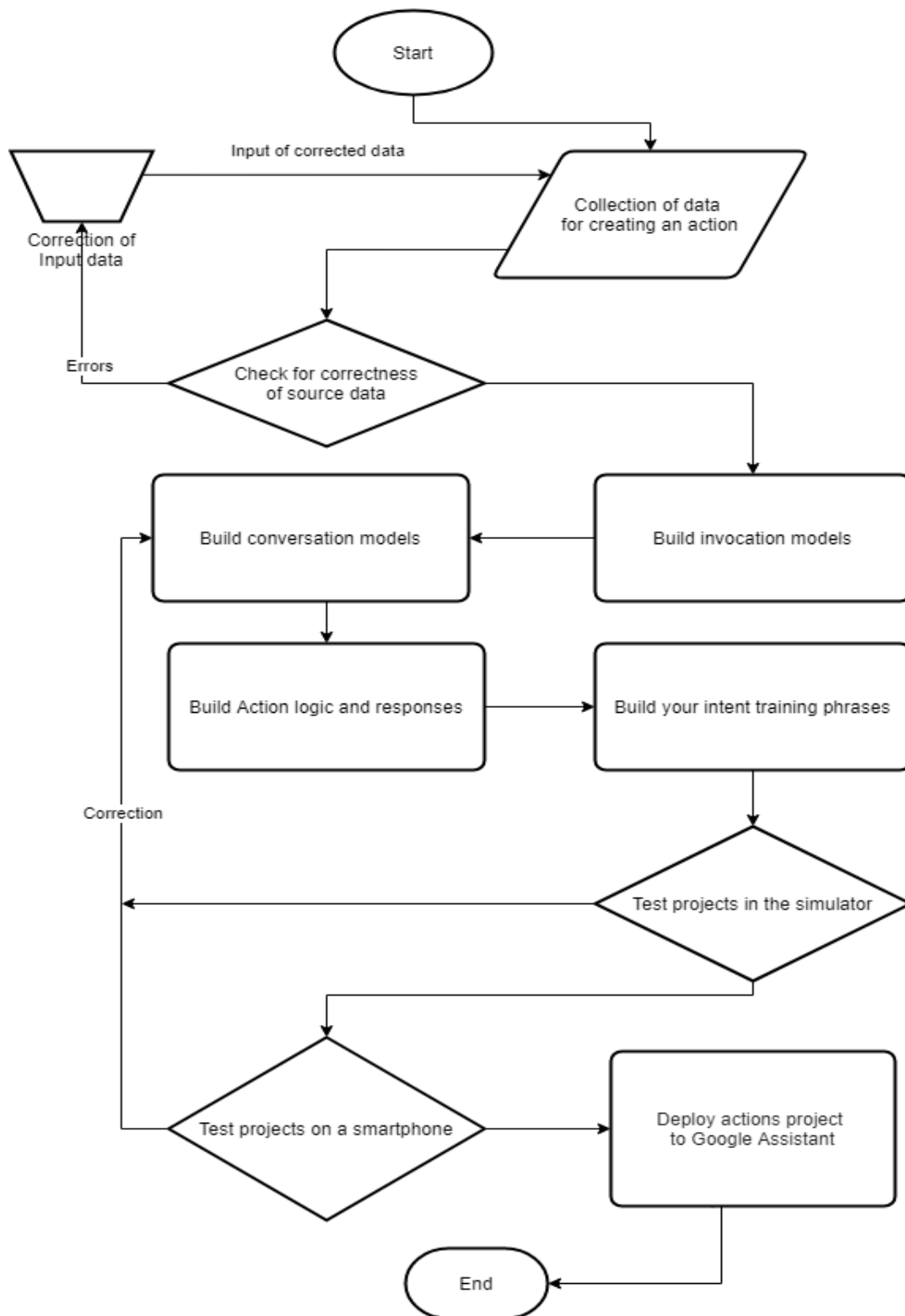


Figure 1. The algorithm of creating the application

You can fine-tune the quiz by specifying the difficulty and category for each question and create your own sound effects or speech files. You can specify difficulty levels for the questions, add a "Difficulty/Grade Level" column on the sheet, then specify a difficulty level or grade level for

each question. You can specify up to three different difficulty levels: "Easy", "Medium" and "Hard". To separate the questions into categories, you will need to add a "Category/Topic" column and then specify a category for each question on the worksheet. You can set up to three custom categories. Before running the quiz, the user will be asked about the desired topic and the level of difficulty for the current round of questions. Additional settings and examples of using them there are in User Guide [16].

Once the action has been created, it can be tested. To do this, it is necessary to select the Test command in the top menu to test the action through the web interface. Figure 2 shows testing application in action console.

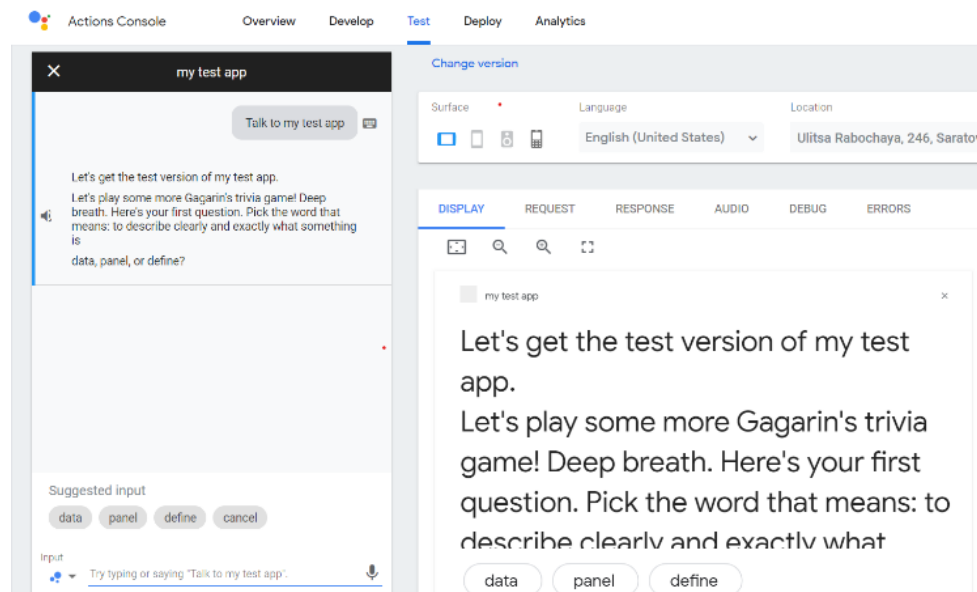


Figure 2. Test application in action console

To test the action in your smartphone, you need to run the Google Assistant and say into the microphone or enter the phrase "Talk to my test app". You can test your action through a Google Home device, the device must be connected to the same developer account. If the action supports more than one language, you should test each language separately.

5. VOICE ASSISTANT FOR ACADEMIC ENGLISH TRAINING

A Trivia Action template was used for developing an application, in the order that was described above. This template allows you to create test quizzes, which can be run on mobile devices with Google Assistant, as well as on smart speakers with the Google Home system.

The action "Gagarin Trivia" was developed by us for vocabulary learning and training academic writing. For questions we used terms recommended by the HSE Centre for Academic Writing [17] and other quality sources [18].

To start the action, it needs to say the phrase "Play Gagarin Trivia game" to the Google assistant. Figure 3 shows the launched application, and the assistant tells you the rules of the quiz. You can choose the number of participants and the number of questions in one round. Then the first question will be read out and you will get three choices for answer. You must either say the

answer or the answer number. If the answer is correct, the program will give the approval reaction, if not, you will get the correct answer and some explanations.

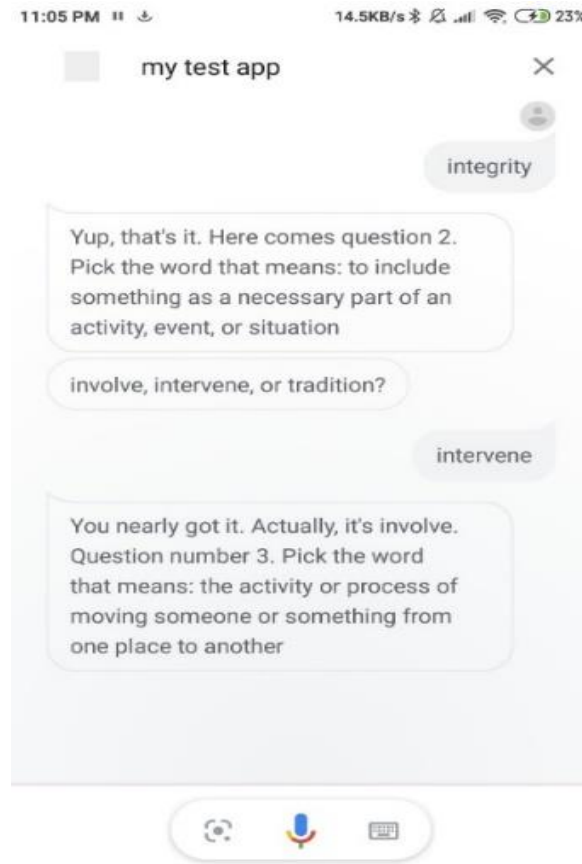


Figure 3. App view on a smartphone

If you do not understand the question, you can ask the assistant to repeat it with the command "Repeat question". If you do not know the correct answer, you can skip the question with the command "Skip question". To finish the game, you should say "Stop". At the end of the round, the assistant will summarize and tell you the number of correct answers. The application is quite fascinating. If you have the high level, you can use it even while walking, communicating with the assistant using headphones.

The second action is "Academic phrase bank Trivia" which allows to improve the knowledge of academic English. The information from English textbooks, specialized websites, and Academic phrase bank was used as a content of the application [19].

As answer to the assistant's question, it is necessary to choose the sentence closest by it meaning to the given phrase in English. As a hint, the assistant gives three options. In order for a phrase to be counted, it is not necessary to pronounce it verbatim, partial correctness is enough.

The exercise allows to improve the skill of scientific English attentive listening, and train pronunciation. In addition, the pronunciation of the templates will help them to memorize well. If you cannot pronounce the suggestions correctly, you can skip the question and proceed to the next one.

6. CONCLUSIONS

English proficiency enhances the quality of scientific research. Many modern scientific works published in highly ranked scientific journals are written in English. The ability to communicate confidently in English enables effective communication with other researchers.

Improving language skills is a long-time process that requires considerable time and financial expenditure. The application uses the artificial intelligence of the Google platform and was designed for self-study of academic English and improving academic speaking skills. The application allows to study new words, to carry out the construction of grammatically correct sentences in English, to memorize templates of academic phrases. Users can train the correct pronunciation of phrases and developing skills of scientific texts listening.

REFERENCES

- [1] N. W. Buddhima and S. Keerthiwansa, "Artificial Intelligence Education (AIEd) in English as a Second Language (ESL) Classroom in Sri Lanka," *Int. J. Conceptions Comput. Inf. Technol.*, vol. 6, no. 1, pp. 2345–9808, 2018.
- [2] S. Popenici and S. Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Res. Pract. Technol. Enhanc. Learn.*, vol. 12, no. 1, p. 22, Dec. 2017, doi: 10.1186/s41039-017-0062-8.
- [3] T. Karsenti, "Artificial intelligence in education: The urgent need to prepare teachers for tomorrow's schools," *Form. Prof.*, vol. 27, no. 1, p. 105, 2019, doi: 10.18162/fp.2019.a166.
- [4] M. J. Timms, "Letting Artificial Intelligence in Education out of the Box: Educational Cobots and Smart Classrooms," *Int. J. Artif. Intell. Educ.*, vol. 26, no. 2, pp. 701–712, 2016, doi: 10.1007/s40593-016-0095-y.
- [5] A. Luczak, "Using Memrise in Legal English Teaching," *Stud. Logic, Gramm. Rhetor.*, vol. 49, no. 1, pp. 141–152, Mar. 2017, doi: 10.1515/slgr-2017-0009.
- [6] G. D. Boca, "The Impact of IT on Knowledge Feedback to Education Design," *Procedia - Soc. Behav. Sci.*, vol. 83, pp. 856–861, Jul. 2013, doi: 10.1016/j.sbspro.2013.06.161.
- [7] P. B. Sing, M. A. Embi, and H. Hashim, "Ask the Assistant: Using Google Assistant in Classroom Reading Comprehension Activities," *Int. J. New Technol. Res.*, vol. 5, no. 7, pp. 39–43, 2019, doi: 10.31871/ijntr.5.7.6.
- [8] Berdasco, López, Diaz, Quesada, and Guerrero, "User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana," *Proceedings*, vol. 31, no. 1, p. 51, Nov. 2019, doi: 10.3390/proceedings2019031051.
- [9] O. Metatla, A. Oldfield, T. Ahmed, A. Vafeas, and S. Miglani, "Voice User Interfaces in Schools," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 2019, pp. 1–15, doi: 10.1145/3290605.3300608.
- [10] O. A. Pynova and R. S. Zaripova, "Technologies of artificial intelligence in education," *Russ. J. Educ. Psychol.*, vol. 10, no. 3, pp. 41–50, Jun. 2019, doi: 10.12731/2658-4034-2019-3-41-44.
- [11] "How software that learns as it teaches is upgrading Brazilian education." <https://www.theguardian.com/technology/2016/jan/10/geekie-educational-software-brazil-machine-learning> (accessed Feb. 24, 2020).
- [12] "School of English - individual learning English at Skyeng School." <https://skyeng.ru/> (accessed Feb. 24, 2020).
- [13] C. Pearl, *Designing Voice User Interfaces by Cathy Pearl Chapter 1 . Introduction A Brief History of VUIs*. O'Reilly Media, Inc., 2016.
- [14] G. Terzopoulos and M. Satratzemi, "Voice assistants and artificial intelligence in education," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3351556.3351588.
- [15] "Mondly - Learn languages online for free." <https://mondly.com/> (accessed Feb. 25, 2020).
- [16] "Create a Trivia Action using templates | Actions on Google Templates." <https://developers.google.com/assistant/templates/trivia> (accessed Feb. 24, 2020).
- [17] "HSE - UniversityAcademic Development - Academic Writing Centre - Podcasts." <https://academics.hse.ru/awc/podcasts> (accessed Feb. 25, 2020).

- [18] “Glossary of terms for academic writing - Wiki - innsida.ntnu.no.” <https://innsida.ntnu.no/wiki/-/wiki/English/Glossary+of+terms+for+academic+writing> (accessed Feb. 25, 2020).
- [19] “Home Page - Academic Phrasebank.” <http://www.phrasebank.manchester.ac.uk/> (accessed Jun. 12, 2020).

AUTHOR

Received degree of Doctor of Economic Sciences in Logistics from Saint Petersburg state university of Economics, Russia in 2011. He has an industrial experience of 10 years in programming, and an academic experience of over 20 years in Yuri Gagarin State Technical University of Saratov, with research in versatile fields of transport, logistics, and computer science.



© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

TECHNIQUE FOR REMOVING UNNECESSARY SUPERIMPOSED PATTERNS FROM IMAGE USING GENERATIVE NETWORK

Kazutake Uehira and Hiroshi Unno

Department of Network Engineering, Kanagawa Institute of Technology, Kanagawa, Japan

ABSTRACT

A technique for removing unnecessary patterns from captured images by using a generative network is studied. The patterns, composed of lines and spaces, are superimposed onto a blue component image of RGB color image when the image is captured for the purpose of acquiring a depth map. The superimposed patterns become unnecessary after the depth map is acquired. We tried to remove these unnecessary patterns by using a generative adversarial network (GAN) and an auto encoder (AE). The experimental results show that the patterns can be removed by using a GAN and AE to the point of being invisible. They also show that the performance of GAN is much higher than that of AE and that its PSNR and SSIM were over 45 and about 0.99, respectively. From the results, we demonstrate the effectiveness of the technique with a GAN.

KEYWORDS

GAN, Auto encoder, Depth map, Pattern removing.

1. INTRODUCTION

3D images have been used for various applications and will attract more attention in the multimedia field as advanced 3D displays for virtual reality, augmented reality, etc. are developed. The depth map represents two-dimensional distance information from a camera to a subject, and it is key information for creating 3D image content [1-3]. There are various ways to obtain a depth map. The pattern projection method is one effective method for obtaining a depth map because it makes it possible to obtain a map even in areas where there are no textures or edges, for which it would be difficult for other methods to obtain the depth. The pattern projection method is divided into two types: the projection of a visible pattern [4, 5] and the projection of an invisible infrared pattern [6, 7]. Visible patterns are used when only a depth map is needed because the projection pattern appears as noise in a captured RGB image. An infrared pattern is used when a depth map and a RGB image are both needed because the projected pattern does not appear in RGB images. This method requires an infrared projector and an infrared camera in addition to a normal RGB camera.

We have been developing a technique for acquiring depth maps that projects visible periodic patterns on a subject [8]. When a periodic pattern is projected onto a subject under certain conditions, a depth map is obtained from the spatial frequency of the projected pattern in the captured image with a camera.

In the applications that we are assuming, both an RGB image and a depth map are used, and depth map is embedded in the RGB image invisibly. To achieve this, first, we project a periodic pattern to obtain a depth map, and after obtaining the depth map, we remove the pattern from the RGB image. The main question was how to remove the superimposed pattern in the image of a subject. Many studies have been reported on periodic noise reduction in images [9-13]. Many of these studies applied the frequency domain approach, as periodic noise cannot be simply separated from the original image in the spatial domain. In the frequency domain, the periodic pattern is concentrated at one point, which makes it easy to remove it. As this kind of method, a method using a notch filter [9, 10] and a method using a median filter in the frequency domain [11] and so on have been reported. However, in our case, the spatial frequency of the pattern depends on the depth of the subject and changes depending on the location, so it has a spread in the frequency domain. Therefore, these conventional methods cannot satisfactorily remove the pattern.

In our previous study, we confirmed the feasibility of a technique that uses a generative adversarial network (GAN) to remove a line and space pattern from a target image [14]. In this study, we continue development on the method of our previous study, verifying the performance of the GAN under various conditions and comparing it with an auto encoder (AE) as a generative network using a deep learning method similar to a GAN. This paper describes the experiments we conducted and clarifies the condition for removing the line and space pattern from captured images.

The remainder of this paper organized as follows. Section 2 overviews the whole study where this study is a part. Section 3 describes the experiment. Section 4 presents results of the experiments and discusses them. Finally, Section 5 presents our conclusions.

2. METHOD OF ACQUIRING DEPTH MAP BY PERIODIC PATTERN PROJECTION

Figure 1 shows the overall configuration of our study, which we are now working on. A periodic pattern is projected onto a subject. A camera captures an image of the subject on which the pattern is projected. If the projector is placed a distance away from the camera in the depth direction as shown in Figure 1 (shown as D in the figure), the spatial frequency of the projected pattern in the captured image depends on the depth of the subject. Therefore, a depth map can be obtained by obtaining the spatial frequency of the projection pattern for each small area in the image captured with the camera.

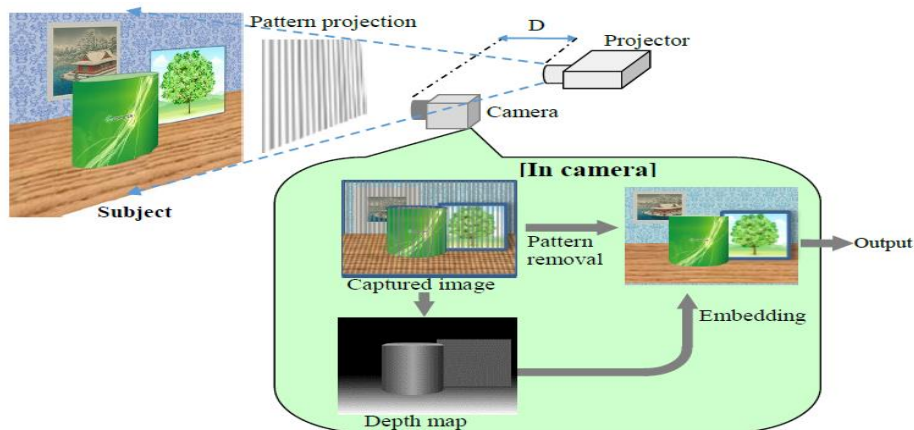


Figure 1. Overall configuration of proposed technology

The contrast of the projection pattern is set high so as to make it easy to obtain the spatial frequency of the pattern. Therefore, the pattern is clearly superimposed in the captured image. When using this image as an RGB image of the subject, this superimposed periodic pattern is obstructive and thus need to be removed. After removing this pattern, the depth map obtained is embedded in an RGB image invisibly. This can be done with technology for hiding information in images used such for as digital watermarking etc. We already confirmed that a depth map could be embedded in the image of a subject and embedded data was retained even if JPEG-compressed [15].

The final output image of this system can be treated as a normal 2D image for transmission and storage, although it includes 3D information in the form of a depth map. If necessary, we can read out the depth map to create a 3D image from the 2D RGB using the depth map.

Removing the superimposed periodic pattern is key for achieving this system. Because the periodic pattern depends on the depth of the subject and differs from place to place, that is, the pattern period is not constant and unknown in the captured image, it is difficult to remove patterns until they disappear with conventional methods mentioned above.

3. EXPERIMENTS

We conducted experiments by simulation. Figure 2 explains how to produce captured images of subjects on which periodical patterns are projected. Figure 3 shows the example images used as subjects. First, we cropped 448 images of 256×256 pixels from the images of subjects. We used the line and space (L/S) pattern as the periodical projection pattern. Assuming that the pattern is projected with blue light, B component images of the cropped images and projection pattern images were multiplied to produce captured images, on which the L/S pattern was superimposed. We used L/S patterns with different amplitudes and spatial frequencies. The averaged brightness was 200, and the amplitude was changed in 5 steps, of which the minimum was 10 and the maximum was 50. These values indicate the grayscale with a maximum of 255. Figure 4 shows magnified L/S pattern and the example image of the simulated captured images when the amplitude was 10 and 50. The spatial frequency was changed in 16 steps from 0.18 to 0.25 lines/pixel.

We generated 7,192 simulated captured images from the combination of 448 subject images and 16 L/S patterns with different spatial frequencies. Of these, 5,600 were used for learning, 700 were used for evaluation in learning, and 700 were used for evaluation in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

Figure 5 shows the configuration of the generative adversarial network (GAN) and Figure 6 shows the configuration of the encoder, decoder and discriminator of the GAN. We mainly used the GAN, and the AE was used as a reference. The encoder and decoder of the generator of the GAN consisted of 8 layers each, and the discriminator consisted of 6 layers. These layers were convolution layers, and a 4×4 kernel was used for convolution. Leaky-relu was used as an activation function in the encoder and the discriminator and relu was used in the decoder. Dropout was used in the decoder and dropout rate was 0.25.

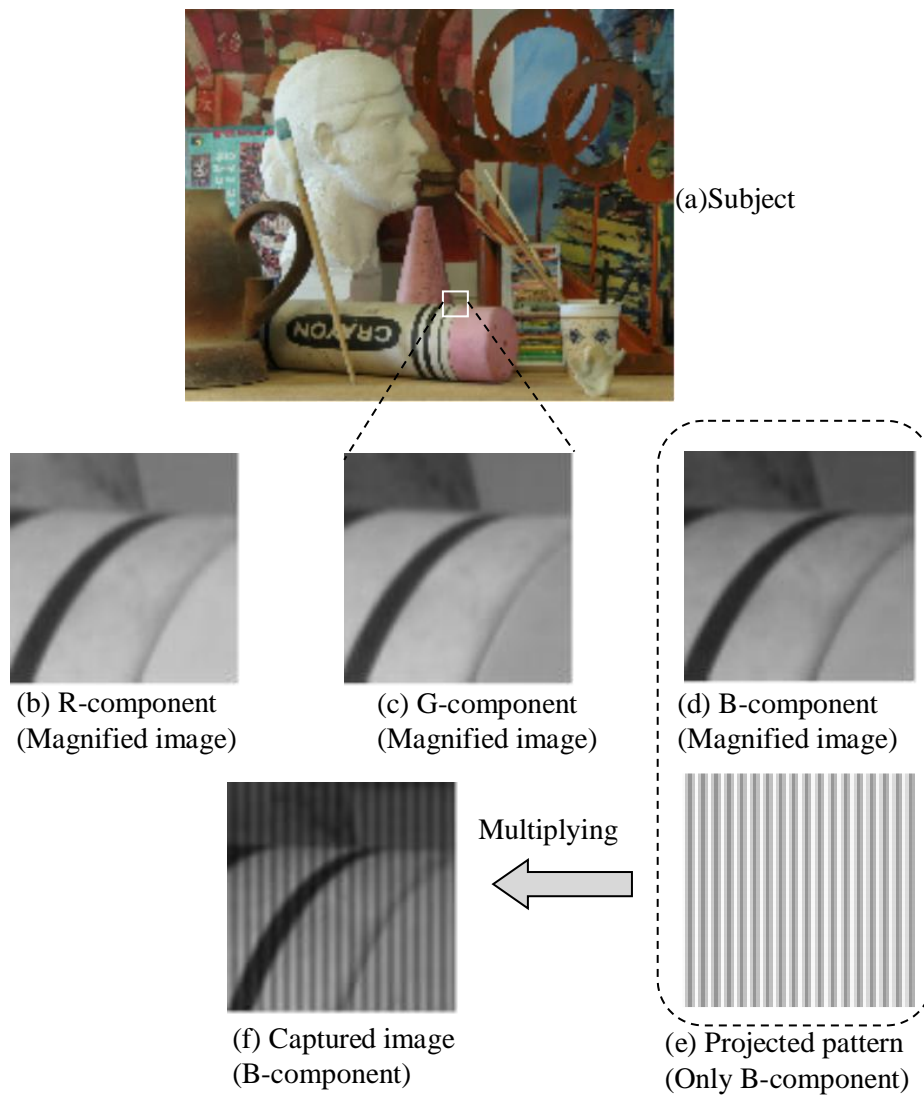


Figure 2. Simulation procedure of generating captured images

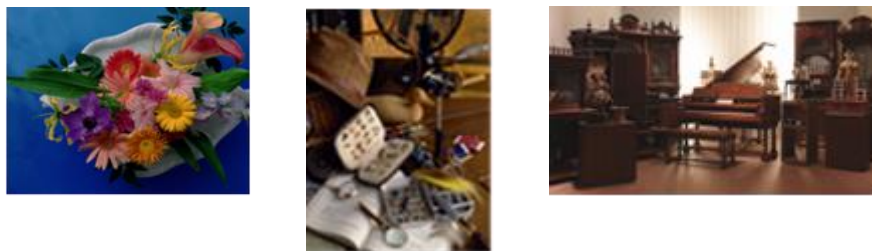


Figure 3. Example of images used as subjects

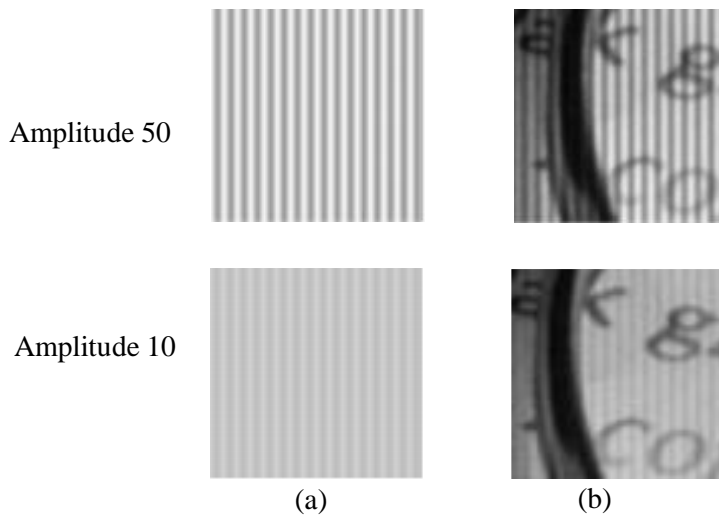


Figure 4. Magnified projected pattern (a) and simulated captured image (b)

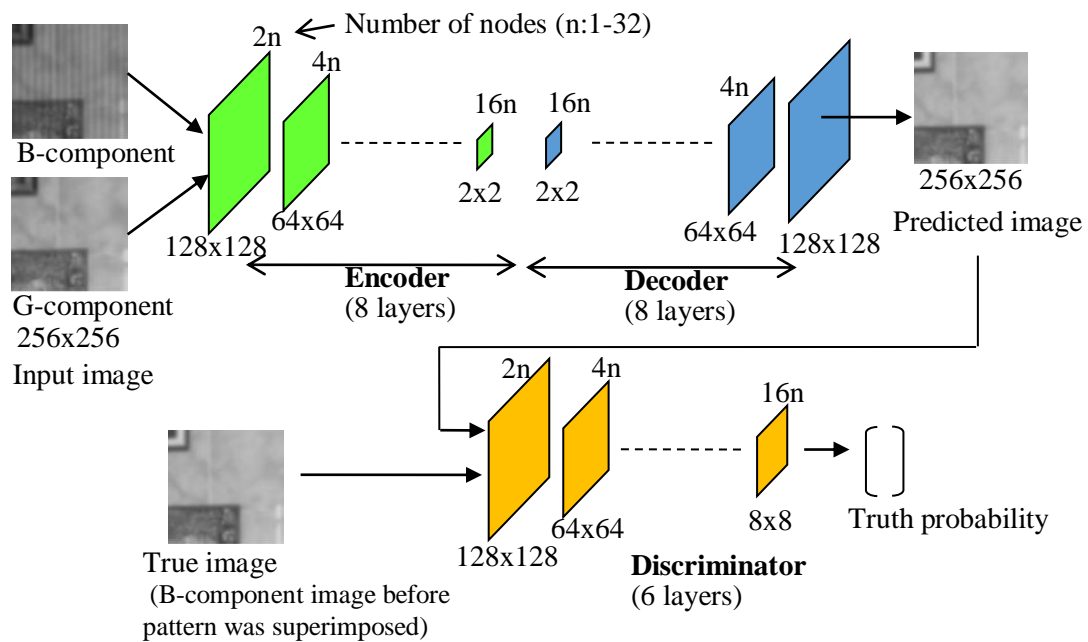


Figure 5. Configuration of the generative adversarial network (GAN)

B component images on which the L/S pattern was superimposed were input to the generator, and the generator output predicted images that were expected to contain no L/S pattern. True images were input to the discriminator as training data.

Figure 7 shows the configuration of the AE and Figure 8 shows the configuration of the encoder, decoder of the AE. The encoder and decoder of the AE consisted of five convolution layers, and a 3 x 3 kernel was used for convolution. Also in the AE, Leaky-relu was used in the encoder and relu was used in the decoder as an activation function.

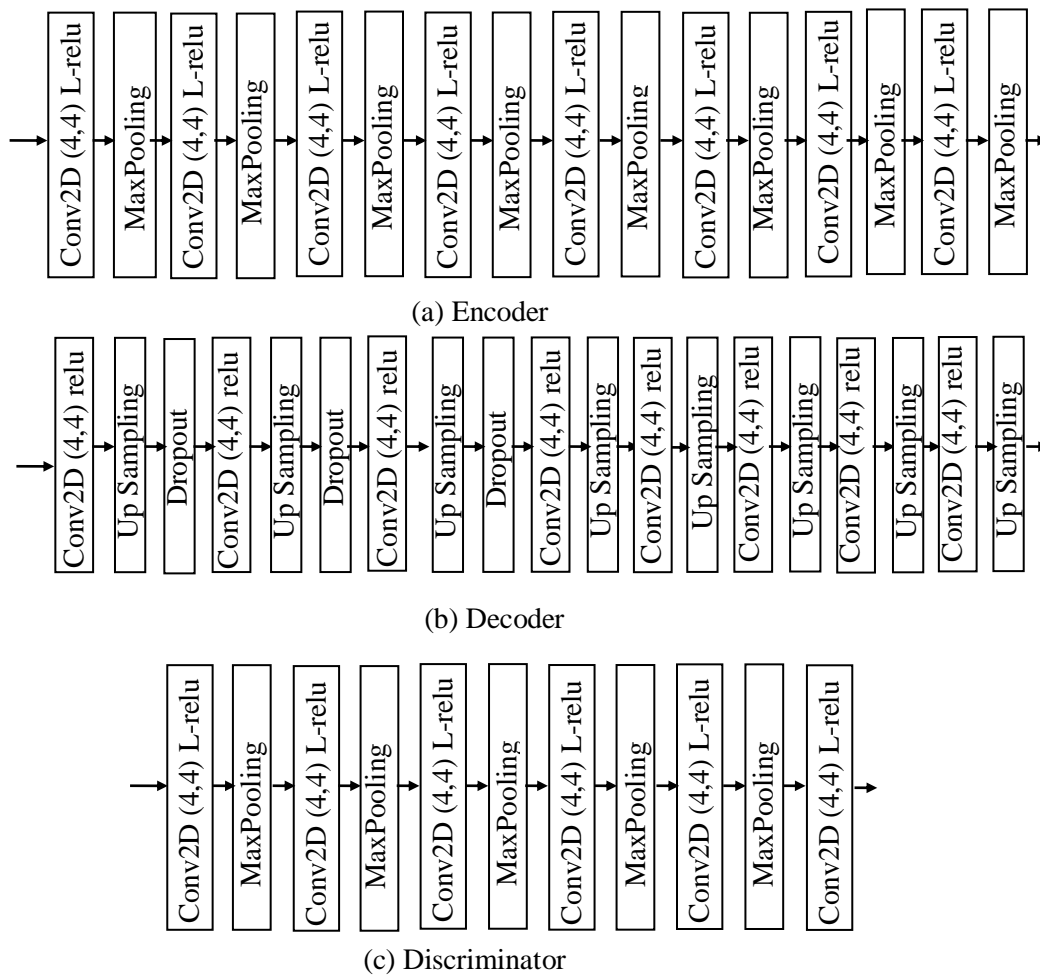


Figure 6 Configuration of the encoder decoder and discriminator of the GAN

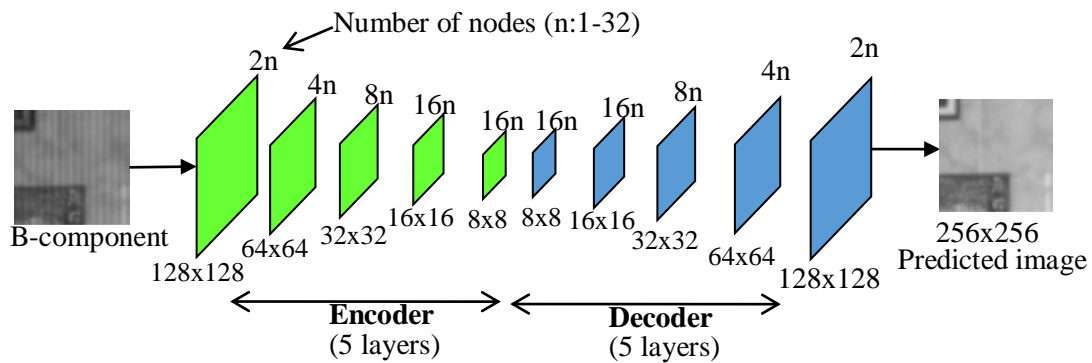


Figure 7. Configuration of the auto encoder (AE)

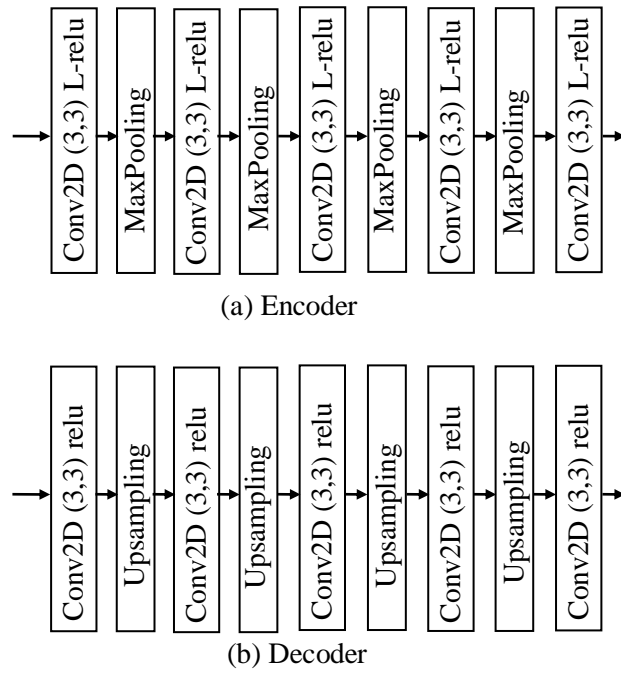


Figure 8. Configuration of the encoder and decoder of the AE

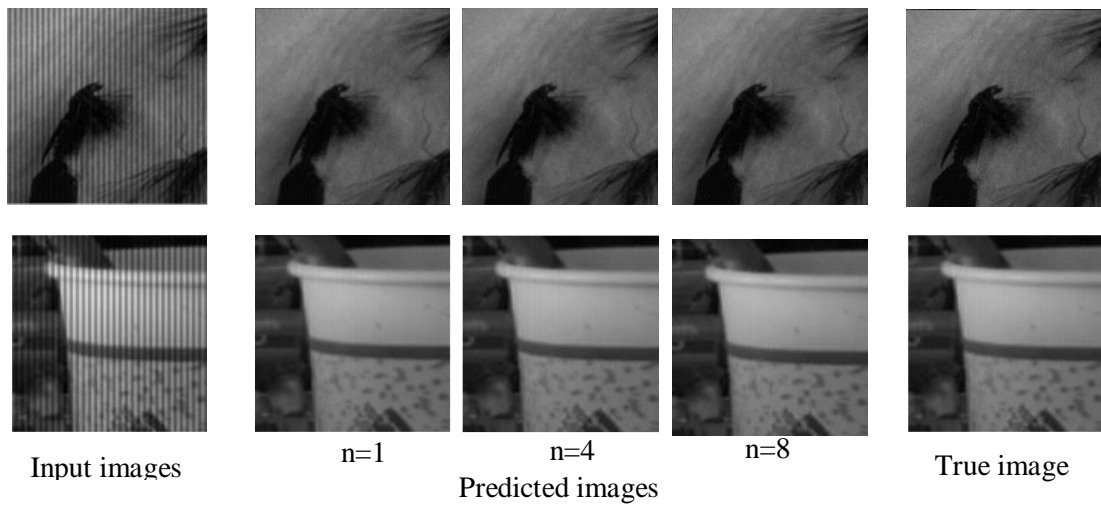


Figure 9. Examples of input, predicted, and true image

4. RESULTS AND DISCUSSION

Figure 9 shows examples of the input, predicted, and true images. As these figures show, we cannot see the L/S patterns at all in all of the predicted images, and the predicted images look almost the same as the true image.

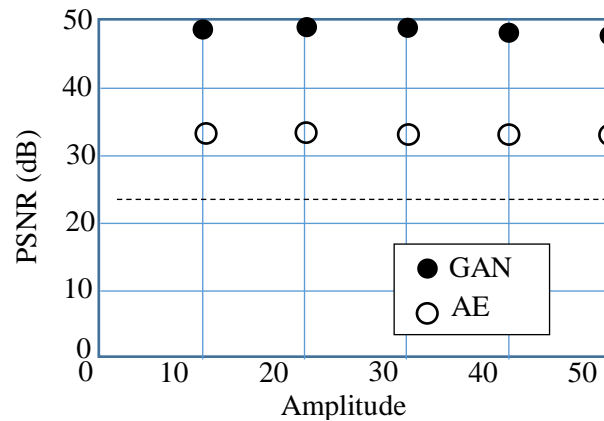


Figure 10. PSNR of image after pattern removal against original image

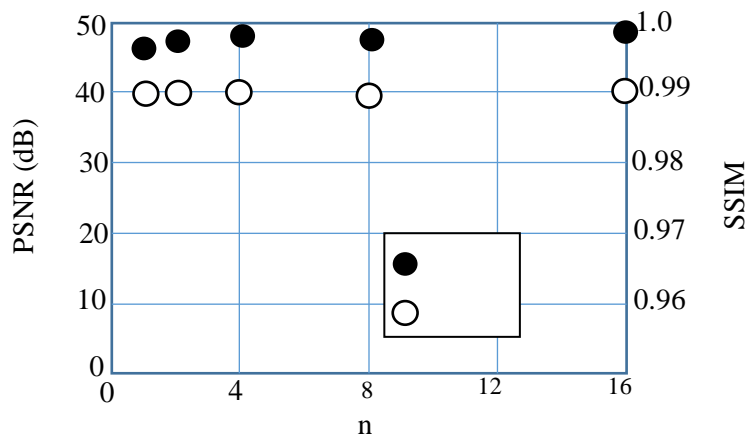


Figure 11. Dependence of PSNR and SSIM on number of nodes in middle layers

Figure 10 shows the PSNR of images after pattern removal using GAN and AE against the original image. Original image means the B component image before the pattern was superimposed. Figure 10 shows the dependence on the amplitude of the L/S pattern. The dashed line indicates the values before removing the pattern. It can be seen that the PSNR improved greatly, especially when the GAN was used. It exceeded 45, and this is a high enough value for the purpose of this study. In this figure, the PSNR was obtained as the dependence on the amplitude of the L/S pattern, and it is also seen that the PSNR decreased as the amplitude of the pattern increased. However, this decrease was slight considering the difference in amplitude of the L/S patterns. This might be because the pattern is not complicated, so the GAN can be easily trained and output an image close to the true image regardless of the amplitude of the L/S. We can see that the PSNR for the GAN was higher than that for the AE. Therefore, a GAN is better than an AE for our system.

Figure 11 shows the dependence of the PSNR and SSIM when using GAN on the number of units of the middle layers. In this figure, n for the horizontal axis is the number shown in Figure 5, and it is proportional to the number of nodes in the middle layers. From Figure 11, it is seen the PSNR and SSIM increased as the number of nodes increased; however, even when n was one, the PSNR exceeded 45, and SSIM was 0.99; both are a high enough for our system.

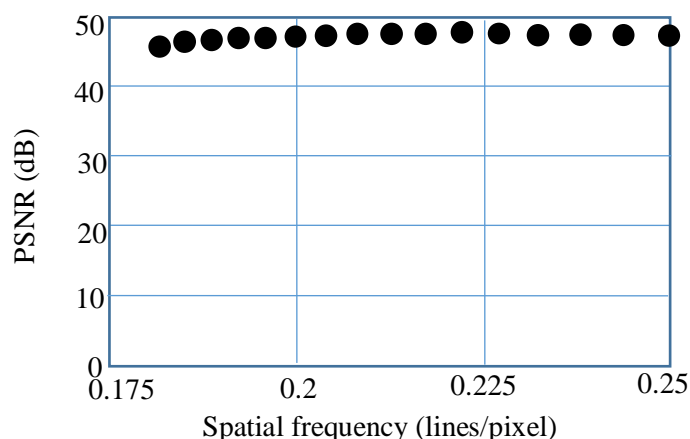


Figure 12. Dependence of PSNR on spatial frequency of L/S pattern

Figure 12 shows the dependence of the PSNR when using GAN on the spatial frequency of the L/S pattern. It shows that they are almost independent from the frequency and high enough at any frequency.

5. CONCLUSION

We studied a technique to remove periodic patterns from the RGB image of the subject taken by the camera. These patterns are projected onto the subject when its image is captured for the purpose of acquiring a depth map of the subject. Since these patterns become unnecessary after acquiring the depth map and it is just noise for RGB images, we attempted to remove these patterns using generative network, GAN and AE. From the experimental, it was shown that these periodic patterns were effectively removed by using GAN and AE to the point of being invisible. They also show that the performance of GAN is much higher than that of AE and that its PSNR and SSIM were over 45 and about 0.99, respectively. From the results, we demonstrate the effectiveness of the technique with a GAN.

In future work, we will perform similar experiments on other periodic patterns such as checkered patterns to confirm the effectiveness of the GAN method.

ACKNOWLEDGEMENTS

This study was carried out with the support of a research grant from Kanagawa Institute of Technology.

REFERENCES

- [1] L. Zhang, C. Vázquez, and S. Knorr (2011) “3D-TV Content Creation: Automatic 2D-to-3D Video Conversion”, IEEE Transactions on Broadcasting, Vol. 57, No. 2, pp372-383
- [2] S. Mathai, P Mathai, and K A Divya (2015) “Automatic 2D to 3D video and image conversion based on global depth map”, Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research
- [3] S.Zingera, L.Do, and P.H.N.de With (2010) “Free-viewpoint depth image based rendering”, Journal of Visual Communication and Image Representation, Vol. 21, Issue, 5-6, pp533-541

- [4] D. Scharstein and R. Szeliski, (2003) “High-Accuracy Stereo Depth Maps Using Structured Light”, Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, vol. 1, pp195-202.
- [5] I. Ishii, K. Yamamoto, K. Doi, and T. Tsuji (2007) “High-speed 3D image acquisition using coded structured light projection”, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp925-930.
- [6] A. Maimone and H. Fuchs (2011) “Encumbrance-Free Telepresence System with Real-Time 3D Capture and Display Using Commodity Depth Cameras”, The 10th IEEE International Symposium on Mixed and Augmented Reality, Washington, DC, United States, pp137-146.
- [7] F. Alhwarin, A. Ferrein, and I. Scholl (2014) “IR Stereo Kinect: Improving Depth Images by Combining Structured Light with IR Stereo”, Proceedings of Pacific Rim International Conference on Artificial Intelligence, pp409-421.
- [8] H. Unno, S. Isaka, Y. Takashima, K. Uehira, (2015) “New Technique for Embedding Depth Information in Captured Images Using Light Beam Containing Invisible High-Frequency Patterns-Design and Preparation of New Experimental Setup With Some Comments”, Journal of Display Technology, Vol. 11, No. 2, pp136-145.
- [9] I. Aizenburg, C. Butakoff, (2008) “A windowed Gaussian notch filter for quasi-periodic noise removal”, Image and Vision Computing, Vol. 26, Issue 10, pp1347-1353.
- [10] F. Sur, M. Grediac (2015) “Automated removal of quasiperiodic noise using frequency domain statistics”, Journal. of Electronic Imaging, Vol. 24, No.1, 013003.
- [11] I. Aizenberg, C. Butakoff (2002) “Frequency Domain Median-like Filter for Periodic and Quasi-Periodic Noise Removal”, SPIE Proceeding, Vol. 4667, pp181-191
- [12] J. Varghese (2016) “Adaptive threshold based frequency domain filter for periodic noise reduction”, International Journal of Electronics and Communications, Vol. 70, Issue 12, pp1692-1701
- [13] S. Ketenci, A. Gangal (2012) “Design of Gaussian star filter for reduction of periodic noise and quasi-periodic noise in gray level images”, Proceedings of the IEEE International Symposium on Innovations in Intelligent Systems and Applications.
- [14] K. Uehira and H. Unno, (2020), “Technique for removing superimposed patterns on objects using GAN,” Proceedings of IEEE ICCE2020
- [15] K. Uehira, (2014), “Technique of Embedding Depth Maps into 2D Images”, Journal of Electronic Science and Technology, Vol. 2, No. 1, pp95-100

AUTHORS

Kazutake Uehira received his B.S. and M.S. in Electronics in 1979 and 1981 and his Ph.D. in 1994 all from the University of Osaka Prefecture, Japan. He joined NTT Electrical Communication Laboratories in Tokyo in 1981. Since then, he has been engaged in research on imaging technologies. In 2001, he joined Kanagawa Institute of Technology, Japan, as a professor and is currently engaged in research on information embedding technology and 3D image.



Hiroshi Unno received the B.S. degree in mathematical science from Tokyo Denki University, Japan, in 1987, and the M.S. degree in information science from the Japan Advanced Institute of Science and Technology, Hokuriku, Japan, in 1996. He joined Chuo University, Japan, in 1992. Since then, he has been engaged in research on software development environments, formal methods, and computer networks. He joined the Kanagawa Institute of Technology, Japan, in 2001 and is currently engaged in research on image processing and 3D displays.



ADAPTIVE FILTERING REMOTE SENSING IMAGE SEGMENTATION NETWORK BASED ON ATTENTION MECHANISM

Cong zhong Wu¹, Hao Dong¹, Xuan jie Lin¹, Han tong Jiang¹, Li quan Wang¹, Xin zhi Liu¹ and Wei kai Shi²

¹Department of Computer Engineering and Information, Hefei University of Technology, Anhui, P.R.China

²Department of Faculty of Information Technology, Macau University of Science and Technology, Macau, P.R.China

ABSTRACT

It is difficult to segment small objects and the edge of the object because of larger-scale variation, larger intra-class variance of background and foreground-background imbalance in the remote sensing imagery. In convolutional neural networks, high frequency signals may degenerate into completely different ones after downsampling. We define this phenomenon as aliasing. Meanwhile, although dilated convolution can expand the receptive field of feature map, a much more complex background can cause serious alarms. To alleviate the above problems, we propose an attention-based mechanism adaptive filtered segmentation network. Experimental results on the Deepglobe Road Extraction dataset and Inria Aerial Image Labeling dataset showed that our method can effectively improve the segmentation accuracy. The F1 value on the two data sets reached 82.67% and 85.71% respectively.

KEYWORDS

Convolutional Neural Network, Remote Sensing Imagery Segmentation, Adaptive Filter, Attention Mechanism, Feature Fusion

1. INTRODUCTION

Remote sensing imagery segmentation is an important part of computer vision tasks. It is widely used in environmental monitoring, urban planning, and rescue of natural disasters such as earthquakes, floods, and mountain fires. Especially in natural disaster rescue, if remote sensing imagery can be segmented faster and more accurately, more rescue time can be obtained and thus damage can be minimized. Roads and buildings are often the objects captured by remote sensing satellites. And backgrounds of these objects are much more complex and diverse. The road image, contains different scenes, such as cities, towns. The building image contains town buildings and sparse country buildings (Figure 1). These factors make it difficult for the network to accurately locate and identify the foreground features of remote sensing imagery.

The research methods for semantic segmentation of remote sensing imagery are mainly two types: traditional methods based on manual feature extraction and deep learning methods based on convolutional neural networks. The traditional segmentation methods include based on region, edge, threshold, etc. These methods can only extract the low-level features of the image. While it cannot fully express the high-level features of the image. With convolutional neural networks David C. Wyld et al. (Eds): ITCSE, ICDIPV, NC, CBIoT, CAIML, CRYPIS, ICAIT, NLCA - 2021 pp. 23-36, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.110903

such as VGGNet [1], GoogleNet [2], ResNet [3], etc. widely used in computer vision tasks, a large number of research works on remote sensing imagery segmentation are based on deep learning methods. At this time, the network can extract the features images faster and more accurately by combining convolution, downsampling, and activation functions. Among these operations, downsampling can reduce the number of parameters and computation of the network. What's more, it can also expand the receptive field of the network. However, it can also arise as aliasing which cause the use information of object to be lost.

In traditional digital signal processing, aliasing refers to the distortion of the sampled signal due to the low sampling frequency, resulting in the inability to recover the original signal. In this case, according to Nyquist's sampling theorem, the sampled signal can recover the original signal completely when the sampling rate must be at least twice the highest frequency of the original signal. In the deep neural network, aliasing also exists due to the downsampling layer. Inspired by the traditional method in which a low-pass filter can recover or reconstruct the original signal, Richard Zhang [4] proposed the concept of filter and applied Gaussian blur layer (filter) before downsampling. We can avoid aliasing by applying a Gaussian filter. However, as the high-frequency noise, such as background, needs to be blurred more compared to the lower frequency edges when using a single Gaussian filter tuned for the noise, the edges are over-blurred leading to significant information loss. To solve this issue, what we need is to apply different Gaussian filters to the foreground and background separately, so that we can avoid aliasing while preserving useful information.

Long proposed a fully convolutional network (FCN) [5] to replace the fully connected layers at the end of the network, making the successful application of deep learning in image segmentation. For example, Shunping Ji [6] successfully applied FCN to building segmentation. Skip-connection proposed by Ronneberger in U-Net [7] was widely used in codec network structure which can help to recover image details and edge information. The U-Net was successfully applied to segment road image by Zhengxin Zhang [8]. Cambridge proposed downsampling index in SegNet [9]. The key component of SegNet is the decoder network which consists of a hierarchy of decoders corresponding to each encoder. Of these, the appropriate decoders use the max-pooling indices received from the corresponding encoder to perform non-linear upsampling of their input feature maps. Chaurasia A proposed to fuse the features between encoder and decoder by pixel summation in LinkNet [10]. While D-LinkNet [11] improved LinkNet and successfully applied it to road segmentation. However, all these networks ignore the aliasing caused by downsampling. Meanwhile, it also ignores the probable semantic gap between the corresponding levels of Encoder-Decoder as proposed by Nabil [12].

For the large-scale variation in remote sensing imagery, pooling or dilation convolution are two effective ways to deal with. Zhao used different sizes of pooling kernels in PSPNet [13] to increase the receptive field of the network and fuse different scale features. By aggregating information from different regions, the purpose of fully mining the global information is achieved. Dilated convolution [14], firstly proposed by Yu et al, can support the exponential expansion of the receptive field without loss of resolution or coverage. The LFE [15] proposed by Ryuhei Hamaguchi uses dilated convolution to effectively segment building remote sensing imagery. However, as the dilated rate increases, the receptive field of the network increases exponentially and there may be redundant information. Hence, the design of dilated rate can affect the performance of the network. In DeepLabv3 [16] and HDC [17] are improved by dilated rate.

The introduction of an attention mechanism is an effective way to improve remote sensing imagery segmentation. The FarSeg [18] proposed by Zhuo Zheng is based on the correlation between the distribution of remote sensing image data, by capturing the different dimensions of

the feature map to highlight the foreground feature information and suppress irrelevant redundant background information. Unlike previous works that capture contexts by multi-scale feature fusion, Dual Attention Network (DANet)[19] adaptively integrate local features with their global dependencies, which model the semantic interdependencies in spatial and channel dimensions respectively. SCAAttNet [20] proposed by Haifeng Li combines spatial attention with

channel attention to segment remote sensing imagery. In summary, the following problems still exist when using deep learning networks to segment remote sensing imagery:

1. Operations such as pooling and downsampling can cause aliasing.
2. When facing larger-scale variance, dilated convolution can expand the receptive field to aggregate multi-scale contextual information but also bring redundant information of background.
3. Semantic gap between the corresponding levels of Encoder-Decoder.

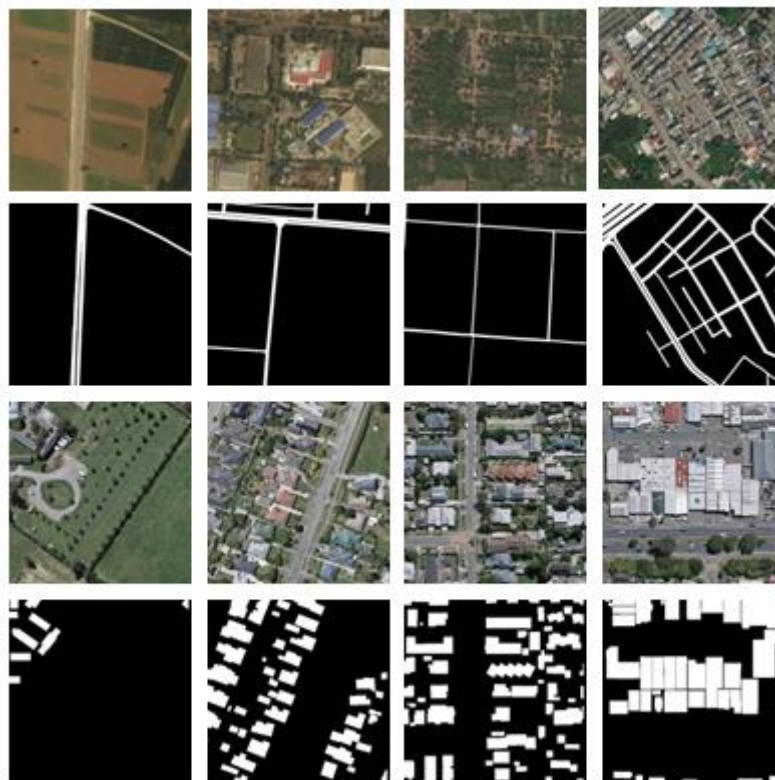


Figure 1. Example road extraction, building detection training images, and corresponding labels

2. RELATED WORK

It is a classification task that makes dense prediction of all pixels in an image. Remote sensing image segmentation methods are mainly divided into two categories. The first kind is based on the traditional artificial feature extraction and segmentation method. The second kind is the image segmentation method based on deep learning. In the traditional segmentation methods, the common image feature extraction includes: based on color feature, based on texture feature, and based on shape feature. Typical feature extraction algorithms include HOG algorithm, SIFT algorithm, and so on. The basic idea of Histogram of Oriented (HOG) is that the detected local object can be described by the distribution of light intensity gradient or edge. It constructs

features by calculating and counting the histogram of gradient direction in the local area of an image. HOG feature combined with SVM classifier is widely used in image recognition. SIFT algorithm obtains features by finding descriptors of feature points in an image and their related dimensions and directions and then carries out image feature point matching. SIFT algorithm is a local feature extraction algorithm, which can maintain invariance in the rotation and scaling of the image, and also can maintain certain stability to noise. The advent of deep learning has revolutionized industries from software to manufacturing. At the same time, it greatly promotes the development of remote sensing image segmentation. The concept of deep Learning originates from the artificial neural networks, which is a research branch in the field of Machine Learning. Through deep learning, low-level features can be combined to form more abstract high-level features to discover the distribution characteristics of data. Among them, Convolutional Neural Network is a typical deep learning model for image segmentation. Based on the classification of training data, deep learning can be divided into supervised learning and unsupervised learning. In recent years, a new semi-supervised learning method has emerged, which combines partially labeled data with partially unlabeled data to train neural networks. In this paper, our approach is based on supervised deep learning.

3. RESEARCH OBJECTIVES

There are a large number of objects in remote sensing images, but the size and shape of the objects are often different. Compared with natural image segmentation, remote sensing image segmentation is more difficult. Especially in the segmentation of small objects and the edge of the object, it is easy to have the situation of wrong segmentation and missing segmentation, which leads to low segmentation accuracy. There are various reasons for low segmentation accuracy, and the common one is the loss of spatial information caused by above and below sampling. However, in the process of subsampling, the sampled signal may be distorted, which makes it impossible to recover spatial details, which is easy to be ignored in remote sensing image segmentation based on deep learning. In addition, this paper proposes a remote sensing image segmentation model based on adaptive filtering.

4. METHOD

4.1. Network Architecture

In this paper, the segmentation of our research is a pixel-level two-class classification. Pixels in the need to be divided into two different parts: foreground object and background. At the same time, each pixel in the foreground is assigned a uniform semantic label. We use pre-trained ResNet-34 as the encoder. The decoder is transposed convolution [21] as for upsampling. The backbone of our network is LinkNet. Firstly, we apply the proposed anti-aliasing module (AFM) before each downsampling operation in the network. Then, we insert the RFM module to eliminate the semantic gap between the corresponding levels of Encoder-Decoder. Thirdly, in the central part of the network, the GAM module is added to aggregate multi-scale contextual information while also avoid redundant information of background. Finally, function of sigmoid is used to classify the output of the network. Figure 2 shows our final model (ARG-Net).

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

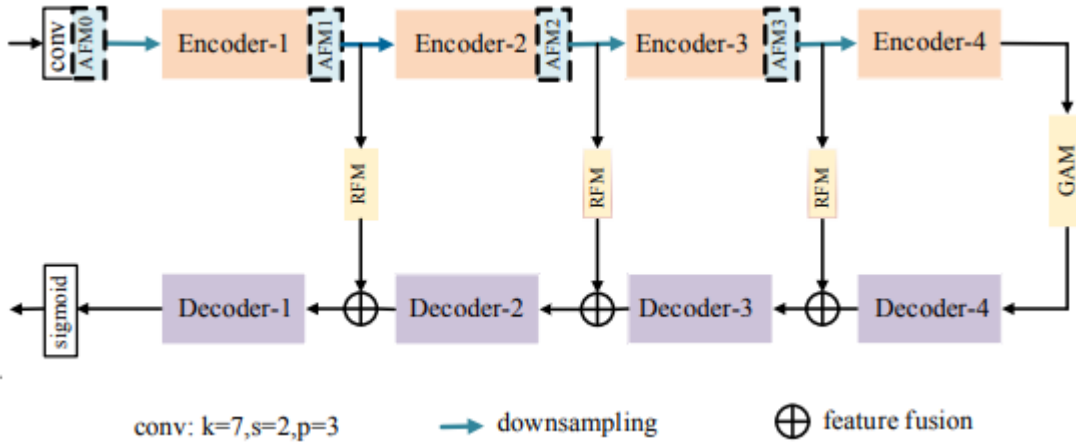


Figure 2. Architecture of ARG-Net. (1)

4.2. Adaptive Filter Model (AFM)

To enable anti-aliasing for ConvNets, we apply the proposed AFM module before each downsampling operation in the network. Inside the module, we first generate filters (a 3x3 conv filter) for different spatial locations and channel groups. Then we apply the predicted filters back onto the input features for anti-aliasing (Figure 3). In remote sensing imagery, low-frequency information tends to be relatively smooth, while high-frequency information tends to have obvious intensity. As frequency components can vary across different spatial locations in an image, the network needs to learn different filters across spatial locations. With the predicted filter, we apply it to input X :

$$Y_{i,j}^g = \sum_{p,q \in \Omega} \omega_{i,j,g}^{p,q} \cdot X_{i+p,j+q}^c \quad (3)$$

where $Y_{i,j}$ denotes output features at location i, j and Ω points to the set of locations surrounding i, j

In this way, the network can learn to blur higher frequency content more than lower frequency content, to reduce undesirable aliasing effects while preserving important content as much as possible.

Different channels of a feature map can capture different aspects of the input that vary in frequency. Therefore, in addition to predicting different filters for each spatial location, it can also be desirable to predict different filters for each feature channel. Motivated by the observation that some channels will capture similar information, we group the channels into k groups and predict a single low-pass filter for each group. Then, we apply a filter to the input X :

$$Y_{i,j}^g = \sum_{p,q \in \Omega} \omega_{i,j,g}^{p,q} \cdot X_{i+p,j+q}^c \quad (3)$$

where g is the group index to which channel c belongs.

Figure 3 shows a filtering process on a channel group. Each channel group has the same number of channels. In a channel group, different filters are applied at different spatial locations on a channel (where the filter size is $k \times k$ and the feature map size is $h \times w$). For this c/g continuous channel, there is a corresponding consistency in the filter at different positions on each channel. Finally, all channel groups are synthesized into a complete filtered feature map channel through concatenate operation.

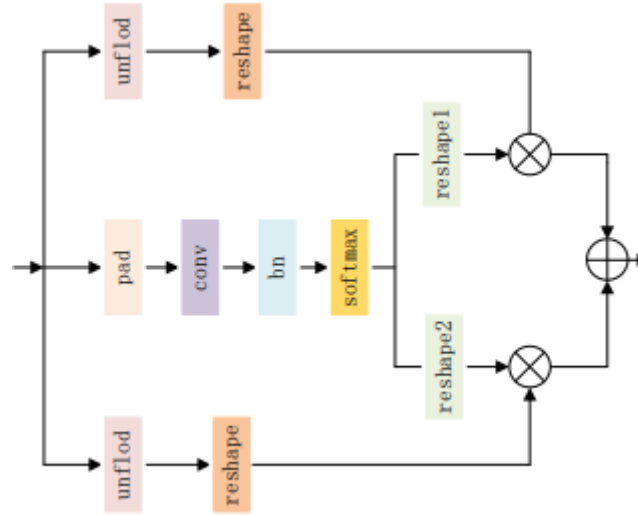


Figure 3. Architecture of AFM.

4.3. Global Attention Model (GAM)

For the large-scale variation in the remote sensing imagery, D-LinkNet used dilated convolution to increase the receptive field of the network, which the output feature map of each dilated convolution contains a larger range of object information. By adopting different dilated rates (where a small dilated rate can extract local information, larger dilated rate extract long-distance information), the network can extract useful feature information of different scales from different receptive fields. Meanwhile, it is helpful for enhancing the learned feature representation ability. However, the local information of the image will be lost when the dilated rate becomes larger and larger. What's more, the data sampled from the input becomes sparser, which is not conducive to the convolutional learning of small objects. And there is interference from redundant information in the larger receptive field information. To reduce this influence and further improve the expressive ability of features, an attention guidance module is proposed, as shown in Figure 5. In the original cascaded dilated convolution of D-LinkNet, we removed the dilated convolution block with $r=8$ and retained the part of $r=1,2,4$. In the spatial dimension, the global information of the foreground is adaptively captured by operations such as global pooling (GP) and 1×1 conv on the first branch. In the second branch, through 1×1 conv, the foreground information is further learned by the improved dilated convolution, while suppressing redundant and irrelevant information. Finally, features on each branch are fused to extract information of different scales.

$$I * k'(i, j) = \sum_h \sum_w I(h, w) k'(i - rh, j - rw) \quad (4)$$

where I is the input of the remote sensing imagery, k' is the size of dilated convolution kernel, h , w is the height and width of the image respectively, and r is the dilated rate.

$$k' = (k-1) \cdot (r-1) + k \quad (5)$$

where k is the size of the ordinary convolution kernel.

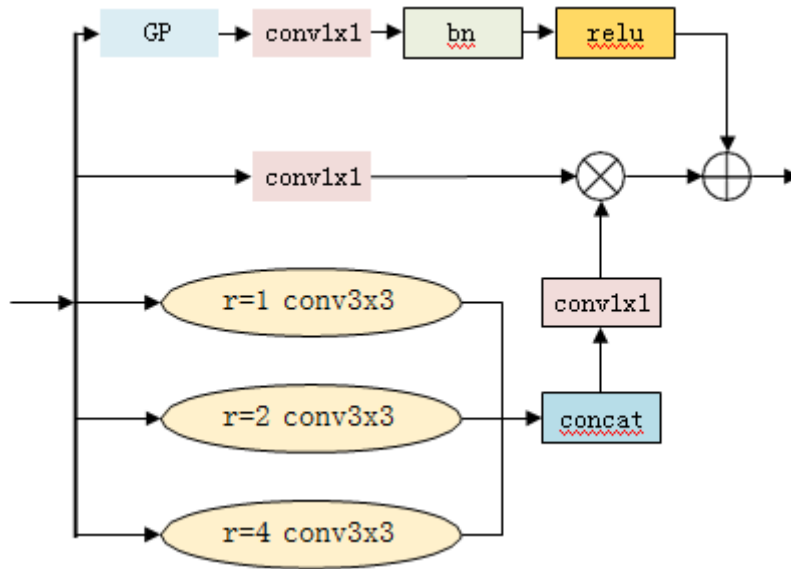


Figure 4. Architecture of GAM.

4.4. Residual Fusion Model (RFM)

D-LinkNet directly used skip-connection to connect the encoder with the decoder, which helps to reduce the loss of spatial information needed to restore the details of the image. The information of the encoder part is of a lower level, while the information of the decoder part is more high-level. There may be a semantic gap between the corresponding levels of Encoder- Decoder. Instead of combining the encoder feature maps with the decoder feature in a straight- forward manner, we pass the encoder features through a sequence of convolutional layers. These additional non-linear operations are expected to reduce the semantic gap between encoder and decoder features. Furthermore, residual connections are also introduced as they make learning easier and are very useful in deep convolutional networks.

The module structure is shown in Figure 5. Firstly, we reduce the number of channels through 1x1 conv to avoid the redundancy of calculations. Then, the low-level feature is learned through two 3x3 convs to reduce the semantic gap between the encoder and decoder.

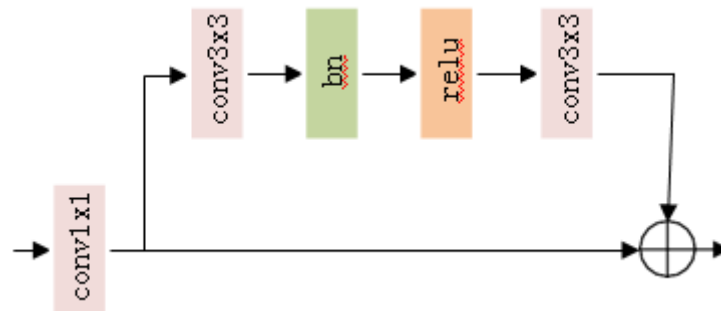


Figure 5. Architecture of RFM.

5. EXPERIMENTS

5.1. Dataset

DeepGlobe Road Extraction (Demir I, etc.) [22] is a road segmentation dataset, including covering cities, towns, suburbs, seashores, tropical rain forests, and other roads in different three countries, such as Thailand, India, Indonesia. The image size is 1024x1024, and the ground resolution is 0.5m. There are 6226 images in the dataset, which are randomly divided into 5226 and 1000 images. Among these images, 5226 images are used for training and the other 1000 images are used for testing. The RGB image of the road is in jpg format, and the corresponding label is in png format. Due to the large size of the original image, we crop all the images to a size of 512x512.

Inria Aerial Image Labeling is a building segmentation dataset, including 187,000 buildings in Christchurch, New Zealand. The image size is 512x512 and the ground resolution is 0.3m. The dataset used in the experiment has a total of 5736 images, of which 4736 are used for training and 1000 are used for testing. The RGB image format of the building is tiff, and the corresponding label format is also tiff. In remote sensing imagery segmentation, there are relatively few public datasets. Even in some public data, the number of images is far from the requirements for training the network. Therefore, we often use data augmentation to optimize training while preventing network overfitting. Image morphological transformation and color transformation are two common ways of data enhancement. In morphological transformation, there are horizontal and vertical flips, rotations of 90 degrees, 180 degrees, 270 degrees, and scale scaling. The color conversion includes the adjustment of saturation, brightness, and contrast.

5.2. Implementation details

All models are trained on the Intel Xeon (R) CPUE5-2640 v4@2.40Hz, which has two graphics cards of GeForceGTX1080Ti with 184.4GB RAM. We use Pytorch as a deep learning framework. We define the initial learning rate of the network as 0.0002. If the loss function does not decrease in 3 training epochs, the learning rate is reduced to 1/5 of the current one. The batch size is set to 8 and the optimizer is Adam[23], where $\alpha=0.9$, $\beta=0.999$, $\text{eps}=1\text{e-}8$.

5.3. Evaluation metrics

To accurately evaluate the segmentation effect of the module, we calculated four quantitative indicators. These indicators are precision, recall, F1 score, and IoU. Among them, P represents the proportion of the number of correctly predicted objects to all predicted objects. R represents the ratio of the number of correct objects to all positive samples and measures the classifier's ability to recognize positive classes. Since accuracy and recall are not conducive to the comparison of ablation experiments, F1 is often used as the harmonic average of them. IoU is another standard metric for segmentation, which represents the ratio of the intersection of the true value and the predicted value.

$$P = \frac{TP}{TP + FP} \quad (6) \quad R = \frac{TP}{TP + FN} \quad (7)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (8) \quad IoU = \frac{TP}{TP + FP + FN} \quad (9)$$

Where TP represents the total number of pixels correctly classified as the foreground object. FP represents the total number of pixels which the background is predicted to be the foreground object. TN represents the total number of pixels which the background is correctly determined as the background. FN represents the total number of pixels where the foreground object is predicted as the background.

5.4. Results and Analysis

We conduct comparative experiments on the DeepGlobe Road Extraction and Inria Aerial Image Labeling datasets.

We use ResNet-18 as our model encoder when we conduct ablation studies on AFM. In this experiment, the size of the filter is set to 3, to match the size of the convolution kernel in the ordinary convolution, which helps to improve the spatial dimension of the filter. As shown in Table 1, through grouping experiments with different channels, it is found that as the number of groups increases, the filtering performance of the network gradually improves. When $g=8$, it reaches the optimum. If the number of groups still increases, the performance may be degraded due to the over-fitting of the network. Of course, there could be other reasons.

The test result of DeepGlobe Road Extraction is shown in Figure 6. From left to right, it is the original image, label, LinkNet34 segmentation result, adding GAM segmentation result, adding GAM, adding RFM segmentation result, and the final ARG-Net (GAM+RFM+AFM) segmentation result. Among them, white represents the road foreground object and black represents the background. In the figure, the background in the first and second original images occupies a larger proportion. The first and third original images have large background differences. Besides, the roads vary in shape. These characteristics increase the difficulty of road segmentation.

As shown in the segmentation results of the first and second rows in the figure, by adding our module, the occlusion caused by the imbalance between background and foreground can be gradually improved. The interference of the redundant information in the complex background can be reduced. As shown by the segmentation results of the third and fourth rows, the contour of the small object can be segmented step by step by adding different modules. In the end, the ARG-Net improves the overall road segmentation and makes the road more connected. To quantitatively verify the road segmentation performance of the model, recall and F1 are used as evaluation indicators. The results are shown in Table 2. Compared with the original LinkNet-34, our final model increases the recall and F1 by approximately 3.2% and 3.6%, respectively.

Table 1. Ablation comparison experiment of different parameters in AFM on the test of DeepGlobe Road Extraction dataset.

Model	P	F1
LinkNet18	0.7720	0.7868
LinkNet18($g=2$)	0.7769	0.7890
LinkNet18($g=4$)	0.7851	0.7905
LinkNet18($g=8$)	0.7924	0.7981
LinkNet18($g=16$)	0.7855	0.7938

Table 2. Comparative ablation on the test of Deep Globe Road Extraction dataset

Model	R	F1
LinkNet34	0.7897	0.7906
LinkNet34+GAM	0.8101	0.8083
LinkNet34+GAM+RFM	0.8137	0.8097
ARG-Net(g=8)	0.8204	0.8267

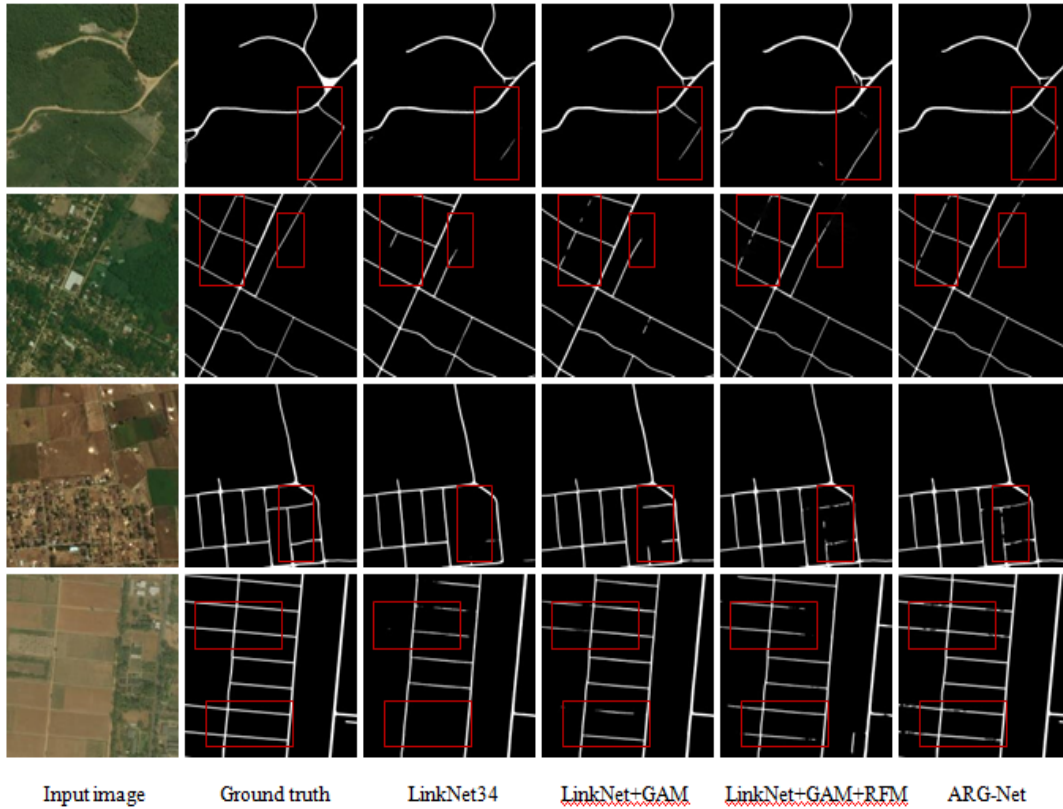


Figure 6. Example results on the test of DeepGlobe Road Extraction dataset.

The test results of Inria Aerial Image Labeling are shown in Figure 7. From left to right, they are the original image, label, LinkNet34 segmentation result, adding GAM test result, adding GAM, RFM test result, and final ARG-Net (GAM+RFM+AFM) test results. Among them, white represents the foreground of the building, and black represents the background.

As shown in the original picture, the size, color, and different backgrounds of the building image increase the difficulty of segmentation. As shown in Figure 7, the first line is an example showing that our module can gradually improve the missing points caused by the low contrast between background and foreground. The second line is an example showing that our method improves the jagged phenomenon that the original network will segment at the edge of the red building (enlarge the image to obtain high-resolution edges). In the last two lines, we can

gradually improve the omission of small target buildings. At the same time, we can also improve the misclassification of some small buildings. Compared with the original LinkNet34, the final ARG-Net model can improve the edge segmentation of small buildings. It can be seen from the last column of Figure 7 that the result segmentation of the building image has more regular,

smooth, and complete. To fully verify the segmentation performance of each module, IoU and F1 are used as the quantitative evaluation index for segmentation. As shown in Table 4, the final network ARG-Net is about 4.3% and 3.2% higher in IoU and F1 than LinkNet34.

Table 3. Results of different models on the test of DeepGlobe Road Extraction dataset

Model	F1
FCN-4s	0.7941
SegNet	0.7818
U-Net	0.7730
LinkNet34	0.7906
ARG-Net(g=8)	0.8267

Table 4. Comparative ablation experiment on the test of Inria Aerial Image Labeling dataset

Model	IoU	F1
LinkNet34	0.7166	0.8251
LinkNet34+GAM	0.7287	0.8436
LinkNet34+GAM+RFM	0.7433	0.8538
ARG-Net(g=8)	0.7579	0.8571

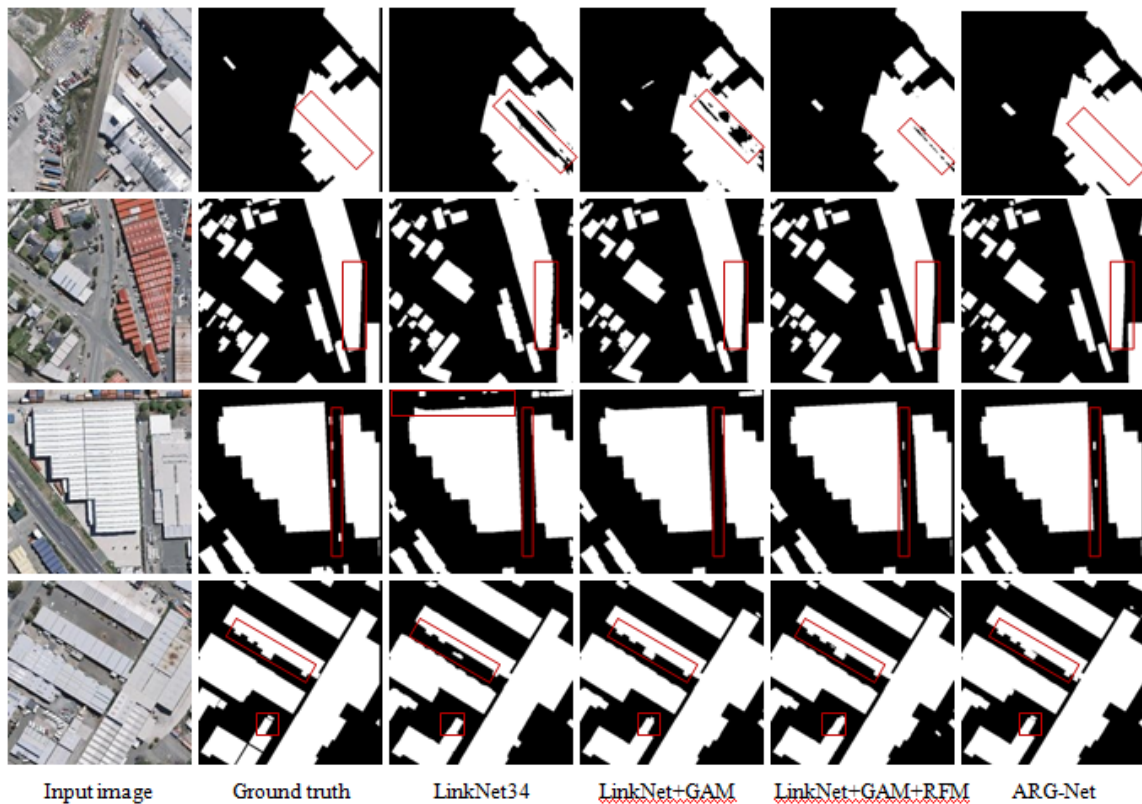


Figure 7. Example results on the test of Inria Aerial Image Labeling dataset.

The limitations or deficiencies observed during testing and evaluation of proposed system are as follows.

It takes a large amount of memory and a long time during training. In the study of the number of channel groups of AFM modules, a large number of network parameters are generated due to the need to predict the filter at each spatial location and each channel group. Therefore, the training will occupy a large amount of memory, at the same time, increase the training time of the network. For the original large remote sensing image, this may not be conducive to network training. The usual solution is to process these images to reduce the number of network training parameters.

Occlusion problem. The ARG-Net proposed by us can reduce the misclassification and missing classification of remote sensing images by the network to a certain extent. However, for some deep occlusion problems (as shown in Figure 6, the road is occluded by trees), the network can not be effectively identified and segmented.

6. CONCLUSION

In this work, we proposed an adaptive filtering layer, which predicts separate filter weights for each spatial location and channel group. This filter can avoid aliasing while preserving useful information. Through a sequence of convolutional during skip-connection between the encoder feature and decoder feature, the semantic gap can reduce. By dilated convolution, the network can ensemble multi-scale features in the center part while avoiding redundant information of background.

Since the production cost of pixel-level image segmentation data sets is relatively high, in the next research, we will focus on adding unlabeled data into the training of the network and using semi-supervised and weakly supervised methods to assist remote sensing image segmentation. In addition, how to balance the relationship between the precision and speed of segmentation network, so as to build an accurate and fast lightweight network model is also the direction of further research in this paper.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China Grant No: 61371156, and Anhui Province Key Scientific and Technological Research Programs Grant No: 201904d07020018. The authors would like to thank the anonymous reviews for their helpful and constructive comments and suggestions regarding this manuscript.

REFERENCES

- [1] Muhammad U, Wang W, Chattha S P, et al. Pre-trained VGGNet Architecture for Remote- Sensing Image Scene Classification[C]//2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018: 1622-1627.
- [2] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [4] Zhang R. Making convolutional networks shift-invariant again[J]. arXiv preprint arXiv:1904.11486, 2019.
- [5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [6] Ji S, Wei S, Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(1): 574-586.

- [7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
- [8] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753.
- [9] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.
- [10] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation[C]//2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017: 1-4.
- [11] Zhou L, Zhang C, Wu M. D-LinkNet: LinkNet With Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction[C]//CVPR Workshops. 2018: 182-186.
- [12] Ibtihaz N, Rahman M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020, 121: 74-87.
- [13] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.
- [14] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.
- [15] Hamaguchi R, Fujita A, Nemoto K, et al. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery[C]//2018 IEEE winter conference on applications of computer vision (WACV). IEEE, 2018: 1442-1450.
- [16] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.
- [17] Wang P, Chen P, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE winter conference on applications of computer vision (WACV). IEEE, 2018: 1451-1460.
- [18] Zheng Z, Zhong Y, Wang J, et al. Foreground-Aware Relation Network for Geospatial Object Segmentation in High Spatial Resolution Remote Sensing Imagery[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 4096-4105.
- [19] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.
- [20] Li H, Qiu K, Chen L, et al. SAttNet: Semantic Segmentation Network With Spatial and Channel Attention Mechanism for High-Resolution Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2020.
- [21] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1520-1528.
- [22] Demir I, Koperski K, Lindenbaum D, et al. Deepglobe 2018: A challenge to parse the earth through satellite images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018: 172-17209.
- [23] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

AUTHORS

Cong zhong Wu, corresponding author, born in 1964, male, professor. His main research direction is computer vision, pattern recognition and image segmentation. E-mail: 2315882652@qq.com



Hao Dong, born in 1994, male, master reading. His main research direction is remote sensing imagery segmentation. E-mail: 2018110990@mail.hfut.edu.cn



Xuan jie Lin, born in 1997, male, master reading. His main research is data mining and machine learning.

Han tong Jiang, born in 1997, male, master reading. His main research direction is data mining and intelligent computation.

Li quan Wang, born in 1994, male, master reading. His main research direction is medical imagery segmentation.

Xin zhi Liu, born in 1994, male, master reading. His main research direction is medical imagery segmentation.

Wei kai Shi, born in 1994, male, master reading. His main research direction is image segmentation.

FRAMEWORK FOR ENTERPRISE LOCAL AREA NETWORK DESIGN: AN OBJECT-CONNECTIVITY APPROACH

Sunil Seneviratne¹ and Rohan de Silva²

¹School of Engineering and Technology Central Queensland University
Melbourne, Australia

²School of Engineering and Technology Central Queensland
University Sydney, Australia

ABSTRACT

Local Area Networks (LANs) provide necessary infrastructure and services required for organizations to conduct their businesses efficiently and securely. While a small LAN could be designed and deployed in ad-hoc fashion, an enterprise LAN should be designed systematically starting from analyzing the business and technical requirements and constraints.

The Top-down network design methodology has been an excellent way to design a new network, however it has disadvantages considering its time-consuming process of designing networks. The organizational structures change in a fast pace in these days. These changes require their networks also to be realigned at the same pace. In this climate, a time-consuming design approach such as the Top-down design approach may not be the best option.

In this paper, we propose a framework of a network design methodology that could be used for quickly designing a LAN for an organization in this landscape while satisfying their business and technical requirements.

KEYWORDS

LAN, Network design, Object-connectivity, Top-down network design.

1. INTRODUCTION

Almost all organizations, big or small, use computers to undertake their day-to-day businesses today. LANs provide necessary infrastructure and services required for organisations to carry out their businesses in the most efficient and secure manner. When designing enterprise LANs, a systematic approach starting from analysing the business and technical requirements is required. In such a design, understanding the performance expected from the network stays at the forefront. The network performance is described by several parameters, including security, scalability, availability, reliability and Quality of Service (QoS).

Larsson [1] discusses the nature and complexity of the network design process, including the theoretical concepts, problems and solutions. The author provides details on programming aspects of combinational algorithms in designing communication networks. Al-shawi and Laurent [2] describe how to perform the conceptual, intermediate, and detailed design of a network infrastructure.

The Top-down network design methodology proposed by Oppenheimer [3], and the design and deployment methods of 802.11 wireless networks by Geier [4], are methods for designing and deploying networks recommended by Cisco Systems Inc.

The organisational structures are changing rapidly due to the market changes, restructures, mergers, forming alliances and flexible work patterns of employees. These changes demand that the organisational networks also be realigned at the same pace. When there is a need to be a merger of enterprise LANs of two different organisations, where the designer has to deal with two enterprises without disrupting the existing businesses while satisfying the evolving business and technical requirements, the success of the Top-down network design methodology is questionable. Also, the trend in modern organisations is to move the services to the cloud service providers while retaining the organisational network. In this climate, an expensive and time-consuming design approach such as the Top-down design approach may not be the best option. As such, our motivation in this research is to develop a modular design framework that can transfer the business and technical requirements into a suitable network design in the shortest possible time.

The rest of the paper is organized as follows. Section II deals with a discussion of relevant literature, including the Top-down network design methodology and Section III discusses the framework of our new modular design approach. We provide concluding remarks in Section IV.

2. BACKGROUND

Gen et al. [5] focus on the aspects of Genetic Algorithms (GAs) and their applications in design of difficult-to-resolve network design problems. In their research they consider a bicriteria LAN architecture design and apply a non-linear programming model to LAN architecture design problem. Their experimental results show that the spanning tree-based GA approach has a good performance on the bicriteria LAN architecture design problem. However, their research focus is limited to an architecture design rather than a complete network design framework or methodology.

Lima et al. [6] compare three Multiobjective Evolutionary Algorithms (NSGA-II, GDE3, and MOEA/D-DE) on Wireless Local Area Networks (WLANs). The authors identify several problem characteristics that need to be addressed in current WLAN configurations (location and coverage, channel assignment, and load balance) and how to mitigate them in practice. Testing is then undertaken to determine which one of the three algorithms performs most efficiently. This research is only concerned about WLAN configuration and not network design. The configurations are only a minor aspect that happen at a lower level in network design. As such, though this information may be useful to be considered in developing a new network design methodology, our focus is on the entire process of network design and not just the configuration part of it.

Giovanni and Surantha [7] utilise the top down network design methodology in order to create a converged network design suitable for business requirements in a government related enterprise. The research identifies the benefits of converged networks over traditional separated networks and examines previously proposed converged network designs. Moreover, the authors detail the methodology used to design the proposed converged network, including details of the testing performed and final results. Their research focuses primarily on converged networks for large-scale enterprises as opposed to smaller consumer networks.

This research is limited in designing a converged network for a single large-scale government enterprise, whereas our research topic requires coverage of enterprise LANs in different business

sectors. Further in their research they do not consider enterprise merges and there was no attempt made to understand the connection between the performance of the designed network and its architecture and components.

Sung et al [8] propose the use of a systematic approach to design enterprise networks which is time consuming and complex in nature. Their research suggests that the complexity of some enterprise networks exceed those of carrier networks, resulting in the need for more efficient and effective approaches for designing these networks. The authors analyse the viability of using a systematic design approach for enterprise networks, with a focus on VLAN design and reachability control through placement of packet filters as two design tasks in their systematic approach in designing enterprise networks. The feasibility of this design approach on large-scale enterprise networks is evaluated using an existing large-scale campus network.

As the authors point out one limitation in their work is, they have validated the performance in their heuristics only on a single network. Furthermore, as mentioned earlier, VLANs belong to the configuration part of a designed network, and hence, are only a very small part in the network design process. As such, their work is also not about a network design methodology but about network configuration.

Rozenblit et al. [9] in their design framework for designing LANs, discuss about an application of knowledge-based system design concepts to design LANs. They propose activities, which include, organizing a family of possible design configurations of the system being designed, inducing appropriate generic experimental frames, pruning the generic frames with respect to system entity structure, use of pruned substructures as skeletons to generate production rules for design models. These design models are evaluated via simulations studies. They further discuss about applying this framework to design LANs.

The authors describe four types of knowledge is needed to construct a design model of a LAN architecture. They propose to apply those details to generate all model structures that satisfy design constraints proceeded by the model construction process in hierarchical manner. Then the final design will be selected after evaluating the simulation studies. An advantage of this method as claimed by the authors is its objectives driven nature. It is arguable that this is an iterative and time-consuming approach. It has its limitations in applying for rapid LAN designs in dynamic environments.

Raza et al. [10] provide an analysis of the implications of network implementation choices on healthcare applications. The authors cite existing applications of network implementations in medical institutions in order to determine how different implementations have impacted on healthcare applications. Their research focuses on medical institutions that have recently implemented network technology for which updated enterprise performance statistics are available.

This research is limited to network implementation choices, and it does not consider network design choices. Further, the research focus is on healthcare applications only. Though implementation is not a part of network design, the network designer should foresee the implementation choices. This research findings may be useful to consider when we investigate the network design requirements of medical enterprise LANs. Furthermore, they analyse several different network applications and the challenges and benefits of each. Among the analysed options in their research is the use of Cisco Medical Grade Network solution which has many network design options for health institutions.

Oppenheimer's Top-down network design methodology [3] is a methodology for designing networks starting from the upper layers of the Open Systems Interconnection (OSI) reference model and then moving to the lower layers. This methodology focuses on applications, sessions, and data transport before the selection of routers, switches, and media that operate at the lower layers. It takes a systems analysis approach and starts the design process with requirements gathering. This design methodology goes through in sequence of analysing business goals, business constraints, and technical goals and trade-offs to create a logical network design. Addressing and numbering for the network and the security strategies are designed at this stage along with selecting routing and switching protocols for the designed network. According to this methodology, the next step is to design the physical network by selecting the technologies and devices for the network followed up with testing, optimising and documenting the designed network. Not only is this methodology time consuming, but it also appears that the biggest problem in this methodology is that there is no clear approach to map the requirements to the devices and the architecture of the network. The phases and the processes a network designer has to go through to design a network using the Top-down network design methodology are as shown in Figure 1 below.

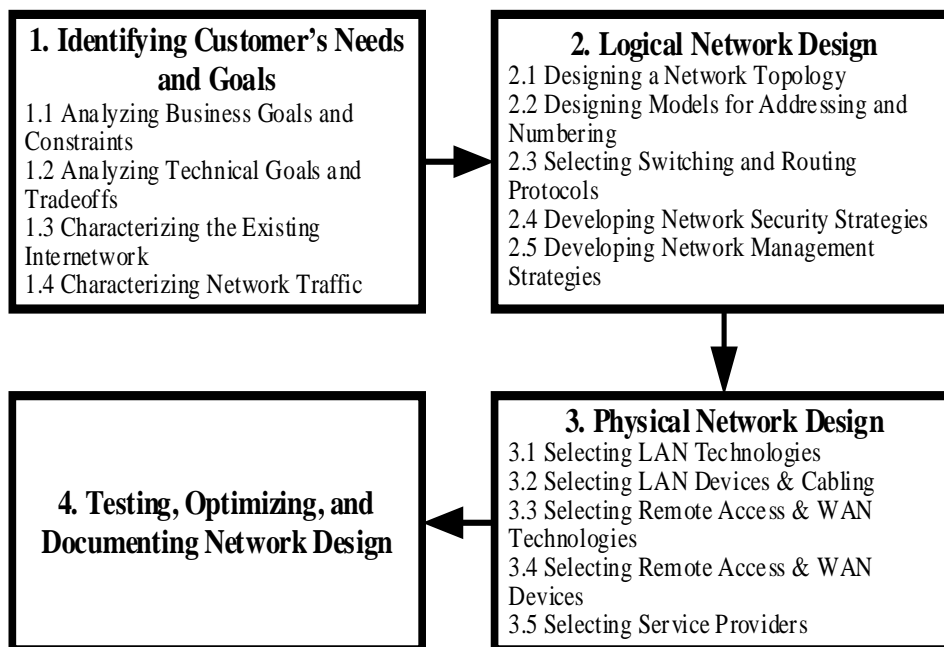


Figure 1. Top-down network design methodology. (Source: extracted from Oppenheimer [3])

As shown in Figure 1, the design process starts with requirement gathering, analysing goals and trade-offs, then designing the logical network followed by the physical architecture. Network designers who follow this methodology may come up with various design options and then need to select the suitable devices that meet the design criteria. It can be seen that the Top-down network design methodology is a tedious and time-consuming process, even though most of the design is undertaken systematically, at the end, the designer is not provided with a specific way of selecting the network components, including the software and hardware for the physical design. At this point, the design can become ad hoc as it is based on the experience, prejudice and perception of the network designer.

Object Oriented Design (OOD) is an approach used in other areas such as software design to solve software design problems. It is possible that OOD could be helpful in designing the new network design methodology to design a LAN in the shortest possible time.

Huda et al. [11] assess the effectiveness factor of OOD and develop an efficient model for effectiveness quantification using OOD constructs. The authors initially describe and analyse effectiveness factor, then use this information to develop a model for effectiveness quantification using OOD constructs.

The OOD constructs used in their analysis are encapsulation (correlated to Data Access Metrics), inheritance (correlated to Measure of functional Abstraction), coupling (correlated to Direct class Coupling), and cohesion (correlated to Cohesion among method), all of which together can be used to quantify the effectiveness of OOD. Their research focuses solely on evaluating the effectiveness factor of OOD using a testability approach. This research finding would be beneficial to our research in investigating the possibility of using OOD approaches to develop a framework of a modular network design methodology.

Georgatsos et al. [12] propose a new networking framework based on object-oriented programming (OOP), dubbed object-oriented networking (OON) in Internet technology and mobility integration. The authors describe the possibility of abstracting network layer resources as objects with attributes and methods, akin to data storage in objects using OOP applications, in order to avoid compatibility issues and create more sustainable and scalable network infrastructures. Their research focuses on the existing Information-centric networking (ICN) approach to network infrastructure, and the benefits of OON in comparison to it. Furthermore, significant documentation is provided explaining the operation of OON and its use of data and information networking layers to seamlessly transfer data.

The literature survey shows that there are no previous design methods or frameworks that can match the business and technical requirements to the finished design in a few days.

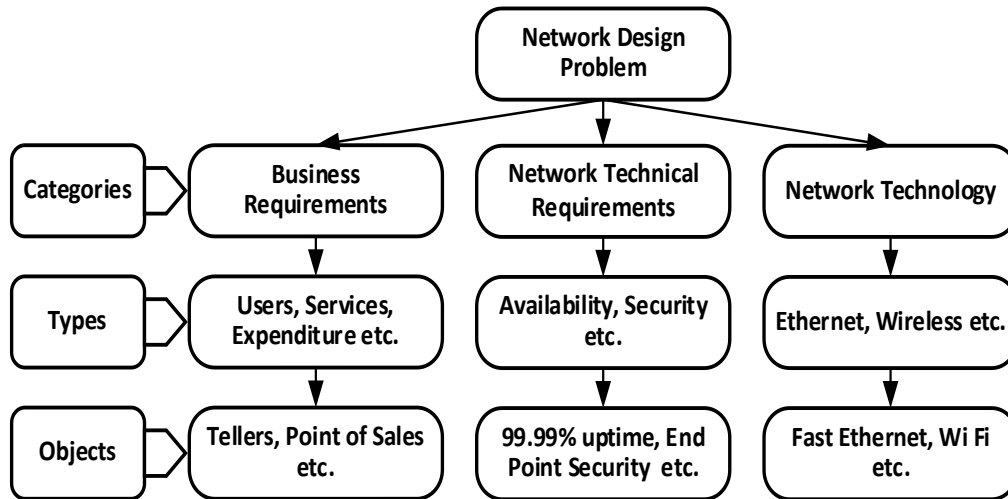
3. PROPOSED FRAMEWORK

Therefore, in this paper we propose a framework as to how categories, types and objects could be used in creating a new network design methodology to transfer the business and technical requirements into a suitable LAN design in the shortest possible time. As opposed to previous work, our approach is to use objects mapping in the entire design process.

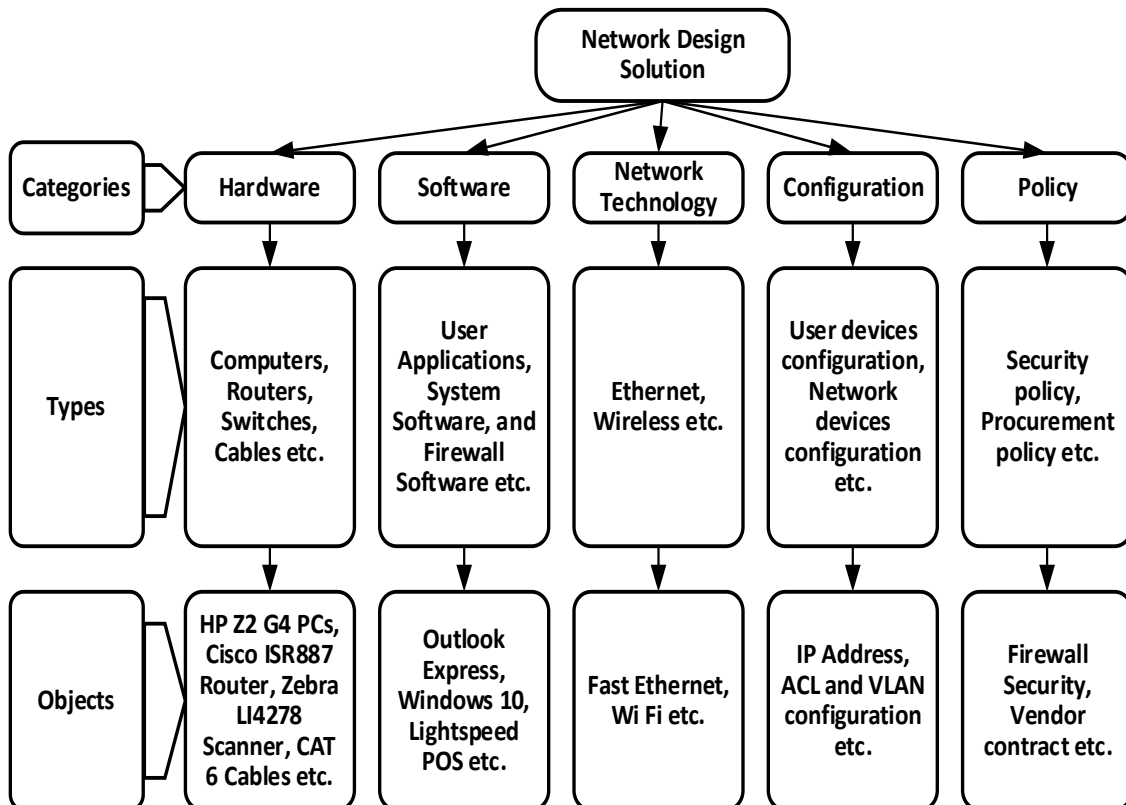
We see the entire network design task as a transformation of a problem to a solution. Here, the problem is the set of requirements and the solution is a functioning network that satisfies the requirements. In our approach, the business and technical requirements as well as the network technology are categories. The designer may not be able to specify the network technology through requirements gathering as the business heads and users, or the system engineers or technologists may not have any special inclination for a particular technology. Thus, it becomes a designer's choice, but we still call it a requirement as it is required for the design. As such business and technical requirements as well as network technology can be called just requirements. Each category contains types, and each type has several objects. Individual requirements can then be converted to objects that have various attributes.

Through the design, we convert the requirements to a number of categories that belong to the network (the solution). These categories are hardware, software, network technology, configuration and policy. Each of these categories contains types. For example, hardware category has switches as one type, and each design of switch is an object that has certain

attributes. Each object has the connectivity as a compulsory attribute. The connectivity attribute indicates the list of other objects to which the object was connected in the design. Examples of categories, types and objects involved in our network design are shown in Figure 2. Note that network technology is a category in the problem as well as in the solution. rules must be followed strictly.



A). Categories, Types and Objects of Network Design Problem



B). Categories, Types and Objects of Network Design Solution

Figure 2. Example of Categories, Types and Objects of the Proposed Framework.

Now, assume that we have a set of requirements for an enterprise LAN design and a completed network design that satisfy these requirements. The network may have been designed using Top-down network design methodology. Since the network satisfies the requirements, the objects in the types falling under the technical and business requirements categories must have been matched to the objects in the types that fall under hardware, software, network technology, configuration and policy categories (Figure 3). These mappings would be many to many. We can repeat this process for a number of different pairs of requirements and finished network designs. Then, we can create a dataset by selecting each object that belongs to the finished design in turn and the corresponding objects that belong to the problem (Table 1). We call an object that belong to the finished design as a network object (NO) and an object that belong to a requirement as a requirement object (RO).

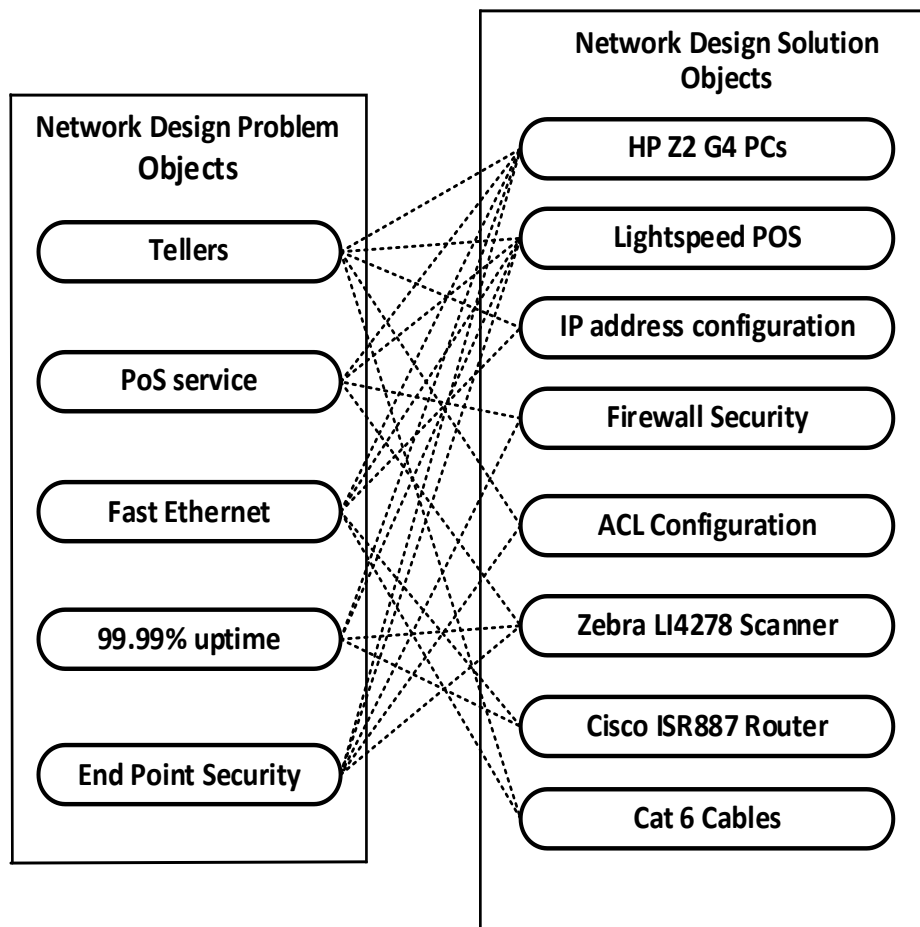


Figure 3. Mapping of Design Problem Objects to Design Solution Objects.

When we have a new network design problem, we can first extract the RO objects and using data mining techniques find NO objects in the finished design corresponding to each of these objects. Due to errors in data mining tools, it may happen that the designed network needs to be tweaked a little. But we hope that we can create a network much faster by this approach.

Instead of using objects, we may undertake the design at a higher level by mapping the classes. In this case, we need to identify recurring modules or patterns using a data mining tool.

Table 1. Sample Dataset.

NO	RO1	RO2	RO3	RO4	RO5
HP Z2 G4 PCs	Tellers	PoS service	Fast Ethernet	99.99% uptime	End Point Security
Lightspeed POS	Tellers	PoS service	Fast Ethernet	99.99% uptime	End Point Security
IP address configuration	Tellers		Fast Ethernet		
Firewall Security		PoS service			End Point Security
ACL Configuration	Tellers				End Point Security
Zebra LI4278 Scanner	Tellers	PoS service		99.99% uptime	End Point Security
Cisco ISR887 Router			Fast Ethernet	99.99% uptime	
Cat 6 Cables	Tellers		Fast Ethernet		

4. CONCLUSIONS

In today's business climate, change of business processes, restructuring of departments, mergers and acquisitions of businesses are inevitable. These changes demand that the networks are designed and implemented quickly to cater for the new business requirements. In general, a well-developed network design methodology such as Top-down network design methodology can be used for designing a new network or upgrading an existing network. However, the design process takes much time and may not be suitable for environments mentioned above.

As such, we have proposed a framework to design a network rapidly. Our approach is to convert the requirements into objects and find the connections or relationships between them, and the objects in the designed network. By repeating this process for many networks, we can construct a dataset or a table containing these relationships. Then, given the requirements of a new network design problem, we would be able to quickly find the corresponding objects of the new network using this dataset. Thus, we can complete the network design task very quickly.

As future work, we will develop this methodology further and create an actual dataset. After testing it on various new design cases, we will report our results in a future paper. We also hope to investigate the possibility of creating modules or design patterns and report our outcomes as well.

REFERENCES

- [1] C. Larsson, Design of Modern Communication Networks Methods and Applications. Burlington: Elsevier Science, 2014.
- [2] M. Al-shawi & A. Laurent, Designing for Cisco Network Service Architectures (ARCH) Foundation Learning Guide. 4th ed., Cisco Press, 2016.
- [3] P. Oppenheimer, Top-down Network Design. 3rd ed., Cisco Press, 2011.
- [4] J. Geier, Designing and Deploying 802.11 Wireless Networks: A Practical Guide to Implementing 802.11n and 802.11ac Wireless Networks for Enterprise-Based Applications, 2nd Edition. Cisco Press, 2016.
- [5] M. Gen, A. Kumar, J. R. Kim, "Recent network design techniques using evolutionary algorithms," In International Journal of Production Economics, vol. 98, Issue 2, pp 251-261, Nov. 2005.

- [6] M. Lima, R. Alexandre, R. Takahashi, and E. Carrano, "A comparative study of Multiobjective Evolutionary Algorithms for Wireless Local Area Network design," in IEEE Congress on Evolutionary Computation (CEC), pp 968-975, 2017.
- [7] Giovanni and N. Surantha, "Design and Evaluation of Enterprise Network with Converged Services," in Procedia Computer Science, vol. 135, pp 526-533, 2018.
- [8] Y. E. Sung, X. Sun, S. G. Rao, G. G. Xie and D. A. Maltz, "Towards Systematic Design of Enterprise Networks," in IEEE/ACM Transactions on Networking, vol. 19, no. 3, pp. 695-708, June 2011.
- [9] J. W. Rozenblit, S. Sevinc, B.P. Zeigler, "Knowledge-based design of LANs using system entity structure concepts," in 18th conference on Winter simulation, pp 858-865, Dec. 1986.
- [10] M. Raza, V. Kumar, A. Nafareih, and W. Robertson, "An Analysis of the Effects of Network Implementation Choices on Healthcare Applications," in Procedia Computer Science, vol. 94, pp 318-323, 2016.
- [11] M. Huda, M. Y. D. S. Arya, M. H. Khan, "Evaluating effectiveness factor of object oriented design: A testability perspective," in International Journal of Software Engineering & Applications (IJSEA), vol.6, no.1, January 2015.
- [12] P. Georgatsos, P. Flegkas, V. Sourlas and L. Tassioulas, "Object-Oriented Networking," in Computer Science - Networking and Internet Architecture, eprint arXiv:1502.07495, 2015.

GLOBAL SYSTEMS PERFORMANCE ANALYSIS FOR MOBILE COMMUNICATIONS (GSM) USING CELLULAR NETWORK CODECS

MaphuthegoEtu Maditsi, Thulani Phakathi,
Francis Lugayizi and MichaelEsiefarienrhe

Department of Computer Science,
North West University, Mahikeng, South Africa

ABSTRACT

Global System for Mobile Communications (GSM) is a cellular network that is popular and has been growing in recent years. It was developed to solve fragmentation issues of the first cellular system, and it addresses digital modulation methods, level of the network structure, and services. It is fundamental for organizations to become learning organizations to keep up with the technology changes for network services to be at a competitive level. A simulation analysis using the NetSim tool in this paper is presented for comparing different cellular network codecs for GSM network performance. These parameters such as throughput, delay, and jitter are analyzed for the quality of service provided by each network codec. Unicast application for the cellular network is modeled for different network scenarios. Depending on the evaluation and simulation, it was discovered that G.711, GSM_FR, and GSM-EFR performed better than the other codecs, and they are considered to be the best codecs for cellular networks. These codecs will be of best use to better the performance of the network in the near future.

KEYWORDS

GSM, CODECS, Cellular Network, G.711, GSM_FR, GSM-EFR.

1. INTRODUCTION

Cellular Networks play an important role in most organizations in the Information Technology (IT) field and countries as it potentially improve the economic growth and it also contributes extremely to human development and employment. Mobile communication networks equally help in sharing and gathering information [1]. With that being said, cellular networks have become very fundamental in our daily lives in a way that we cannot live without them, hence, the number of mobile subscribers to networks increases on regular basis. The evolving technology such as the 5G mobile network, makes today's network monumental in size and complexity. The consumption of mobile network data is so high that it puts a lot of strain on the structure of the network due to the limited radio spectrum. This may lead to performance degradation that causes lower Quality of Service (QoS) or poor network performance [2]. Moreover, poor services to the customers include dropped calls, insufficient bandwidth, and slow response time for data downloads to mention but a few. Good quality of the network will retain customers whereas a poor network service will cause a high level of churn for the cellular network operator. It is the duties of the Network Management System (NMS) to make the network as efficient as possible. The NMS is composed of Fault Management, Configuration Management, Accounting Management, Performance Management, and Security Management (FCAPS) functional areas and as well as traffic prediction and congestion control [3]. There are various mobile

communication standards such as GSM, Code Division Multiple Access (CDMA), Long Term Evolution (LTE), e.t.c. The motive for various communication standards was introduced to improve technical performance leading to an acceptable level of QoS in cellular networks/mobile communication networks [4]. GSM is the most reliable network coverage that has maximum capacity and confidentiality and it is a network used for transmitting mobile voice and data services. During this transmission of voice and data, the network may encounter various obstacles like phone echoes and loss of signals that may be caused by network latency, insufficient bandwidth, delay, jitter, destructive interference, congestion in the towers, and other external factors. This research work focuses on performance analysis of the Global System for Mobile Communications (GSM) network with the view to recommend the best cellular network codecs that will provide good performance and to determine features that need to be improved in the structure and operation of the network. The performance of the network will be analyzed through simulation where various codecs shall be compared and this is the basic building block for efficient transmission of data. The different codecs in the simulation to be used includes G.711, G.729, G.723, GSM-FR, and GSM-EFR. G.711 does not use compression at all and produce the best quality of call while G.729 at a low bit rate of 8kbps gives a good level of quality. This means one would be able to get more calls through bandwidth if you were to use the G.711 Codec [5]. GSM-FR is the first digital speech coding standard used in the GSM digital mobile phone system and operates at a bitrate of 13kbps [5]. GSM-EFR has better quality of speech and better robustness to network impairments [5]. G.723 is for speech and uses extensions of G.721 that provide voice quality using adaptive differential pulse code modulation [5].

This work is arranged as in the following: Section 2 explains the basic functions of a GSM network, Section 3 explains various methods for performance analysis, Section 4 presents the proposed methods, and section 5 presents the results.

2. THE GSM NETWORK

The GSM is a standard that was developed by European Telecommunication Standard Institute (ETSI) to describe various protocols for the second-generation digital cellular networks used by mobile devices such as mobile phones and tablets [6]. It was first deployed in Finland in December 1992 to meet certain criteria like support for a wide range of latest services, the security of transmission, compatibility with the fixed voice network, and the data networks are improved concerning the existing first-generation systems [7].

GSM as a network consists of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) where FDMA is about dividing the frequency band into different ones and TDMA allows the same channel of frequency to be used by different subscribers [8]. This makes every user have their time slot to use a specific frequency allocated to them. GSM is different in that it utilizes TDMA approaches to subdividing carrier frequency that is further segmented into separate time slots. GSM composes of three major systems namely, the Switching System (SS), the Base Station System (BSS), and the Operation and Support System (OSS).

2.1. The Switching System

The importance of the switching system is to perform call processing and functions of the related subscribers. The following are the functional unit of these switching systems:

Home Location Register (HLR), a database that is used to permanently store subscriber's data that includes a subscriber's service profile, location information, and activity status. The user is registered in the HLR of the operator that is buying a subscription from.

Mobile Service Switching Centre (MSC), is responsible for performing telephony functions of the system like routing calls and other services like FAX and conference calls.

Visitor Location Register (VLR) is responsible for storing temporary information about subscribers that is required by the MSC to service visiting subscribers. When a mobile station roams into a new area of MSC, the VLR that is always integrated into that MSC will request data about the mobile station from the HLR. In case the mobile station makes a call then the VLR will have the information that is required for setting up the call without the need to request it from the HLR each time.

Authentication Centre (AUC) provides security for network users by providing authentication and encryption capabilities.

Equipment Identity Register (EIR), stores information about the identity of mobile equipment that prevents calls from stolen, defective mobile stations.

2.2. The Base Station System (BSS)

The BSS consists of the Base Station Controller (BSC), Base Transceiver Station (BTS), and where all the functions related to radio are performed. The BSC is primarily responsible for all the control functions and physical links between the MSC and BTS.

2.3. The Operation and Support System (OSS)

All the equipment in the switching systems is connected to the operation and maintenance center. The main responsibility of the OSS is to service customers with support that is cost-effective for all operations and maintenance activities in the the GSM network. Figure 1 shows the GSM network structure [7].

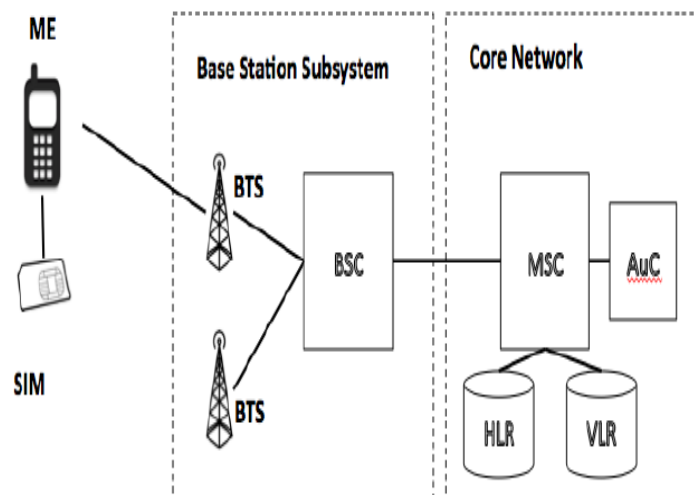


Figure 1. The basic structure of the GSM network

3. THE RELATED WORK

3.1. Literature Review

To establish the research gap from literature and to better understand the efforts of the work studied, a systematic review was carried out. As according to authors in [8], it is the critical and verifiable summary arising from the various publications that address the aim of this research done by various researchers globally related to the field of study.

Several researchers have proffered various solutions to performance analysis of GSM using cellular network codecs that are equally robust and challenging. Here the review of literature is provided to gain insight into what has been achieved thus far and to indicate how our work differs from previous work. The authors in [6] proposed performance analysis based on GSM Key Performance Indicators (KPIs) where they described the importance of pre-selecting relevant KPIs to focus on when analyzing and monitoring the network performance. The authors as above proposed the use of generating measurement data from a live network with the sole aim of analyzing and evaluating the results for optimizing the performance of the GSM network. Their work was proposed to contribute to the systematic examination of GSM mobile networks performance and end-user experience using four NCC KPIs metrics.

Moreover, the authors in [7] analyzed the performance of several audio encoding schemes using Resource Reservation Protocol (RSVP) based on Wireless Local Area Networks (WLAN). The codecs G.711, G.723, and G.729 were compared to come up with results using the RSVP services. On the other hand, voice codecs were used by the authors in [8] to examine the performance of mobile Worldwide Interoperability for Microwave Access (WiMAX). The simulation results showed that the performance of audio codecs stays fixed for random waypoint and group mobility manner while this is in contrast with our work because the performance of the codecs is not fixed for random ER_LAN call application of different nodes.

The most important KPIs from the author's view in [6] include Bit Error rate (BER), Frame Erasure Rate (FER), Bit Error Probability (BEP), and Mean Opinion Score (MOS). These are main KPIs that NCC used for rating the quality of service of cellular networks in Nigeria [8]. Wireless communication is expanding rapidly as most users are switching to it. The capacity of cells in the existing digital cellular mobile networks like GSM needs to meet the requirements that come with these changes in the network. The issue faced is how to increase the capacity of an existing network without causing performance or QoS degradation while our work focuses on a modeled network with increasing nodes that doesn't affect the service provided.

The authors in [9] proposed that the GSM band should be increased, along with increasing the number of serving channels or frequencies in an area. Although the overall spectrum of GSM is limited and it is divided between two or three network operators, leaving a spectrum of not more than 10 MHz for each operator. Another solution that was proposed by the author in [9] was to deploy more base stations or to introduce hierarchical structures like micro and pico-cells [10,11]. This method is limited to a denser base station grid resulting in increased interference. This limits the quality and capacity in terms of soft blocking.

Also, in literature, several analytical models based on continuous-time Markov chains have been proposed for studying the performance problems in GSM networks. The authors in [12] evaluated the impact of reserving channels for data and multimedia services on the circuit switch GSM network performance. Under a given GSM call characteristics, the authors in [13] developed a model called the Markov model that was used to analyze the performance of GPRS. The authors

in [14] also developed a Markov model to derive average data and restricting probabilities rates for GPRS in GSM networks.

In [15] the performance of a personal communication network based upon microcells was analyzed to find some fundamental phone traffic parameters such as blocking probabilities of new calls and handovers. A wireless system that has traffic scenarios based on Poisson time-dependent process is described using the fluid model [16]. The techniques to reduce dropped calls in progress due to failures of handovers are also proposed by the author in [17]. This approach describes several priority schemes, however, our work provides the use of codecs to maximize the QoS for cellular communications.

The authors in [18] analyzed the performance of a hierarchical cellular system based on microcells and overlaying macrocells and also the advantage of introducing “tier handovers”. These are handovers among cells that are from various hierarchical levels. As cellular networks collect a vast amount of measurement information that can be utilized to measure the performance of a network and QoS, the author in [19] studied the application of different data-analysis methods for processing the available measurement information. It is studied to produce more adequate methods for optimizing the performance while our study focuses on selecting the best performing codec through network simulation to maintain a better QoS.

The authors in [19] propose expert-based methods for monitoring and analyzing multivariable cellular network performance data. This method enables the analysis of performance bottlenecks that impacts performance indicators in multiple networks. Also, methods for more advanced failure diagnosis have been proposed to identify the causes of performance bottlenecks. The authors in [19] studied the use of measurement information in the identification of relevant optimization actions that leads to good network performance and good QoS, whereas our research focus is on analyzing performance measures to identify relevant audio codecs that give the best performance.

4. METHODS AND MATERIALS

The preceding research methods were applied in this work:

4.1. Simulation Process

Simulation is a process that models the behavior of a network by measuring the interaction between different network nodes [13]. The primary purpose of simulations is identifying problems that exist in a network or troubleshooting for unexpected interactions with one that has not been yet constructed. Simulation is mainly used in performance analysis, comparison, or even management and also for determining how a network would behave in a real-life situation. The generated results of the simulation aids in identifying the performance [12].

Discrete event simulation is a method used to model real-world structures, it is used to simulate the performance and behavior of a network [14]. The performance of codecs G.711, G.723, G.729, GSM-FR, and GSM-EFR is simulated where no change in the models were assumed during packets transmission [14]. This is done to address issues related to cellular traffic generated and handovers between neighboring calls, on a certain network capacity (number of BS, MSC, and cell towers). The network will be configured with the same properties including mobility model, type of protocol, application method, application type, QoS class, simulation duration, and various codecs. Different network codecs will be simulated to select the best one

suitable for the networks. The codecs that produce a good QoS will be recommended for use in the network.

Table 1. Parameters

Parameters	Configuration
Mobility Model	Random walk
Protocol	GSM
Application Method	Unicast
Application Type	ER_Lang Call
QoS service type	UGS
Codec(s)	G.711, G.723, G.729,GSM-FR and GSM-EFR
Simulation Duration	100 s
Priority	High

In this study, the QoS measures used to ensure that quality is optimized in the network are throughput, delay, and jitter. Performance measures refer to factors or parameters that are utilized to measure network performance. Each network varies in nature and design, therefore there is various way to measure it.

Jitter- these are voice packets that arrive at regular intervals to be intelligible. Jitter measures the degree of variation in packets interval which can be caused by improper queuing and configuration errors. The equation 1 below is used to calculate jitter where J is jitter, PJ is packet jitter, and N is the total amount of packets [15].

$$J = \frac{PJ}{(N-1)} \quad (1)$$

Throughput- network throughput is the rate at which voice packets are successfully transmitted over a network channel. It is the sum of all received packets by all nodes [15]. The equation 2 for throughput where S is the transmission time, T is the throughput, P is the average packet size and N is the total number of packet received, is as follows

$$T = \frac{(N*P)}{S} \quad (2)$$

Delay- delay in voice communications is defined as “the time it takes for voice packets to be transmitted from source to destination” which may lead to delay and echo [16]. Delay is measured in two ways, one direction, and a round trip. In one direction, the transmission of packets is unidirectional while in round trip voice packets travel to and from the destination and back to the source. Delay is measured in milliseconds (ms) . The equation 3 shown below is used to compute delay where μ is the number of packets per second, D is the delay, and λ is the average rate at which packets arrive [17].

$$D = \frac{1}{\mu - \lambda} \quad (3)$$

5. RESULTS AND DISCUSSION

5.1. In this chapter, the results of the network model simulation from the previous chapter are presented. The two different scenarios with different codecs will be compared. The quality of service is as well analyzed based on the measures, jitter, throughput, and delay. The best codec that provides good quality and performance will be recommended for cellular network usage. Each scenario will be fixed for unicast application and they will be analyzed based on QoS performance measures.

5.1.1. G.711 And G.729

Figure 2 shows the throughput for application one that was generated during the G.711 simulation. As the time for transmitting the data increases, the throughput in the network also increases. The graph shows that close to 0.0114 Mbps voice is generated at 707 ms and after that remains constant. Figure 3 shows throughput results for G.729 simulation. Initially, there was a sharp increase in throughput that reached 0.0084 Mbps for the first 10000 ms, which gradually slows down from 20000 ms. As it, reaches 0.009 Mbps, the throughput starts to increase very slowly towards the end of the simulation.

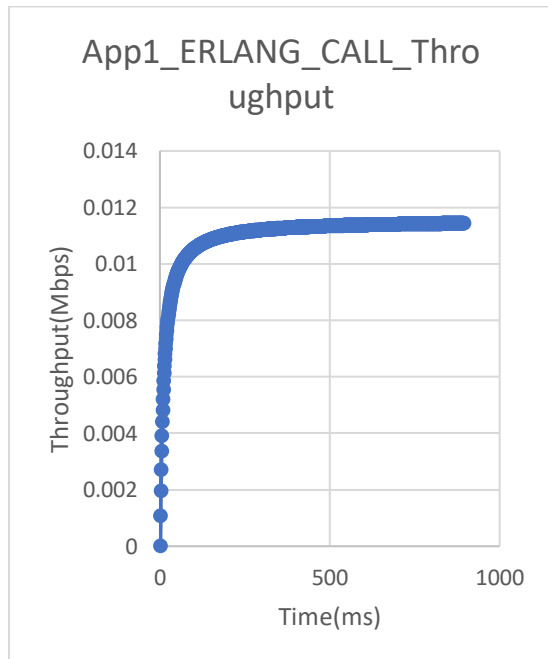


Figure 2 The throughput for G.711

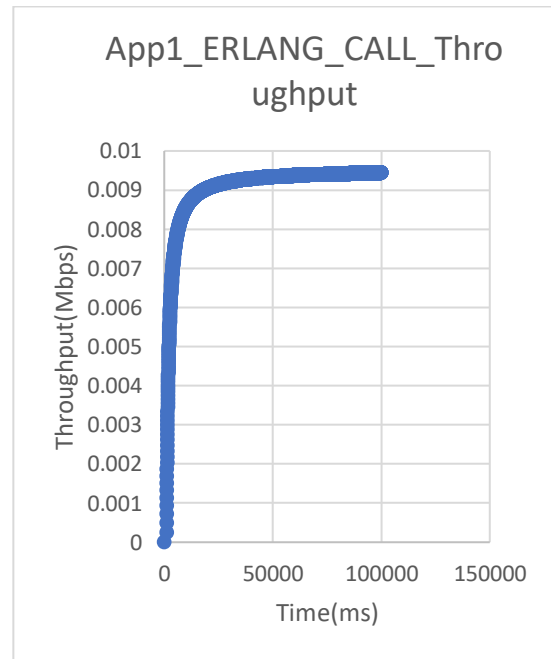


Figure 3 The throughput for G.729

5.1.2. G.723.gsm-fr and gsm-efr

In figure 4 the first 1000ms of simulation shows a sharp increase in throughput reaching 0.007 Mbps. The next 9000 ms shows a slow constant increase of 0.0079 Mbps until it reaches 0.007839 Mbps at the end of the simulation. Figure 5 illustrates the throughput for GSM-FR. Initially, there is a sharp increase in the throughput reaching 0.0055 Mbps. At the peak where simulation time was close to 19 990 ms, it begins to increase slowly from 0.007 Mbps until it reaches a constant throughput of 0.008 Mbps at 40000 ms of simulation time. More packets are transmitted in a lesser period, which results in good QoS, making codec G.723 desirable for use in a cellular network. The GSM-EFR figure 6 shows that throughput started to be transmitted at

0.0113 Mbps and begins to increase gradually with an increase in time from the beginning of the simulation.

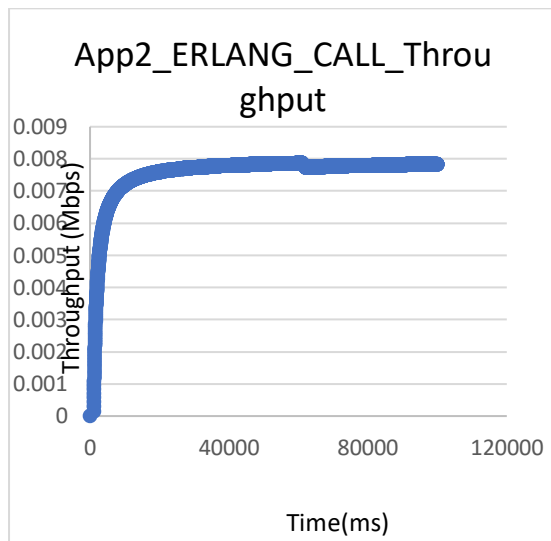


Figure 4. The throughput codec G.723

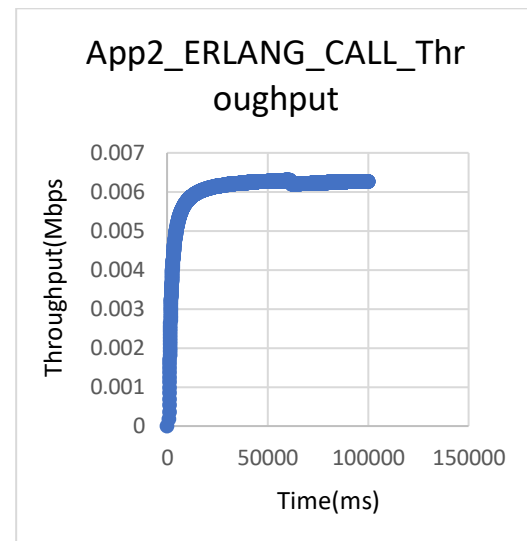


Figure 5. The throughput for GSM-FR

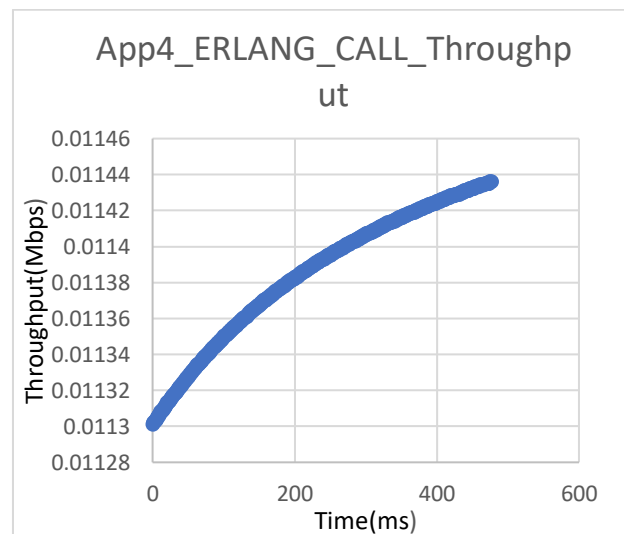


Figure 6. The throughput for GSM-EFR

5.2. 10-Nodes Scenario

A network model was created with 10 mobile stations. The results for throughput are shown in the scatter plot for each network codec.

5.2.1. G.711 and G.729

The following Figure 7 with an increased number of nodes illustrates that Initially, there was no data sent until 60000 ms where the transmission started, and then after increase rapidly with an increase in time up until 0.0044 Mbps. More time is required to transmit data. Figure 8 illustrates a rapid increase in the throughput initially, and the time it takes to send data starts to increase

after 0.003 Mbps. At 10000 ms, the achieved throughput is 0,008 Mbps. This gives a better performance of the codec.

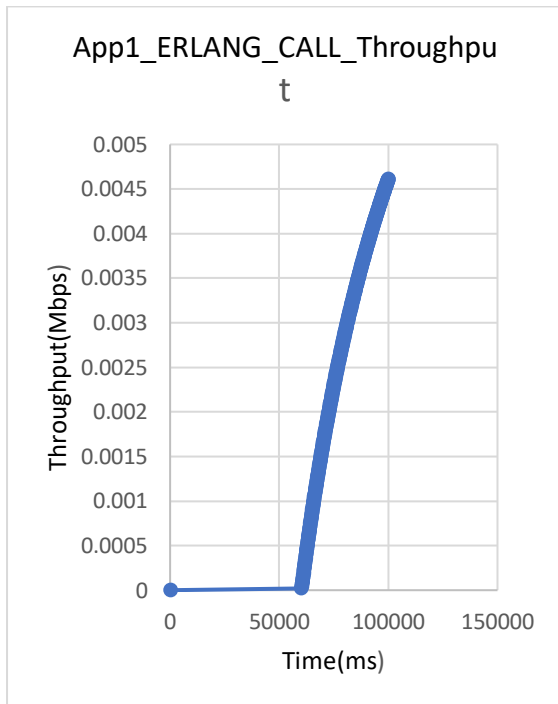


Figure 7. The throughput for GSM-EFR

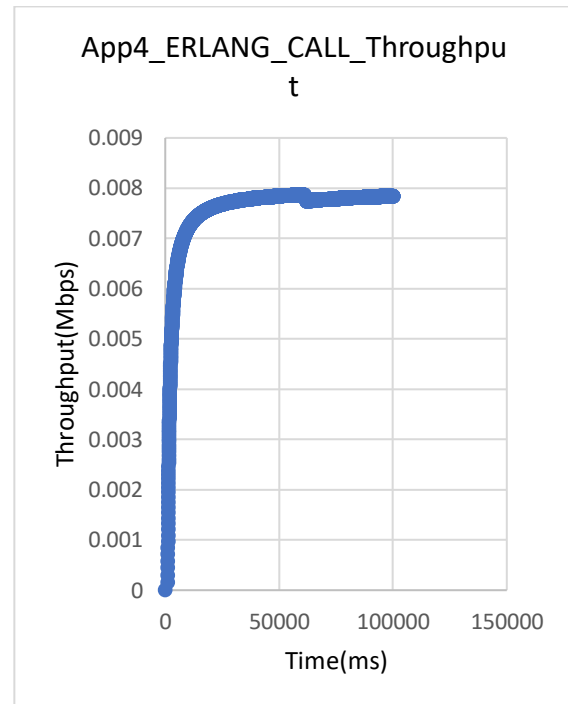


Figure 8. The throughput for G.729

5.2.2. G.723, GSM-FR and GSM-EFR

In Figure 9 the throughput for G.723 illustrates that the data sent increases as time increases. There is a gradual increase in throughput initially. After 10000 ms the throughput was 0.006 Mbps and begins to increase very slowly at 20000 ms of simulation time. There is a slight decrease of 0.0061 Mbps in throughput just after 60000 ms and continues with a slow increase of 0.0062 Mbps for the rest of the simulation time. Figure 10 shows that the throughput increases fast at the beginning of the transmission and slows down a bit to 0.004 Mbps after the first 10000 ms of simulation time. It gradually reached a peak of 0.0087 Mbps. Figure 11 shows the throughput for codec GSM-EFR increases gradually until it reaches 0.008 Mbps after the first 10000 ms of simulation time. The throughput begins to slow down from 0.0084 Mbps as time increases. It starts to be constant at 60000 ms with a throughput of about 0.0088 Mbps.

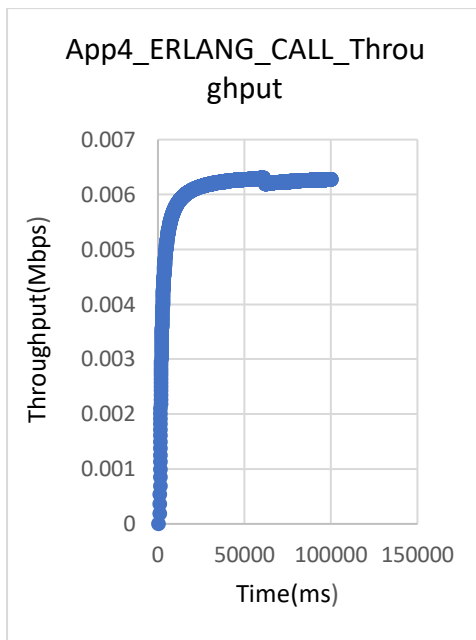


Figure 9. The throughput for G.723

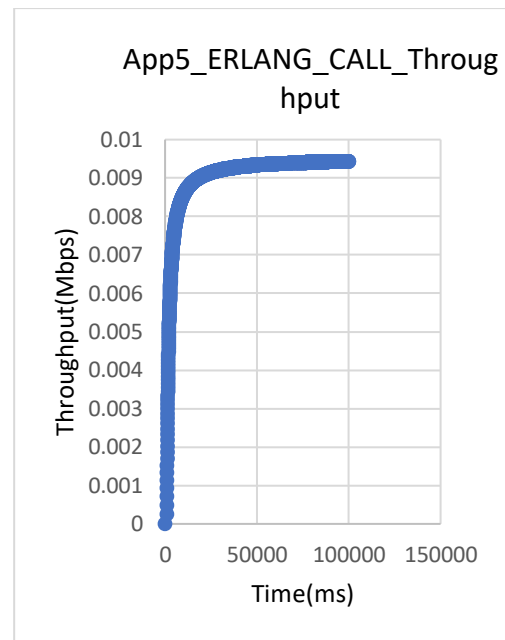


Figure 10. The throughput for GSM-FR

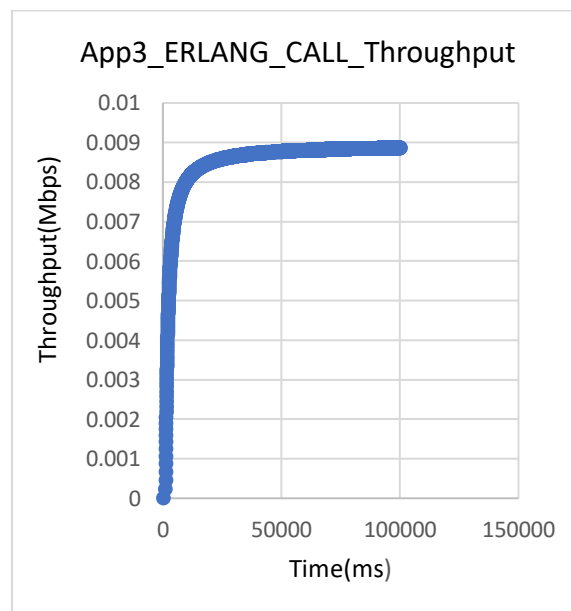


Figure 11. The throughput for GSM-EFR

5.3. Jitter and Delay

Jitter is the variation of the time between the packets that are arriving caused by the network congestion or route changes while throughput is the actual bandwidth that is measured with a specific time unit used to transfer data of a certain size [16].

5.3.1. G.711 AND G.729

The Figure 12 below for codec G.711 shows that there is a 31932.403879 μ s difference in delay for the 5-Nodes scenario and 10-Nodes scenario. According to the graph, there is no jitter experienced during packet transmission using this codec. Figure 13 for codec G.729 illustrates that the delay for 5-Node is less than that of the 10-Node scenario with 3.509242 μ s. The jitter increases with the increase in the number of nodes. The graph shows that the jitter for 5-Nodes is less than the jitter for the 10-Nodes scenario.

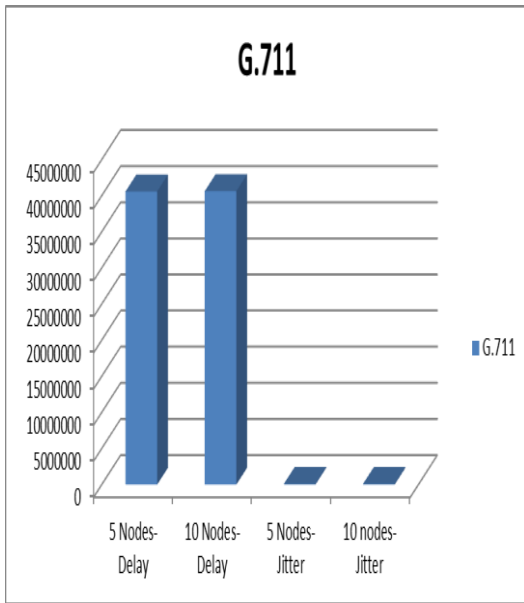


Figure 12. Variation in jitter and delay for G.711

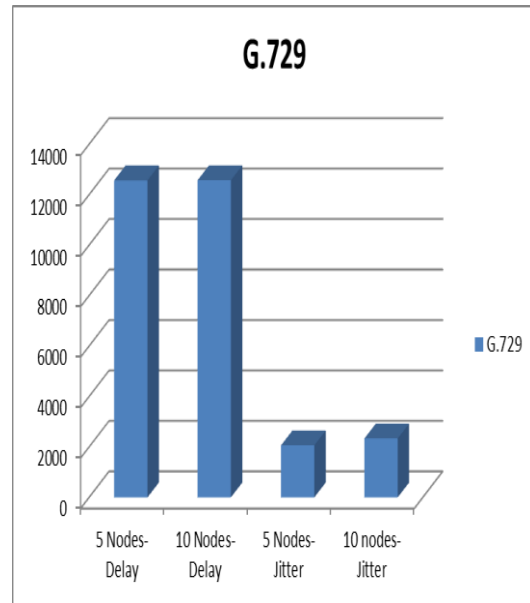


Figure 13. Variation in jitter and delay for G.729

5.3.2. G.723, GSM-FR and GSM-EFR

The following Figure 14 clearly shows that the delay experienced during transmission of packets using the G.723 codec is the same for both scenarios. The number of nodes does not have an impact on delay. There is high jitter of 12569.741364 μ s for 5-nodes and 12566.232122 μ s for 10-nodes connected to the network. Figure 15 for codec GSM-FR shows that for 5 nodes and 10 nodes, there is a delay of 14000000 μ s. There was no indication of jitter during the transmission of packets according to the results simulated for both scenarios. Desired data is transmitted but at an undesirable time. Figure 16 for codec GSM-EFR indicates that both scenarios have a delay of 14000000 μ s. During the transmission of voice, no data is lost and this leads to better QoS.

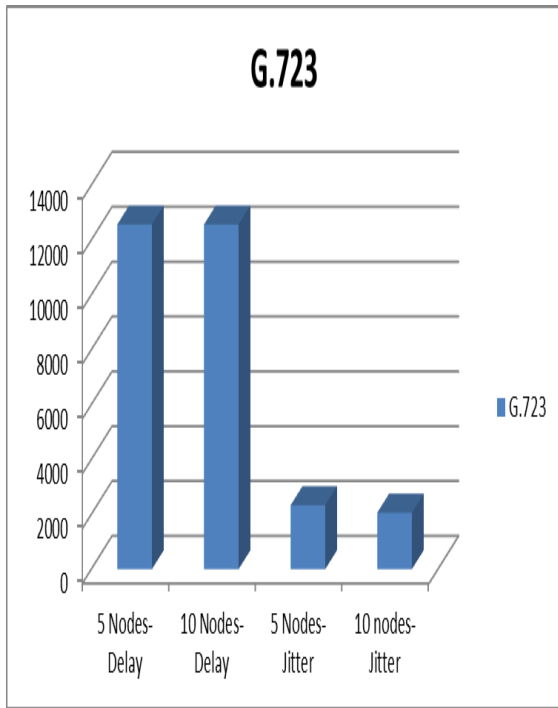


Figure 14. Variation in jitter and delay for G.723

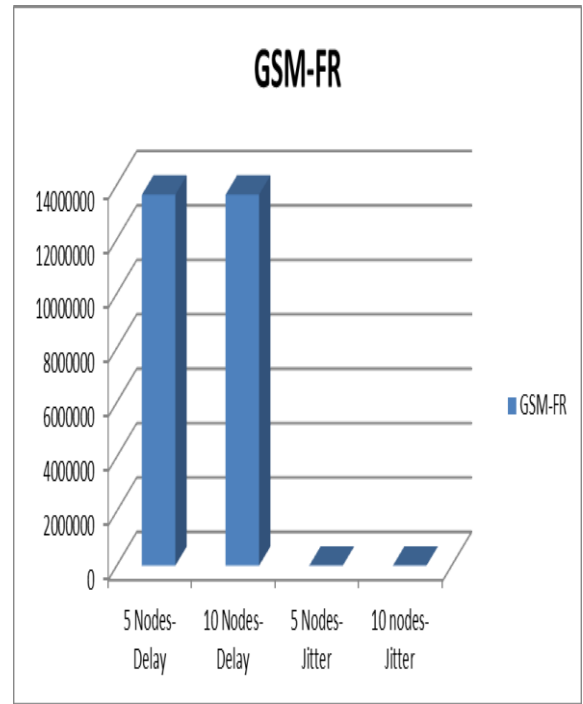


Figure 15. Variation in jitter and delay for GSM-FR

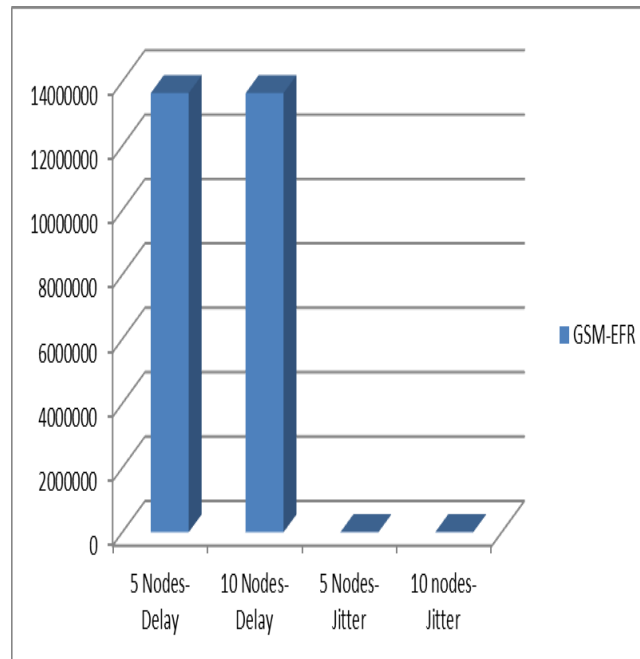


Figure 16. Variation in jitter and delay for GSM-EFR

5.4. Discussion

The variance in delay, jitter, and throughput among different codecs was simulated. G.729 operates differently compared to other codes. The lesser the nodes the higher the jitter and vice versa. More jitter is experienced for 10 nodes and less for 5 nodes. The delay that occurred in G.723 has a slight difference between the scenarios, it is almost the same and more jitter is experienced with fewer mobile stations. The codec G.711, GSM-FR, and GSM-EFR have a constant delay and does not experience any jitter for both scenarios, the number of mobile stations doesn't matter in this case. Therefore G.711, GSM-FR, and GSM-EFR will best suit the cellular network.

6. CONCLUSIONS

It can be concluded that the codec G.711, GSM-FR, and GSM-EFR showed the best performance for both 5-Nodes and 10-Nodes scenarios. G.723 and G.729 performed poorly for both scenarios. This was in terms of throughput, delay, and jitter. To maximize the value of QoS, it is very fundamental to appropriately utilize codecs in analyzing GSM network performance. The aim of this paper is shown from the simulation results that show a selection of G.711, GSM-FR, GSM-EFR as they produce significant results for the performance of the GSM cellular network. These codecs have no jitter and delay during transmission of data compared to other codecs like G.723, and G.729. This means that data packets will be sent in the desired period, and good quality of service will be provided to users using the selected codec for the network. As a results G.711, GSM-FR and GSM-EFR will positively impact the network performance.

ACKNOWLEDGEMENTS

This research work was supported by the research entity MaSIM of the FNAS, North-West University and our partners at TETCOS, India.

REFERENCES

- [1] O. A. Osahenwemwen and J. Emagbetere, "Traffic analysis in mobile communication in Nigeria," *Journal of Emerging, Trends in Engineering and Applied Sciences*, vol. 3, no. 2, pp. 239-243, 2012.
- [2] Yemini Y. "A critical survey of network management protocol standards", *Telecommunications Network Management into the 21st Century* (S. Aidarous and T. Plevyak, Eds), 1994; IEEE Press.
- [3] Barco, R., Canete, F., Diez, L., Ferrer, R., and Wille, V. (2000). *Analysis of mobile measurement-based interference matrices in GSM networks*.
- [4] B. Haider, M. Zafrullah and M. K. Islam, "Radiofrequency optimization & QoS evaluation in operational gsm network," in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1, pp. 1-6, 2009.
- [5] M. Hicham, N. Abghour and M. Ouzzif, "device to device(D2D) communications under LTE advanced network, *International Journal of Wireless & Mobile Networks (IJWMN)*, vol. 8, 2016, pp. 11-22.
- [6] Garg VK, Smolik K, Wilkes JE. *Applications of Mobile Agent in Wireless/Personal Communications*. Prentice Hall PTR, Upper Saddle River NJ (1997).
- [7] Priyanka Luthra, and Manju Sharma, "Performance Evaluation of Audio Codecs using VoIP Traffic in Wireless LAN using RSVP", *International Journal of Computer Applications*, pp. 15-21, Vol. 40, no. 7, February, 2012.
- [8] M. Tariq, M. Azad, and S. Rizvi, "Effect of Mobility Patterns on VoIP QoS in Mobile WiMAX", *International Journal of Computer Science and Telecommunications*, pp. 1-7, Volume 4, Issue 1, January, 2013.
- [9] K. Begain, G. I. Rozsa, A. Pfening, & M. Telek. "Performance Analysis of GSM Networks with Intelligent Underlay-Overlay". Dept. of computing university of Brandford BD7 1DP. February

2002. Available: <https://www.cse.iitb.ac.in/~varsha/allpapers/wireless/performance-analysis-of-gsm.pdf>
- [10] R. Steele, M. Nofal, Teletraffic Performance of Microcellular Personal Communication Networks, IEEE Proceedings-I Vol. 139, No. 4, August 1992.
- [11] S. Marano, C. Mastroianni, A Hierarchical Network Scheme for Multi-layered Cellular Systems, Proc. of the VTC 1997 Phoenix, Arizona, USA, May 1997.
- [12] M. Ajmone Marsan, S. Marano, C. Mastroianni, and M. Meo, Performance Analysis of Cellular Mobile Communication Networks Supporting Multimedia Services, Mobile Networks and Applications (MONET) 5. 167- 177, 2000.
- [13] R.J. Boucherie, R. Litjens, Radio Resource Sharing in a GSM/GPRS Network, Proc. 12th ITC Specialist Seminar 011 Mobile Systems and Mobility, Lillehammer, Norway, 2000.
- [14] M. Ermel, K. Begain, T. Muller, J. Schiiler, M. Schweigel, Analytical Comparison of Different GPRS Introduction Strategies, Proc. 3rd ACM Int. Workshop 011 Modeling, Analysis and Simulation of Wireless and Mobile Systems, Boston, MA, 3- 10, 2000.
- [15] Li Zheng, Liren Zhang and Dong Xu, "Characteristics of network delay and delay jitter and its effect on voice over IP (VoIP)," *ICC 2001. IEEE International Conference on Communications. Conference Record (Cat. No.01CH37240)*, 2001, pp. 122-126 vol.1, doi: 10.1109/ICC.2001.936286.
- [16] P. Lehtimäki. April 2008." Data Analysis Methods For Cellular Networks Performance Optimization". Public examination and debate. Available: <http://lib.tkk.fi/Diss/2008/isbn9789512292837>
- [17] Mishra, A. R., "Fundamentals of cellular Network planning and Optimisation", John Wiley and Sons, 2004.
- [18] Researchgate May, 2008. Discrete event Simulation [online]. Available: https://www.researchgate.net/publication/47705981_Discrete_Event_Simulation_of_Wireless_Cellular_Networks
- [19] R. Beraldi, S. Marano and C. Mastroianni, Performance of a reversible hierarchical cellular system, Wireless Networks 4(1) May, 1997.
- [20] Mohd Pardi, Mohd Baba, and Muhammad Ibrahim, "Analysis of handover performance in mobile WiMAX networks", IEEE conference on Control and System Graduate Research Colloquium (ICSGRC), pp. 143 – 149, 27-28 June, 2011.

SEVERAL TYPICAL PARADIGMS OF INDUSTRIAL BIG DATA APPLICATION

Hu Shaolin, Zhang Qinghua, Su Naiquan and Li Xiwu

Guangdong University of Petrochemical Technology,
Maoming, Guangdong, China

ABSTRACT

Industrial big data is an important part of big data family, which has important application value for industrial production scheduling, risk perception, state identification, safety monitoring and quality control, etc. Due to the particularity of the industrial field, some concepts in the existing big data research field are unable to reflect accurately the characteristics of industrial big data, such as what is industrial big data, how to measure industrial big data, how to apply industrial big data, and so on. In order to overcome the limitation that the existing definition of big data is not suitable for industrial big data, this paper intuitively proposes the concept of big data cloud and the 3M (Multi-source, Multi-dimension, Multi-span in time) definition of cloud-based big data. Based on big data cloud and 3M definition, three typical paradigms of industrial big data applications are built, including the fusion calculation paradigm, the model correction paradigm and the information compensation paradigm. These results are helpful for grasping systematically the methods and approaches of industrial big data applications.

KEYWORDS

Industrial Big Data, Paradigms, Big Data Fusion, Model Correction, Information Compensation.

1. INTRODUCTION

Since the concept of big data was accepted by the Chinese in 2013, a big wave of big data has emerged across the country. On the one hand, the government has raised big data into a national development strategy and established a series of big data research institutes or big data centers^[1]. On the other hand, for ordinary people, big data appear in communication in the way of common language in daily life. As mentioned in [1], the field of big data plays a vital role in various fields. But, in many different fields, the big data is just a term for massive data sets having large amounts of data, more varied and complex structure with the difficulties of analysing, storing and visualizing for further processes or results^[1].

Although there are still many fundamental problems to be solved, big data technologies and applications are still in progress. In fact, the Big data approach and its applications are very extensive in various fields. Paper [1] presented an explanation of these applications such as telecommunication, business process management and human resources management. Paper [2] introduced and analysed the application trend of big data in power industry. In this paper, the concept definition and several typical use paradigms of big data in petrochemical industry and other industries will be discussed. It is worth mentioning that these basic problems, such as what is big data and how to process practical big data, affect the enthusiasm of technical personnel to innovate big data technology.

Over the years, although the concept of big data has been receiving increasing attention, some basic issues have not been solved until now^[3,4], for example, what is big data, how to measure the big data, how to analyse and process the big data, and how to apply the big data. Especially in the industrial field, such as petrochemical industry, although the term "big data" has been widely mentioned, but there is no specific details that can be easily accepted by technical personnel. It is even more difficult for technical personnel to grasp the occasions and how to make full use of these "big data".

Up to now, the big data definition widely adopted by almost everyone is the descriptive definition of "4V" (Volume-large quantified volume, Velocity-fast velocity, Variety-diversified variety, Value-low value density, short as 4V). However, parts of 4V are inappropriate or ambiguous^[5-6]. For one example, in a long-running actual petrochemical system, a single-dimensional long-term sequence sampling data collected by a sensor for a long time, although it may be very large, is not enough to constitute the so-called "big data" in the field of big data research. For the other example, low value density is a concept that is very difficult to understand. Maybe some big data does not have much value for big data. The key points depend on which perspective you look at, but it's not that big data necessarily has a low density of specific data.

In order to overcome the limitations of 4V definition of big data, and to more fully and accurately describe the characteristics of big data, the second section of this paper will present a new set of descriptive definitions of big data from a new perspective, and explore the measurement of big data volume. Based on the new definition of big data, the third section of this paper will briefly describe several typical paradigms of big data applications. These typical paradigms will help us understand the actual use of big data and how to implement these practical applications. At the end of this paper, several conclusions will be refined.

2. MEASURABLE DEFINITION FOR INDUSTRIAL BIG DATA

As we all know, "big" is a vague concept without clear boundaries. In other words, the "Big" is a vague qualitative descriptive adjective. What is "big" data and how to measure whether the "big" data is big or not big enough? The "4V" definition for big data does not rigorously tell us what the big data is, and how to determine whether the data is big data or not. The fuzziness of big data concept directly affects the exploration and application of big data by engineering and technical personnel.

In order to overcome some limitations about the 4V definition for big data^[2,4], this section presents a new measurable definition that can be used to measure a data set is big or not.

Generally, the manufacturing or production process may continue for a period of time, the production process may be repeated again and again, and the products may be affected by various internal and external factors such as raw materials, environment, and processing disturbances, etc. So, the data sources of big data are extensive and have the certain time spans. In other words, the big data is a data cluster made up of many different data as well as data set:

$$S(\omega \in \Omega, t \in T, n \in N) \quad (1)$$

where the set Ω is the collection set of various sources, the set T is the collection set of various time passages, the set N is the collection set of various data dimensions.

In recent years, the concept of cloud has been widely accepted. Industrial big data is like a data cloud floating over factories. The cluster of big data described above can be represented graphically, as shown in Fig. 1.

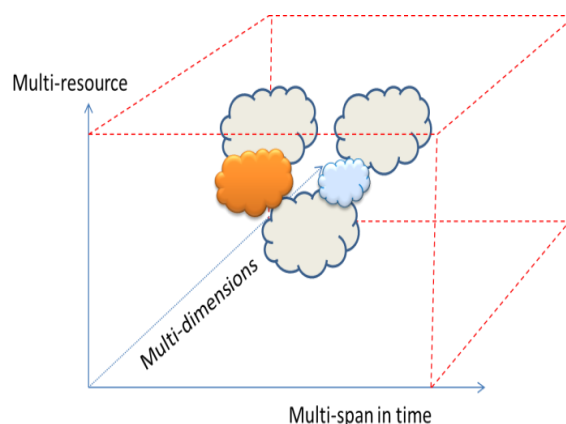


Figure 1. Big Data Cluster Diagram

Figure 1 shows that the big data cluster is very like several clouds floating in the air. Some of the clouds can be separated, and some clouds may be partially overlapped with other clouds. And, it can be seen from Figure 1 that big data has “3M” features: multi-source, multi-dimension and multi-span in time.

- (1) Multiple sources. In the field of industrial production and processing, data can come from different scenarios, different production periods, different production links, different objects, and different incentives. For example, in the petrochemical production process, data can come from the working conditions of petrochemical equipment, the production process of petrochemical products, the external environment of the petrochemical production process, the internal state of petroleum refining, the quality of petroleum raw materials, as well as personal factors, etc. All these data are important for the safety management of petrochemical processes. All these data come from different sources, just like a colourful cloud floating in the air, and they come together to form big data of petrochemical industry. In the petrochemical production management process, each cloud has its own rationality and necessity.
- (2) Multi-dimension: When we observe the real thing, we usually examine it from different perspectives and levels. Similarly, a petrochemical production process is taken as an example. In the production process, a large number of sensor devices are set up. Each sensor is like a person's eyes, observing a certain side of the chemical process, and obtaining information which display the states of the process partially. Observation information obtained from all different sides is brought together to form a set of high-dimensional data (even ultra-high-dimensional data). Although single-dimensional data is not enough to describe the overall change, the multi-dimensional data flow can be used to completely describe the changing process of petrochemical production.
- (3) Multi-span in time. Industrial production is an ongoing process. In a sense, actions performed at different times are repeated before or after other time points. The repetitive nature of process fragments is very useful for us to analyse process changes, judge working conditions, and diagnose abnormal working conditions. Data fragments at different periods or at different time points are also like colourful clouds floating in the data space. The

advancement of data over time is an important feature of industrial big data and an important aspect of big data.

Based on the Cloud diagram description and 3M characteristics of big data clusters, this section builds a set of measurement index and calculation methods to quantitatively measure the size of big data. For big data composed of multiple clouds, the size of the big data clusters is equal to the sum of the size of each piece of data cloud:

$$\|S(\Omega, T, N)\| = \sum \|S_i(\Omega_i, T_i, N_i)\| \quad (2)$$

So, if the number of clouds is large enough, or at least one cloud is large enough, the industrial data clusters can be called as big data.

In order to describe the size of a single cloud, we can reasonably assume that each cloud $S_i(\Omega_i, T_i, N_i)$ has a compact cube denoted by C_i , and the cloud can be embedded in the cube compactly. The volume of a cube C_i is equal to the norm of three sides.

$$\|C_i\| = \|\Omega_i\| \times \|T_i\| \times N_i \quad (3)$$

Equation (3) is computable and easy to calculate. As long as the formula (3) is used to calculate the size of each cloud, then the formulae (2) can be used to estimate the size of the "big data". In this way, we can give an intuitive measurement of the size of the "big data".

3. SEVERAL TYPICAL PARADIGMS FOR INDUSTRIAL BIG DATA CLUSTER

In the industrial fields, we should not only be able to reasonably estimate the size of the data cloud, but also grasp the approaches of solving technical issues with the data cloud. Generally, the big data clusters are often widely used in three different kinds of occasions: fusion calculation so as to improve accuracy for calculations and to use all usable information for statistics inference; model correction so as to provide a more basis for prediction and support decision making; information compensation so as to bridge over gaps between fragment information. The role of big data cluster is different in some different application fields^[7-9]. Correspondingly, the paradigm is also different.

3.1. Paradigm of Fusion Calculation

In industrial production, there are quite a large number of data information forming a piece of data cloud floating over the factory, such as the changing production process, raw material ratio data, production environment monitoring data, working condition data, and the data collected by various sensors continuously. These multi-source heterogeneous data form various types of data clouds. If these data clouds overlap, the overlapping data clouds can give us valuable measurement information of objects from different perspectives. Making full use of overlapping data clouds is helpful for us to improve the accuracy of calculation results.

Data fusion is one of the important ways of big data application. Intuitive understanding is that different data can bring different information. Combing together these data, quite a lot of information can be integrated together, which is helpful to eliminate errors and to correct prejudices. In this way, we get more and more accurate results and approximate correct inferences. In other words, data fusion technology is an information processing technology which is used to analyse various observations under certain criteria so as to complete the required

decision- making and evaluation tasks. Data fusion has achieved amazing development in the past ten years and has entered many different application areas.

In a big data environment, not every floating data cloud can participate in fusion. Data fusion is a purposeful activity. The data cloud that can participate in the fusion must be consistent in time, space, or object connotation. At the very least, if a data cloud can participate in fusion, it must be able to overlap after time traversal or space translation.

There are quite a lot of approaches for big data fusion. For example, for data layer fusion, the more widely used methods include least squares adjustment, etc.; for information layer and decision layer fusion, if all sheets of the floating data clouds $\{S_i\}$ are independent, we may use the Bayesian inference and stochastic decision making:

$$P(B|S) = \frac{\sum_i P(BS_i)}{\sum_i P(S_i)} = \sum_i \frac{P(S_i)}{\sum_i P(S_i)} P(B|S_i) \quad (4)$$

where B is the event to be inferred.

3.2. Paradigm of Model Correction

Because of its large size, big data makes it difficult for conventional statistical methods and conventional computer data processing software to work. The combination of machine learning, big data and artificial intelligence is a measure to solve the difficult problem of big data application.

This section describes a data modelling logic that combines big data with transfer learning shown in Fig. 2:

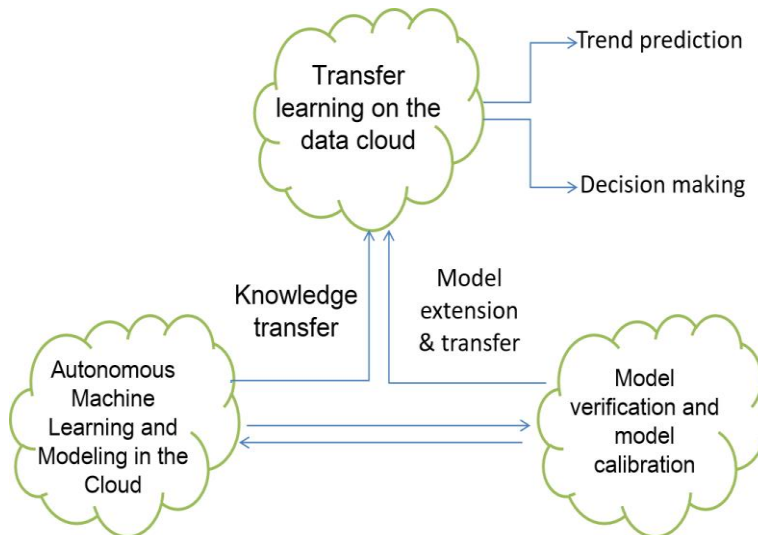


Figure 2. Prediction and Decision Based on Big Data Transfer Learning

The advantage of the above paradigm based on big data transfer learning is that the learning process is completed in the cloud chip. Due to the relative consistency of the data structure in the cloud chip, the learning process is relatively simple and easy to implement; the data of other

cloud chips is used to verify the learning model And perfection to ensure the inheritance and usability of knowledge and models.

Applying big data analysis technology to specific industrial big data, the above paradigm can ensure that the big data application process is understandable and interpretable

3.3. Paradigm of Information Compensation

Big data is a cluster of large amounts of data. Big data has a wide range of sources, that is, multiple sources. The biggest advantage is that it can give us the opportunity to understand objects from different perspectives.

Regular-scale data gives us insight into several sides of things or objects, rather than overall impressions, similar to six blind people touching an elephant. Everyone just touches the shape of one side of the elephant. Multi-source heterogeneous data big data can be used to fill the information gaps of the untouched parts.

Based on this consideration, the fourth typical application paradigm of big data is the filling and repair of information gaps. Specifically, a map Ψ is suitably constructed from the data cloud to the feature subspace:

$$\Psi: S \rightarrow \Theta \quad (4)$$

This map has the following properties: if the big data cloud sheet S_i contains the information of interest, then $\Psi(S_i) \subset \Theta$; otherwise, $\Psi(S_i) = \Phi$, an empty set. What we want to do is to find all the cloud sheets mapped to the non-empty sets from the big data cloud sheet shown in Fig.3.

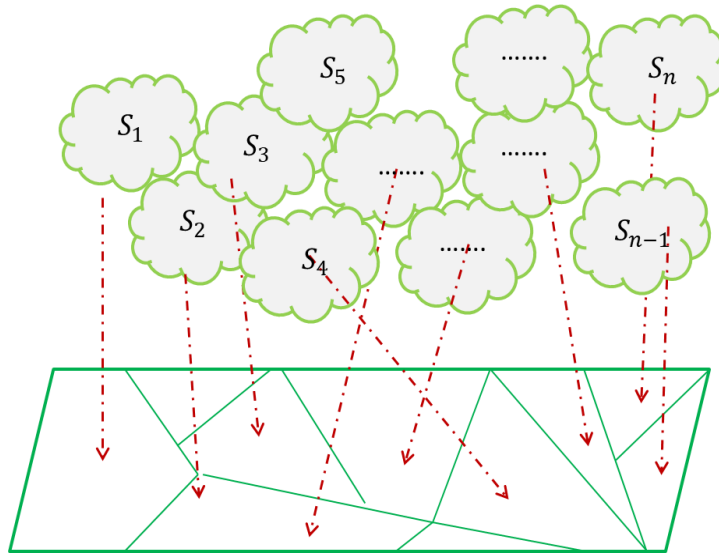


Figure 3. Mapping from big data to feature space

Obviously, if the phase space union of the big data clouds can cover the entire feature space of interest, we can use big data to achieve a complete understanding of the feature of interest. Otherwise, even if the volume of big data is large enough, it will not be possible to fully understand the changes in features from the big data cloud.

This paradigm stated above is helpful for us to adopt a problem-driven approach to delete or reduce unnecessary data blocks to achieve big data compression.

4. CONCLUSIONS

This paper intuitively proposes the concept of big data cloud and the 3M definition of cloud-based big data. 3M definition of big data is more suitable for industrial big data than 4V, which can fully reflect the time persistence, spatial distribution and source diversity of industrial big data. Moreover, 3M definition creates conditions for quantifying the "big" of big data. Based on 3M definition, this paper establishes a big data measure index based on the cloud compact cube volume for industrial big data, which has guiding significance and reference value for quantifying the size of big data.

On the basis of proposing big data cloud and measurement index, this paper establishes three typical application paradigms around typical applications of industrial big data, including the fusion calculation paradigm, the model correction paradigm and the information compensation paradigm, etc. These three typical application paradigms cover the basic form of the big data applications in the industrial field, which is helpful for systematically grasping the methods and approaches of industrial big data application.

ACKNOWLEDGEMENTS

This paper is financially supported by the National Nature Science Foundation of China (No.61973094, No.61933013).

REFERENCES

- [1] Loubna R., Nouredine F., Abdlekbir A., Belaid B. *Big Data Approach and its applications in Various Fields: Review*, Procedia Computer Science 155 2019,599–605.
- [2] Qian Zhang, Liyu Xia, Wangren Feng. *Analysis of the application trend of big data in power industry*. Journal of Physics: Conference Series, Volume 1616, 3rd International Symposium on Big Data and Applied Statistics 10-12 July 2020, Kunming, China
- [3] Li Ning. *Artificial Intelligence Paradigm in Big Data Era*. China Computer & communication, 2019, 8:104-105
- [4] Jarosław W, Jarosław J, Paweł Z, *Generalised framework for multi-criteria method selection*. Omega, 2019, 86:107–124
- [5] Manyika J, Chui M, Brown B, Bughin J., et al. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2017
- [6] Abbass H. A, Leu G, Merrick K.A *Review of Theoretical and Practical Challenges of Trusted Autonomy in Big Data*. Theoretical Foundations for Big Data Applications: Challenges and Opportunities, 2016
- [7] Ammu N, Irfanuddin M. *Big data challenges*. International Journal of Advanced Trends. Computer Science and Engineering, 2013, 2(1), 613-615
- [8] Kuchipudi S, Reddy T. *Application of Big data in Various Fields*. International Journal of Computer Science and Information Technologies, Vol. 6 (5), 2015, 4629-4632
- [9] Mukherjee S, Shaw R, *Big Data-Concepts, Applications, Challenges and Future Scope*. Wikipedia, https://en.wikipedia.org/wiki/Google_Cloud_Platform, 2016

AUTHORS

Hu Shaolin, professor, PhD supervisor, senior member of China Automation Society as well as China Digital Simulation Union Council; member of the editorial board of Journal of System Simulation and Journal of Ordnance and Equipment Engineering; director of the academic committee and technical director of "engineering technology center for monitoring and evaluation of smart city infrastructure" of Guangdong province. He has been committed to process monitoring, safety control, statistical learning and big data technology and application research of dynamic systems with complex structures for a long time. He has presided over 8 national natural science foundation projects and over 5 major/key pre-research projects. He has published 5 books and more than 210 papers. His recent research interests are artificial intelligence and big data technology, situational awareness and fault diagnosis, process monitoring and system safety.



© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

AN INTELLIGENT SYSTEM TO ENHANCE VISUALLY-IMPAIRED NAVIGATION AND DISASTER ASSISTANCE USING GEO-BASED POSITIONING AND MACHINE LEARNING

Wenhua Liang¹, Ishmael Rico² and Yu Sun³

¹St.Margaret's Episcopal School, San Juan Capistrano, CA 92675

²University of California, Berkeley, CA, 94720

³California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

Technological advancement has brought many the convenience that the society used to lack, but unnoticed by many, a population neglected through the age of technology has been the visually impaired population. The visually impaired population has grown through ages with as much desire as everyone else to adventure but lack the confidence and support to do so. Time has transported society to a new phase condensed in big data, but to the visually impaired population, this quick-pace living lifestyle, along with the unpredictable natural disaster and COVID-19 pandemic, has dropped them deeper into a feeling of disconnection from the society. Our application uses the global positioning system to support the visually impaired in independent navigation, alerts them in face of natural disasters, and remind them to sanitize their devices during the COVID-19 pandemic.

KEYWORDS

Geo-based navigation assistance, machine learning, mobile computing

1. INTRODUCTION

The advancement of technology has long been praised and cheered on as a significant accomplishment or representation of progress made for the convenience of the entire human race. However, as stairs are built toward the fancy and quick-paced future that we all anticipate to witness, little did we notice that inconveniences are left behind to some neglected populations like the elderly population and the population with disabilities. Indeed, while technologies have contributed to the many's well-being and revolutionized the future, it would be unfair to ignore how it has also widened the technological generation gap, where the elderly population falls behind to adopt to new technologies, and the [1] capacity-ability gap, where the population with disabilities feel challenged in face of technologies. [2] According to the Centers for Disease Control and Prevention's *Disability impacts All of Us infographic*, 61 million adults in the United States live with some type of disability, which is, in fact, 1 in every 4 adults in the United States. Among them, about 3 million experience blindness and difficulty with seeing. Most importantly, as technologies enhance in this rapidly-evolving world, the visually-impaired are put in face with more problems in their daily life; they have trouble navigating around the ever-changing city, looking for reading materials in Braille, and most importantly, gaining a sense of independence. Even worse, [4] a research conducted by Yan Wu's team through questionnaires shows that while as much as 77% of survey participants appreciate the value provided by

technologies, only 23 % of responders own smartphones (due to restricted accessibility and insufficient financial supports) compared to the 76% of UK adults who own their own smartphones. This clearly indicates that despite the immense technological advancement in this society, the visually impaired population experience way less from this advancement and even lack the opportunity to get hands on these technologies that many claim to be the success. Therefore, to respond to the [3] World Health Organization's call to "leave no one behind", our application aims to guide the visually impaired population toward independence.

Some of the systems that have been proposed to help the visually impaired with achieving a sense of independence are the mobility tools like white canes and the guide dogs. The visually-impaired people usually use the white cane to feel the changes on the surface of the pathway ahead of their steps. [5] By swinging the cane from one side to another in rhythm with their feet, the users will be able to find and be aware of the obstructions ahead of them. However, the effectiveness of the cane can be restricted when the cane breaks during a navigation, when poor weather or unfamiliar landmarks negatively influence its performance and, most importantly, when the person has to cross an intersection, considering that the white cane remains ineffective toward detecting vehicles in motion. In contrast, the guide dog, a trained pet for leading the visually impaired around obstacles, is aware of the danger imposed by the speeding vehicles and is thus a safer choice when it comes to crossing intersections. Indeed, according to *The Benefits of Guide Dog Ownership* by L. Whitmarsh, [6] guide dog is preferred by many for its mobility and companionship. However, a guide dog can be an unconvincing choice for many because it requires additional cares, training, and expenses on goods to feed it. While guide dog and white cane are the most common systems employed by the visually impaired, they only provide assistance in independent navigation but lack assistance in other forms. Another proposed system that has helped the visually impaired to achieve independence is the subscription to real-time agents in needs of assistance. The idea behind this is that the visually impaired population can receive assistance immediately through a touch of a button or a call through applications on smartphone or laptop. While this definitely provides more forms of assistance to the users other than on independent navigation, the expenses can be higher and the assistance would not be there to be accompanied with the user twenty-four-seven. In fact, some platforms would not work without internet access, providing inconveniences and restrictions in navigation when the user travels to locations with no internet and thus access to support. [7] Electronic glasses is another proposed system to help with not only the navigation of the user but in many other forms. In fact, its principle builds upon the capturing of real-time images through a camera and the displaying of these motions to the users' sight, which basically allows the users to live in a world with clear vision. However, this piece of device is costly and not covered by medical insurances, so though many hope to experience vision with such advanced technology, the majority of them cannot afford it. Moreover, this device only works for those who are not completely blind, as it functions upon enhancing one's vision by displaying processed images.

In this paper, we follow the same line of research by assisting the visually impaired population to achieve a sense of independence through integration of software, hardware, and machine-learning. We have three goals to fulfill through this application:

1. Save data of the number of obstructions detected by the device in a one hour walk done by a visually impaired person (frequency) with its corresponding longitude and latitude into a database. With such database and through machine-learning, we will be able to predict frequency at similar categorizes of location using the characteristics of the specific location and to understand the relationship between frequency and different interior layout better for designing the most traveling-friendly layout for the visually- impaired in the future.

2. Use frequency, longitude, and latitude to determine whether the user is indoor (in an enclosed space) or outdoor (at an open space) so that instructions can be sent to the users who are at an unsafe type of location, either an enclosed space or an open space, during the specific natural disaster.
3. When reached a specific frequency, the user will be alerted to sanitize the device to avoid potential contact with Covid-19.

Our method is inspired by our aims to understand what's best for the visually impaired population but to also support this community during natural disasters and the COVID-19 pandemic. There are some significant features on our application. First, our application is voiced by computer instead of human agents, allowing for a greater scale of flexibility and accessibility everywhere at any time. Second, our application covers functions outside of independent navigation like an alerting system during natural disasters, which contributes toward the independence of the visually impaired through different aspects of their life. Third, as a technology itself, our device acts as a potential bridge to close the capacity-ability gap between the visually impaired population and the society, and it also makes the positivity within the advancement of technology apparent to the visually impaired population. Therefore, we believe that this application will connect the visually impaired population closer to the society and encourage them to be more independent and confident.

Experiments are performed to calculate and compare accuracies for determining the most appropriate machine learning model through the implementation of support vector learning (svm) and regression models. In the experiment, we tested different models through adjusting the regression model, polynomial parameters, and inputted data sets.

The rest of the paper is arranged as follows: Section 2 provides the details on the challenges that the visually impaired population encounters; Section 3 proposes solutions in response to the challenges mentioned in Section 2; Section 4 discusses the relevant details about our process of experimentation, following by a presentation on related work in Section 5. Finally, Section 6 offers the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

A few technical challenges have been identified and listed as follows. We will present the solution in Section 3.

2.1. Challenge 1: Unsafe and Exclusive Urban Design

In today's society, the layout of cities and the interior design of many buildings are planned and implemented to closely match with the major population's aesthetic perception and their desire toward a sense of freshness. While the aesthetic within these designs brings pleasure to a great majority, a building's exterior appearance in the society's vision has unreasonably dominated over the many other factors that should've been more significant or, to be specific, safety and convenience. To the many who are blessed with clear vision and have the ability to navigate freely in spaces, the domination of appearance over safety and convenience in designs does not seem to be a problem. However, to the visually impaired population who have nothing to rely upon when navigating in an unfamiliar space and only spatial memory to rely upon in a familiar space, a simple yet safe interior layout design saves them time and decreases their risk of injury. The unsafe and inconvenient city design has been the major factor which discourages the visually impaired population to enjoy the same quality of life and to travel freely around a city, so a solution has to be proposed to create a more friendly environment for the independent navigation of the visually impaired population.

2.2. Challenge 2: Lack of Preparedness and Communication in Emergency Situation

Humans are usually alerted about dangerous presences through their five senses - sight, hearing, smell, taste, and touch. During unexpected natural disasters, those nearest to the situation observe the danger to make judgement about what to do next while those who live farther away yet will possibly be impacted receive warnings on smart devices to prepare ahead of time. The visually impaired population, however, lack information to make judgments in face of the sudden onset of natural disasters or to prepare for such situations due to their vision loss, so a solution has to be proposed to take their community's safety into consideration during such crises.

2.3. Challenge 3: Sanitization During and After the Pandemic

The emergence of COVID-19, also known as the coronavirus disease, is a disease that leads to symptoms like dry cough, fever, and tiredness. Though this pandemic has introduced challenges to everyone, newly enforced rules like social distancing and the adjustments newly made to many markets have been particularly tough for the visually impaired population to adapt. [8] According to an article from *The Conversation*, a public media source in the United Kingdom, the changes in operation in markets has created difficulties for the visually impaired individuals to navigate using a spatial map and that social distancing has been difficult for them to employ in some situation due to their vision loss or inability to notice when people are walking overly close to them. [9] Moreover, the World Health Organization (WHO) mentions that the visually impaired population are potentially at a greater risk to contract COVID-19 due to their need to frequently touch items near them for directions during navigation. To limit the spread of COVID-19 in the visually impaired community and the whole society for a more immediate end to this pandemic, a solution must be proposed to guide the visually impaired population about the best method of navigation during this time and to develop procedures for limiting the chance of being exposed to COVID-19.

3. SOLUTION

3.1. Overview to the Solution

The application has been implemented using MIT App Inventor, and it is carefully developed to serve as a multi-functional platform to support the visually impaired population in navigations, during natural disasters, and in the midst of the COVID-19 pandemic. The application intends to take all aspects into consideration when it provides features like QR code login, locative marker placement, vibration when detected obstacles, alert in face of disasters, GPS-frequency database, and sanitization reminder.



Figure 1. Overview of the solution

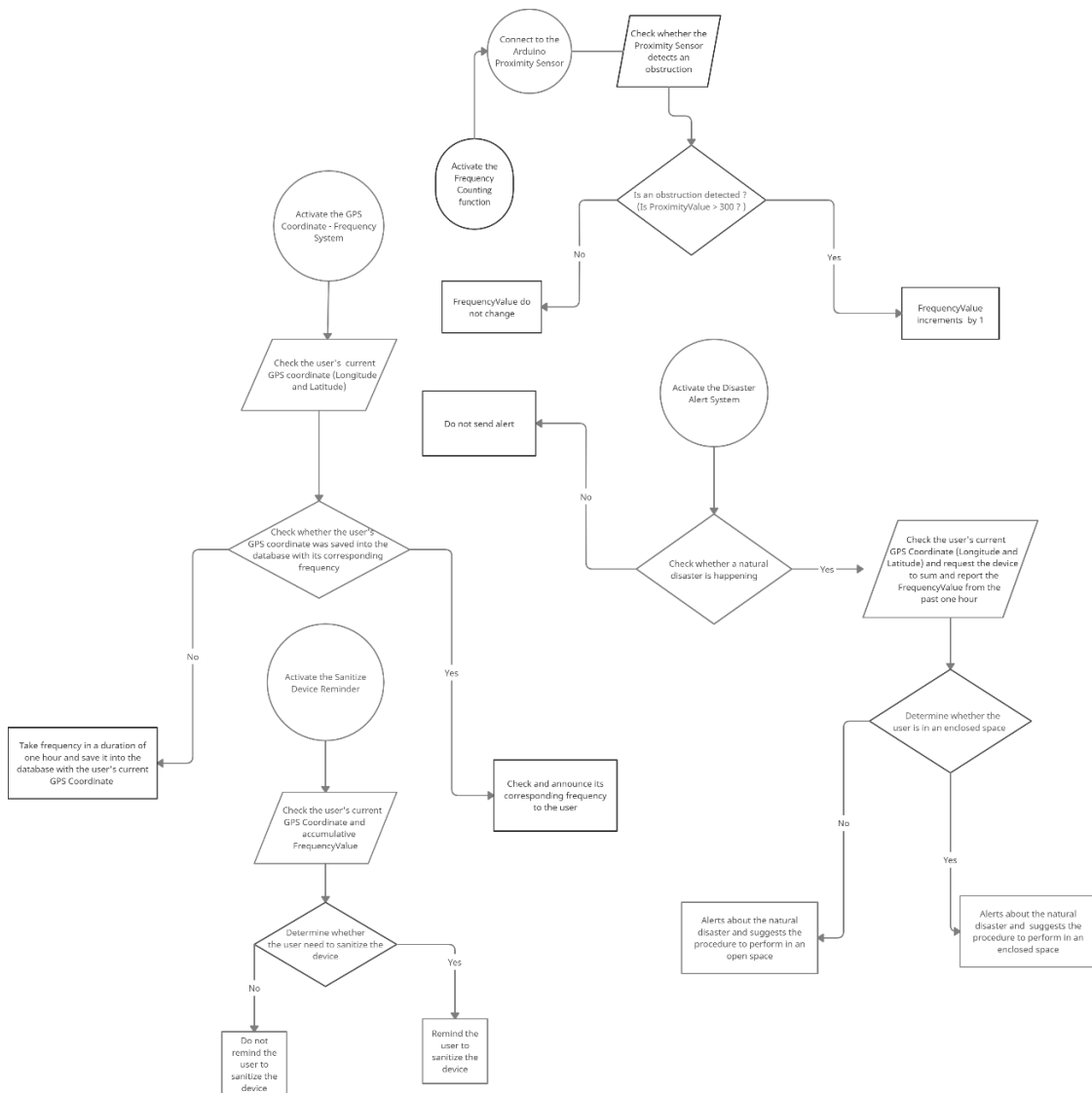


Figure 2. Overview of the solution (Flowchart)

3.2. QR Code Login System

Since the traditional login procedure of typing username and password can provide inconvenience to the visually impaired users and leads to reduced user experience, our application intends to prevent such inconvenience by allowing logins through QR codes. After the user signs up for an account, they can shake the smartphone to activate the QR code login system, which allows them to directly scan the code to login. The system also comes with voice guidance, which will provide feedback in response to successful login, allowing communication to happen directly between the application itself and the visually impaired users.



Figure 3. QR Code Login System

In our application, the Global Positioning System (GPS) is employed and used to display the user's current location on a map and in text. Longitude, latitude, and a detailed line of address are also displayed on screen after occurrences of new rounds of syncing. Through Bluetooth connection, this application can also connect to a smart cane which performs vital tasks like vibration in detection of obstructions and marker placement. The device -smart cane- itself is built using a proximity sensor, some wires, and a button from the Arduino kit. Shown in Figure 4, a proximity sensor is mounted near the tip of the smart cane to detect the surrounding or, specifically, the path before the user. Whenever the proximity sensor finds an obstruction that can distract and change the user's planned path for navigation, it will send a signal to the vibration motor to vibrate and through it, alert the user to change direction. Shown in Figure 2, the detection of an obstruction will also increment the frequency value displayed in the application by one, keeping a record of the frequency or how many times a visually impaired population would have its device detect an obstruction ahead of steps at different locations. Last but not least, the device has a button mounted on the very top. Whenever the button is pressed, a tiny red pin will be placed on the map. This can help the many users on the platform to mark and share locations that are inconvenient or unpleasant to navigate near at.

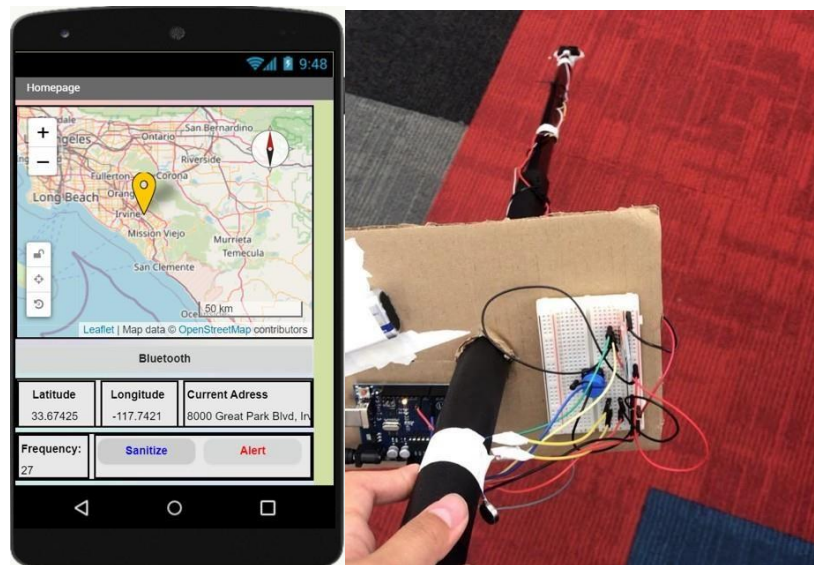


Figure 4. Global Positioning System

3.3. GPS Coordinate

Our application encourages the users to navigate freely and, indeed, safely. With the global positioning system, the application accompanies the user on their road and requests consent to save their GPS coordinates with the device's corresponding acquired frequency at a specific location into a database. This will eventually produce a large database that can map out the locations that are more friendly for the visually impaired population to navigate nearby and, most importantly, raises the society's awareness toward making the environment more convenient and welcoming to every population, which would thus encourage entrepreneurs in the future to learn from those locations that succeed in making their spaces inclusive and accepting to everyone.

3.4. Natural Disaster Alert System

The frequency, along with the current longitude and latitude of the user, perform vital roles in the alert system for natural disasters. During a natural disaster, requests will be sent to the users' device to request for their data on frequency and GPS coordinate. With such information, services behind the calls will be able to make announcements and assign more personalized procedures to self-protect based on where the user is at and what resources are accessible near them.

3.5. COVID-19 Sanitization Reminder

Tracking frequency within the application also assists with the function of reminding the user to sanitize the device during the time of COVID-19 pandemic. Whenever a specific value of frequency is reached that the device has a high risk of having the virus on its surface, the sanitize button will light up on the application's home page while a voice will be played to remind the user to sanitize the device for minimizing the chance of contracting COVID-19.

3.6. Machine-learning

Dummy data were generated for machine learning in our program for testing and figuring out the most appropriate machine learning model through the use of svm and regression algorithms for the following three cases:

- use latitude, longitude, and the three categories of locations (home, store, park) to predict frequency.
- use longitude, latitude, and frequency to predict between the three categories of locations (home, store, park).
- use frequency to predict whether a user should sanitize the device.

4. EXPERIMENT

To evaluate the accuracy of the application's algorithms, we experimented with the different models, parameters, and data set features to find the most accurate machine learning model. The first experiment was conducted to find a machine learning model that best describes the relationship between the user's GPS Coordinates and the frequency of which an obstruction will be detected by the user's device in a one hour walk. This experiment evaluates and compares the accuracy of the three machine learning models created with 1000 datasets with different polynomial parameters. The second experiment, otherwise, is conducted with the intake of 1000 datasets with different set features to find a machine learning model that best predicts the user's presence in one of the three categories of locations (home, park, store). Lastly, the third experiment also consumes 1000 data sets with different set features for the purpose of predicting whether the user should sanitize their device at times. In each of the three experiments, the accuracy of different models is calculated and compared among each other to find the best representation in each that executes the most accurate prediction.

For experiment 1, machine learning models with different polynomial parameters were applied to the same data set to find out which model would produce the most accurate algorithm. A linear model, along with polynomial models with parameter 2, 5, and 7, are compared together to evaluate the model that produces the most accuracy.

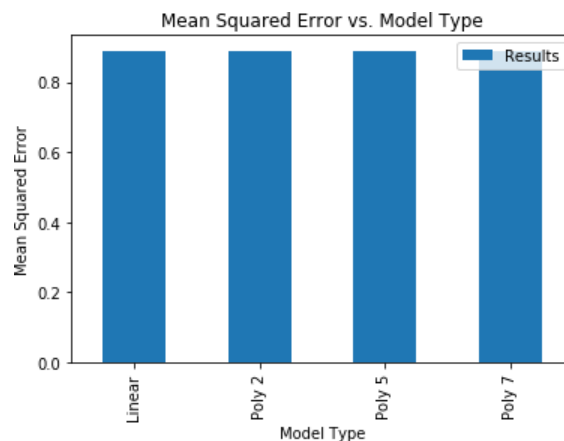


Figure 5. polynomial parameters influence on the model

The results of experiment 1 reveals neither being linear or polynomial nor changing the polynomial parameters influence the accuracy of the model. Shown in Figure 5, the accuracy of the linear model, along with that of the polynomial model with parameters of 2, 5, and 7, are all at a value of 0.87 or a percentage of 87%.

In experiment 2, three machine learning models that differ by their data sets are compared among each other for determining the one that is most accurate at predicting. The three models take in datasets of longitude, latitude, and frequency in different combinations, with the first model intaking three inputs (longitude, latitude, frequency), the second model intaking two inputs (longitude, latitude), and the third model intaking a single input (frequency).

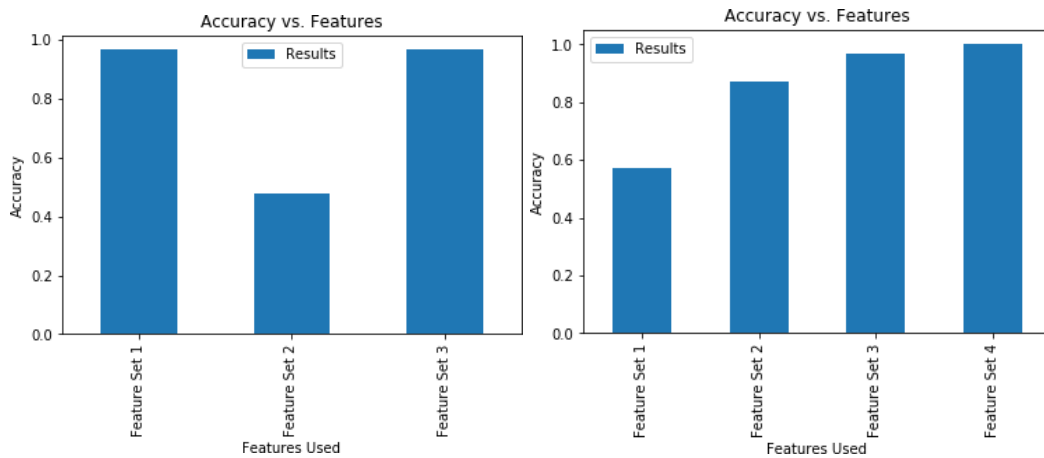


Figure 6. data sets influence on models. Figure 7: different data sets influence on four models.

The results of experiment 2 reveals that the accuracy for intaking the full data set which contains longitude, latitude, and frequency is slightly lower than that when intaking a single dataset of frequency, as the accuracy for model 1 is shown to be at a value of 0.94 or a percentage of 94 % while that of model 3 is a value of 0.95 or a percentage of 95% . Amongst the three models, however, the model that intakes two inputs (longitude and latitude) differ the most and produces the least accuracy to a value of about 0.48 or 48%. [Figure 6]

A total of four machine learning models with different data sets are compared against each other in experiment 3, in which the first model is created with two inputs (longitude, latitude), the second model with one single input of the three categories of location (home, park, or store), the third model with one single input of frequency, and the fourth with a full dataset of four inputs (longitude, latitude, frequency, categories of location).

The results of experiment 3 reveals that, like what happened in experiment 2, an increasing amount of intake dataset positively influences the accuracy of its model only in some cases. To support, Figure 7 displays the model with the most input (Figure 7 Feature Set 4) - longitude, latitude, frequency, and categories of location - as having the highest accuracy in a decimal of 0.98 and a percentage of 98 %. However, it is also significant to notice that the model producing the lowest accuracy is the one with intake of two inputs (Figure 7 Feature Set 1) instead of intake of one single input. To be specific, the model with longitude and latitude as the only inputs produces an accuracy in decimal of approximately 0.57 and in a percentage of 57%, which is significantly lower than that produced by the models with one single intake in inputs or, to emphasize, an accuracy of greater than a value of 0.8 or percentage of 80%. Lastly, the comparison of accuracy between model 2 and model 3 in Figure 7 emphasizes the difference in

the accuracy of the same amount of input but different input or, in other words, between predicting based on a single input of the three categories of location from model 2 and based on a single input of frequency from model 3. To compare, the machine learning model with frequency as its input has a higher accuracy in decimal of 0.96 or percentage of 96 % than the machine model with categories of location as its single input, which has a lower accuracy of 0.84 in value and 84% in percentage. [Figure 7]

5. RELATED WORK

J. M. Loomis [10] emphasized the system of spatial display, including but not limited to body pointing, HPI-speech, virtual-tone display, and virtual-speech display, to be an effective system for navigational purpose of the visually-impaired population and proposed that virtual-speech display, regardless of its blockage toward environmental sounds, performs the best in leading to the fastest mean travel times and is rated subjectively as the top among the virtual displays used during the study.

A. Helal [11] proposed a navigation system called Drishti for the visually impaired population through the integration of hardware like wearable computers and software built with a web of technological components such as spatial database and map server. Similar to our application, Drishti provides a significant amount of guidance to the user through voice support and addresses their needs through the active use of a digital map with GPS. Drishti readily involves the navigational features to support the visually impaired population, but our application extends support to disaster alert and the prevention of COVID-19.

I. Ulrich [12] proposed a navigation device called GuideCane to guide the visually impaired population during independent navigation. GuideCane is built with a cane on a wheel, and it implements a steering servo motor with an ultrasonic sensor for the main functionality to operate, which is to guide the users away from hazards and obstacles through the steering commands. While GuideCane performs its navigational feature on a wheel, our device operates on its own and is lighter and can be more portable for the users to bring everywhere for navigation.

6. CONCLUSION AND FUTURE WORK

In conclusion, our application integrates with a Bluetooth-enabled device to encourage the visually impaired population to navigate independently, alert and guide them in natural disasters, and remind them to sanitize their devices during the pandemic. In our experiments, we tested trials of different machine learning models which differ by regression model, polynomial parameter, and inputted data sets. The first experimentation results show that since adjusting the regression model or polynomial parameter does not change the accuracy of the prediction at all, it's proper to just use a linear regression model to train and test for making predictions on frequency based on longitude, latitude, and the three categories of locations (home, park, store). The second and third experimentation results show that the machine learning model that intakes one single data set of frequency will predict among the three categories of locations the best and that the machine learning model that takes in the full data set of frequency, longitude, latitude, and the three categories of location will predict most accurately in whether or not to sanitize the device. The accuracy of each experiment when using the most appropriate machine learning model when making predictions all exists above 90% or a value of 0.90, suggesting that the application will run well and accurately to ensure a pleasant and convenient experience for the visually impaired population.

While our application exceeds the navigational feature and remains unique from other prototypes in that it encourages independence in other aspects, it also has some limitations. Since the application interacts with the user solely through voice, it cannot provide the population that is both visually impaired and hearing-impaired the same quality of support. In the future, we plan to improve the accuracy of the system by boosting the signal connections between the device and API, strengthen the practicality by decreasing the size of the bracelet, and enhance the optimization by adding a “help” button. We strongly believe that by following these measures, our alert system will be able to reach its full potential.

REFERENCES

- [1] Tracy L Mitzner, PhD, Jon A Sanford, March, Wendy A Rogers, PhD, Closing the Capacity-Ability Gap: Using Technology to Support Aging with Disability, *Innovation in Aging*, Volume 2, Issue 1, January 2018, igy008, <https://doi.org/10.1093/geroni/igy008>
- [2] “Disability Impacts All of Us Infographic.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 16 Sept. 2020, www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html.
- [3] “Assistive Technology.” World Health Organization, World Health Organization, who.int/news-room/fact-sheets/detail/assistive-technology.
- [4] Yan, Wu, et al. Digital Media Usage of Sensory Impaired Users in Wales 2018 Report. Swansea University and RNIB Cymru, 2018.
- [5] Attia, I., and D. Asamoah. “The White Cane. Its Effectiveness, Challenges and Suggestions for Effective Use: The Case of Akropong School for the Blind”. *Journal of Education, Society and Behavioural Science*, Vol. 33, no. 3, May 2020, pp. 47-55, doi:10.9734/jesbs/2020/v33i330211.
- [6] L. Whitmarsh (2005) The Benefits of Guide Dog Ownership, *Visual Impairment Research*, 7:1, 27-42, DOI: 10.1080/13882350590956439
- [7] Andrew Zaleski, special to CNBC.com. “Amazing Electronic Glasses Help the Legally Blind See, but They Are Costly.” CNBC, CNBC, 20 Aug. 2018, [www.cnbc.com/2017/09/20/these-amazing-electronic-glasses-help-the-legally-blind-see.html#:~:text=Amazing electronic glasses help the legally blind see, but they are costly,-Published Wed, Sep&text=Smart glasses, called eSight3, help,price tag of about \\$10,000.](http://www.cnbc.com/2017/09/20/these-amazing-electronic-glasses-help-the-legally-blind-see.html#:~:text=Amazing%20electronic%20glasses%20help%20the%20legally%20blind%20see,%20but%20they%20are%20costly,-Published%20Wed,%20Sep&text=Smart%20glasses,%20called%20eSight3,%20help,%20price%20tag%20of%20about%20$10,000.)
- [8] Senior Lecturer in Psychology. “The Pandemic Is Undermining Visually Impaired People's Independence – Here's How to Fix This.” *The Conversation*, 14 Jan. 2021, theconversation.com/the-pandemic-is-undermining-visually-impaired-peoples-independence-heres-how-to-fix-this-142209.
- [9] “Coronavirus.” World Health Organization, World Health Organization, www.who.int/health-topics/coronavirus#tab=tab_1.
- [10] Loomis, Jack M et al. “Personal Guidance System for People with Visual Impairment: A Comparison of Spatial Displays for Route Guidance.” *Journal of visual impairment & blindness* vol. 99,4 (2005): 219-232.
- [11] A. Helal, S. E. Moore and B. Ramachandran, "Drishti: an integrated navigation system for visually impaired and disabled," *Proceedings Fifth International Symposium on Wearable Computers*, Zurich, Switzerland, 2001, pp. 149-156, doi: 10.1109/ISWC.2001.962119.
- [12] I. Ulrich and J. Borenstein, "The GuideCane-applying mobile robot technologies to assist the visually impaired," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 31, no. 2, pp. 131-136, March 2001, doi: 10.1109/3468.911370.

ADOPTION FACTORS OF ENABLING I4.0 TECHNOLOGIES AND BENEFITS IN THE SUPPLY CHAIN

José Carlos Franceli and Silvia Novaes Zilber Turri

Universidade Federal do ABC, São Paulo, Brazil

ABSTRACT

Industry 4.0 technologies represent a new paradigm of integration of cyber-physical systems, information and communication solutions, with applications in many different domains. This topic has to date been vaguely explored in the realm of the social sciences; hence this study attempts to bridge this gap by investigating the challenges of innovation adoption, based on I4.0 technologies, more specifically the factors affecting adoption decisions. This paper, based on previous adoption literature, aims to identify the barriers and benefits generated in the Supply Chain. Given the nature and novelty of the technology, whose adoption is the primary theme of this study, a systematic literature review was developed. The results present a framework that connects adoption factors, enabling technologies of I4.0, and benefits to the Supply Chain. The Model can be easily adapted to serve as a tool in the evaluation and selection of technological innovations to be adopted.

KEYWORDS

IoT, Supply Chain, Digital Transformation, I4.0, Adoption.

1. INTRODUCTION

The new Fourth Industrial Revolution, also known as Industry 4.0 (I4.0), is disrupting the way companies do business all around the world, where traditional manufacturing methods and production are undergoing digital transformation. The concept of Industry 4.0 was initially introduced in Germany in 2011 [19], referring to the integration of physical objects, human actors, intelligent machines, production lines and processes across organizational boundaries, with the aim of creating a system where processes are integrated and information is shared in real time [14]. The emergence of I4.0 is deeply rooted in the Third Industrial Revolution, characterized by rapid developments in information technology (IT), electronics and digitalization, where Advanced Manufacturing Technologies (AMTs) are at their core. AMTs can be described as computer-assisted technologies used to control and monitor manufacturing activities, providing greater flexibility, shorter production cycles, faster responses to changing market demands, better precision, and control of production processes [7]. IoT is revolutionizing the manufacturing industry and the consumption of goods; products that once were exclusively composed of mechanical and electrical parts are now becoming complex systems combining hardware, data storage, sensors, microprocessors, software, and connectivity in various formats. This is how the Internet of Things (IoT) is becoming widespread. Smart and connected products, enabled by major improvements in the processing of device miniaturization and wireless connectivity, are currently unleashing a new era of competition [25]. However, the wide spectrum of applications for the same technology can be, at the same time, an incentive and an obstacle to its adoption [5].

The adoption of innovation is a complex process especially when the innovation or technology in question is incipient [26] and several authors talk about the benefits it can generate when overcoming technological challenges. However, none of these authors discussed the challenges of adoption in different environments and market segments where solutions may be applied. The lack of a model for studying the adoption of this technology led us to the elaboration of a structure that contemplates the several dimensions related to the adoption decision. Therefore, this study, through a Systematic Literature Review (SLR), aims to identify the factors that lead to the adoption of Enabling Technologies of Industry4.0 (I4.0 technologies) and the benefits that they bring to the Supply Chain (SC). The following sections comprise the research method, followed by SLR results, the proposition of a conceptual framework and final considerations.

2. METHOD

2.1. Research Planning and Keyword Identification

This article adopts the methodology proposed in the stages of planning, execution and presentation of results obtained through this SLR [31]. The keywords selected represent the terms available in the literature of Industry 4.0 Technology Adoption in SC. The search for relevant articles was done through a keyword-combination process to identify the main study themes. The initial search utilized the keyword combination “Digital AND Transformation AND Adoption” to look for articles in the scope of our study objective. Subsequently, to expand the range of studies related to I4.0 technology adoption, a second search was carried out using the keyword combination “I4.0 AND Adoption”, resulting in a group of relevant articles primarily focused on the topic of adoption in the field of I4.0 technologies. Finally, a third search was undertaken using the keyword combination “IoT AND Supply AND Chain AND Adoption” resulting in a broad search, combining keywords that brought us the results outlined in the study objectives.

2.2. Conducting the Search for Articles

To systematically evaluate the presented theme in the literature, this analytical review process requires relevant articles on the topic [22]. Following this process, it is possible to identify the main relevant publications, trends on the topics being discussed and researched, as well as to evidence the gaps present in the literature [30]. This session presents the research protocol used as part of the strategy to identify relevant studies and the criteria for inclusion and exclusion of previously selected documents [17]. Initially, from reading the “abstracts” of the selected articles, it was found that those who had “adoption” as the main theme resulted in a group of relevant articles for the study. We applied a second selection process on this group of articles to identify the most relevant articles according to content (adoption barriers and benefits), results (adoption framework) and type of organizations (manufacturing companies, transport services and retail).

3. RESULTS

3.1. Results Presentation

To have a detailed presentation of the study results, this section is structured in the following manner:

- Descriptive Statistics Analysis
- Content Analysis
- Summary of Results

3.2. Descriptive Statistics Analysis

Twenty-three (23) relevant articles were selected for this study after an analysis of ninety-three (93) abstracts from the articles found through the search using the three-keyword combination procedure. The 226 citations found in the 23 articles represent 41% of the total citation quote (556) between 2018 to 2020 (Figure 1). The evolution of citations for articles with “adoption” as a subject for the same period is show in Figure 2.



Figure 1. Relevant papers according to number of citations

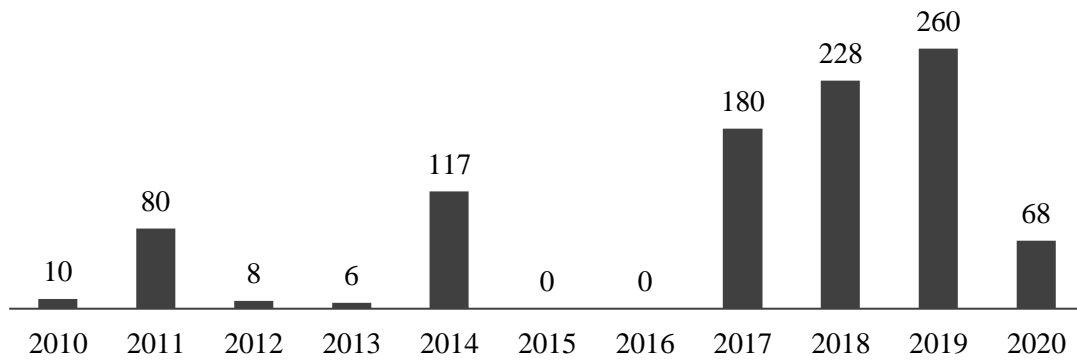


Figure 2. Number of citations in articles with “Adoption” as a theme over time

After a ranking process, the last prioritization selection was conducted to identify the ten (10) most relevant articles to the theme of this paper (Table 1) following the criteria below:

- a. Articles that present adoption frameworks.
- b. Articles that present factors and benefits of adoption.
- c. Articles that reference companies in the manufacturing, logistics services and retail sectors.

Table 1. Selected articles and criteria

Criteria	Number of Articles
a. Articles that present adoption frameworks.	4
b. Articles that present factors and benefits of adoption.	10
c. Articles that reference companies in the manufacturing, logistics services and retail sectors.	10

Figure3. outlines the complete research and selection methodology of articles used in this study.

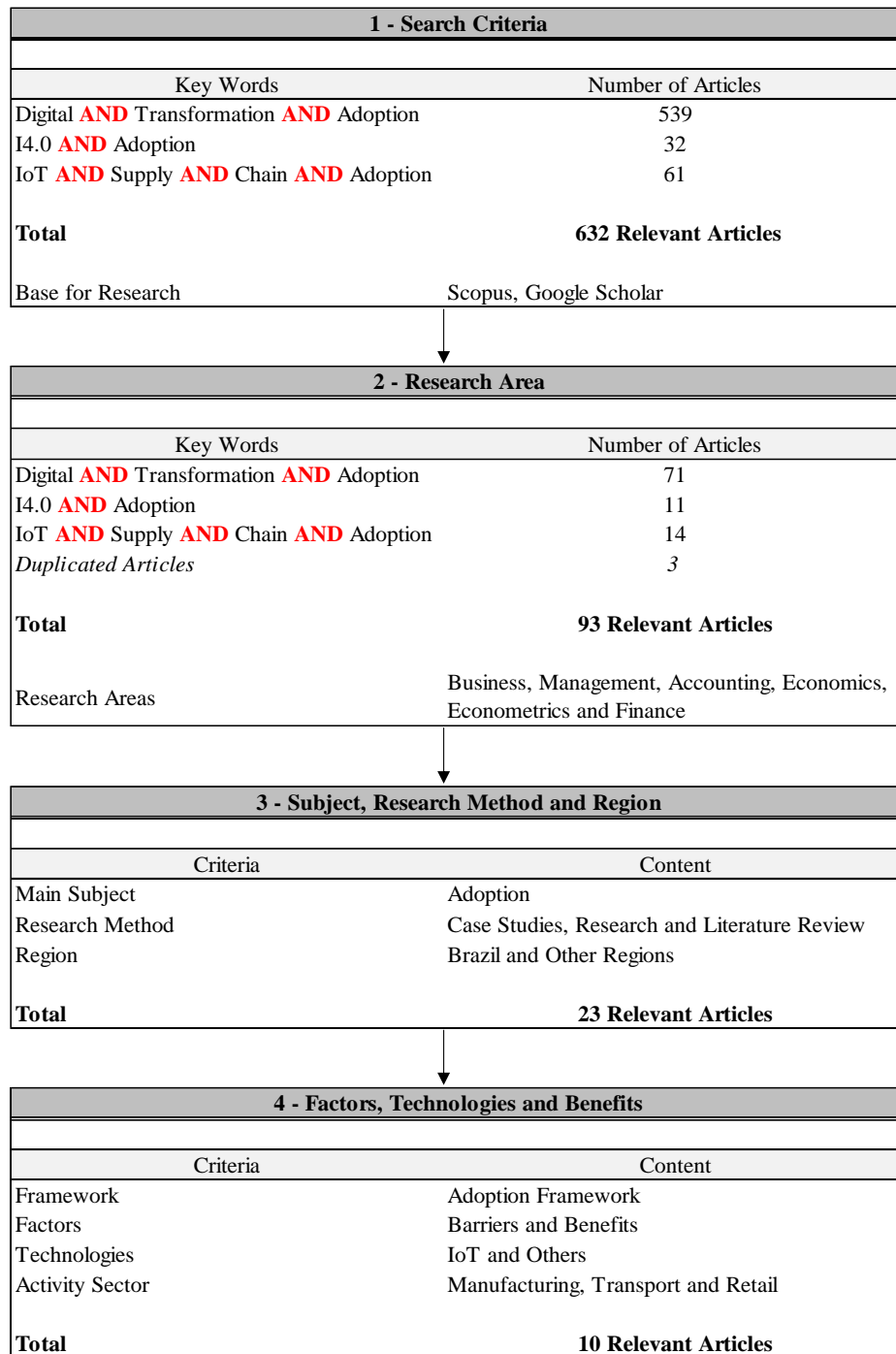


Figure 3. Research methodology

3.2.1. List of Articles According to Keyword Combinations

The list below presents the 93 articles found in the research process.

Article	Title	Author	Source
1	Moving towards digitalization: a multiple case study in manufacturing	Zangiacomì, A., Pessot, E., Fornasiero, R., Bertetti, M., Sacco, M.	Production Planning and Control 31(2-3), pp. 143-157
2	Organizational learning paths based upon industry 4.0 adoption: An empirical study with Brazilian manufacturers	Tortorella, G.L., Cawley Vergara, A.M., Garza-Reyes, J.A., Sawhney, R.	International Journal of Production Economics 219, pp. 284-294
3	Bringing it all back home? Backshoring of manufacturing activities and the adoption of Industry 4.0 technologies	Dachs, B., Kinkel, S., Jäger, A.	Journal of World Business 54(6),101017
4	Providing industry 4.0 technologies: The case of a production technology cluster	Dalmarco, G., Ramalho, F.R., Barros, A.C., Soares, A.L.	Journal of High Technology Management Research 30(2),100355
5	The adoption of Industry 4.0 technologies in SMEs: results of an international study	Agostini, L., Nosella, A.	58(4), pp. 625-643 Management Decision
6	Organizational and managerial challenges in the path toward Industry 4.0	Agostini, L., Filippini, R.	European Journal of Innovation Management 22(3), pp. 406-421
7	The impacts of Industry 4.0: a descriptive survey in the Italian manufacturing sector	Zheng, T., Ardolino, M., Bacchetti, A., Perona, M., Zanardini, M.	Journal of Manufacturing Technology Management Article in Press
8	A comparison on Industry 4.0 and Lean Production between manufacturers from emerging and developed economies	Tortorella, G.L., Rossini, M., Costa, F., Portioli Staudacher, A., Sawhney, R.	Total Quality Management and Business Excellence Article in Press
9	Fashion 4.0. Innovating fashion industry through digital transformation	Bertola, P., Teunissen, J.	Research Journal of Textile and Apparel 22(4), pp. 352-369
10	New Model for Identifying Critical Success Factors Influencing BIM Adoption from Precast Concrete Manufacturers' View	Phang, T.C.H., Chen, C., Tiong, R.L.K.	Journal of Construction Engineering and Management 146(4),04020014
11	Exploring the growth challenge of mobile payment platforms: A business model perspective	Jocovski, M., Ghezzi, A., Arvidsson, N.	Electronic Commerce Research and Applications
13	Digital transformation in supply chain, challenges and opportunities in SMEs: a case study of Al-Rumman Pharma	Faridi, M.R., Malik, A.	Emerald Emerging Markets Case Studies
14	Digital systems and new challenges of financial management – fintech, XBRL, blockchain and cryptocurrencies	Mosteanu, N.R., Faccia, A.	Quality - Access to Success
15	How enterprises adopt agile forms of organizational design: A multiple-case study	Gerster, D., Dremel, C., Brenner, W., Kelker, P.	Data Base for Advances in Information Systems
16	Analysis of barriers in implementation of digital transformation of supply chain using interpretive structural modelling approach	Agrawal, P., Narain, R., Ullah, I.	Journal of Modelling in Management
17	Fostering digital transformation of SMEs: a four levels approach	Garzoni, A., De Turi, I., Secundo, G., Del Vecchio, P.	Management Decision
18	Digital transformation priorities of India's discrete manufacturing SMEs – a conceptual study in perspective of Industry 4.1	Dutta, G., Kumar, R., Sindhvani, R., Singh, R.K.	Competitiveness Review
19	Digital initiatives for access and quality in higher education: An overview	Ahmad, S.	Prabandhan: Indian Journal of Management
20	PERCEPTION or CAPABILITIES? AN EMPIRICAL INVESTIGATION OF THE FACTORS INFLUENCING THE ADOPTION OF SOCIAL MEDIA and PUBLIC CLOUD in German SMEs	Hassan, S.S., Reuter, C., Bzhilava, L.	International Journal of Innovation Management
21	Organizational learning paths based upon industry 4.0 adoption: An empirical study with Brazilian manufacturers	Tortorella, G.L., Cawley Vergara, A.M., Garza-Reyes, J.A., Sawhney, R.	International Journal of Production Economics
22	Digitalization of construction organisations—a case for digital partnering	Aghmieni, D., Aigbavboa, C., Oke, A., Thwala, W., Moripe, P.	International Journal of Construction Management
23	What matters in implementing the factory of the future: Insights from a survey in European manufacturing regions	Pessot, E., Zangiacomì, A., Battistella, C., (...), Sala, A., Sacco, M.	Journal of Manufacturing Technology Management
24	Digitalization of world trade: Scope, forms, implications	Strelets, I.A., Chebanov, S.V.	World Economy and International Relations
25	The impact of Industry 4.0 implementation on supply chains	Ghadge, A., Er Kara, M., Moradlou, H., Goswami, M.	Journal of Manufacturing Technology Management
26	Big data and HR analytics in the digital era	Dahlbom, P., Siikanen, N., Sajasalo, P., Jarvenpää, M.	Baltic Journal of Management
27	Conceptualising digital transformation in SMEs: an ecosystemic perspective	Pelletier, C., Cloutier, L.M.	Journal of Small Business and Enterprise Development
28	Tortoise, not the hare: Digital transformation of supply chain business processes	Hartley, J.L., Sawaya, W.J.	Business Horizons
29	Readiness factors for information technology adoption in SMEs: testing an exploratory model in an Indian context	Nair, J., Chellamy, A., Singh, B.N.B.	Journal of Asia Business Studies
30	Digital passengers: A great divide or emerging opportunity?	Mayer, C.	Journal of Airport Management
31	A systematic literature review of big data adoption in internationalization	Dam, N.A.K., Le Dinh, T., Menvielle, W.	Journal of Marketing Analytics
32	The change of pediatric surgery practice due to the emergence of connected health technologies	Niemelä, R., Pikkariainen, M., Ervasti, M., Reponen, J.	Technological Forecasting and Social Change
33	To be or not to be digital, that is the question: Firm innovation and performance	Ferreira, J.J.M., Fernandes, C.I., Ferreira, F.A.F.	Journal of Business Research
34	Industry 4.0 and capability development in manufacturing subsidiaries	Szalavetz, A.	Technological Forecasting and Social Change
35	Do German Works Councils Counter or Foster the Implementation of Digital Technologies? : First Evidence from the IAB-Establishment Panel	Genz, S., Bellmann, L., Matthes, B.	Jahrbucher fur Nationalökonomie und Statistik
36	Technology adoption for the integration of online-offline purchasing: Omnichannel strategies in the retail environment	Lobo, F., VASCONCELLOS, E., & Guedes, L. V	International Journal of Retail and Distribution Management
37	Readiness of upscale and luxury-branded hotels for digital transformation	Lam, C., Law, R.	International Journal of Hospitality Management
38	Blockchain technology and its relationships to sustainable supply chain management	Saberi, S., Kouhizadeh, M., Sarkis, J., Shen, L.	International Journal of Production Research
39	Disruptions of account planning in the digital age	Zimand Sheiner, D., Earon, A.	Marketing Intelligence and Planning
40	Industry 4.0 technologies: Implementation patterns in manufacturing companies	Frank, A.G., Dalenogare, L.S., Ayala, N.F.	International Journal of Production Economics
41	Exploring the digital innovation process: The role of functionality for the adoption of innovation management software by innovation managers	Huesig, S., Endres, H.	European Journal of Innovation Management
42	Opportunities and challenges in the e-commerce of the food sector	Vargas, V.M., Budz, S.	Quality - Access to Success
43	Fintechs: A literature review and research agenda	Milán, E.Z., Spinola, M.D.M., Carvalho, M.M.D.	Electronic Commerce Research and Applications
44	Transformation of accounting through digital standardisation: Tracing the construction of the IFRS Taxonomy	Troshani, I., Locke, J., Rowbottom, N.	Accounting, Auditing and Accountability Journal
45	The effective factors of cloud computing adoption success in organization	Yoo, S.-K., Kim, B.-Y.	Journal of Asian Finance, Economics and Business
46	The impact of digital transformation on formal and informal organizational structures of large architecture and engineering firms	Bonomini, M.M., Hall, D.M., Staub-French, S., Tucker, A., Talano, C.M.L.	Engineering, Construction and Architectural Management
47	Digital transformation in the Spanish agri-food cooperative sector: Situation and prospects [La transformación digital en el sector cooperativo agroalimentario español: Situación y perspectivas]	Vázquez, J.J., Cebolla, M.P.C., Ramos, F.S.	CIRIEC-España Revista de Economía Pública, Social y Cooperativa

48	Bridging the gender digital gap	Mariscal, J., Mayne, G., Aneja, U., Sorgner, A.	Economics
49	Disruptive technology adoption, particularities of clustered firms	Molina-Morales, F.X., Martínez-Cháfer, L., Valiente-Bordanova, D.	Entrepreneurship and Regional Development
50	Bitcoin distribution in the age of digital transformation: Dual-path approach	Lee, W.-J., Hong, S.-T., Min, T.	Journal of Distribution Science
51	Determinants of information and digital technology implementation for smart manufacturing	Ghobakhloo, M.	International Journal of Production Research
52	Embracing artificial intelligence and digital personnel to create high-performance jobs in the cyber economy	Lobova, S.V., Bogoviz, A.V.	Contributions to Economics
53	The impacts of Industry 4.0: a descriptive survey in the Italian manufacturing sector	Zheng, T., Ardoino, M., Bacchetti, A., Perona, M., Zanardini, M.	Journal of Manufacturing Technology Management
54	Blockchain: A paradigm shift in business practices	Kizildag, M., Dogru, T., Zhang, T., (...), Ozturk, A.B., Ozdemir, O.	International Journal of Contemporary Hospitality Management
55	Extremes of acceptance: employee attitudes toward artificial intelligence	Lichtenthaler, U.	Journal of Business Strategy
56	Fashion 4.0. Innovating fashion industry through digital transformation	Bertola, P., Teunissen, J.	Research Journal of Textile and Apparel
57	The impacts of digital transformation on the labour market: Substitution potentials of occupations in Germany	Dengler, K., Matthes, B.	Technological Forecasting and Social Change
58	IT consumerization and the transformation of IT governance	Gregory, R.W., Kaganer, E., Henfridsson, O., Ruch, T.J.	MIS Quarterly: Management Information Systems
59	Loosely Coupled Systems of Innovation: Aligning BIM Adoption with Implementation in Dutch Construction	Papadomikolaki, E.	Journal of Management in Engineering
60	New technologies and the transformation of work and skills: a study of computerisation and automation of Australian container terminals	Gekara, V.O., Thanh Nguyen, V.-X.	New Technology, Work and Employment
61	Active seniors perceived value within digital museum transformation	Traboulsi, C., Frau, M., Cabiddu, F.	TQM Journal
62	Insights on the adoption of social media marketing in B2B services	Buratti, N., Parola, F., Satta, G.	TQM Journal
63	Digital technologies and the modernization of public administration	Todorut, A.V., Tselentis, V.	Quality - Access to Success
64	Digital transformation at carestream health	Smith, H.A., Watson, R.T.	MIS Quarterly Executive
65	IT Governance Mechanisms and Contingency Factors: Towards an Adaptive IT Governance Model	Levstek, A., Hovelja, T., Pucihar, A.	Organizacija
66	The impact of digital technology on relationships in a business network	Pagani, M., Pardo, C.	Industrial Marketing Management
67	Making Sense of Africa's Emerging Digital Transformation and its Many Futures	Ndemo, B., Weiss, T.	Africa Journal of Management
68	Work 4.0 — Digitalisation and its Impact on the Working Place [Arbeiten 4.0 — Folgen der Digitalisierung für die Arbeitswelt]	Klammer, U., Steffes, S., Maier, M.F., (...), Bellmann, L., Hirsch-Kreinsen, H.	Wirtschaftsdienst
69	How transformational leadership facilitates e-business adoption	Alos-Simo, L., Verdu-Jover, A.J., Gomez-Gras, J.-M.	Industrial Management and Data Systems
70	Tackling the digitalization challenge: How to benefit from digitalization in practice	Parviainen, P., Thinen, M., Käiriäinen, J., Teppola, S.	International Journal of Information Systems and Project Management
71	Are millennials transforming global tourism? Challenges for destinations and companies	Veiga, C., Santos, M.C., Águas, P., Santos, J.A.C.	Worldwide Hospitality and Tourism Themes
72	Introducing data driven practices into sales environments: examining the impact of data visualisation on user engagement and sales results	Magee, B., Sammon, D., Nagle, T., O'Raghallaigh, P.	Journal of Decision Systems
73	Evaluating data driven practices in sales environments: user engagement and sales results	Magee, B.	Journal of Decision Systems
74	The effects of rfid applied in ground handling system on aircraft turnaround time: A simulation based analysis	Khumboon, R., Isaradech, B.	Academy of Strategic Management Journal
75	Exploring the relevancy of Massive Open Online Courses (MOOCs): A Caribbean university approach	Dyer, R.A.D.	Information Resources Management Journal
76	Paradoxical digital worlds	Munar, A.M.	Tourism Social Science Series
77	The role of ICTs in conflict transformation in Egypt	Richardson, J.W., Brantmeier, E.J.	Education, Business and Society: Contemporary Middle Eastern Issues
78	Analysis of emerging technology adoption for the digital content market	Jim, B.-H., Li, Y.-M.	Information Technology and Management
79	Student attitudes and behaviors towards digital textbooks	Weisberg, M.	Publishing Research Quarterly
80	Fundamentals of H.324 desktop videoconferencing	Herman, Mort	Electronic Design
81	Technology embracing by 3pl service providers in india: Tuticorn port trust — a case study	Senthil, M., Ruthramathi, R., Gayathri, N.	International Journal of Scientific and Technology Research 9(3), pp. 138-144
82	Managing the food supply chain in the age of digitalisation: a conceptual approach in the fisheries sector	Coronado Mondragon, A.E., Coronado Mondragon, C.E., Coronado, E.S.	Production Planning and Control
83	Modeling the internet of things adoption barriers in food retail supply chains	Kamble, S.S., Gunasekaran, A., Parekh, H., Joshi, S.	Journal of Retailing and Consumer Services 48, pp. 154-168
84	The Internet of Things (IoT) in retail: Bridging supply and demand	Raman, S., Patwa, N., Niranjan, I., (...), Moorthy, K., Mehta, A.	Journal of Retailing and Consumer Services 48, pp. 154-169
85	Adoption of Internet of Things in India: A test of competing models using a structured equation modeling approach	Mital, M., Chang, V., Choudhary, P., Papa, A., Pani, A.K.	ITechnological Forecasting and Social Change 136, pp. 339-346
85	Impact of big data on supply chain management	Raman, S., Patwa, N., Niranjan, I., (...), Moorthy, K., Mehta, A.	International Journal of Logistics Research and Applications 21(6), pp. 579-596
86	Factors influencing the adoption of the internet of things in supply chains	Yan, B., Jin, Z., Liu, L., Liu, S.	Journal of Evolutionary Economics
87	Real-Time business data acquisition: How frequent is frequent enough?	Townsend, M., Le Quoc, T., Kapoor, G., (...), Zhou, W., Piramuthu, S.	Information and Management 55(4), pp. 422-429
88	The effect of "Internet of Things" on supply chain integration and performance: An organisational capability perspective	Tsang, Y.P., Choy, K.L., Wu, C.H., (...), Lam, H.Y., Koo, P.S.	Australasian Journal of Information Systems 22
89	Internet of Things (IoT) in high-risk Environment, Health and Safety (EHS) industries: A comprehensive review	Thibaud, M., Chi, H., Zhou, W., Piramuthu, S.	Decision Support Systems 108, pp. 79-95
90	An IoT-based cargo monitoring system for enhancing operational effectiveness under a cold chain environment		International Journal of Engineering Business Management
91	Examining potential benefits and challenges associated with the Internet of Things integration in supply chains	Haddud, A., DeSouza, A., Khare, A., Lee, H.	Journal of Manufacturing Technology Management
92	Impact of RFID technology on supply chain decisions with inventory inaccuracies	Fan, T., Tao, F., Deng, S., Li, S.	International Journal of Production Economics 159, pp. 117-125
93	Integrated billing mechanisms in the Internet of Things to support information sharing and enable new business opportunities	Uckelmann, D., Harrison, M.	International Journal of RF Technologies: Research and Applications 2(2), pp. 73-90

The 10 articles below form the basis of the SLR selected by prioritization criteria:

Article Number	Title	Author	Source
1	The adoption of Industry 4.0 technologies in SMEs: results of an international study	Agostini, L., Nosella, A.	Management Decision
2	Organizational and managerial challenges in the path toward Industry 4.0	Agostini, L., Filippini, R.	European Journal of Innovation Management 22(3), pp. 406-421
3	Perception or Capabilities? An Empirical Investigation of the Factors Influencing the Adoption of Social Media and Public Cloud in German SMEs	Hassan, S.S., Reuter, C., Bzhalava, L.	International Journal of Innovation Management
4	The impact of Industry 4.0 implementation on supply chains	Ghadge, A., Er Kara, M., Moradlou, H., Goswami, M.	Journal of Manufacturing Technology Management
5	To be or not to be digital, that is the question: Firm innovation and performance	Ferreira, J.J.M., Fernandes, C.I., Ferreira, F.A.F.	Journal of Business Research
6	Technology adoption: Factors influencing the adoption decision of the internet of things in a business environment	LOBO, F., VASCONCELLOS, E., & GUEDES, L.	International Association for Management of Technology
7	Disruptive technology adoption, particularities of clustered firms	Molina-Morales, F.X., Martínez-Cháfer, L., Valiente-Bordanova, D.	Entrepreneurship and Regional Development
8	Modeling the internet of things adoption barriers in food retail supply chains	Kamble, S.S., Gunasekaran, A., Parekh, H., Joshi, S.	International Journal of Production Research
9	Adoption of Internet of Things in India: A test of competing models using a structured equation modeling approach	Mital, M., Chang, V., Choudhary, P., Papa, A., Pani, AK	ITechnological Forecasting and Social Change 136, pp. 339-346
10	Factors influencing the adoption of the internet of things in supply chains	Yan, B., Jin, Z., Liu, L., Liu, S.	Journal of Evolutionary Economics

3.2.2. Statistics of Adoption Factors, Technologies and Benefits

As shown in Table 2, the results from the SLR found nineteen (19) factors impacting the adoption process.

Table 2. Adoption factor per analyzed study (authors)

AUTHORS	Hassan, S.S., Reuter, C., Bzhalava, L.	Kamble, S.S., Gunasekaran, A., Parekh, H., Joshi, S.	Agostini, L., Nosella, A.	Mital, M., Chang, V., Choudhary, P., Papa, A., Pani, A.K.	Ghadge, A., Er Kara, M., Moradlou, H., Goswami, M.	Agostini, L., Filippini, R.	Ferreira, J.J.M., Fernandes, C.I., Ferreira, F.A.F.	Molina-Morales, F.X., Martínez-Cháfer, L., Valiente-Bordanova, D.	Yan, B., Jin, Z., Liu, L., Liu, S.	Lobo, F., VASCONCELLO S, E., & Guedes, L. V
ADOPTION FACTORS										
Perception of Importance of Use	x			x						
Perception of Easy of Use				x						
Encouragement of Use by Other Users				x						
Adopter's Profile							x			
Employee Skills	x	x	x			x				
Innovation and Digital Culture	x				x			x		
Security and Privacy	x	x								
Acquisition and Operating Costs	x	x			x				x	
Social Capital			x		x	x				
Management Support			x		x	x				
Absorptive Capacity			x					x		x
Sector Activity	x						x			
Market Demand		x								
Government Policies and Regulations		x			x		x	x		
Standards and Validations		x			x					x
IT Infrastructure		x			x					
Systems Architecture, Integration and Compatibility		x			x					x
Research and Development					x					
Data Management Quality					x					

Following a structural adaptation from the adoption framework by Lobo *et al* [18], the 19 factors were grouped into four (4) dimensional clusters according to their nature: “Human”, “Organizational”, “Political-Market” and “Technological”. Figure 4 shows the 4 clusters and their respective adoption factors according to the frequency in which they appear in the studies.

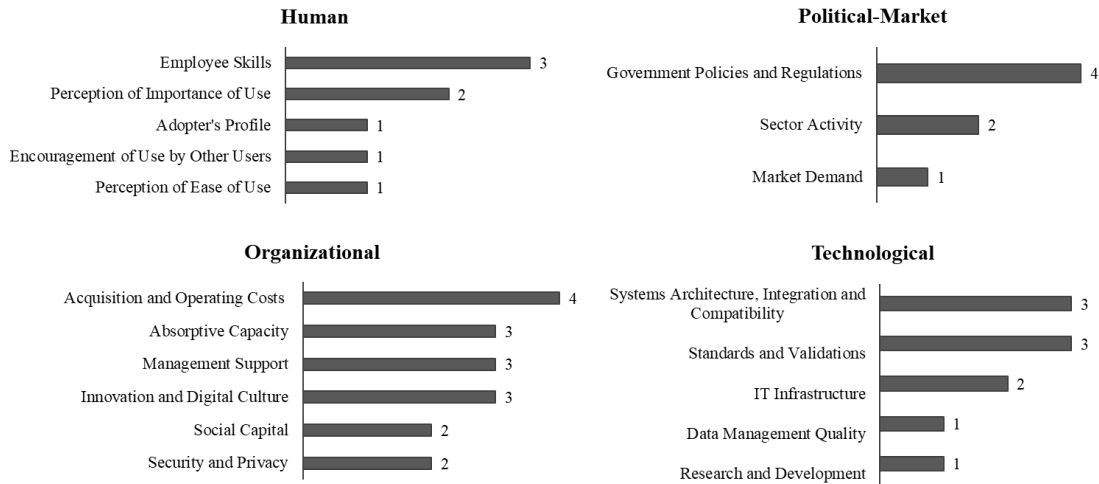


Figure 4. Frequency of factors per dimension

3.3. Content Analysis

In this section of the study, a content analysis is undertaken on the main adoption factors and benefits of I4.0 technology adoption based on the results drawn from the 10 articles of the SLR.

3.3.1. Human Factors

3.3.1.1. Perception of Importance of Technology Use

The “Perception of Importance of Use” acts as a high-relevance factor in technology adoption [28], as decision-makers tend to adopt innovations that are deemed useful and consistent with the organizational configurations processes [29].

3.3.1.2. Perception of Ease of Technology Use

The Technology Acceptance Model (TAM) [7] shows that individuals’ Perceived Usefulness (PU) of the technology and its Perceived Ease of Use (PEOU) impact the Behavioral Intention (BI) to use the technology. PEOU proved to be even more impactful than PU in the technology adoption process [31].

3.3.1.3. Encouragement of Technology Use by Other Users

The studies using the TRA (Theory of Reasoned Action) Model and the “Physical and Behavioral Concept” [3], identified that the encouragement of technology use by other users, who also share the same perception, impacts the intended adoption and use of technologies [31].

3.3.1.4. Adopter's Profile

Factors related to personal characteristics of entrepreneurs, *i.e.* factors of non-economic nature, explain the adoption behavior of companies regarding digital processes [11]. More experienced entrepreneurs and managers have a lower propensity to adopt digital technologies; while female managers, graduates and university students are more prone to adopting these technologies [10]. Therefore, the “Adopter's Profile” plays an important role in the adoption of I4.0 technologies.

3.3.1.5. Employee Skills

In the study conducted by Kamble *et al* [16], 12 categories were identified to impact I4.0 technologies adoption, more specifically IoT. Among these 12 categories, one of the most relevant was human skills, which can influence the entire structure of the adoption process [16]. IoT systems require highly trained professionals to develop and implement practical applications [28]. Qualified employees in innovation processes, who have advanced innovation skills and understand the use of lean methods and Information and Communication Technologies (ICT), pave the way for the adoption of I4.0 technologies [1]. Moreover, training and professional development are essential in the initial stages of the transition to digitalization [1].

3.3.2. Organizational Factors

3.3.2.1. Innovation and Digital Culture

Digital technology is changing business practice and organizational culture around the world. The convergence of modern digital technology, *e.g.* I4.0 technologies, is widely discussed as the next source of innovation and productivity in organizations. These are enabled by organizations' capabilities and innovation resources [29]. The change in business ecosystems due to innovation profoundly influences operational models and structures, and management strategies to adapt and explore new challenges in this evolving landscape [12].

Early adopters of I4.0 technologies, who are highly oriented toward implementing early-stage technologies, that work in clusters, can also multiply the likelihood of adoption by 473.7 per cent. The grouping process in clusters generates tacit knowledge and exchange of high-quality information, hence promoting innovation [23].

3.3.2.2. Security and Privacy

Given the existing threats in the virtual world, “Security and Privacy” are two of the most critical factors in the decision to adopt I4.0 technologies. Security is in danger in a network-based system due to threats of substitution by false data, access to confidential data and many other unauthorized intrusions that can paralyze networks [24]. Generally, companies internalize their security and privacy specifications, despite there being a lack of qualified competences and diversified capabilities related to the implementation of technologies beyond the scope of organizations' operations [29]. Moreover, the uncertainties associated with security and privacy lead decision-makers in small and medium enterprises (SMEs) to be reluctant to adopt I4.0 technologies or simply to ignore the potential benefit of these technologies for their businesses [8].

3.3.2.3. Acquisition and Operating Costs

“Acquisition Costs” and un foreseen “Operating Costs” of I4.0 technologies are some of the most relevant factors in the adoption decision, mainly to SMEs, and in many cases can lead to a

rejection of the technology [21]. Operating costs should be considered as a high priority despite the upfront uncertainty of the magnitude of these costs, *i.e.* generally related to energy consumption, operations, and a long payback period.

There is a critical risk related to the financial loss and irreversibility of investment [9]. Additionally, the implementation of IoT solutions requires advanced techniques and infrastructure support [16]. Therefore, SMEs due to limited financial resources are generally cautious with hidden costs and unaccounted expenses, especially in the absence or shortage of skilled labor to operate new technologies [29].

3.3.2.4. Social Capital

The successful adoption of I4.0 technologies often requires reliable cooperation, not only within the boundaries of the organization but also with external stakeholders [12]. Internal social capital represents the shared beliefs and values that allow individuals within an organization to work toward a common purpose, whereas external social capital involves the shared principles among external stakeholders along the SC. The impact of “Social Capital” in the adoption process is structured in 2 phases: digitalization of the business and transformation of the business network [17]. These 2 phases demand changes in the corporate culture, management of strategies and business model, as well as internal and external operations and relationships of the company.

At a SC level, collaboration is fundamental in the implementation of I4.0 technologies that have vertical and horizontal integration at their core [17]. Internal social capital has a high impact on the decision-making process of I4.0 technology adoption among the different adopter levels: “beginners”, “adopters” and “non-adopters”, whereas external social capital has a relevant impact only on “beginners” and “non-adopters” [1]. This is in line with the vision that the ability to effectively work in teams and leverage social contexts can contribute to the successful adoption of I4.0 technologies. To have a complete digital automation of the manufacturing process, covering both vertical and horizontal dimensions, employees are required to have an overall understanding of the organizational processes and information flows [12].

3.3.2.5. Management Support

Organizational and managerial practices in the SC and company processes directly affect the adoption of I4.0 technologies. “Management Support” acts as a moderating factor in the relationship between investment, social capital, and adoption, where social capital acts as a catalyst in the adoption process [1]. According to the aforementioned 2-phase framework, where digitalization of the business and transformation of the business network are the desired outcomes, management support is required to facilitate this process through changes in the corporate culture, management of strategies and business model, and organization of internal and external operations and relationships [12]. To support adoption, managers should encourage employees to make sense of the benefits of taking on new responsibilities [15], support decision-making, facilitate the reduction of hierarchical tiers and increase employee autonomy [1].

3.3.2.6. Absorptive Capacity

“Absorptive Capacity” moderates the relationship between social capital and the adoption process [1]. When the absorptive capacity” is high in organizations, it directly affects the adoption of I4.0 technologies [23]. Moreover, the “Exploratory Absorptive Capacity”, *i.e.* the ability of companies to profit from the external knowledge flows, enable companies inserted in a certain cluster to adopt disruptive technologies before other companies that do not have this capacity. Exploiting is seen as a company's ability to improve, expand, and use its existing routines, skills, and technologies to create something new based on transformed knowledge [23].

3.3.3. Political-Market Factors

3.3.3.1. Activity Sector

The “Activity Sector” factor has an impacting role in the adoption of I4.0 technologies, where generally the service sector is more susceptible to adoption than the manufacturing sector [11]. The scale of the organization also plays a role, since the adoption of certain technologies among SMEs is slower than in larger organizations, due to their limited resources [29].

3.3.3.2. Market Demand

The “Market Demand” factor has a high impact on the adoption process since the perception of value held by end customers generates a positive demand spiral for new benefits that consequently become part of the standard service offered by the company. This, in turn, generates competitive pressure in the market. The more devices that generate data on the corporate network and connect themselves to cloud solutions, the greater the possibility of creating value to customers; hence developing a differentiating factor in the market [16].

3.3.3.3. Government Policies and Regulations

The “Government Policies and Regulations” factor works as a key driver in the adoption process because legal information systems are needed in order to support the development and expansion of IoT in logistics and in Supply Chain Management (SCM); thereby enhancing the security standards to regulate operations. Government, institutions and organizations must work together to promote and support technological initiatives and solutions [16].

3.3.3.4. Standards and Validations

The existence of market standards and the ability for solutions from multiple vendors to work interchangeably can facilitate the adoption decision [11]. Validations play an important role in the adoption process [16], due to the scarcity of research on multiple applications of I4.0 technologies in industries. Due to the lack of standards, very few IoT technologies show clear returns on investment across the industry, which discourages SMEs in the adoption of innovative and disruptive technology [28].

3.3.4. Technological Factors

3.3.4.1. IT Infrastructure

“IT Infrastructure” is directly linked with the quality and data generation in the SC. It is therefore one of the main factors impacting the decision-making process of I4.0 technology adoption. Poor IT infrastructure and internet connectivity prove to be substantial barriers to digital transformation or adoption of these technologies [12]. The “IT Infrastructure” factor forms the core of I4.0 technologies use. The implementation of IoT solutions requires advanced skill sets and infrastructural support [16], which are scarce and proves to be a critical challenge for adoption [20].

3.3.4.2. Systems Architecture, Integration and Compatibility

The “Systems Architecture, Integration and Compatibility” factor influences the adoption of I4.0 technologies, hence its volatility requires special attention from adopters [18]. Integration and

compatibility issues have a significant impact on the adoption process of IoT technologies [16], as challenges in integrating IoT with existing legacy systems act as barriers to adoption [5].

3.3.4.3. Research and Development of Technologies & Data Management Quality

The “Research and Development” and “Data Management Quality” factors show significant impact on adoption since insufficient research and development practices in I4.0, absence of IT infrastructure, low-quality data, lack of digital culture, and distrust from partners create barriers to those organizations trying to innovate [12].

3.4. Benefits

The benefits of I4.0 technologies adoption herein presented are derived from the articles comprised in the SLR. Figure 5 displays the different technologies against their respective frequency of appearance in the 10 articles considered for this study.

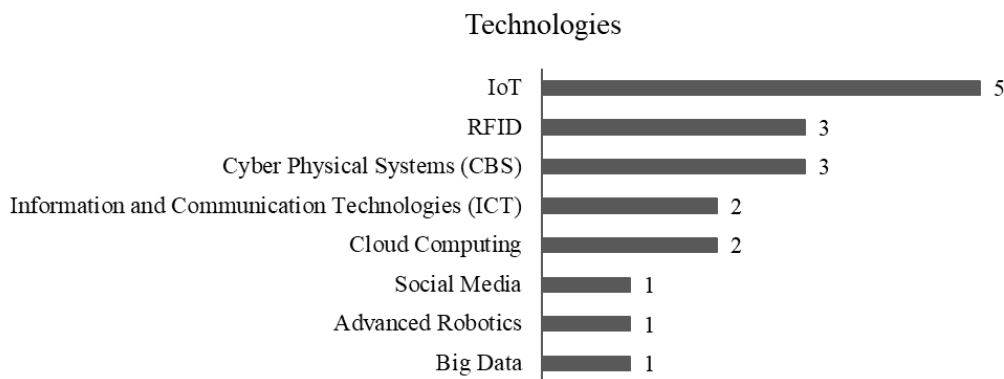


Figure 5. Frequency of technologies in studies

In the context of adapted SC, the adoption of radio-frequency identification (RFID) and cloud computing, which enable greater visibility and agility throughout the SC, leads to cost reductions, stabilization of inventory levels and increase of order fulfillment. Due to efficiency gains, the adoption benefits associated with I4.0 technologies in the SC are irrefutable, brought about by cyber-physical systems, RFID, IoT technologies, cloud computing, big data analytics, and advanced robotics [12].

Through absorptive capacity, innovation becomes a joint action between members, where the different relationships between organizations promote not only trust and other shared norms and values, but also the transmission of tacit knowledge [23]. Moreover, the adoption of IoT in operations and in the SC offers commercial benefits, including improved operating processes, low risks and costs, and increased productivity. IoT facilitates the search for new organizational capabilities, from a management and control perspective [16]. The data gathered from the IoT systems provide decision-makers with new ideas about value proposition and value creation, thereby helping to strengthen the bond with customers and to adopt more efficient policies and effective business practices [27]. Through enhanced visibility, transparency, adaptation, flexibility and virtualization in SC [32], companies with the greatest innovation capacity experience increased global competitiveness, reduction in costs, increased market share and increased quality of services [11]. Figure 6 illustrates the benefits and their respective frequency of appearance in the SLR.

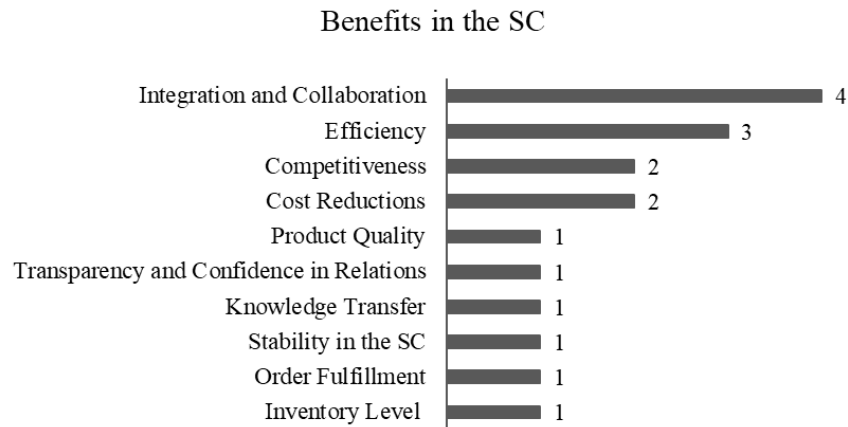


Figure 6. Frequency of benefits in studies

3.5. Results

The Framework below (Figure 7) displays the relationship, in a systematic fashion, between adoption factors, I4.0 technologies and the benefits yielded from adoption in the SC.

Factors such as “Perception of Use”, “Perception of Ease of Use” and “Acquisition and Operating Costs” in AMTs directly affect the intention to adopt I4.0 technologies. Both “Internal and External Social Capital” are mediated by the “Management Support” and “Absorptive Capacity”. Additionally, “Systems Architecture”, “Standards and Validations”, “Integration and Compatibility”, and “Security and Privacy” are highly volatile and have an impact on one other, which may affect the adoption of technology. “Government Policies and Regulations” together with “IT Infrastructure” are determining factors of this adoption structure, forming the basis of integration of processes and performance.

The benefits of I4.0 technology adoption in organizations lead us to reduced operating costs through increases in efficiency in the manufacturing and logistics processes. With a high level of process and IT integration in the SC, organizations can improve their market competitiveness. Finally, they can foster innovation along the SC due to collaboration and knowledge transfer.

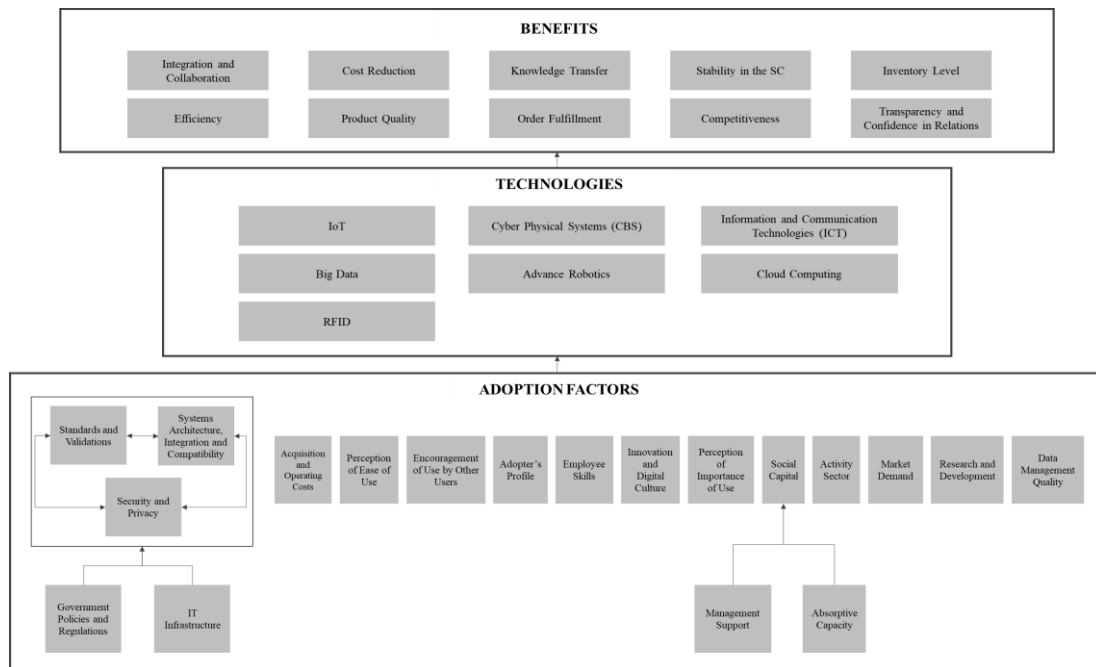


Figure 7. Adoption framework

4. CONCLUSION

The SLR demonstrated, through the results of relevant studies, that the existence or absence of impacting factors, as well as their intensity, affects the adoption of I4.0 technologies. As proposed in the objective of this work, the results from this study contribute to the academic literature related to I4.0 technology adoption by presenting a framework that integrates adoption factors, I4.0 technologies and benefits to the SC. The framework proposed in this study can be easily adapted to serve as a tool in the assessment and selection of technological innovations. This study paves the way for organizations in the adoption process of I4.0 technologies to understand the challenges related to the adoption factors and introduce the potential benefits that can add value to the SC, thereby guiding entrepreneurs in their digital transformation journey.

5. LIMITATIONS AND FURTHER STUDIES

This study has the following limitations:

First: Despite the study showcasing a broad scope of technologies; several other enabling I4.0 technologies did not take part in this study, *e.g.* Extended Reality, 3D Printing and Simulations. Further studies should also try to encompass these technologies.

Second: Due to the broad approach of this study, the adoption process for specific technologies was not deeply explored. Further studies should steer the focus toward a specific technology.

REFERENCES

- [1] Agostini, L., & Filippini, R. (2019). Organizational and managerial challenges in the path toward Industry 4.0. *European Journal of Innovation Management*.
- [2] Agostini, L., & Nosella, A. (2019). The adoption of Industry 4.0 technologies in SMEs: results of an international study. *Management Decision*.
- [3] Ajzen, I., Fishbein, M. (1973). Attitudinal and normative variables as predictors of specific behavior. *J. Pers. Soc. Psychol.* 27 (1), 41.
- [4] Atzori, L., Iera, A. e Morabito, G. (2010). Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.
- [5] Bughin, J., Chui, M., Manyika, J. (2015). An executive's guide to the Internet of Things. *McKinsey Q.* 2 (9), 89–105.
- [6] Caputo, A., Marzi, G., Pellegrini, M. M., Al-Mashari, M., & Del Giudice, M. (2016). The internet of things in manufacturing innovation processes: development and application of a conceptual framework. *Business Process Management Journal*.
- [7] Davis, FD (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319. DOI: 10.2307/249008.
- [8] Dixon, AT, B Thompson and P Mcallister (2002). Report for small business serviceresearch programme the value of ICT for SMEs in the UK: A critical literaturereview, September. <http://centaur.reading.ac.uk/24204/>.
- [9] Ericsson (2016). “Consumer view on future wearables beyond health and wellness”, Ericsson Consumer lab Wearable technology and Internet of Things. Available at: <https://www.ericsson.com/en/press-releases/2016/6/ericsson-consumerlab-personal-safety-to-drive-wearables-market-beyond-health-fitness>.
- [10] Fernandes, C., Ferreira, J., & Raposo, M. (2013). Drivers to firm innovation and their effects on performance: An international comparison. *International Entrepreneurship and Management Journal*, 9, 557–580.
- [11] Ferreira, J. J., Fernandes, C. I., & Ferreira, F. A. (2019). To be or not to be digital, that is the question: Firm innovation and performance. *Journal of Business Research*, 101, 583-590.
- [12] Ghadge, A., Kara, M. E., Moradlou, H., & Goswami, M. (2020). The impact of Industry 4.0 implementation on supply chains. *Journal of Manufacturing Technology Management*.
- [13] Ginsberg, A., & Venkatraman, N. (1985). Contingency perspectives of organizational strategy: A critical review of the empirical research. *Academy of management review*, 10(3), 421-434.
- [14] Hozdić, E. (2015). Smart factory for industry 4.0: A review. *International Journal of Modern Manufacturing Technologies*, 7(1), 28-35.
- [15] Kagermann, H., Helbig, J., Hellinger, A. and Wahlster, W. (2013). Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group. *Forschungsunion*.
- [16] Kamble, S. S., Gunasekaran, A., Parekh, H., & Joshi, S. (2019). Modeling the internet of things adoption barriers in food retail supply chains. *Journal of Retailing and Consumer Services*, 48, 154-168.
- [17] Liao, Y., Deschamps, F., Loures, E. D. F. R. and Ramos, L. F. P. (2017). Past, present and future of Industry 4.0-a systematic literature review and research agenda proposal. *International Journal of Production Research*, 55(12), 3609-3629
- [18] Lobo, F., VASCONCELLOS, E., & Guedes, L. V. (2018). Technology adoption: factors influencing the adoption decision of the internet of things in a business environment. *Towards sustainable technologies and innovation*.
- [19] Lu, Y. (2017), “Industry 4.0: a survey on technologies, applications and open research issues”, *Journal of Industrial Information Integration*, Vol. 6, pp. 1-10.
- [20] Luthra, S. and Mangla, S.K. (2018), “Evaluating challenges to Industry 4.0 initiatives for supply chain sustainability in emerging economies”, *Process Safety and Environmental Protection*, Vol. 117, pp. 168-179.
- [21] Mital, M., Chang, V., Choudhary, P., Papa, A., & Pani, A. K. (2018). Adoption of Internet of Things in India: A test of competing models using a structured equation modeling approach. *Technological Forecasting and Social Change*, 136, 339-346.

- [22] Morioka, S. N., Iritani, D. R., Ometto, A. R., & Carvalho, M. M. D. (2018). Revisão sistemática da literatura sobre medição de desempenho de sustentabilidade corporativa: uma discussão sobre contribuições e lacunas. *Gestão & Produção*, 25(2), 284-303.
- [23] Molina-Morales, F. X., Martínez-Cháfer, L., & Valiente-Bordanova, D. (2019). Disruptive technology adoption, particularities of clustered firms. *Entrepreneurship & Regional Development*, 31(1-2), 62-81.
- [24] Navajo, M., Ballesteros, I., D'Elia, S., Sassen, A., Goyet, M., Santaella, J. (2010). Draft report of the task force on interdisciplinary research activities applicable to the Future Internet. European Union Task Force Report. Available at: http://www.eurosfair.prd.fr/7pc/doc/1265811651_report_future_internet_content_v4_1.pdf.
- [25] Porter, M.; Heppelmann, J. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, November, p. 65-88.
- [26] Rogers, E. (2003). *Diffusion of Innovations*, 5th Edition. New York: The Free Press.
- [27] Roy, S.K., Balaji, M.S., Quazi, A., Quaddus, M. (2018). Predictors of customer acceptance of and resistance to smart technologies in the retail sector. *J. Retail. Consum. Serv.* 42, 147–160.
- [28] Ryan, P.J., Watson, R.B. (2017). Research challenges for the Internet of Things: what role can OR play? *Systems* 5(1), 1-32.
- [29] S. HASSAN, S. O. H. A. I. B., REUTER, C., & BZHALAVA, L. (2020). PERCEPTION OR CAPABILITIES? AN EMPIRICAL INVESTIGATION OF THE FACTORS INFLUENCING THE ADOPTION OF SOCIAL MEDIA AND PUBLIC CLOUD IN GERMAN SMEs. *International Journal of Innovation Management*, 2150002.
- [30] Tornatzky, LG, M Fleischer and AK Chakrabarti (1990). *The Processes of Technological Innovation*. Lexington, MA: Lexington Books. <http://www.worldcat.org/title/processes-of-technological-innovation/oclc/20669819>.
- [31] Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence informed management knowledge by means of systematic review. *British journal of management*, 14(3), 207-222.
- [32] Yan, B., Jin, Z., Liu, L., & Liu, S. (2018). Factors influencing the adoption of the internet of things in supply chains. *Journal of Evolutionary Economics*, 28(3), 523-545.

AUTHORS

José Carlos Franceli

Student of the Master's Program in Production Engineering at the Federal University of ABC (UFABC) with thesis theme “Role of the Top Management in the IoT adoptions in OEM Suppliers and the Supply Chain benefits”. Professional in the Automotive Industry with more than 30 years experience in Supply Chain Management in Brazil, USA and Germany.



Silvia Novaes Zilber Turri

Adjunct professor of the Undergraduate course in Management Engineering and Coordinator of the Master's Program in Production Engineering at the Federal University of ABC (UFABC). Collaborator professor at FEA / University of São Paulo and a full professor at the Postgraduate Program in Management (PPGA) at Universidade 9 de Julho. Evaluator of national and international journals. With published articles in several international journals in the area of Innovation and Digital Transformation of the Value Chain, Business Models for Innovation and Innovation in Small Companies.



FINDTHATQUOTE: A QUESTION-ANSWERING WEB-BASED SYSTEM TO LOCATE QUOTES USING DEEP LEARNING AND NATURAL- LANGUAGE PROCESSING

Nathan Ji¹ and Yu Sun²

¹Portola High School, Irvine, CA, 92618

²California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

The digital age gives us access to a multitude of both information and mediums in which we can interpret information. A majority of the time, many people find interpreting such information difficult as the medium may not be as user friendly as possible. This project has examined the inquiry of how one can identify specific information in a given text based on a question. This inquiry is intended to streamline one's ability to determine the relevance of a given text relative to his objective. The project has an overall 80% success rate given 10 articles with three questions asked per article. This success rate indicates that this project is likely applicable to those who are asking for content level questions within an article.

KEYWORDS

Deep learning, question-answer engine, natural-language processing.

1. INTRODUCTION

The main topic concerning our project is Natural Language Processing (NLP) [1] [15]. NLP is defined as the relationship between computers and human language, but more specifically, in our case, NLP is how a computer interprets human language in textual form. We utilize NLP within the context of the answer engine by allowing the program to interpret the sentence in the text and compare that information to the information within the question.

Natural Language Processing is an extremely important field of computer science because in order for us to be able to make any progression in AI development, we must first be able to understand language. An AI's understanding of linguistics is an important requirement in being able to fully automate the workforce and will most likely be one of the final steps necessary to create fully functional and lifelike AI [2].

Natural Language Processing as well as Question Answer Engines [3] are the staple of future AI and deep learning programs [4] as they are the most vital algorithms that are going to be required for successful integration of AI. If an AI is unable to interpret natural linguistics, then any hope of being able to act more autonomously based on a user's command would be impossible. The entire premise of AI is to make the code more condense and make the actions more broad, which can only be done if the AI can successfully interpret what needs to be done

without the information being hard coded within its software. Question Answer algorithms are vital tests for deep learning and AI programs to test retention as well as the ability to process linguistics. Our ability to determine the successfulness of software without the need to deploy it is also equally important, and thus it is important that Question Answer Engines be developed to test the effectiveness of strategies that will be incorporated in deep learning and AI programs.

Currently, there are a few approaches for combining NLP and question and answer algorithms. So far, the most successful algorithm is a deep learning [5] or AI approach. A good example of how successful this algorithm is the IBM Watson [6] [14], an advanced deep learning question and answer engine that also uses NLP to interpret its inputs. IBM Watson is the modern approach to this problem; it uses a deep learning software that is more accurate, memory efficient, and resistant to time complexity issues. This methodology is most likely the ideal approach to question and answering problems that require NLP, as it is much more efficient and also less naive.

The other approaches used are built upon hardcoding the data derived from deep learning algorithms and the storing of such information in a library--a method that therefore suffers from even greater time and memory issues.

There are multiple libraries, each with their own strengths and weaknesses. Word2Vec [7], the vectors developed by Google, are objectively the most reliable vector set. By combining this with a language comparison model like Gensim [8], its combination of NLP and question and answer can effectively return correct answers a majority of the time.

Another option used is Spacy [9], which was developed by Stanford. The approach using Spacy is to match questions with possible answers using its NLP capability. A massive flaw this approach has is that the question and answer part of this approach is particularly weak due to Spacy's lack of applicable data that can be derived using its software. Compared to the method above, it is less effective and also less accurate. NLTK--Python's Natural Language Toolkit--is a library that is essentially a staple built-in NLP processor used by Python [10]. Much like with Spacy, there is a lack of robust question and answer capability. However, NLTK uses vectors instead of comparison, making it perhaps a little more accurate, but the effects are almost negligible.

My project uses a vector system provided by Google, which is then used by a package called Genism that has a library full of possible English structures. This approach is a very brute force approach, which requires us to hard code all the possibilities, meaning packages like these are extremely large and rather ineffective for commercial use. Although our project works, it doesn't have a decent time complexity nor is it memory.

Initially, this project used Spacy in a naive approach to first interpret the question and the text and then compare the two results in another naive approach to question and answer. This yielded around a 60% accuracy rate, which, compared to the current 80, is much less. This can mostly be attributed to the amount of corner cases Spacy could not accurately interpret and compare, making the more broad option of using vectors and Gensim far better than Spacy. NLTK also would derive similar results to Spacy, making Gensim and Word2vec the best non-AI approach.

Additionally, it should be noted that a deep learning approach that doesn't need to derive its information from vast libraries would easily trump the success of all three naive approaches. Comparatively, an AI algorithm makes the most sense in this scenario given the inherent complexity involved.

In testing our algorithms, there were ten articles with three separate questions, each meant to derive a specific part of text within the article. Each time, we would give the algorithm a question and an article, and it would then return a sentence. If the sentence matched the sentence that was interpreted to be correct, the algorithm would have been successful. The algorithm was successful in doing this around 80% of the time.

2. CHALLENGES

In order to create an application that can streamline one's ability to determine the relevance of a given text relative to his objective, a few challenges have been identified as follows.

2.1. Challenge 1:

When I first approached this problem, I initially used SVOs to match the sentences. However, this approach returned a multitude of errors when I encountered several special cases. Some sentences are structured in OVS, SVVO, and SVOO formats, which make them difficult to read. Furthermore, some sentences had multiple SVOs in them in convoluted patterns like SSVOO or SVVVO. There were too many deviations from the original SVO that the overall accuracy was rather low. Additionally, I was ignoring the prepositional phrases, which further decreased the accuracy. Ultimately, I switched to using Gensim, whereby I changed the approach, and rather than highlighting key words, I used all of the words to find the best match.

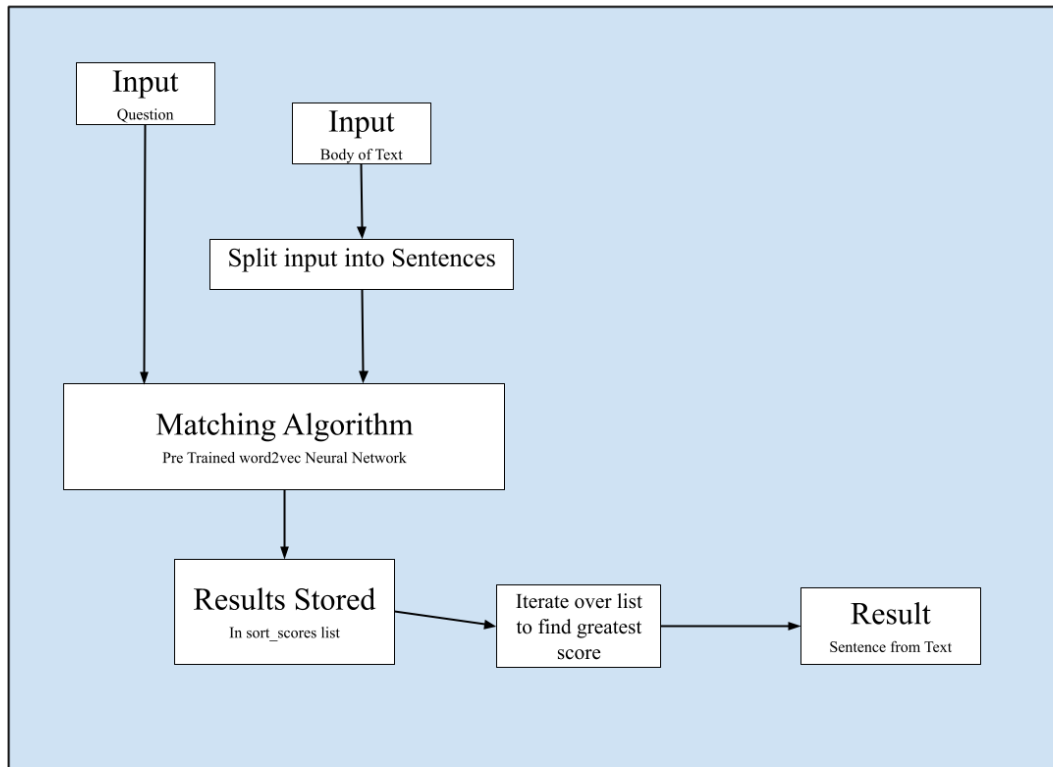
2.2. Challenge 2:

After I figured that a model for matching is the best way to go about answering my goal questions, I needed to use either a vector model or a NLP model to match the words. However, each matching algorithm was extremely different, and all were equally convoluted. Matching singular words didn't work because the context that each model each word was in was different, which ultimately made it impossible for me to only match one word at a time, meaning that it was more difficult to find a model that didn't define each word individually. Eventually, I found Gensim through trial and error and ended up using that.

2.3. Challenge 3:

The 3rd challenge I faced was a natural consequence of challenge 2: the inability to match words when using the model. After determining the best model, there was a lot of difficulty using that model to actually successfully match the words. The inputs for the words were difficult because splitting sentences was extremely difficult. Using Gensim solved this problem because we were able to match entire sentences with vectors in order to establish context. Overall, Gensim proved the best option, as it matched entire sentences.

3. SOLUTION



The main goal of this solution is to match the question given with each sentence within the test to find the answer to the question. When working to achieve this goal, the first step is splitting the sentences within the test using Spacy's sentence splitter. This is important because I couldn't split sentences based on periods because of instances with a title like "Mr." or an abbreviation. Spacy does this by reading through the sentence and establishing the context in which a period is found. If the period is found within a title, which involves a proper noun, it would skip over it. Spacy does this for a multitude of cases, which ultimately divides the sentences correctly. Step two is to input each sentence with the question to my spacyMatching file, which would naturally lead to execution of step three. Step three involves first initializing vectors so that Gensim has all the necessary parameters. We do this by downloading the vectors from the IDE, which then gives us a list of vectors, which we can't interpret but Gensim can. These vectors are a specific list of a series of numbers that define a word based on context, which helps Gensim piece together the meaning of the sentence. These numbers are different for different vector lists, but Gensim can read Google's list accurately. Step four is then to pass each sentence, question, and vector to Gensim, which then returns a number based on how closely these match one another. The greater the number, the greater the similarity. This occurs by comparing the various vectors in the sentence as well as the position of those vectors to achieve a number from 0 to 1, which we can then interpret as similarity. Step five is to rank these numbers with the relevant sentence and then return the requested number of top sentences in the order of similarity.

I used the following imports in my code: Math, Gensim, En_core_web_sm (spacy model), Scipy, numpy,

Flask, flask_cors, JSON, and Spacy.

Step 1 is splitting the sentences within the text. This is achieved by a list comprehension done by SPACY. Essentially, we run a for loop for each phrase that SPACY deems a sentence.

```
doc = self.nlp(text)
#splits text into sentences
sentences = [sent.string.strip() for sent in doc.sents if len(sent) > 2]
# print(sentences)
```

Step 2 is sending the input sentences to spacyMatching. First, I initialized scores for later use as I needed to rank the sentences in step 5. This also helped me store the return values of step 4, which is the similarity ratings for the sentences. What we pass into the function of spacyMatching is our question, sentences, and a variable called wv, which stands for word vectors [11]. This is all the parameters needed for steps 3 and 4.

```
scores = {}
for sentence in sentences:
    scores[sentence] = spacyMatching.matching(self.nlp(question), self.nlp(sentence), self.wv)
#sort scores
sort_scores = sorted(scores.items(), key=lambda x: x[1], reverse=True)
#check for no match
# if sort_scores[0][1] <= 0.5 or math.isnan(sort_scores[0][1]):
if math.isnan(sort_scores[0][1]):
    return scores, ['there was no match']
else:
    result_list = []
    for i in range(numofanswers):
        # sentence_index = sentences.index(sort_scores[i][0])
        result_list.append(sort_scores[i][0])
```

Step 3 is initializing vectors for the matching algorithm. We do this simply by first interpreting the parameter that was imported in the file within step 2 and step 1, which we then use for the next two lines. The next two lines is where step 4 comes in and matches the two numbers. The next line gives us the actual similarity from the two that we use to match the sentences as the best way to quantify similarity is using an imported cosine function that helps derive the average difference between the two numbers.

```
def matching(a, b, word2vec):
    index2word_set = set(word2vec.wv.index2word)
    s1_afv = avg_feature_vector(a, model=word2vec, num_features=300, index2word_set=index2word_set)
    s2_afv = avg_feature_vector(b, model=word2vec, num_features=300, index2word_set=index2word_set)
    sim = 1 - spatial.distance.cosine(s1_afv, s2_afv)
    return sim
```

Step 4 is the actual matching, which was called in step 3. This step first splits the sentence into words and for each of those words, it first checks if it is our set of vectors (index2word_set). Then, if it is, we pass it through the model and add it to our feature vector variable and if not, we get rid of it. We then return the feature vector back to step 3.

```
def avg_feature_vector(sentence, model, num_features, index2word_set):
    stop_words = ['a', 'an', 'the']
    # words = sentence.split()
    feature_vec = np.zeros((num_features,), dtype='float32')
    n_words = 0
    words = [token.text for token in sentence if not token.is_stop]
    for word in words:
        # if word in index2word_set and not word in stop_words:
        if word in index2word_set:
            n_words += 1
            feature_vec = np.add(feature_vec, model[word])
    if (n_words > 0):
        feature_vec = np.divide(feature_vec, n_words)
    return feature_vec
```

Step 5 is ranking the sentences. This is done by simply first sorting our list of scores using lambda and then making a return list to make it easier to return in a readable form.

```
scores = {}
for sentence in sentences:
    scores[sentence] = spacyMatching.matching(self.nlp(question), self.nlp(sentence), self.wv)
#sort scores
sort_scores = sorted(scores.items(), key=lambda x: x[1], reverse=True)
#check for no match
# if sort_scores[0][1] <= 0.5 or math.isnan(sort_scores[0][1]):
if math.isnan(sort_scores[0][1]):
    return scores, ['there was no match']
else:
    result_list = []
    for i in range(numofanswers):
        # sentence_index = sentences.index(sort_scores[i][0])
        result_list.append(sort_scores[i][0])
```

4. EXPERIMENT

My solution uses a pre-trained neural network that identifies similarity between two different sentences, in order to answer the question given within a piece of text. The first few approaches all utilized subject verb object pairs, only reaching up to a 60% success rate with fewer test cases, while the newer approach managed to reach an 80% success rate with more varied and difficult test cases. This newer version is better because we used a pre-trained neural network that was far more efficient as well as accurate due to the heightened abilities of AI.

The results of my algorithm isolated a certain body of text given a question, hence increasing the speed at which someone can analyze the text by allowing them to ignore a large majority of an article, which they didn't need. Although the success rate was only 80%, many of those 20% of cases were extreme corner cases, which we wouldn't see people normally asking.

Summary: 30 Test Cases yielded in an 80% success rate.

5. RELATED WORK

Chapter from Capturing Intelligence, "From search engines to question answering systems—the problems of world knowledge, relevance, deduction and precisiation" by Lotfi A. Zadeh.

The first very clear instance of question answering engines in the world was the application of search engines, the most prominent one being Google. However, Google uses much more advanced algorithms to do this, not only using text relevance, but also systematical deduction, and other logical theories. Although the bases of both projects are similar, Google instead matches searches based on a vast network of data rather than a sole question. This allows Google's results to be far more precise and accurate. Google has perfected its advanced Question Answer design, however one flaw is the vast amount of data this engine would require, obtainable by no more than two or three extremely large corporations [12].

"Improving chronological ordering of sentences extracted from multiple newspaper articles" by Naoaki Okazaki, et al.

This research paper explores a different approach to Question Answering Engines using Information retrieval models, which are an alternative to NLP and matching. Since this is an entirely different approach, while it has the same input and output as my algorithm, it uses chronological ordering and probability to derive its results. It is likely that this algorithm is just as effective, or even more effective, than NLP. Indeed, according to the abstract, "ABRIR uses both a word index and a phrase index formed from combinations of two adjacent noun words. The effectiveness of these two methods was confirmed according to the NTCIR-4 Web test collection," demonstrating that they most likely achieved a rather high accuracy result rate [13].

"Extracting Radiotherapy Treatment Details Using Neural Network-Based Natural Language Processing" by D.S. Bitterman, et al.

This research paper explores the application of neural networks as well as natural language processing in extracting data from cancer treatments, specifically radiation therapy. Due to the lack of a robust NLP system that exists for this task, these researchers developed a neural network to extract data from reports and conglomerate them. This research displays that much

like my research, many subjects and documents can be greatly enhanced through the applications of NLP in conjunction with neural networks, where my design helps find quotations for fields like debate and general writing, while their research creates a concise way to analyze radiation therapy data. Comparatively, their research application goes more specific and in depth than mine as they noted that, "neural networks achieved reasonable performance on RT [Radiation Therapy] detail extraction despite the small dataset and the highly specialized language," whereas my research provides a general summary that will be unable to interpret more specialized datasets [16].

"Semantic Convolutional Neural Network model for Safe Business Investment by Using BERT" by Maryman Heidari, Setareh Rafatirad

This research paper explores the applications of neural networks and semantic analysis in the realm of finances in both a response to the 2008 financial crisis and future real estate investment. Researchers used a semantic convolutional neural network to predict rent to offer a safe real estate investment, ideally to better those attempting to find affordable housing. In this instance, these researchers used semantic analysis contrary to my natural language processing, which although greatly related, semantic analysis is more geared to analysis of financial markets with a binary. Additionally, their research pulls from a much larger data pool, as they used a new public data set with over 5 million houses, allowing for their results to be more accurate and precise [17].

6. CONCLUSION AND FUTURE WORK

Essentially, I have created an algorithm that matches sentences and a question by similarity in order to answer the questions provided by the user. This can be comfortably integrated into any person's day who seeks to quickly find the answer within a large article. The effectiveness of this algorithm is around 80%, which makes it usable, but not ideal. Overall, the algorithm works and helps people accurately identify sections of a passage that may contain the answers to their questions, which can help reduce time wasted on reading lengthy articles.

The main problem with the algorithm is its lack of accuracy. A user would expect a very high level of accuracy to consistently use this, which poses a problem for this specific algorithm. From a practical standpoint, the runtime is also fairly long and rather inconsistent. Given that there is no cap to the data that can be currently inputted, it is plausible that a long document would render the algorithm untenable due to its inefficient runtime. Further work could possibly include an even greater level of AI incorporation where I would train a neural network with just the text and sentence desired to increase accuracy.

REFERENCES

- [1] Chowdhury, Gobinda G. "Natural language processing." *Annual review of information science and technology* 37.1 (2003): 51-89.
- [2] Nilsson, Nils J. *Principles of artificial intelligence*. Morgan Kaufmann, 2014.
- [3] Harabagiu, Sanda, et al. "Falcon: Boosting knowledge for answer engines." *TREC*. Vol. 9. 2000.
- [4] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.
- [5] Goodfellow, Ian, et al. *Deep learning*. Vol. 1. No. 2. Cambridge: MIT press, 2016.
- [6] High, Rob. "The era of cognitive systems: An inside look at IBM Watson and how it works." *IBM Corporation, Redbooks* 1 (2012): 16.
- [7] Goldberg, Yoav, and Omer Levy. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." *arXiv preprint arXiv:1402.3722* (2014).
- [8] Řehůřek, Radim, and Petr Sojka. "Gensim—statistical semantics in python." Retrieved from genism.org (2011).
- [9] Srinivasa-Desikan, Bhargav. *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018.
- [10] Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc.", 2009.
- [11] Lam, Maximilian. "Word2bits-quantized word vectors." *arXiv preprint arXiv:1803.05651* (2018).
- [12] Zadeh, Lotfi A. "From search engines to question answering systems—the problems of world knowledge, relevance, deduction and precisiation." *Capturing Intelligence*. Vol. 1. Elsevier, 2006. 163-210.
- [13] Okazaki, Naoaki, Yutaka Matsuo, and Mitsuru Ishizuka. "Improving chronological ordering of sentences extracted from multiple newspaper articles." *ACM Transactions on Asian Language Information Processing* 4.3 (2005): 321–339.
- [14] Lally, Adam, and Paul Fodor. "Natural language processing with prolog in the ibm watson system." *The Association for Logic Programming (ALP) Newsletter* 9 (2011).
- [15] Liddy, Elizabeth D. "Natural language processing." (2001).
- [16] Bitterman, D.S. et al, (2020, November 1). *Extracting Radiotherapy Treatment Details Using Neural Network-Based Natural Language Processing*.
- [17] Heidari, M., & Rafatirad, S. (2020, December 14). *Semantic Convolutional Neural Network model for Safe Business Investment by Using BERT*. *IEEE Xplore*.

ENHANCEMENT OF CONSISTENT DEPTH ESTIMATION FOR MONOCULAR VIDEOS APPROACH

Mohamed N. Sweilam^{1,2,*} and Nikolay Tolstokulakov²

¹Faculty of Computers and Information, Suez University, Egypt.

²Department of Mechanics and Mathematics,
Novosibirsk State University, Russia.

ABSTRACT

Depth estimation has made great progress in the last few years due to its applications in robotics science and computer vision. Various methods have been developed and implemented to estimate the depth, without flickers and missing holes. Despite this progress, it is still one of the main challenges for researchers, especially for the video applications which have more difficulties such as the complexity of the neural network which affects the run time. Moreover to use such input like monocular video for depth estimation is considered an attractive idea, particularly for hand-held devices such as mobile phones, nowadays they are very popular for capturing pictures and videos. Here in this work, we focus on enhancing the existing consistent depth estimation for monocular videos approach to be with less usage of memory and with using less number of parameters without having a significant reduction in the quality of the depth estimation.

KEYWORDS

Monocular video, monocular depth estimation, deep learning, geometric consistency, lightweight network.

1. INTRODUCTION

Depth estimation of a monocular video presents an attractive point of research for computer vision, and is important for Robotics to provide the distance information needed for different applications, 3D reconstruction of scenes, augmented reality, and object detection. Nowadays, most of the research works are focusing on the unsupervised monocular depth estimation as most of the techniques produce a prediction of depth as a supervised problem and it requires a lot of ground truth depth data for training even for the depth estimation for a single image such as Eigen et al.[1,2] their technique as results have dense pixel depth estimation using a two deep neural network have trained on images and their corresponding depth values, Karsch et al. [3] tried to have a consistent image predictions by taking a copy from the whole depth images from a training data set. The problem in that technique is that it requires the whole training set to be available at the test time. All previous methods and the other supervised methods require a high quality, pixel aligned, ground truth depth data at the training time. But here we perform our work using a single depth estimation network and apply it on the video frames but as an unsupervised method as it needs a stereo color image, instead of ground truth depth during the training time. The Deep3D network of Xie et al. [4] is an unsupervised technique aiming to produce the corresponding right view from an input as a left image to be the context of binocular pairs. The

right image pixels as the results are a combination of the pixels on the same scan line from the left image which was the input, weighted by the probability of each disparity. The disadvantage of this technique is that increasing the number of disparity values leads to increase in memory consumption, which makes it difficult to apply on bigger data or bigger resolution. For depth estimation from a video is a challenging problem recently because of moving objects and camera pose ,that's why the video depth estimation technique suffering from poorly textured areas, occlusions and repetitive patterns. The existing techniques for that purpose rely on motion segmentation and explicit motion modeling for the moving objects like [5]; Moreover now the one of the easiest ways to capture a video is by hand held camera phones which leads to more challenges such as high noise level ,lightening ,motion blur and shaking that's why the existing method produce some errors in the depth estimation such as missing regions which make some white holes in the depth, in addition to it's consider as an inconsistent geometry depth and flickering depth such as in figure [1] ,that's why the geometrically consistent approaches have the best results and accurate ones but suffering from complexity and long test time as the The produced depth is flicker free and geometrically consistent throughout the input monocular video. For these reasons we have decided to enhance the Consistent Video Depth Estimation approach [6] by making that approach use less memory and lighter depth estimation network.

2. CONSISTENT VIDEO DEPTH ESTIMATION

The Consistent Video Depth Estimation approach produces a single image depth estimation and further improves the geometric consistency values of the depth estimation on the videos. It contains two phases :

2.1. Pre-processing

In that phase the approach performs a traditional Structure from Motion (SfM) reconstruction using the open source software COLMAP [7] and using Mask R CNN [8] to detect people segmentation and remove these regions to make it more reliable for keypoint extraction and matching. This phase is important to provide accurate intrinsic and extrinsic camera parameters in addition to a sparse point cloud reconstruction.

2.2. Test-time Training

In that phase what happens is that the approach takes two frames randomly and then have their depth images after processing them to a single depth estimation network and usually the results of that network will have some flickers. Moreover, it calculate the camera pose using COLMAP [7] , then assume we have a point on an image then the approach will find the corresponding point in the other frame using Optical Flow and re project the two point in a 3-D scene and here this process break down into two components, 1) re project the a point to another camera and calculate the distance on image plane and 2) re project a two points along the Z axis and compute the difference. Because of the depth is inconsistent there will be a distance between the two point in the 3-D scene which called geometric losses as Spatial Loss for distance between the two points in the screen space and Disparity Loss for the distance between the two points in the depth space.

For these losses the depth is inconsistent so the approach takes the two losses and fine tune the initial single depth estimation network by back propagation at the test time for all pairs of frames. Finally the approach will estimate a sharper and geometric consistent depth which will be very accurate and better than before.

2.3. Geometric Losses

This approach has two geometric losses as the following if we assume we have a given frame pair(i,j) so:

$$L_{i \rightarrow j}^{spatial}(x) = \|p_{i \rightarrow j}(x) - f_{i \rightarrow j}(x)\|_2 \quad (1)$$

Where the flow displaced point is, $f_{i \rightarrow j}(x)$ the depth reprojected point $p_{i \rightarrow j}(x)$. so the image space loss (1) indicates the distance in the image space between the flow displaced point and the depth reprojected point.

$$L_{i \rightarrow j}^{disparity}(x) = |u_i |z_{i \rightarrow j}^{-1}(x) - z_j^{-1}(f_{i \rightarrow j}(x))| \quad (2)$$

Where u_i is the frame's focal length, the z components are scalar z components from a 3D point in the frame's camera coordinate system.

The Total loss is a combination of both losses for all pixels:

$$L_{i \rightarrow j} = \frac{1}{|M_{i \rightarrow j}|} \sum_{x \in M_{i \rightarrow j}} L_{i \rightarrow j}^{spatial}(x) + \lambda L_{i \rightarrow j}^{disparity}(x) \quad (3)$$

Where $\lambda=0.1$ is a balancing coefficient.

2.4. Optimization

That approach takes the geometric loss between the frames and fine-tunes the initial depth estimation network using the standard back propagation.

Having the parameters of this network using a pre-trained network for depth estimation allows the approach to transfer the knowledge to produce the depth map on the images that are already considered as challenging for traditional geometric based reconstruction. This approach fine tune using 20 epochs for all experiments.

3. PYD-NET

It's a lightweight network called Pyramidal Depth Network (PyD-Net) [9], when we train it in an unsupervised method, it represents a high accuracy in that field.

If we compare that model after training to the others, this model is about 94% smaller as it can even work on CPUs without reducing the accuracy slightly, it requires limited resources only. Moreover, the PyD-Net can be deployed even on embedded devices, such as the Raspberry Pi 3, allowing to have depth with low number of parameters using less than 150 MB memory available at test time because the other approaches count a huge number of parameters and thus require a large amount of memory For Example with the VGG model [19], counts 31 million parameters, however in the Pyd-net number of parameters reach to 2 Millions of parameters. Which makes Pyd-net more efficient for low power devices or CPUs.

3.1. The Architecture

The PyD-Net architecture in Figure 2, as it contains a pyramid of features coming from the input image and at each level of that pyramid a network build depth. The features which they proceed are sampled to the upper level to refine the estimation, up to the highest one.

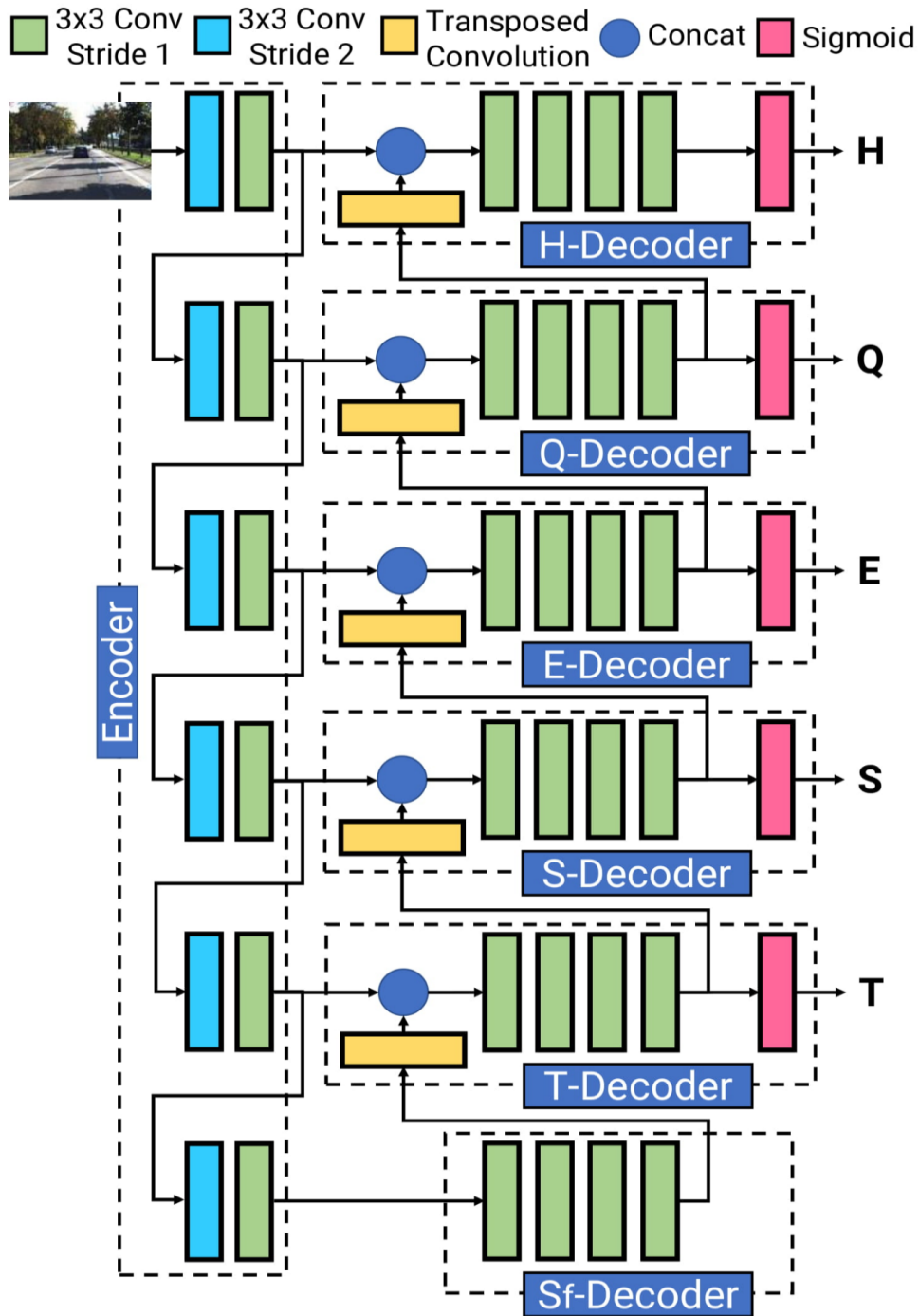


Figure 2. Pydnet architecture

3.2. Pyramidal Features Extractor

It's a small encoder made by [10], made of 12 convolutional layers. At full resolution, the first level of the pyramid is produced by the first layer by applying convolutions with stride 2 followed by a second convolutional layer. By having the same technique for each level the resolution is reached to the lowest resolution at the highest level, the total number of levels is 6 levels, from L1 to L6, corresponding respectively to the image resolution from 1/2 to 1/64 of the original input size.

Every down sampling module builds a number of extracted features, respectively 16, 32, 64, 96, 128, and 192, and each convolutional layer deploys 3×3 kernels and is followed by a ReLU with $\alpha = 0.2$.

3.3. Depth Decoders and Upsampling

There is a decoder at the highest level of the pyramid, the decoder made of 4 convolutional layers, producing respectively 96, 64, 32 and 8 feature maps.

This decoder has two purposes: 1) to produce a depth map at the current resolution, by means of sigmoid operator, and 2) to pass the features which processing to the next level in the pyramid, by means of a 2×2 deconvolution with stride 2 which increases by a factor 2 the spatial resolution.

The next level matches the features extracted from the input frame with the features which are sampled and processes them with a new decoder, repeating this procedure up to the highest resolution level.

Each convolutional layer uses 3×3 kernels, leaky ReLU activations, just the last one followed by a Sigmoid activation for normalizing the outputs. This design makes at each scale the PyD-Net to learn to produce depth at full resolution.

4. MONODEPTH

A new training method made by C. Godard, O. Mac Aodha, and G. J. Brostow[11] for enabling the convolutional neural network to learn to make a single image depth estimation, with the absence of ground truth depth data. Using the epipolar geometry constraints, the method generates disparity images by training the network with an image reconstruction loss. That is why it is considered as a new training loss which enforces consistency between the disparities estimated according to the left and right images, which leads to improve the performance and robustness compared to the existing techniques. This method has state of the art results for monocular depth estimation when it is trained on the KITTI driving dataset, and even better than the supervised methods which have been trained on ground truth depth.

4.1. Sampling Strategies

Sampling strategies for backward mapping as in figure [3] here which originally in [11]. With naive sampling the CNN generates a disparity map of the right image from the left image and the disparity map aligned with the right image which is the target.

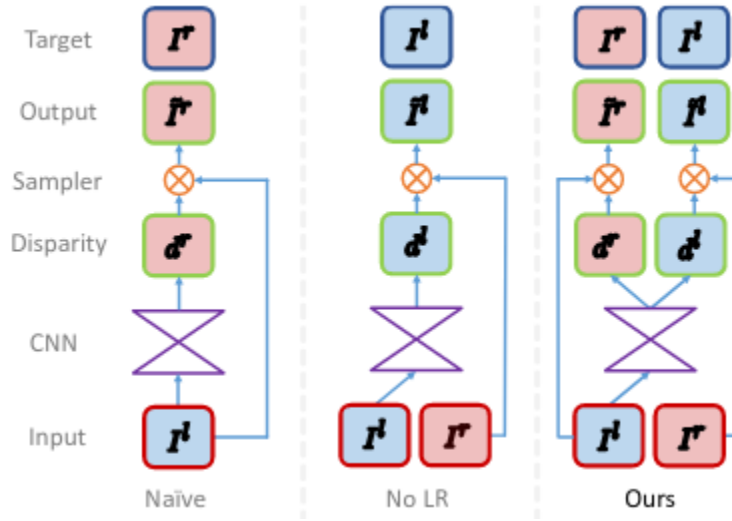


Figure 3

But the output should be the disparity map of the left image which considered as the input therefore the network has to sample it from the right image that's why the second strategy which is No LR is to train the network to produce the left view from the right image and creating a disparity map aligned to left view but at that strategy there are some errors like exhibit 'texture-copy' artifacts at depth discontinuities and as a solution for that the approach has to be trained to produce disparity maps for both right and left views by sampling it from the opposite input image. It requires only one image during the test time as the left view but in training time needs the right view image too .Enforcing the consistency for both disparity maps like that leads to better accuracy.

4.2. The Network Architecture

The architecture here by Disp-Net [12], with some modifications to train with the absence of ground truth data .The network contains two main parts: an encoder and decoder. The decoder uses a technique called skip connections [13] from the encoder's activation blocks, it helps for high resolution details. The output predictions contain four scales (disparity 4 to disparity 1); it is going to be doubled in spatial resolution at every subsequent scale. Moreover, it only takes a single image as input, the network produces two disparity maps at every output scale left to right and right to left.

4.3. Training Loss

The loss C_s here calculated at each outer scale s , the C_s has a three parts

$$C_s = \alpha_{ap} (C_{ap}^l + C_{ap}^r) + \alpha_{ds} (C_{ds}^l + C_{ds}^r) + \alpha_{lr} (C_{lr}^l + C_{lr}^r) \quad (3)$$

Where C_{ap} to make the reconstructed image to similar to the corresponding training input image , ap C_{ds} for the smooth disparities, and C_{lr} to make the produced left and right disparities consistency.

$$C_{ap}^l = \frac{1}{N} \sum_{ij} \alpha \frac{1-SSIM(I_{ij}^l, \hat{I}_{ij}^l)}{2} + (1 - \alpha) \|I_{ij}^l - \hat{I}_{ij}^l\| \quad (4)$$

this equation (4) for reconstruction error calculating the difference between the original image I and the warped one \hat{I} by using SSIM [14].

$$C_{ds}^l = \frac{1}{N} |\partial_x d_{ij}^l| e^{-\|\partial_x d_{ij}^l\|} + |\partial_y d_{ij}^l| e^{-\|\partial_y d_{ij}^l\|} \quad (5)$$

Equation (5) for disparity smoothness, it tries to make the disparities to be locally smooth on the disparity gradients ∂ .

$$C_{lr}^l = \frac{1}{N} \sum_{ij} |d_{ij}^l - d_{ij+d_{ij}}^l| \quad (6)$$

Equation (6) represents the Left Right Disparity Consistency Loss as the approach trying to produce more accurate disparity maps, the training of the network to produce both the left and right image maps, however the only input is the left view to the network.

At test time, the disparity for the left image d^l , it has the same resolution as the input image. While estimating the right disparity d^r during training, it is not used at test time. Using the camera baseline and focal length from the training set, the approach converts the disparity map to a depth map.

5. EXPERIMENT

Our Experiment is to enhance Consistent Video Depth Estimation approach by changing the initial depth estimation for single image in that approach to be more lighter and use less memory, so we decided to use a lightweight architecture for network which we chose to be the Pyd-net architecture[9] as it has the ability to enable such an accurate and unsupervised monocular depth estimation with very limited resource requirements, but it needs a framework to train with, therefore we used Monodepth framework[11] for training as it has better results even than the supervised methods that have been trained with ground truth data. After testing the pretrained network we succeeded to integrate the network in the Consistent Video Depth Estimation approach and run the model in test time and observe the fine tuning process as it used 20 epochs for fine tuning, then we evaluated the results, tested on hand-held videos and compared.

5.1. Dataset

KITTI dataset [15] as it has been recorded from a moving platform while driving in and around Karlsruhe, Germany. Using KITTI Split which contains contain 30,159 images almost 175 GB, we keep 29,000 for training and the rest for evaluation .we used unlabeled stereo pairs of images as according to the approach we are using for training, we need at the training phase to have right and left view, but in test time we need just one image.



Example from the KITTI dataset of a stereo image the upper one is right view and the below one is the left view

5.2. Implementation

We implemented our work in Pytorch [16], therefore we had to reimplement the whole architecture of Pynet into Pytorch as it is official as tensorflow. We also had to implement the framework of Monodepth in Pytorch and train it. We use the specifications of Novosibirsk State University for training, we used a GeForce GTX TITAN X GM200 as GPU, trained using 200 epochs with a batch size of 12 and using Adam optimizer [17] and the loss during training shown in figure [4].

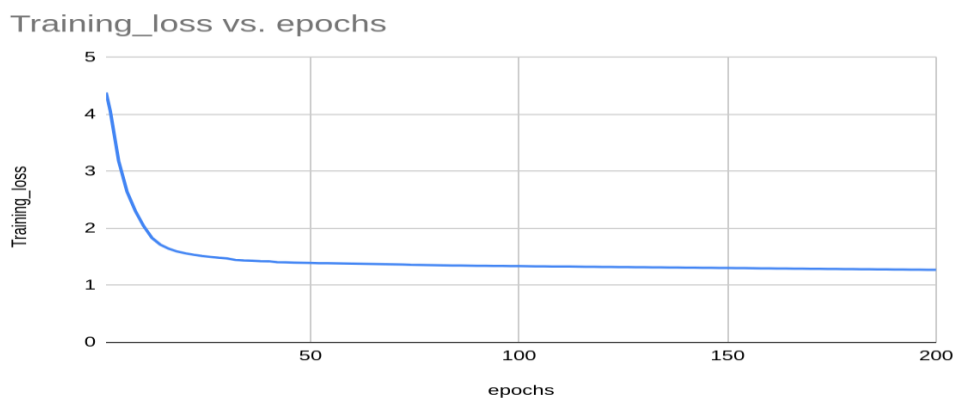


Figure 4. Train loss during training for 200 epochs.

5.3. Evaluation Metrics

To compare with the previous methods we use the popular Evaluation metrics for comparison several errors from prior works such as [18] and we used these metrics: SQ_REL : Relative squared error, ABS_REL : Relative absolute error, $RMSE$: Root mean squared error of the inverse depth and $RMSE$ (log)

5.4. Results

For the visual results we didn't notice a significant difference between our results after the modification and the previous results even with our videos which are shown in figure [5], figure [6].



Figure 5. a test video frames from the previous work , the upper frame and its depth is using the existing approach , the below frame and its depth is using the approach after enhancement.

For Quantitative results we compared using the evaluations metrics between our modification and two from the previous models for depth estimation approaches ,shown in table [1]:

Table 1: comparison of depth Models which are all trained and evaluated on KittiRAW

Model	ABS_REL	SQ_REL	RMSE	RMSE (log)
<i>Ours</i>	0.149	1.250	5.464	0.228
Monodepth2	0.132	1.042	5.138	0.210
Fast depth	0.317	13.325	10.207	0.384

We noticed that our modification has very close numbers to the Monodepth 2 [21] the better than fast depth [20], however it have a significant difference in the memory usage as we have done an experiment to compare the memory usage, we have noticed that Monodepth2 using 0.95 GB of memory at the test time and Monodepth without any modifications using 2.14 GB at test time, in the other hand our modification with Pyd-net uses less than 150 MB for all experiments. Which is considered as a one step to make the Consistent Video Depth Estimation faster and lighter so it can be applied on cell phones or low power devices. Moreover we observed the geometric loss during the test time in that approach after modifying it and we noticed that the geometric loss values before and after modifications were in the same range between 0.2 and 0.8, which means that the new initial depth network did not cause more geometric loss than before.

6. CONCLUSION

In this research, we have proposed a new modification for the Consistent Video Depth Estimation approach which uses a huge of memory at the test time, therefore we have reduced that amount by changing the initial depth estimation network for a single image in that approach with a new one enhanced by a lightweight architecture can be used for low power devices and mobile phones which is Pyd-net. After testing, the results showed that there is no significant difference in the depth quality, however there is a significant difference in the memory usage at the test time. The future work is to try to focus more on the geometric consistency to make it less complex and lighter which can make the approach in future lighter to work on mobile phones or low power devices.

ACKNOWLEDGEMENT

We gratefully acknowledge the support of Stream Data Analytics and Machine Learning laboratory at Novosibirsk State University for providing us the specifications we needed to complete this research.

REFERENCES

- [1] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. InICCV, 2015
- [2] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. InNIPS, 2014.
- [3] K. Karsch, C. Liu, and S. B. Kang. Depth transfer: Depth extraction from video using non-parametric sampling.PAMI,2014.

- [4] J. Xie, R. Girshick, and A. Farhadi. Deep3d: Fully automatic 2d-to-3d video conversion with deep convolutional neural networks. In ECCV, 2016
- [5] Vincent Casser, Soeren Pirk, Reza Mahjourian, and Anelia Angelova. Depth Prediction without the sensors: Leveraging structure for unsupervised learning from monocular videos. In AAAI Conference on Artificial Intelligence (AAAI), 2019.
- [6] Luo, Xuan and Huang, Jia Bin and Szeliski, Richard and Matzen, Kevin and Kopf, Johannes. Consistent Video Depth Estimation. In ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH), 2020.
- [7] Johannes L Schonberger and Jan-Michael Frahm. Structure from motion revisited. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [8] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In International Conference on Computer Vision, 2017.
- [9] M. Poggi, F. Aleotti, F. Tosi, and S. Mattoccia, "Towards real-time unsupervised monocular depth estimation on cpu," in IROS, 2018.
- [10] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," arXiv preprint arXiv:1709.02371, 2017.
- [11] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in CVPR, vol. 2, no. 6, 2017, p. 7.
- [12] N. Mayer, E. Ilg, P.H. Ausser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In CVPR, 2016.
- [13] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. PAMI, 2016
- [14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.
- [15] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
- [16] Paszke, Adam and Gross, Sam and Chintala, Soumith and Chanan, Gregory and Yang, Edward and DeVito, Zachary and Lin, Zeming and Desmaison, Alban and Antiga, Luca and Lerer, Adam. Automatic differentiation in PyTorch. 2017.
- [17] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations, 12 2014
- [18] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In NIPS, 2014.
- [19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [20] Wofk, Diana and Ma, Fangchang and Yang, Tien-Ju and Karaman, Sertac and Sze, Vivienne. "FastDepth: Fast Monocular Depth Estimation on Embedded Systems", IEEE International Conference on Robotics and Automation (ICRA) .2019.
- [21] Godard and Oisin {Mac Aodha} and Michael Firman and Gabriel J. Brostow. "Digging into Self-Supervised Monocular Depth Prediction" .The International Conference on Computer Vision (ICCV).2019.

AN INTELLIGENT MOBILE APP TO DETECT DROWSY DRIVING WITH ARTIFICIAL INTELLIGENCE

Thomas Xiao¹ and Yu Sun²

¹Yorba Linda High School, Yorba Linda, CA 92886

²California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

Drowsy driving is lethal- 793 died from accidents related to drowsy driving and 91000 accidents related to drowsy driving occurred [1]. However, drowsy driving and accidents related to drowsy driving are preventable. In this paper, we address the problem through an application that uses artificial intelligence to detect the eye openness of the user. The application can detect the eyes of the user via computer vision. Based on the user's eye openness and frequencies, the sleepy driving condition can be inferred by this application. We applied our application to actual driving environments on the highway, both day and night, as well as within a normal control situation using a qualitative evaluation approach. The result shows that it is 88% effective during the day and 75% effective during nighttime. This result reveals effectiveness and accuracy of detection during daytime application under controlled testing, which is more flexible and efficient comparing to previous works. Effectiveness and accuracy for nighttime detection and detections with the presence of other distractions can be further improved.

KEYWORDS

Drowsy driving, Mobile Application, Artificial Intelligence, Driving Safety.

1. INTRODUCTION

Drowsy driving, as simple as it sounds, is driving while sleepy [10, 11, 12]. Drowsy driving is an acute problem. It can occur with any driver of any age group around the world, affecting millions even if they are not behind the wheel. The Centers for Diseases control and Prevention (CDC) once estimated that 1 in 25 adult drivers report having fallen asleep while driving in the previous 30 days [1]. Drowsy Driving often ends in accidents with a variety of effects: car crash, injury, destruction of interstates and roads, and in the worst-case scenarios, death. In fact, the National Highway Transportation Safety Authority estimated that there were 72,000 crashes, 44,000 injuries, and 800 deaths related to drowsy driving in 2013, and some believe even this is underestimated [2]. In California alone, in 2016, 2% of traffic accident deaths were caused by drowsy driving [3]. Each one of these numbers represent human lives, and there are ways to prevent all these avoidable injuries and deaths. Those who are injured badly by car accidents often end up in the ICU, which is even more dangerous during the Coronavirus Pandemic. When people are sleepy behind the wheel, they tend to display certain eye patterns that are recognizable. With a functioning app that can detect and notify the user of drowsy driving, many lives could be saved.

There are not many technological techniques to detect drowsy driving. However, the most advised technique for drivers to prevent drowsy driving is to rest well before driving and stop driving immediately when feeling drowsiness; this method has many flaws since many people do not have enough conscientiousness to first consider rest. For example, most drivers would not want to rest in the middle of a trip. Other examples would be commercial drivers whose job is to drive for most of their days. There are different methods developed by universities, such as the research conducted by H.J Dijkers and M.A. Spaans from Delft University of Technology [6, 13]. Their research conducted on drowsy driving detection depended on facial expression detection. Facial expression detection requires three steps of detection, which are: Facedetection, Facial expression data extraction, and facial expression classification [4]. Although this method can be extremely accurate, it was tested on a computer with an old-style recording camera. The results were accurately collected but have never been projected onto the UI of a mobile application. Their method has never been tested on a modern-day cell phone and is too complex to run in real time on a phone, so is therefore not suitable as a mobile phone app. Also, the algorithm required facial expression detection, which means that facial hair or face coverings on the subject's face may produce inaccurate or false results, which is extremely unhelpful during Covid-19.

In this paper, we present a new approach to detect and curb drowsy driving. Our goal is to develop an application with an algorithm that would use the detection of eye openness to determine if the driver is driving under drowsy conditions. The method we have developed relies on the collaboration of Google Firebase's eye detection algorithm. This algorithm was written in Dart language and utilizes the Flutter Camera package plug in to access the phone's face cam [14, 15]. When the user clicks the start button of the application, the algorithm is programmed to automatically take 10 photos per second. All the pictures taken are analyzed with Google Firebase's eye openness detection. The Google firebase eye detection can detect the eye openness of a subject and returns a value from 0 to 1 based on how much their eyes are open (0 means closed, while 1 means open). Based on the information returned by Google firebase, we can make calculations of the value returned by Google Firebase to determine if the driver is driving while drowsy. The algorithm determines when to notify the user that they are sleepy through a specifically designed calculation, which will be discussed in the next section. There are some good features present in the algorithm of this current app. First, the algorithm detects eye openness without relying on facial expression. This would be extremely helpful because facial hair and other face coverings may affect the accuracy of results or return a false result to the user. Second, this method has a complex algorithm (a multiple step calculation shown in the third section) to determine if the driver is drowsy when driving. This ensures the data is returned accurately and there are no false alarms [9].

To evaluate the accuracy of our method, we tested the accuracy of the application's design within different situations that drivers can experience in real life. The factors that we decided to test include Day time without glasses vs with glasses, nighttime without glasses vs with glasses, shade without glass and with glass, location of the phone glass vs without glass. Within each trial, we observed if the detection would trigger a warning. The results of these real-time detection proves to be effective overall.

The rest of the paper is organized as follows: Section Two gives the details on the challenges that we met during the experiment and while designing the sample; Section Three focuses on the details of our solutions corresponding to the challenges mentioned in Section Two; Section Four presents the relevant details of the experiment, followed by the related work in Section Five. Finally, Section Six gives concluding remarks, as well as the future work of this project.

2. CHALLENGES

A few challenges arose while developing this application, as follows.

2.1. Which Part of the Face to Detect

There are multiple parts of our face we can monitor for drowsy driving, such as eyes, mouth, facial expression, or all of them. Each detection has its own benefits and downsides. We can use eyes because if we close our eyes too frequently or our eye openness becomes too small, we can assume the user is tired. We can also use facial expression because when people are tired, they tend to have certain facial expressions. It would be an extremely accurate detection if we could use all the mechanisms concurrently, yet we are developing an application that runs on a cell phone so the algorithm cannot be overly complicated such as could be run on a computer. Each type of detection has its downside, e.g., you cannot determine the drowsiness of a user through eye detection if the driver is wearing sunglasses [4].

2.2. Unprecedented Challenges

An algorithm that can accurately detect drowsy driving is the key to developing a successful application. But there are not many previous works or methods from which we may gain insight. All previously published works involved data analyzed on a computer, not a smartphone. These previous apps were never intended to be an algorithm used by a phone application. We are trying to develop an application that runs on a phone in which the app's algorithm correctly detects and warns the user of drowsy driving, so we are coming up with our own unique algorithm suitable to run on a phone. We were also faced with the challenge of how the app would be able to gather data constantly and automatically from the user through the phone's camera.

2.3. Designing an Effective Algorithm

As stated earlier, the core of the application is an effective and accurate algorithm. There are many different algorithms we can implement for this app. However, since this app is going to be used on real roads by real drivers, there are many unpredictable factors that we need to consider. We also need to determine what is the most effective way to gather data from the user and how to calculate a value to determine if the user is sleepy. We will need an extremely effective and accurate way of collecting data on the user's face movements. We also need to consider the most accurate algorithm to determine when to notify the user.

3. METHODOLOGY

An overview of the system is presented in Figure 1. The user would first have to choose a sound (input) they would like to play for the detection (input by user). Then when the user starts the detection, the phone's face camera would gather the value for each of the eyes with the help of Google firebase eye detection. After gathering the data of the user's eyes, the algorithm would then determine if the value collected is considered closed or open. Then the value would be passed on to two calculations. One for the eye opened, one for the eye closed. The calculated value would then determine when the notification sound would be played. At the same time, a timer also records how long the user has been detecting. While detecting, the user can choose to pause the detection at any time. When the detection is done, the user interface would display the overall status of the user- in this case sleepy or not- and the time they have been using the detection.

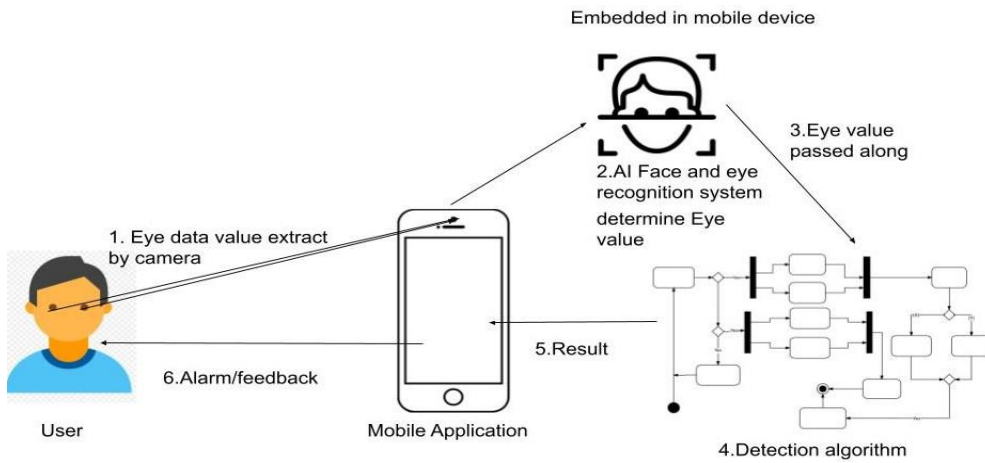


Figure 1. Schematic of system

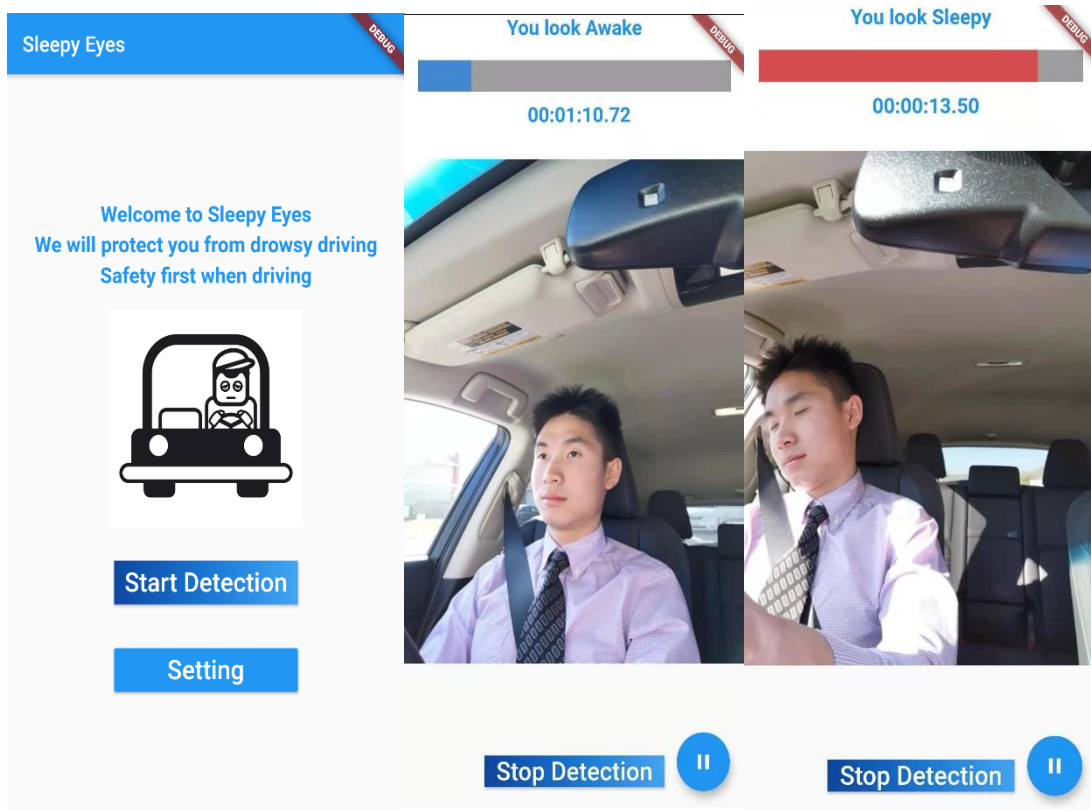


Figure 2. System in use

[image 1] Note the progress bar, first calculation.

[image 2] Note that nothing shows in the progress bar, eyes are open, second calculation is conducted.

[image 3] Progress bar more than halfway, which means “closed_to_total” is greater than 0.5, notification sound is played.

3.1. Camera

To be able to collect the user's eye value, we need to use the user's face cam to get the value. We were able to do this by first implementing a camera package. First, we add camera to the pubspace.yaml file. Then we import the camera package. Finally, we set up the camera, so the camera usage of this app is through the face camera.

```
dependencies:
  flutter:
    sdk: flutter
  camera: ^0.5.8+2
  path_provider: ^1.6.4
  path: ^1.6.4
  fi: ^0.1.3
  image: ^2.1.12
  permission_handler: ^5.0.1+1
  firebase_ml_vision: ^0.9.6+2
  firebase_core: ^0.5.0
  flutter_ringtones_player: ^2.0.0
  flutter_animation_progress_bar: ^1.0.5
  audioplayers: ^0.16.1
  stop_watch_timer: ^0.6.0+1
  fluttertoast: ^7.1.1
```

Figure 3. Set up

```
Future<CameraController> initCameras() async {//setting up camera

  // Obtain a list of the available cameras on the device.
  final cameras = await availableCameras();//asking phones for available camera

  return availableCameras()
    .then((camera) async {
      _controller = CameraController(//let app operate the camera
        // Get a specific camera from the list of available cameras
        cameras[1],
        // Define the resolution to use.
        ResolutionPreset.medium,
      );

      await _controller.initialize();

      runDetection();

      return _controller;
    });
}
```

Figure 4. Set up code

3.2. GoogleFirebase (ML core)

We implemented Google firebase face detection. The face detection can return values for the user's eyes positions and eye openness. For this application, we need the data collection of the eye openness value. We allow the app to retract eye openness values of the user by adding

Google-service.json file to the app's source (src) folder, which would enable us to use its face recognition service. We need another class file specifically in the lib folder along with other codes in order for the Google-service.json file to work perfectly alongside of other aspects of the app and to implement a face detector. After adding the Google-service.json file, we were able to utilize and design an AI eye detection system with the help of basic Google firebase face recognition system. The picture below is how we utilized the face detection code originally given and added in some of our codes to retrieve the eye value. With this code, the app can work perfectly with other parts of the program:

```

static EyeInfo findEyesOpen(List<Face> faces) {
    print("faces: " + faces.length.toString());
    if(faces.length == 0) {
        return EyeInfo(
            leftEye: null,
            rightEye: null
        );
    }

    double largestFaceSize = -1.0;
    Face largestFaceValue = null;

    for(var f in faces) { //prevent too many people, assume the biggest face is the drive
        double size = (f.boundingBox.right-f.boundingBox.left)*(f.boundingBox.bottom-f.boundingBox.top);
        if(size > largestFaceSize) {
            largestFaceSize = size;
            largestFaceValue = f;
        }
    }

    return EyeInfo(
        leftEye: largestFaceValue.leftEyeOpenProbability,
        rightEye: largestFaceValue.rightEyeOpenProbability
    );
}

```

Figure 5. Adding code to retrieve eye values

The AI face detector will be used alongside later in many places. The image above is the basic codes of the face detector.

3.3. Detection Algorithm

The most important part of this app is the detection algorithm. We show all the logic gates that would trigger the system to notify the user being sleepy. The previous two sections (Camera and Google Firebase) are just setting up the basics for this detection. This part is when we put everything together. To be able to run the detection well, we need to first set the camera to be able to continuously take pictures and for the Google firebase to detect the face. The camera is automatically set to take one picture every minute during detection. The detection begins with the algorithm taking one image of the user. If nothing goes wrong with taking the image, the algorithm will go on and construct a path to store the image in a temporary directory and then use the path-provider plugin (plugin for finding commonly used locations on the filesystem) to locate it [5]. Then, the new image path taken would be stored and the old image path deleted. After the image has been temporarily stored in the path, the face detector mentioned previously would determine if there are any faces. If there are, then it would retrieve the user's eye value. Then, the

eye value retrieved would be entered into the following calculation. We defined any value below 0.3 as sleepy. If the eye value is smaller than 0.3, a variable named “closed_to_total” starting with an initial value of 0 would be going through this formula:

$$\text{closed_to_total} = \text{closed_to_total} * (3/4) + 1/4;$$

The calculation is done this way because we would like the progress bar (shown below) to increase not linearly or exponentially. This is designed to quickly notify the user if the detection detects that he/she is sleepy. If the retrieved eye value is greater than 0.3, then the following calculation would be conducted:

$$\text{closed_to_total} = \text{closed_to_total} * (15/16);$$

This would decrease the progress bar much slower than when the bar is increasing. It is done this way to let the user go to rest quickly and let the notification sound keep playing. The “closed_to_total” variable determines when the app would play a notification sound. If the “closed_to_total” value is greater than 0.5, then the notification sound would be played. The sound would keep playing (as a loop) unless closed_to_total is smaller than 0.5. The image below shows the user interfaces when the first calculation is conducted and when the second calculation is conducted. It also shows when the notification sound is played.

```
void runDetection() async {
  while(true) {
    if(!onPage) return;
    // Take the Picture in a try / catch block. If anything goes wrong,
    // catch the error.
    if(isRunning) {
      try {
        // Construct the path where the image should be saved using the path
        // package.
        final path = join(
          // Store the picture in the temp directory.
          // Find the temp directory using the `path_provider` plugin.
          (await getTemporaryDirectory()).path, //gets the name of the image taken
          'currentImage.png',
        );
        if(await File(path).exists()) {
          await File(path).delete();
        }

        // Attempt to take a picture and log where it's been saved.
        await _controller.takePicture(path); //saves the picture from the source

        List<Face> faces = await fd.detectFaces(path); //detects face only
        //1 picture every second, delete the previous pictures taken
        await ec.addEyes(AbsFaceDetector.findEyesOpen(faces)); //detect open eyes
        sleepyness = ec.sleepyness();
        result.add(sleepyness); //replaces old values in the result stream
        totalPicTaken++;
        status.add(ec.status()); //get the user's statuses & puts it into a string
      } catch (e) {
        print(e);
      }
    }
    await pause(const Duration(milliseconds: 1000));
  }
}
```

Figure 6. code for “closed_to_total” values and sound notification

3.4. AI and algorithm integration

As shown above, we used artificial intelligence eye and face recognition to first detect the face currently in the camera's frame and the eye values on the face. After the AI face recognition retrieves the eye's data, we use these data and put them through the calculation and decide when to notify the user of being sleepy.

4. EXPERIMENT

To evaluate the accuracy of our method, we decided to conduct experiments in real-life driving situations. We decided to test the accuracy of this application's design in many situations that drivers can experience in real life. The factors that we decided to experiment include Day time Without Glasses vs Glasses, Nighttime Without Glasses vs Glasses, Shade Without glass and with glass, location of the phone glass vs without glass. For the first three experiments, the phone is set up at 70 cm from the tester (just as the image below) while for experiment four, the phone is set up at 87cm from the user. Within each trial, we are going to see if the detection would trigger a warning. The accuracy of each detection would be determined as follow: if the test subject's eyes are closed and no alarms have sounded, then the accuracy is 0%. But if the test subject's eyes are open and the alarm goes off, then the accuracy level is 100%. For example, if the detection is done accurately three times out of the four times, it would be a 75% accuracy. The following is the result experiment accuracy data is as follows: 88% effective during day time and 75% effective during night time. Here is a picture of the experiment set up:



Figure 7. Experiment set up

4.1. Experiment on Daytime Facing Sun Without Glasses Vs With Glasses

For this experiment, the glass used in the experiment is an ordinary correctional lens. We are facing the sun during this experiment and sitting in the car. We decide the accuracy of the algorithm through a simple test: if I close my eyes and no alarms, then the accuracy is 0%. But if I close my eyes and the alarm goes off, then the accuracy level is 100%. The following is the result experiment accuracy data is as follow:

Experiment 1(day time face sun)		Trial 1	Trial 2 (with glass)
Run1		100%	100%
Run2		100%	100%
Run3		100%	0%
Run4		100%	100%
Average of Trial		100%	75%
Total Average:		88%	

Figure 8. Data table for Experiment 1

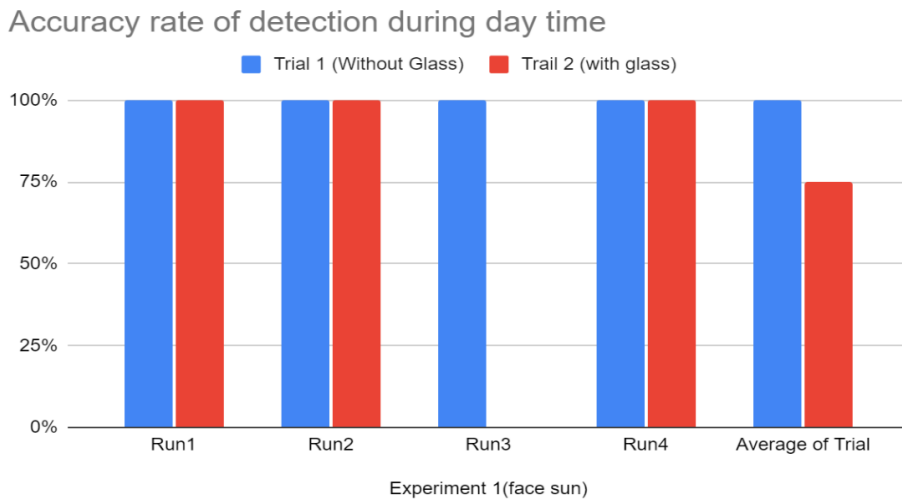


Figure 9. Accuracy rate of detection: Daytime

Based on the results, we can see this algorithm has a 100% accuracy when detecting during daytime with no glasses and a 75% accuracy rate with glasses. The overall accuracy rate for daytime detection when facing the sun is 88%.

4.2.Experiment on night without Glasses vs Glasses

For this experiment, the glass used in the experiment is an ordinary correctional lens. We conducted this experiment during nighttime and sitting in the car. We decide the accuracy of the algorithm through a simple test: if I close my eyes and no alarms have sounded, then the accuracy is 0%. But if I close my eyes and the alarm goes off, then the accuracy level is 100%. The following is the result experiment accuracy data is as follow:

	Trial 1	Trial 2 (with glass)
Run1	100%	100%
Run2	100%	0%
Run3	100%	0%
Run4	100%	100%
Average of Trial	100%	50%
Total Average:	75%	

Figure 10. Data table for Experiment 2

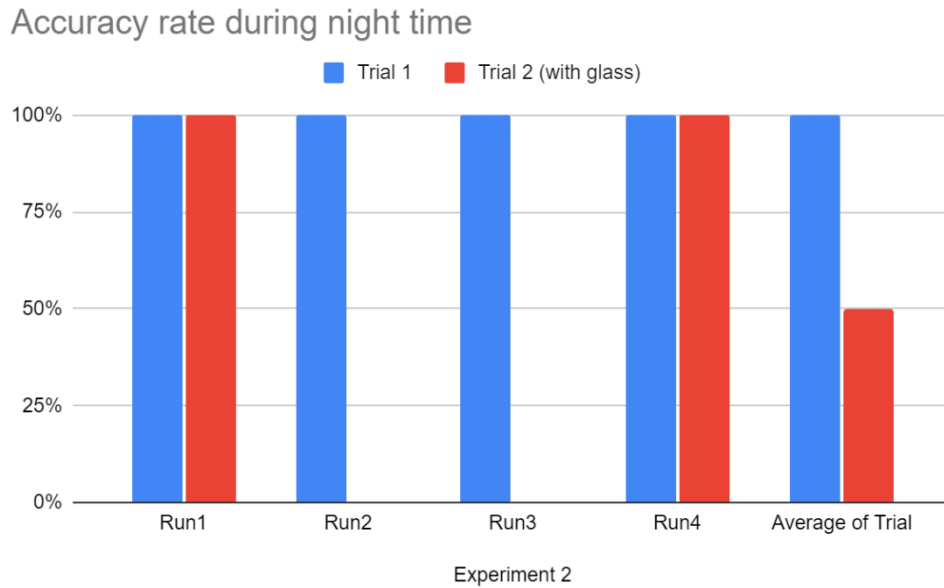


Figure 11. Bar chart of nighttime accuracy

Based on the results, we can see this algorithm at night has a 100% accuracy with no glasses and a 50% accuracy rate with glasses. The overall accuracy rate at night is 75%.

4.3. Experiment on Sunshade with and Without Glass

For this experiment, the glass used in the experiment is an ordinary correctional lens. Our backs are facing the sun during this experiment and sitting in the car. We decide the accuracy of the algorithm through a simple test: if I close my eyes and no alarms have sounded, then the accuracy is 0%. But if I close my eyes and the alarm goes off, then the accuracy level is 100%. The following is the resulting experiment accuracy data:

Experiment 3(back sun)		
	Trial 1 (w/o glass)	Trial 2(w glasses)
Run1	100%	100%
Run2	100%	100%
Run3	100%	0%
Run4	100%	0%
Average of Trial	100%	50%
Total Average	75%	

Figure 12. Data table for Experiment 3

Accuracy rate of detection during day time(not facing sun)

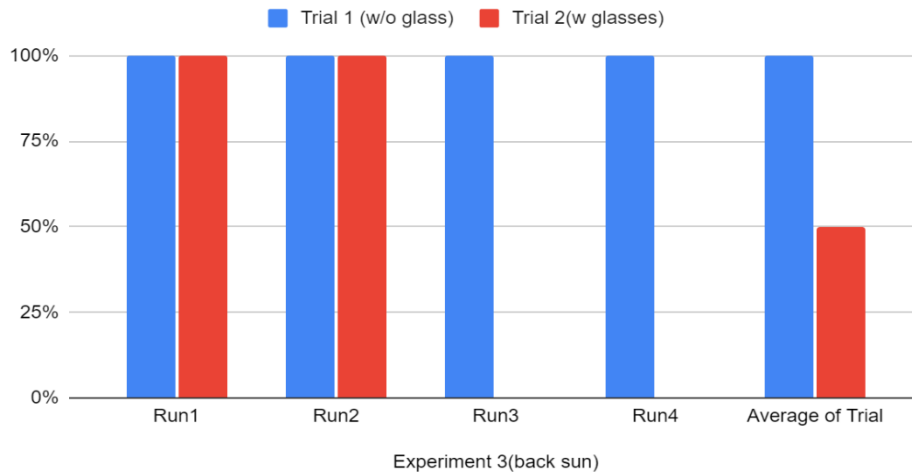


Figure 13. Bar chart of daytime accuracy (not facing sun)

Based on the results, we can see the algorithm has a 100% accuracy when detecting during the daytime with no glasses and a 50% accuracy rate with glasses. The overall accuracy rate for daytime detection when our back is facing the sun is 75%.

4.4.Experiment on Phone at a Further Distance and Glasses

For this experiment, we have decided to place the phone 87 cm from the user’s eyes. (In the previous three experiments, the phone was placed 70cm from the user.) This experiment is conducted under daytime conditions with the sun facing from the back and with variables of wearing and not wearing glasses. The experiment is conducted in a car. We decide the accuracy of the algorithm through a simple test: if I close my eyes and no alarms have sounded, then the accuracy is 0%. But if I close my eyes and the alarm goes off, then the accuracy level is 100%. The following is the resulting experiment accuracy data:

Experiment 4	w/o glasses	with glasses
	Trial 1 (long distance)	Trial 2(long distance)
Run1	0%	100%
Run2	100%	0
Run3	100%	0
Run4	100%	0
Average of Trial	0.75	0.25
Total Average	0.5	

Figure14. Data table for Experiment 4

Accuracy rate of detection when phone is placed far from the user

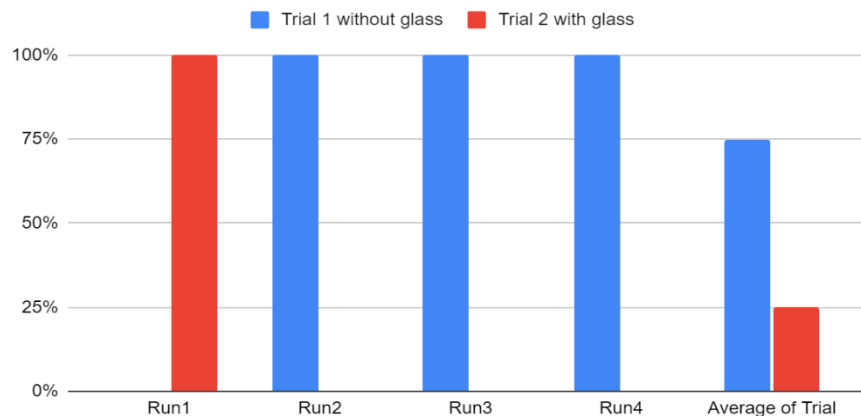


Figure 15. Bar chart of daytime accuracy (phone 87cm from user)

Based on the results, we can see the algorithm has a 75% accuracy when placed at 87 cm from the eyes with no glasses and a 25% accuracy rate with glasses under the same conditions. The overall accuracy rate for daytime detection when the phone is placed at 87cm is 50%.

In conclusion, our method proves to be effective overall. But under certain conditions, the algorithm is not effective. We can improve the algorithm's accuracy under certain conditions, especially nighttime detection. One of the uncontrollable factors is light reflection on the user's glasses. Another uncontrollable variable that can improve the efficiency is setting the phone closer to the user; this is up to the user. In general, our approach to solving this problem is proven to be effective within different experimental conditions.

5. RELATED WORK

Dijkers, et al. present using Face Recognition system for Driver Vigilance Monitoring. Dijkers, et al. use facial expression to detect drowsy driving. They use many similar methods as ours, including data extraction. One aspect that was different was that they tested their algorithm based on facial expressions. Their methods have not been tested in real-life situations and haven't proved to be runnable on mobile applications [6].

Assari, M.A. and M. Rahmati present using infrared lights to first clear out the visual interference such as darkness or light reflections and use facial expression detection algorithms on the face in the video frame. This method of using infrared light is smartly done. They have also tested their model in real-life situations [7].

Xu, L. et al. presented a solution using the percentage of eyelid closure to detect drowsy driving. They relied on factors including blink time and blink rate. The methodology contains many other factors compared to ours and employs different calculations to determine what is considered sleepy. Their methodologies have been tested in real-life situations and on a mobile application [8].

6. CONCLUSION AND FUTURE WORK

In conclusion, we were looking for a solution for drowsy driving through an app that uses an algorithm that can accurately and effectively detect drowsy driving or sleepiness based on the user's eyes that is optimizable on a modern-day smartphone. We decided to use Google firebase's face recognition algorithm and tweak it to extract the eye data from the user. We would then use the eye detector algorithm we developed to extract data from the user's eye. When the value is returned, it is passed down into a chain of calculations to determine if the alarm should go off. We applied this method to four different experiments that are closely related to four scenarios drivers would most likely encounter in real-life driving situations. The experiment results show promising accuracy for most conditions but there are also places for improvement. Therefore, our model can apply to everyday life and benefit a lot of people. Anyone from any age group can use our solution to prevent drowsy driving. In the future, we would improve the accuracy for detection when wearing glasses. The method stated in the paper can also be applied in interdisciplinary studies relating to drowsy driving, including behavioral science. Other scientists can utilize this model to further improve the current method of drowsy driving detection.

REFERENCES

- [1] <https://www.nhtsa.gov/risky-driving/drowsy-driving>
- [2] <https://www.cdc.gov/sleep/features/drowsy-driving.html>
- [3] <https://sci-hub.se/https://ieeexplore.ieee.org/abstract/document/6785367>
- [4] <https://ieeexplore.ieee.org/abstract/document/1400934>
- [5] https://pub.dev/packages/path_provider
- [6] Dijkers, H. J., et al. "Facial recognition system for driver vigilance monitoring." *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*. Vol. 4. IEEE, 2004.
- [7] Assari, Mohammad Amin, and Mohammad Rahmati. "Driver drowsiness detection using face expression recognition." *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2011.
- [8] Xu, Lunbo, et al. "Sober-Drive: A smartphone-assisted drowsy driving detection system." *2014 International conference on computing, networking and communications (ICNC)*. IEEE, 2014.
- [9] Pandit, Prajwal, et al. "VIVIFY: DRIVER'S DROWSINESS DETECTION AND ALARMING SYSTEM." *International Journal of Advanced Research in Computer Science* 11 (2020).
- [10] Vanlaar, Ward, et al. "Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors." *Journal of safety research* 39.3 (2008): 303-309.
- [11] Stutts, Jane C., Jean W. Wilkins, and Bradley V. Vaughn. "Why do people have drowsy driving crashes." *Input from drivers who just did* 202.638 (1999): 5944.
- [12] Tefft, Brian C. "Asleep at the wheel: The prevalence and impact of drowsy driving." (2010).
- [13] Spaans, M. A., and H. J. Dijkers. "Facial recognition system for driver vigilance monitoring." *Res. Rep. No. LSS 169.03* (2003).
- [14] Bracha, Gilad. *The Dart programming language*. Addison-Wesley Professional, 2015.
- [15] Mikolaj, Miroslav. "Using Flutter framework in multi-platform application implementation."

Applications of SKREM-like symmetric key ciphers

Mircea-Adrian Digulescu^{1,2}

¹Individual Researcher, Worldwide

²Formerly: Department of Computer Science, Faculty of Mathematics and Computer Science, University of Bucharest, Romania

Abstract

In a prior paper we introduced a new symmetric key encryption scheme called Short Key Random Encryption Machine (SKREM), for which we claimed excellent security guarantees. In this paper we present and briefly discuss how some other cryptographic applications besides plain text encryption can benefit from the same security guarantees. We task ourselves with and succeed in showing how Secure Coin Flipping, Cryptographic Hashing, Zero-Leaked-Knowledge Authentication and Authorization and a Digital Signature scheme which can be employed on a block-chain, can all be achieved using SKREM-like ciphers, benefiting from their security guarantees. We also briefly recap SKREM-like ciphers and the core traits which make them so secure. The realizations of the above applications are novel because they do not involve public key cryptography. Furthermore, the security of SKREM-like ciphers is not based on hardness of some algebraic operations, thus not opening them up to specific quantum computing attacks.

Keywords: Symmetric Key Encryption, Provable Security, One Time Pad, Zero Knowledge, Cryptographic Commit Protocol, Secure Coin Flipping, Authentication, Authorization, Cryptographic Hash, Digital Signature, Chaos Machine

1 Introduction

So far, most encryption schemes able to serve Secure Coin Flipping, Zero-Knowledge Authentication and Digital Signatures, have relied on public key cryptography, which in turn relies on the hardness of prime factorization or some algebraic operation in general. Prime Factorization, in turn, has been shown to be vulnerable to attacks by a quantum computer (see [1]). In [2] we introduced a novel symmetric key encryption scheme, which does not rely on hardness of algebraic operations for its security guarantees.

1.1 Prior work

The SKREM cipher was introduced in [2]. To the best of our knowledge no such system has been proposed before or since. Nevertheless, we, the author, strongly suspect that non-public research, by researchers such as Marius Zimand (see [3] and [4]) and Leonid Levin (see [5]) might include something similar to SKREM. Nevertheless, to the best of our knowledge, such research, if it exists, is still not public. Chaos Theory and namely Chaos Machines, best described by Czyzewski in [6], can be employed as a black-box subroutine in SKREM. The symmetric key cipher introduced in [2] is conjectured to be *provably* unbreakable. This conjecture is intended to be an educated statement, not a mere shot in the dark and we strongly believe a formal proof exists.

When benchmarking our current approach, we considered popular, well established schemes and methods, such as RSA [7] for public key cryptography, AES (Rijndael) [8]

for symmetric key encryption, Tang et. al [9] for cryptographic commit protocol, Diffie-Hellman key exchange for secure coin flipping [10], El Gamal [11] and Lamport [12] for digital signatures. Additionally, we considered Elliptic-curve cryptography [13] and the Kerberos authentication and authorization scheme [14].

All of the above, except Lamport and Kerberos which are general purpose methods which can work over symmetric key ciphers, rely on the hardness of algebraic operations such as prime factoring or discreet logarithm which are known [1] and respectively suspected by state actors like the NSA to be vulnerable to quantum computing attacks [15].

The AES cipher itself relies on the hardness of a different algebraic operation, namely the field inverse. While it is not presently known to be vulnerable to quantum attacks, its security claims remain unproven even with regards to a classical computer. Successful practical attacks on reduced versions of AES have been developed [16]. Furthermore, revelations connected with Edward Snowden suggest state actors like the NSA explore using tau statistic to break the full AES itself [17] - and these might be successful.

The security of the methods proposed in this paper relies on the security of SKREM. This in turn does not rely on hardness of any algebraic operation. Instead it relies on the properties of sequences sampled in a truly random fashion to be algorithmically random with high probability under a range of transformations. Concretely, it employs a technique called entropy enhancement to increase the length of the secret key by absorbing bits from a random sequence. It does so employing heavily the operation of indirection (memory dereferencing) which is not algebraic: the implied transformation functions involved in SKREM are thus different for each source sequence and expected to be incompressible. Furthermore, it is conjectured that none of them admits any regular structure, except with negligible probability - being sampled essentially at random.

1.2 Overview of this paper

The rest of the paper is organized as follows. In Section 2 we recap the SKREM cipher, its security claims and assumptions thereof, as well as formalize what is meant by SKREM-like ciphers. In Section 3 we discuss how SKREM-like ciphers can be applied to Secure Coin Flipping and Cryptographic Commit Protocol. In Section 4 we describe how such ciphers can be used to compute secure cryptographic hashes. In Section 5 we present Zero-Leaked-Knowledge Authentication and Authorization protocols over public channels, based on SKREM-like ciphers. In Section 6 we discuss how such can be used to generate digital signatures. In Section 7 we draw the conclusion and present avenues for further research. Section 8 offers the brief Vitae of the author. We conclude the paper with Acknowledgments in Section 9.

2 Recap of SKREM-like symmetric key encryption ciphers

Virtually all present day ciphers proceed from the premise that Encryption / Decryption are two functions $ENCR : P \times K \rightarrow C$ and $DECR : C \times K \rightarrow P$, where P is the universe of plain texts, K is the universe of secret keys and C is the universe of cipher texts. Usually $P = \{0, 1\}^n$ and $C = \{0, 1\}^n$, while $K = \{0, 1\}^k$, for some fixed constant key size k and plain text length n .

SKREM like ciphers [2] introduce the novelty of taking an auxiliary input, which is neither plain text, nor key - it is a large master table of truly random bits. By truly random it is understood that they are to be harnessed from nature (eg. from radio/solar noise, mouse movements or quantum computers) from a truly random distribution, are

fully independent and uniformly distributed. This is opposed to them being generated by a classical computer from a short seed. The encryption scheme is thus $ENCR : P \times K \times M \rightarrow C$ and $DECR : C \times K \rightarrow P$. P can be arbitrary $P = \{0, 1\}^n$. $K = \{0, 1\}^k$ is a short key, which has a minimal length dependent on the processing power of the adversary. Namely, a key of length k is conjectured to offer provably unbreakable symmetric key encryption against an adversary with a compute power of $O(2^k)$ - namely one which has enough resources to run this many operations on a classical computer within the relevant attack timeframe. The master table $M = \{0, 1\}^m$ is a special input which is required to consist of a truly random bit sequence. Its length $m = O(n)$ is linear in the size of the plain text; however the constant for the original SKREM cipher is around 100,000. Further refinements and simplifications can be made to better practicalize the cipher, with their security being the topic of active research.

The idea behind SKREM is to obtain a random permutation of m elements, based on the secret key k and the master table M . Since the key is short, it does not have enough entropy to generate a permutation which can be argued to be secure. Instead, the very contents of the grand master table M is used to gradually enhance the length of the secret key, in a manner that is uniform and fully unpredictable (and conjectured to be provably so) to any adversary with computing power less than $O(2^k)$. We call this technique entropy enhancement. A subsequence of this permutation is then used to alter the grand master table M , in order to encode the plain text.

In [2] we presented a full SKREM-like cipher and a simplified version - SKREMS. The method is however general and many variations and adaptations can be developed. The following describes the pseudo code of SKREM-like ciphers in general.

Algorithm 1. SKREM-like encryption. *Short key random encryption machine. Input: $k \in K$, $m \in M$ and $p \in P$. Output: $c \in C$.*

- 1: Split the bits of k into u groups and seed $u \geq 2$ CSPRNGs, $S_1 \dots S_u$, with a seed of length $z > \log(M)$. It is recommended that $u \cdot z = k$.
- 2: **while** u is still too small or not enough rounds have completed **do**
- 3: Double u , creating u new, inactive - yet unseeded CSPRNGs. Intertwine them alternately with the original ones in the sequence S : $s_1, s_{u+1}, s_2, s_{u+2}, \dots$
- 4: **while** there exists a CSPRNG which does not have all the seed-bits filled out **do**
- 5: Sample values v_1 and v_2 from two successive, active CSPRNGs in S , starting at index i . Then increment i by 2.
- 6: Use values v_1 and v_2 to determine two uniformly distributed, random locations, l_1 and l_2 in M , from those which have not yet been visited.
- 7: Use the values at $M[l_1]$ and $M[l_2]$, if they are different, to generate a new uniformly distributed random bit b .
- 8: Distribute the random bit b to some index of a new CSPRNGs which does not yet have its seed completed. Distribute it such that each old CSPRNGs contributes an equal number of bits to each new CSPRNG's seed.
- 9: If $M[l_1] = M[l_2]$ then use then use v_1 and v_2 to permute some $O(1)$ elements/chunks of M , as well as of S .
- 10: When a priory inactive new CSPRNG has gathered enough bits for a full seed, mark it as active and start using it.
- 11: Mark the positions l_1 and l_2 of M as visited.
- 12: **end while**
- 13: Consider the concatenated seeds of all CSPRNGs in the sequence $S_1 \dots S_u$ as a single $u \cdot z$ bit number x . Use a bijective (preferably one way) function $f : \{0, 1\}^{u \cdot z} \rightarrow \{0, 1\}^{u \cdot z}$, and set $x \leftarrow f(x)$. Then consider back x as a sequence of u CSPRNGs

with seeds z -bits long.

14: **end while**

15: When u is large enough (eg. $u \geq 2 \cdot P$) and enough iterations have passed, repeat steps 5-11 for the n bits of the input, with the difference that the bit produced at step 8 is not used for new CSPRNGs, but instead is a bit of the plain text. When encrypting, $M[l_1]$ and $M[l_2]$ can be exchanged conveniently, if needed, to represent the desired bit.

Complexity. The original SKREM and SKREMS presented in [2] had a space and time complexity linear in the size of the plain text - $O(n)$ for an n -bit plaintext. However, the constant factors were rather large - on the order of 100,000. This can cause incurring a significant running time penalty when the grand master table M is too big to fit in RAM memory. Further sophistication in the implementation of the CSPRNGs (eg. using a Chaos Machine with many iterations per PULL operation), in that of shuffling the locations in M and S , and in the execution of the function f in step 13, could add to the running time of some SKREM-like ciphers - however most often just in terms of constant factors, not of asymptotic complexity. SKREM-like ciphers also require as input a series of truly randomly generated bits, on the order of $O(n)$ - again, with a significant constant factor, on the order of 2-300,000.

The method proceeds to incrementally add new CSPRNGs, by seeding them with the added entropy it gets from the master table M . The order in which the locations of this master table are visited is then conjectured to be a cryptographically secure, random permutation of length m . This conjecture relies on the following assumptions.

Assumption 1. *An arbitrary subset of sufficient length of a truly random sequence m , is itself truly random, except with negligible probability.*

Discussion. This says that if you remove some bits from a truly random sequence m , what remains is still truly random, except in astronomically improbable cases (eg. you chose a subset which consists precisely of the true bits).

Assumption 2. *An arbitrary permutation of truly random sequence m , is itself truly random, except with negligible probability.*

Discussion. This says that if you choose to shuffle the elements of a truly random sequence, the result is still truly random - except in astronomically improbable cases (eg. you chose a permutation which sorts m).

The above assumptions are quite natural and probably admit proofs, given the fact that truly random sequences are expected to be incompressible and thus algorithmically random, except with negligible probability: if the above were not true, constructive martingales could potentially be found which succeed on such sequences.

From the above assumptions, the following conjecture is derived.

Conjecture 1. *A member of the set of permutations producible for a given key k by the Algorithm 1 is a random and unpredictable permutation, except with negligible probability.*

Discussion. The conjecture is quite natural and is suspected to admit a formal proof. This stems from the two facts. Firstly, any suffix of the permutation is fully independent from its prefix - the prefix is derived from the secret key k , and a sampling of the positions of M which are then never reused in the suffix. While any such prefix is uniquely determined by the secret key k (for a fixed M), by induction such a prefix is random and unpredictable. Secondly, the bits which are used to extend the prefix are uniformly distributed by construction, and based on the sequence determined by some sampling from the outstanding portion of M . By Assumptions 1 and 2 these are themselves truly

random except with negligible probability - this means the constructed sequence itself is random and unpredictable, except with negligible probability: otherwise a successful martingale could be constructed.

Finally, based on Conjecture 1, the security of SKREM-like ciphers can be argued to be at least 2^{2z-1} : an attacker will not be able to gain any meaningful insight from deducing (or guessing) some locations which together encode some bit of plain text - except so as in to guess the bits which were used to generate the seed for the CSPRNGs which generated them. Since only a small number of bits are used from each prior-round CSPRNG for each new CSPRNG (preferably only 1 - which is always possible in later rounds), in order to speculate any known plain-text advantage, the attacker will need to guess the full seed of some CSPRNG - z bits. Then he will need to validate if it is plausible, by guessing at least another $z - 1$ bits from the resulting new CSPRNG in order to determine if 1 bit of the plain-text is producible by such a priory guessed seed. In practice he will probably need another $\approx z$ bits to gain any meaningful level of confidence for successful guesses.

Attacks are further complicated by the unpredictable mixing of bits in steps 1, 9 and 13 which are intended to mask the correlation between prior-round bits and current round ones, while maintaining statistical properties.

The choice of Cryptographically Secure Pseudo-Random Number Generators (CSPRNGs) can be Chaos Machines [6] which present the advantage that a small variation in source input results in a large variation of output, which is claimed unpredictable except by knowing the seed. Alternatively they can be any CSPRNG such as those based on PRNGs with proven statistical properties [18]. The outputs of CSPRNGs need to be normalized in steps 6 and 9 before they can be used. The original paper on SKREM [2] included a method for converting any distribution to uniform binary and also one for sampling a number in an arbitrary interval $0..p - 1$ based on some uniformly distributed bits. It entails using a larger number of bits than strictly required by the desired output.

For the transformation in step 15, alternatively to a large CSPRNG, an algebraic transformation such as the modular or field inverse, followed by appropriate distribution conversion can be used instead. Other various can be sought up, for example using a time-expensive classical one-way function - such as the one described in [5] (which is one-way contingent on the existence of such functions).

Ultimately, the security of SKREM-like ciphers rests in the fact they rely essentially on indirection operation over some large chunk of randomly initialized memory. So long as the chunk is algorithmically random (which being truly random entails except with negligible probability) and the algorithm does not introduce patterns, the result of these operations cannot be predicted at all. The hardness of reversing indirection relays not on difficulty of solving some mathematical problem such as those of modular algebra, but instead on the unpredictability of truly random sequences - their property that no constructive martingale can succeed over them. While for a regular cipher, the transformation function is fixed, for a SKREM-like cipher there are 2^M such functions, one for each original master table, the vast majority of them being secure - not even theoretically breakable, as conjectured.

3 Cryptographic Commit Protocol and Secure Coin Flipping

Consider that Alice and Bob have access to a public master table M , known to be the result of a truly random sampling. While this table is assumed to be public and known to everyone (not only Alice and Bob) without compromising the scheme, it needs to be truly random - it cannot be arbitrary. Now suppose Alice wants to commit to some k -bit value v . Consider the following protocol.

Protocol 1. Cryptographic Commit Protocol over SKREM-like ciphers. *Input: Both Alice and Bob have access to a common, immutable, public, truly random master table M .*

- 1: Alice chooses the k -bit value v to which she wants to commit. She also chooses a k -bit truly random secret key s .
- 2: Alice computes $c \leftarrow \text{DECR}_n(M, v \oplus s)$, for a maximal $n < m$, such that M cannot encrypt more than n bits. Then she takes $a \leftarrow c[n/2 - k \dots n/2 + k]$ to represent the $2k + 1$ bit number comprised of the middle bits of c .
- 3: Alice sends Bob a .
- 4: Bob acknowledges receipt of a .
- 5: Alice sends Bob s .
- 6: Bob acknowledges receipt of s .
- 7: Alice and Bob do whatever other interaction, given that Alice has committed to some value v .
- 8: Alice sends Bob $w = v$.
- 9: Bob repeats the computations Alice did in step 2, with $v = w$ to get his version of c and a . If Bob's a computed in this round is equal to what he received in step 3 from Alice, he confirms the committal as truthful, otherwise he rejects.

Correctness. Starting with a random master table M , for each key k there is an equal probability of $\frac{1}{2}$ that some arbitrary bit resulting from decryption is set to a particular value. As such, for any t -bit sequence of decrypted output (chosen arbitrarily), there are expected to be $\frac{2^k}{2^t} = 2^{k-t}$ keys which produce it. By sampling $2k + 1$ bits, the probability that there does not exist another value except v which can serve as a secret key to decrypt the same sequence is $(1 - \frac{1}{2^{2k+1}})^{2^{k-1}}$. This tends very quickly to 1 as k tends to infinity. Bob has confidence of at least k -bits security that Alice cannot cheat and that indeed she committed to the value v before step 3 of the algorithm. Furthermore, contingent on the security of SKREM, Bob cannot deduce the secret key v for the “known plain text” a , before Alice reveals it in Step 8.

Complexity. The space/time complexity of the Protocol is dominated by the application of two SKREM decrypt operations, making it linear in the size of the master table M used. In terms of network complexity, only $O(k)$ bits of information are exchanged, making the method robust. The method also requires $O(k)$ truly random bits, on top of the $O(M)$ required by the SKREM-like cipher.

Discussion. Note that if Alice can secretly manipulate the random grand master table M , she could deliberately encrypt a with some key $w \neq v$ chosen by her, and then, in step 8 she would have the choice to send either w or v as her committal value. This is why it is important that the master table M is in fact random, and not chosen or altered by Alice. Conversely, if Bob is allowed to choose M instead, he could theoretically make it that several values decrypt the same sequence (eg. by choosing M to be all zeros). And, as such, he can contest Alice's committal at step 8, by showing a counter example. This is why it is important that the master table is in fact actually random and not chosen by either of the players independently.

The usage of an auxiliary secret key s is because in order for the security claims of SKREM to hold, encryption needs to happen based on a secret key chosen in a truly random fashion, which Alice's committal value v may not be.

Adjusting the above Protocol 1 to support random coin flipping is straightforward.

Protocol 2. Secure Coin Flipping over SKREM-like ciphers. *Input: Both Alice and Bob have access to a common, immutable, public, truly random master table M .*

- 1: Alice and Bob perform Protocol 1 above up to step 7, for some arbitrary random value v_A chosen by Alice.
- 2: Then they again perform the Protocol 1 up to step 7, with Bob committing himself this time to some arbitrary random value v_B chosen by him.
- 3: They both complete the protocols 1 (in arbitrary order) with the respective confirmations of the truthfulness of the committal. They now both possess both v_A and v_B .
- 4: They take $v_A \oplus v_B$ to represent their common random value. If this value is over too many bits, they can just ignore a suffix of them.

Correctness. If at least one of them chose randomly, the common random value will be random, since XORing a random number with an arbitrary number maintains the randomness.

Complexity. The protocol essentially consists of two applications Protocol 1 above, thus having the same asymptotic performance characteristics.

Discussion. Note that Alice and Bob could efficiently commit to a large random sequence by using the protocol repeatedly. Note that the values to which either of them commits cannot be over fewer bits than the minimal key size k , required for the security of the SKREM-like cipher used.

4 Cryptographic Hashing

Suppose Alice has some, potentially long, message r , to which she wants to compute a potentially short cryptographic hash. As before, consider a public master table M exists, known to be the result of a truly random sampling, to which Alice has access.

If $|r| = k$, then r could be used directly as a secret key, to obtain a digest by the same method used in Protocol 1 above by Alice to commit to some value, with the secret key s appended to the digest.

If r is longer than k bits it could be used directly in Step 1 of Algorithm 1 to, while keeping z fixed (chosen beforehand), simply start with more CSPRNGs upfront. The stopping condition in Step 2 needs however ensure a reasonable number of rounds happen (at least 3 is recommended) and that, additionally, all original CSPRNGs, at the very least, are used to determine bits at least as many bits so as to exhaust their entropy (namely r in total) - all this before the final round occurs in Step 15. Alternatively, r could be divided into k -bit sequences and the process repeated, as in the case of block ciphers, with the resulting digests appended or combined by some other method. This approach requires that r be truly random however, since SKREM requires truly random input keys.

There is however a more appealing alternative.

Algorithm 2. Cryptographically Secure Hashing over SKREM-like ciphers. *Input:* There exists a public, well known truly random master table M . Alice wants a secure cryptographic hash of message r .

- 1: Alice chooses truly randomly a k -bit value s .
- 2: Alice computes $c \leftarrow DECR_n^r(M, v \oplus s)$, for a maximal $n < m$, such that M cannot encrypt more than n bits. Then she takes $a \leftarrow c[n/2 - r/2 - k \dots n/2 + r/2 + k]$ to represent the $2k + r + 1$ bit number comprised of the middle bits of c . The function $DECR_n^r$ is the same as that of Algorithm 1, except that, in Step 8, before an obtained bit b is used, it is XORed with the next unused bit from r , until all such are exhausted.
- 3: The secure hash is $\langle h, s \rangle$.

Correctness. There are $2^r \times 2^k$ combinations of potential messages times secret keys. The probability there does not exist another pair to generate the same fixed $r + 2k + 1$ bits is $(1 - \frac{1}{2^{r+2k+1}})^{2^{r+k}-1}$. This again ensures security of at least k bits, for sufficiently large r and k . The underlying implicit assumption is that the probability for a <message, key> pair to produce a certain sequence of bits over M is uniform. This is natural given Conjecture 1.

Complexity. The space/time complexity of the method is given by the application of a single SKREM decrypt operation, making it linear in the size of the master table M used. The method also requires $O(k)$ truly random bits, on top of the $O(M)$ required for the master table M .

Discussion. Note that the size of the message r cannot be arbitrary large, but needs to be small enough for all of its bits to be used by the altered $DECR_n^r$ function. Choosing M so that $|r| \approx n$, where n is defined as in step 2 of Algorithm 2, should prove adequate for the SKREM-like ciphers proposed in [2].

Note that by using Algorithm 2, the obtained bits of the hash are essentially fully independent from those of r . In fact, such a hash would be different for different original master tables M .

In case the hash digest length of $r + 2k + 1$ is too great, it could be reduced (preferably to k) by using a classical cryptographic hash function, such as Whirlpool [19], SHA-2 [20] or SHA-3 [21] as the underlying hash of a Merkel Tree [22], whose root, together with the secret key s give the digest. The added advantage in this situation is that the cryptographic hash function is computed over an $r + 2k + 1$ bit sequence which is chosen truly randomly, and whose bits are not related to that of the original message r . The original bits of r are only used, together with those of s and with a discarded portion of M , to select which random sequence of $r + 2k + 1$ bits from the outstanding portion of M is chosen. By Assumptions 1 and 2 and Conjecture 1, this will be a truly random sequence, except with negligible probability - even in cases when r is very regular, such as all zeros.

Alternatively, the hash length could be reduced to some arbitrary length x by taking fewer bits around the middle of c in step 2 of Algorithm 2, with the drawback of increased risk of collisions.

Yet a better option is to use a Merkel Tree [22] but with this very Algorithm 2, with a reduced digest of size k taken with the idea above, instead of a classical cryptographic hash function such as SHA or Whirlpool. We recommend this third option in practice.

Note that the master table M is required to verify a hash. It is assumed to be immutable, public and available over the lifecycle of the generated hashes.

5 Zero Knowledge Authentication and Authorization

Suppose Alice and Bob are secret agents of the same agency, who have never met before and don't know each other, but were given a shared secret key k by their common HQ beforehand. Consider that Alice and Bob have access to a public master table M , known to be the result of a truly random sampling.

Now suppose Alice and Bob want to establish to one-another that they are part of the same organization. However, they want to achieve this without giving any knowledge about the shared secret key k , neither to their counterparty, nor to any member of the public who didn't already know it in advance (such as Mallory). They also want to defend against a public, non-insider actor, such as Mallory using the conversation she witnessed to potentially deceive Alice or Bob or someone else in the future into believing they also knew the secret.

Furthermore, suppose Alice and Bob may want to be able to reveal their membership of the secret agency to the other only if the other also does the same - namely Alice wants to reveal to Bob she is a secret agent *iff* Bob reveals the same to her during the execution of the protocol.

Protocol 3. Zero Knowledge Secret Agent Authentication over SKREM-like ciphers. *Input: Both Alice and Bob have access to a common, immutable, public, truly random master table M . Alice and Bob belong to the same secret organization, Alpha, and like any such member they were given the same private secret key k . They want to mutually authenticate each other over a public channel, leaking zero knowledge about k .*

- 1: Alice and Bob both start with an $x = 0$ and a security confidence parameter $p = 1$.
- 2: During rest of the Protocol, Alice and Bob take turns. Alice starts first.
- 3: $\{no_rounds$ is a sufficiently large, even integer}
- 4: **for** $i = 0 \dots no_rounds$ **do**
- 5: Alice and Bob agree and commit to number of z public, shared, random k -bit values $v_1 \dots v_z$, using the Coin Flipping Protocol 2 of Section 3.
- 6: For each v_i , in order, Both Alice and Bob compute $c = DECR_n(M, v_i \oplus k)$, for a sufficiently large $n < m$, such that M cannot encrypt more than n bits. They then take $a \leftarrow c[n/2 - 8 \dots n/2 - 1]$ and $b \leftarrow c[n/2 \dots n/2 + 7]$ to represent two bytes around the middle bits of c . Then they set their respective $x \leftarrow x \oplus b$ and append $a \oplus x$ to a fresh internal sequence s they maintain (sequences s are discarded from round to round).
- 7: Whose ever turn it is (say Alice) computes a new sequence s_A by taking the internal sequence s and replacing a number of $z - p$ randomly chosen elements in it with random values and then sends it to the counterparty (say Bob).
- 8: Both Alice and Bob count the number of correct answers in the transmitted sequence, by comparing each to the members of their internal sequence s . Say this number is y .
- 9: Both Alice and Bob validate that $y \geq p$. Otherwise authentication is rejected, and protocol continues with $p = 0$ immutable.
- 10: If authentication did not fail above, both Alice and Bob set $p \leftarrow 1 + (y \bmod z)$. Also, if $p > tr$ for some specific threshold $tr < z$, authentication is considered successful and the protocol continues with $p = 0$ immutable.
- 11: The roles of Alice and Bob are switched for the next round.
- 12: **end for**
- 13: If the authentication was never declared successful in line 10, then it is declared failed now.

Correctness. If both parties know the secret key k , the value of p will monotonically increase by +1 each round, up until $\min\{z, no_rounds\}$. If at least one of them doesn't, by the security of SKREM the best they can do is guess a member of the target sequence. They guess correctly with probability $1/256$, meaning, by linearity of expectation, that the expected number of correct values in this case is $\frac{1}{256} \cdot z$. The exact probability that, by chance, the actual number of correct guesses strays too far from this expectation decreases very fast. For illustration, if we chose $z = 10 \cdot 256$ and threshold $tr = z$, the probability of false positives is $\frac{1}{256}^{10 \cdot 256} = 2^{-20480}$ which is negligible. We recommend choosing $z \geq 256$, for there to be positive expected value for the number of correct hits by chance. The exact value of tr depends on the desired maximal probability of deceit, resulting in a false positive authentication. Finding a good value of tr is left as an exercise for further research. It can be computed algorithmically for a given false positive probability. We expect it be less than 62 for $z = 10 \cdot 256$. This value should also be used as no_rounds .

Complexity. The running time complexity of the method is given by z applications of a SKREM decrypt operation for each round - of which there are z in total maximum, making it $O(z^2 * M)$. Since the master table M is not changed, the space complexity is just $O(M)$. The method also requires $O(z^2)$ truly random bits, on top of the $O(M)$ required for the master table M . Since z is expected to be a small constant, these are asymptotically excellent characteristics.

Discussion. Note that Alice and Bob reveal each others' knowledge of the shared secret with increasing probability, in tandem. If Bob wants to withhold his knowledge of the shared secret, the authentication will fail for both sides. He will at most learn that Alice had the ability to guess +1 more than he revealed about his own ability to guess. If the expected number of correct guesses by chance is large enough (say more than 10 above expected value), this added information should be insufficient for a confident appraisal. Therefore, in order to learn about Alice's membership to the organization he must reveal his own as well.

An outside observer will not be able to as much as decide which from the values he sees were not chosen at random in step 7, much less be able to determine a full sequence of such. Even if they were to learn the exact sequence, by security of SKREM they would be unable to determine the encryption key $k \oplus v_i$ for such a "known plain text". Also, since the values v_i are chosen at random via a cryptographically secure coin flipping protocol, they will not occur again in future instances of the protocol's execution - not even individually, much less so as the entire sequence for all the rounds -, except with negligible probability.

Protocol 3 offers a way for Alice and Bob to (almost) simultaneously prove to each other, over a public channel, that they have knowledge of a shared secret, without leaking any information about the secret itself, not even theoretically, to the rest of the channel participants, who did not already know it. Do note however that any other participant to the channel who did know the secret can determine if authentication of Alice and Bob was successful by listening in. In order to prevent this, they should communicate over a secured, bilateral channel, at least when committing to the v_i sequences.

Note that a knowledgeable third party listening in on Alice's and Bob's chatter cannot determine for sure that they are actually secret agents: they could both NOT be and simply replay a conversation overheard priory between actual secret agents over the same master table M . In order for authentication to be genuine, it must occur over a "never before seen" truly random master table M , or occur interactively.

Protocol 3 opens the door way to a host of interesting related applications. Suppose now that Alice and Bob have determined they are both secret agents, they want to find out who has the highest rank, so they know who gives and who takes the orders.

Suppose the ranking hierarchy of the secret organization Alpha is a linear chain, with the members of each rank r being given, upon promotion, not only the secret key k_r corresponding to their own rank, but also another one, k_{r+1} corresponding to the immediately superior rank, whom they must obey. Also, the secret agency does not want to reveal to all agents how many ranks there are exactly in the organization. Also, when authorizing themselves, agents do not want to reveal their exact rank, but only that they are of superior rank to the counterparty, if that is so.

This can be achieved by following the following protocol.

Protocol 4. Zero Knowledge Secret Agent Authorization over SKREM-like ciphers. *Input: Both Alice and Bob have access to a common, immutable, public, truly random master table M . Alice and Bob belong to the same secret organization, Alpha. Alice's rank is r and she knows keys $k_1 \dots k_{r+1}$ and Bob's rank is q and he knows secret keys $k_1 \dots k_{q+1}$. They want to mutually authenticate each other over a public channel, leaking zero knowledge and also to determine who of them is of higher rank.*

- 1: Let $last \leftarrow 0$, $left \leftarrow 0$ and $right \leftarrow MAX$, with MAX a value known to be greater than the number of ranks existing in the secret agency.
- 2: **while** $left \leq right$ **do**
- 3: Alice and Bob set $mid \leftarrow (left + right)/2$.
- 4: Alice and Bob try to authenticate each other over rank mid , with shared secret key k_{mid} . If one of them is of rank lower than k_{mid-1} , he or she answers randomly to the challenges of the rank, thus causing authentication to fail.
- 5: If authentication succeeds above, they both set $last \leftarrow mid$ and $left \leftarrow mid + 1$.
- 6: If authentication fails in step 4 above, they both set $right \leftarrow mid - 1$.
- 7: **end while**
- 8: To determine who is in charge both Alice and Bob compare $last$ to their own rank. If it is greater, than the other party is their superior, otherwise they are.
- 9: In order to obscure the actual number of rounds the protocol took, step 4 is performed an additional number of times, in order to bring the total to $\log(MAX)$ - the results of these extra rounds are ignored.

Correctness. Essentially, the parties binary search to find the highest rank about which they both know. The person who knows of a higher rank than that is clearly the superior. The party of inferior rank is only able to confirm that the counterparty is her superior, but cannot establish his exact rank.

Complexity. The method consists of a number of applications of a Protocol 3 above, which is logarithmic in the number of ranks in the Alpha secret organization. Thus, the running time is $O(\log(MAX) * z^2 * M)$, the space complexity remains unchanged at $O(M)$ since the same master table is used for all applications. There is a need for an additional $O(\log(MAX) * z^2)$ extra truly randomly generated bits, except the $O(M)$ used by the master table.

Discussion. The protocol can be performed not only bilaterally but also with regard to a third party, say Claire. Claire may be an automated weapons system - such as a Poseidon or Minuteman strategic nuclear article, and Alice and Bob two competing secret agents who want to give conflicting orders to Claire. By having both Alice and Bob perform the protocol with her (not with each other), Claire can decide to whom to listen. This also gives no indication to the other party as to the actual rank of the counterparty, or any information that could help it in the future to pass a similar authorization protocol.

There is one added bonus for using Protocols 3 and 4: **stenography**. Since only a small fraction of the actual elements of the sequences exchanged are required to have fixed values, the rest can be used to **“piggyback a secondary transmission on the same carrier wave”**. This allows for example a double agent Alice to perform an authentication protocol for secret agency Alpha with Bob, while at the same doing an authentication protocol for secret agency Omega with the same Bob. This way, any member of agency Alpha listening in on the conversation will believe that Alice and Bob simply authenticated each other for agency Alpha, when instead they might have also established that they both belong to agency Omega also. Furthermore, in case Bob does not belong to agency Omega (or does not wish to reveal his belonging), Alice’s attempt for mutual authentication for Omega will remain a secret: neither Bob nor other members of Omega on the public channel will be able to ascertain that Alice attempted such.

Clearly such a protocol would have been useful to the conspirators of the Lodge of Perfect Equality in the time of the 1789 French Revolution, to authenticate one-another while seemingly simply authenticating that they are simple members of the Freemasonry, not also part of some conspirator group within it.

When the organization Alpha is a well-known public organization, such as an Internet market place, authenticating against it may seem natural to the observers. This gives the

pretext to piggyback a secondary message (for all intents and purposes seemingly random) as part of the authentication protocol. This can be used not only to authenticate within Omega, but also to send and receive encrypted information which is indistinguishable from random (eg. encrypted with a SKREM-like cipher). Such information will be short over a single transmission, but not of trivial length. It can include specific encrypted orders, or can form portions of a longer message transmitted over several sessions.

Protocols 3 and 4 could be adapted to function with other symmetric key ciphers. However, their current formulation presents the advantages of the security guarantees of SKREM-like ciphers in general. Existing alternatives rely on hardness of some algebraic problem such as discrete logarithm (like [23]) - many of which are clearly vulnerably to quantum computing attacks -. Or they employ the trapdoor approach over NP-complete problems (like [24]) - which approach, in many cases, has been shown to be breakable in practice.

Unlike Kerberos [14], the presented protocols do not require a trusted third party (authentication server) and can be enacted bilaterally. On the other hand, Kerberos itself can be extended to use Protocols 3 and 4 to authenticate and respectively authorize a new client when issuing him a TGT key. Just as well, a SKREM-like cipher can be used as the symmetric key cipher for Kerberos.

6 Digital Signatures

Suppose Alice wants to be able to digitally sign some arbitrary message r , by producing a digital signature $sig(r)$, such that no one else is able to produce such a signature except her and that all members of public are able to verify and be confident that it is indeed her who signed it.

Alice can proceed as follows. Firstly, she generates a secret, truly random grand master table M . Then she has several options. One of them is for her to employ the Lamport digital signature scheme [12], and use the cryptographic hashing Algorithm 2 from Section 4 to compute the cryptographic hashes involved. Then she can publish M unmodified, together with her public key. This has the advantage that she can use any master table M , including some potentially naturally occurring ones. However, the public key can be used only once.

There is, however an even better possibility.

Algorithm 3. Digital Signatures over SKREM-like ciphers. *Input: Alice posses some secret, truly random master table M . She wants to be able to digitally sign messages of length r bytes.*

- 1: Alice computes some k -bit truly random values $v_{i,j,k}$ and secret keys $k_{i,j,k}$, for $i = 1..n$, $j = 1..r$ and $k = 0..255$, with n being the number of messages she plans to sign using the public key.
- 2: Alice computes the following plain texts $m_{i,j,k} = \langle v_{i,j,k}, j, k, i, n, hash(\langle v_{i,j,k}, j, k, i, n \rangle) \rangle$, for $i = 1..n$, $j = 1..r$ and $k = 0..255$. The hash function can be any cryptographically secure hash function with a short digest, such as Algorithm 2 from Section 4 over M .
- 3: Alice sets $C \leftarrow ENCR(m_{i,j,k}, k_{i,j,k}, M)$, encrypting all $256 \cdot n \cdot r$ messages in the same cipher text, using the methodology described in [2] which leverages Universal Perfect Hashing. If the function used above is indeed Algorithm 2 over M , the locations touched by any of the hashing operations must also be protected, the same way as when encrypting.
- 4: Alice reveals C as her public key.

- 5: When Alice wants to sign the i -th public message, consisting of bytes r_1, r_2, \dots, r_r , she publishes $\langle k_{i,j,r[j]} \rangle$ for all $j = 1..r$.
- 6: Someone wishing to verify her signature checks that $DECR(C, v_j) = \langle x, j, r[j], i, n, hash(\langle x, j, r[j], i, n \rangle) \rangle$ for the r values $v_1..v_r$ published by Alice as signature. This involves checking the hash of the deciphered output and checking that the second and third values in the tuple are indeed j and $r[j]$ as expected. The verification succeeds *iff* the check holds for all $j = 1..r$.

Correctness. Similar to Lamport signature, for an attacker to be able to forge Alice's signature over even a 1 byte long message b , he would need to find at least some key k such that $DECR(C, k) = m_{i,1,b}$ for some arbitrary i . But the probability of a naturally occurring such $m_{i,1,b}$ outside those purposefully minted by Alice is less than the probability of hash collisions for the hashing function. This in turn is negligible if Algorithm 2 from Section 4 is used: less than 2^{-k} , which should exceed the compute power available to an adversary.

Complexity. As it is presented, the Algorithm involves encrypting $O(n * r)$ plaintexts within the same master table M . The size of each plain-text is dominated by that of the digest produced by the hash function in step 2. Using Algorithm 2 from Section 4, that is on the order of $O(k)$. This bounds the running time complexity to $O(n * r * k)$. Computing each hash, using the raw Algorithm 2 from Section 4 involves a full $O(M)$ SKREM decrypt operation for each, bringing the total to $O(n * r * k * M)$, which is rather large - even if incurred only one time, during the generation of the public part of the signature. A practical implementation, will most likely adjust the hash algorithm to involve only up to $O(k)$ steps, thus making signature generation take overall $O(n * r * k)$ time. Signing is linear in the message size, namely $O(r)$. Verifying a signature takes one SKREM decrypt operation and one hash operation for each byte of message. The latter takes $O(M)$ using the raw Algorithm 2 and just $O(k)$ using the suggested practicalization. The overall complexity of verification can thus reach $O(r * k)$. Note that besides the master table M , Alice needs another $256 * n * r * k$ bits for the secret keys $k_{i,j,k}$ and a roughly similar amount for the SKREM encrypt operations.

Discussion. Note that the total size of all encrypted messages is $O(n * r)$. This can grow large for large n or large r . In case the message to be signed is very large, she can instead sign a cryptographic hash digest of it. Note that in fact the maximum values for j can be made to vary from one i to another. This means she could "set aside" some of the signatures from the n allotted, to sign longer messages.

In case Alice runs out of signatures, she could resort to signing a special (potentially long) message containing the digest of a fresh public key C_{new} , consisting of a suitably altered new truly random master table M . This should be reserved only for the last $i = n$ of the signatures she uses, and could be understood by the public as such. Note that while she can sign a digest for a new public key which is longer than the usual messages she signs, that digest can never be as long as her original public key. So, if the number of available signatures and security guarantees for the new public key are to be maintained, only a digest of such can be signed. Computing this digest with Algorithm 2 of Section 4 over her original public key C , with the idea to use a Merkle Tree over the same algorithm to shorten the digest should prove enough for ensuring security. No maleficent party can create a fresh public key even after seeing C_{new} , thanks to the security guarantees of the Algorithm 2.

While the signature scheme described is not based on public key cryptography, it could nevertheless be used on a block-chain [25] to sign transactions. A public digest of the private key could be published beforehand as part of the incoming transaction (similar to how hashes of public keys current work on the BitCoin block-chain) and the full public

key revealed and used to sign an outgoing transaction. Since transactions on the BitCoin block chain, do not generally reuse public keys, the number of messages that would need to be signable by a single public key is as low as $n = 1$. In order to support spending inputs of large-valued transactions we can take $n = 2$ or $n = 3$. Thus, the total size of a public key would be reasonably small (linear in the size of the digest). The security of such an approach relies of course on the security of the hashing digest function. The one described in Section 4 for example, was argued to be secure and not reliant on hardness of prime factoring or of any algebraic operation in general.

The described scheme can also be used in secure authorization: A client with a certain level of clearance (perhaps proven via Protocol 4 of Section 5) can formulate certain types of requests, including a user ID and timestamp with them. An Authorization Server could receive such request from an authenticated client, authorize them and then send back a digital signature of such, signed with its own private key (the Authorization Server's). The client could then directly send such requests to a Serving Server which could then simply check the validity of the digital signature. Note that the Serving Server could already have the Authorization Server's public key C , so the two do not need to be connected live at the moment of serving.

This allows for scenarios like the following: Say there are number of underwater mines (say Poseidon strategic articles) deep behind enemy lines, or within neutral waters. They are all programmed with the same public key C of the authorization server. Then, a dully authorized party wants to issue an order (eg. attack / self-destroy / do not attack and remain dormant for at another week / no operation) to one (including its ID in the request) or all of them. The authorization server could be located deep below ground in a bunker in the Urals and be totally disconnected from any of the mines. The authorized party can compose the order message, get it signed by the secure authorization server and then transmit it to the mine (eg. via satellite link, marine mammals, submarines or surface ships). The mine will execute the order *iff* it is authorized and without the need to communicate with the authorization server. Finally, what is interesting is that even if the enemy were to capture one mine, and then even if they were able to reverse engineer it and discover the Authorization Server's public key C - and even if they were able to intercept a message addressed to that particular mine -, they would still be unable to fool any of the other mines into executing any orders.

The ideas in this paper can also be combined to secure a top secret algorithm which is embedded in some forward-deployed hardware and which is activated only on command: It can be encrypted with SKREM, and its encryption key be sent over as part of the authorized request. Since SKREM supports multiple plain-texts over the same cipher note how this could allow export-versions of military hardware to be turned into strategic-versions, by the simple issuing of an authorized command.

7 Conclusions And Further Research

The ideas in this paper have demonstrated the usefulness of SKREM-like ciphers over a range of applications.

We also recapped SKREM-like ciphers and argued the extraordinary promise they present: provably unbreakable symmetric key encryption, with short secret keys. While the fact their security is provably unbreakable remains a conjecture presently, arguments have been presented why this is very likely to be provable shortly. SKREM-like ciphers present other advantages also: they are not block ciphers, but instead operate with the entire plain text, however long it may be. This means there are no vulnerabilities associated with the choice of some chaining method, like XTS, needed to turn constant-sized block ciphers

like AES into stream ciphers. Furthermore, SKREM is not based on algebraic operations. At its core, it leverages the operation of indirection - thus even a small, 1 bit change in its source can result in a fully independent, unpredictable result. This makes SKREM able to counter quantum computing attacks which are known and respectively suspected or feared for some other symmetric key and public key ciphers.

For all these benefits, SKREM only requires that it be provided with a rather large - yet linear - number of truly randomly generated bits. Furthermore, both encryption and decryption are asymptotically optimal.

Nevertheless, SKREM as introduced in [2] is slow. The constant factors of about 100,000 are still too large for a large range of practical applications. An avenue of immediate further research is to further perfect SKREM-like ciphers, reducing the constant factors to something more manageable, without compromising the security guarantees.

Additionally, proving the Conjectures and/or Assumptions on which the security guarantees of SKREM-like ciphers rely is expected to be a major milestone in cryptography since this essentially solves symmetric key encryption, even in the post-quantum computing era. In doing that, a natural first step is to fully implement a SKREM-like cipher and analyze its operation statistically. It is conjectured that it produces output indistinguishable from random, except with negligible probability - so the cipher texts it produces should pass all randomness tests with flying colors.

Further beyond, expanding the applications of SKREM-like ciphers into related areas such as Public Key Cryptography and Turing-complete functional encryption is desired and could be the subject of further research as well.

One aspect which was taken rather for granted with SKREM is the generation of truly random numbers. The approach proposed involved harnessing such randomness from hardware sources - these can include natural phenomena, such as inbound solar radiation or FM noise, a quantum computer or a human user moving the mouse repeatedly. Alternatively, they could be sampled from a sequence produced by a third party trusted for the use case (such as random.org perhaps). For truly sensitive applications however, further research is required into how acceptable master tables can be generated, verified and distributed securely. We expect a lot of practical attacks on SKREM to focus on the low quality and improper reuse of master tables.

Finally, both to allow further analysis of SKREM-like ciphers, and “ideal candidates” therefrom, an area of direct further development is to provide a library implementing SKREM, and the methods based on SKREM above. This can then be integrated into existing private or open source packages, like VeraCrypt [26], making the promise of provably unbreakable security with short secret keys - and its applications - widely available and generally adopted.

We conclude this paper here.

In “good old fashioned” tradition of publications in cryptography we offer 20\$ to the first 14 people who show an attack on SKREM. Furthermore, we offer an additional 28\$ to the first 9 people who invalidate Conjecture 1 or Assumptions 1 and 2.

8 Acknowledgments

The ideas behind SKREM-like ciphers was inspired by the author’s exposure to chaos theory as a child, while visting the Paris science museum Cité des Sciences et de l’Industrie. Specific details were inspired by examining the field and some elementary results pertaining to Kolmogorov extractors by Zimand, M. (see [3] and [4]). The formalization of Chaos machines in [6] is both used as a potential subroutine in SKREM and served as a catalyst for faster organization of the present results in written form.

The secret agent authentication and authorization schemes are based on ideas the author had in the summer of 2015, for a “protocol for mutual recognition of synonymous”. A synonymous is a human person, of female gender, semantically equivalent to love.

Warm thanks for their existence and patient understanding to the few beautiful persons who inspired the author to strive to make the world a better place for them, and furthermore served as an inspiration for some of the constants used in the algorithms. This paper would never have existed without them.

9 Authors

Mircea Digulescu is an independent computer science researcher, software engineer, entrepreneur, military and intelligence enthusiast and amateur writer.

He has a PhD (ABD) at the University of Bucharest, Department of Computer Science of Faculty of Mathematics and Computer Science, from which he also obtained both a Bachelors and a Master’s degree in Computer Science. He gained recognition by being awarded medals and prizes at international contests and Olympiads in Computer Science, including 1 Bronze Medal at CEOI and 2 prizes at ACM SEERC. He is a Div1 coder on Codeforces.com. He also has over 15 years of experience in the software industry, creating systems that currently run in production and scale to billions of transactions.

His research interests include Cryptography, Game Theory, Complexity Theory as well as Algorithms and Data Structures.



References

- [1] Shor, P.W., “Algorithms for quantum computation: discrete logarithms and factoring”, *Proceedings 35th Annual Symposium on Foundations of Computer Science.*, 1994, 124-134.
- [2] Digulescu, M., 2019, *Hiding Data in Plain Sight: Towards Provably Unbreakable Encryption with Short Secret Key and One-Way Functions*. *ResearchGate*. DOI: 10.13140/RG.2.2.34319.94887.
- [3] Zimand, M., “On the topological size of sets of random strings”, *Mathematical Logic Quarterly*, **32(6)** (1986), 81-88.
- [4] Zimand, M., 2009, *Extracting the Kolmogorov complexity of strings and sequences from sources with limited independence*. *arXiv*. arXiv:0902.2141.
- [5] Levin, L.A., “The tale of one-way functions”, *Problems of Information Transmission*, **39(1)** (2003), 92-103.
- [6] Czyzewski, M.A., “Chaos Machine: Different Approach to the Application and Significance of Numbers”, *IACR Cryptol. ePrint Arch.*, 2016, 468.
- [7] Rivest, R.L., Shamir, A. and Adleman, L., “A method for obtaining digital signatures and public-key cryptosystems”, *Communications of the ACM*, **21(2)** (1978), 120-126.
- [8] Daemen, J. and Rijmen, V., “The block cipher Rijndael”, *In International Conference on Smart Card Research and Advanced Applications*, 1998, 277-284.
- [9] Tang, C., Pei, D., Liu, Z. and He, Y., “Non-Interactive and Information-Theoretic Secure Publicly Verifiable Secret Sharing”, *IACR Cryptol. ePrint Arch.*, 2004, 201.
- [10] Diffie, W. and Hellman, M., “New directions in cryptography”, *IEEE transactions on Information Theory*, **22(6)** (1976), 644-654.
- [11] Taher ElGamal, “A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms”, *IEEE Transactions on Information Theory*, **31(4)** (1985), 469-472.

- [12] Lamport Leslie, "Constructing Digital Signatures from a One Way Function", *SRI International (CSL-98)*, 1979.
- [13] Koblitz, N., "Elliptic curve cryptography", *Math. Comput.*, **48** (1987), 203-209.
- [14] Steiner, J.G., Neuman, B.C. and Schiller, J.I., "Kerberos: An Authentication Service for Open Network Systems", *Usenix Winter*, 1988, 191-202.
- [15] *National Security Agency, US, 2015, Elliptic curve cryptography. Commercial National Security Algorithm Suite.*
- [16] Biryukov, A., Dunkelman, O., Keller, N., Khovratovich, D. and Shamir, A., "Key recovery attacks of practical complexity on AES-256 variants with up to 10 rounds", *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 2010, 299-319.
- [17] "Inside the NSA's War on Internet Security", 2014.
- [18] Ryabko B., and Zhuravlev V., "Pseudo-random number generators with proven statistical properties", *The 9th Workshop on current trends in Cryptology*, 2020.
- [19] Barreto, P.S.L.M. and Rijmen, V., "The Whirlpool hashing function", *First open NESSIE Workshop, Leuven, Belgium*, 2000.
- [20] Penard, W. and van Werkhoven, T., "On the secure hash algorithm family", *National Security Agency. Tech. Rep.*, 2008.
- [21] Dworkin, M.J., "SHA-3 standard: Permutation-based hash and extendable-output functions", *Federal Inf. Process. STDS*, 2015.
- [22] Merkle, R.C., "A digital signature based on a conventional encryption function", *Conference on the theory and application of cryptographic techniques*, 1987, 369-378.
- [23] Chaum, D., Evertse, J.H. and Van De Graaf, J., "An improved protocol for demonstrating possession of discrete logarithms and some generalizations", *Workshop on the Theory and Application of Cryptographic Techniques*, 1987, 127-141.
- [24] Blum, M., "How to prove a theorem so no one else can claim it", *Proceedings of the International Congress of Mathematicians*, **1** (1986), 2.
- [25] Damle, A., Bangera, M., Tripathi, S. and Meena, M., "Blockchain Technology: An Overview", *SAMRIDDI: A Journal of Physical Sciences, Engineering and Technology*, **12(SUP 1)** (2020), 243-247.
- [26] Bursać, M., Vulović, R. and Milosavljević, M., "Comparative Analysis of the Open Source Tools Intended for Data Encryption", *International Conference on Information Technology and Development of Education - ITRO 2017*, 2017.

Secure Cloud Key Management based on Robust Secret Sharing

Ahmed Bentajer¹, Mustapha Hedabou², Sara Ennaama¹, and Abderrahim Tahiri¹

¹SIGL LAB., ENSA of Tetouan, University Abdelmalek Essaadi Tetouan, Morocco

²UM6P Benguerir, Morocco

Abstract. The aim of this paper is to propose a model to strengthen the security of key management in cloud computing, where the model is shared or entirely controlled by a non-trusted third party provider. Key management is not a straightforward matter for IT-teams, in addition to critical issues related to properly managing and securing the keys on providers' infrastructures, they have to deal with concerns specific to multi-cloud key management. Hardware Security Module (HSM) solution that offers a secure on-premise encryption key management turned out to be impracticable for widespread cloud deployment. HSM as a Service seems to be the best approach for key management in multi-cloud, but the service is wholly owned and managed by another cloud provider. In This paper, we present an efficient and secure cloud key management that fulfills the requirements of multi-cloud deployment. The proposed design splits the key into a blinded version of n shares that will be stored in encrypted format at the cloud provider side. To demonstrate the efficiency of the proposed design, we implement a fully featured prototype and evaluate its performance. Results analysis shows that the proposed design is highly efficient and can serve as a groundwork for using secret share as a way to protect keys in a multi-cloud environment.

Keywords: Key Management Security, Secret sharing, MultiCloud, Cryptography, Security and Privacy

1 Introduction

According to a study by the International Data Group, 81% of organizations have at least one application or a portion of their computing infrastructure in the cloud [1]. This is due to the economic nature of cloud computing, which can reduce the cost and complexity of owning and managing internal infrastructure in an on-demand and pay-as-you-go metric. However, its adoption lead to data control loss to an unreliable Cloud Service Provider (CSP). The main concerns are about confidentiality, integrity and privacy of the outsourced data [2, 3]

CSPs are leveraging cryptography as lever for mitigating security concerns in order to strength confidence of end users on their services. Cryptography can be involved in two major levels of security, namely secure storage [11, 4–7] and secure

computation [12–15]. Recently, some commercial offers of secure storage services with encrypted data were implemented in cloud infrastructures. The most known secure storage services, which encrypt data on the client side prior to outsourcing it, are Spideroak and Dropbox.

As for On-premise infrastructure, protecting digital assets and secure communication depend mainly on cryptography, the increasing of cyber attacks led the management of cryptographic keys more important than the key itself. This means that companies need to be more vigilant in key management at the cloud level. Depending on the type of cloud service in use, most key management functions are partially or fully controlled by the cloud providers. For PaaS and SaaS service delivery, the major part of the key management is processed internally by the cloud providers. Even for IaaS model, keys used for signing virtual machine template are internally managed [8].

Key management encompasses operations like keys generation, storage, archiving, distribution and destruction at the end of their life cycles. Due to their sensitivity, keys must be handled with care. Keys must be generated in a random way, stored in a very safe place and exchanged via secure protocols [8, 9]. Very likely, this is done by making use of hardware facilities. For particular users, smart card or TPM [20] can be applied, whereas HSM can fit more companies and government needs. Needless to say that HSM has been developed before the advent of cloud computing paradigm, therefore they must go along with a key management system as it occurs on-premise infrastructures.

To alleviate cloud users from managing keys, which are their main goal of embracing cloud services, cloud providers offer HSM as a service. With AWS CloudHSM, Amazon provide HSM appliances in data centers as a service to users [16]. Undoubtedly, the physical HSM limitations related to lack of elasticity and operability have been addressed by HSM as-service, still there is need for software infrastructure, owned and procured by cloud services providers, to drive HSM as service. In a nutshell, HSM as-service has brought some desired security and easy management properties, but the HSM technology was not originally developed for cloud and still presents limitations from cloud users perspective.

Recently, another approach achieving effective cloud key management, based on the use of homomorphic encryption, was put forward. It was dedicated expressly to cloud services and was designed in such a way to meet the five characteristics of the cloud computing paradigm including elasticity, On demand self service and availability. The solution is already integrated with Web Services (AWS) and RedHat, but it works with any cloud platform.

Security can only proved in semi-honest model. The solution provider, namely Porticor, must be trusted to implement the protocol as specified and cloud providers' platforms executing the implementation to have a neutral behavior. Semi-honest models are believed to be non-trivial task and thus may undermine the security

gain provided by the solution. The lack of detailed technical information about the solution and about the span of its adoption by final cloud users make the final statement very difficult.

This paper introduces a new design for secure and efficient software cloud key management system. Based on (t, n) robust secret sharing mechanism, the proposed design splits up the master key on n servers hosted on cloud computing providers side that communicate on asynchronous private and authenticated channels. The design tolerates up to $(n - t - 1)$ faulty servers. In addition, the storage of public information, namely the Lagrange coefficients, speeds up the computation of the secret (master key) reconstruction making the design more efficient. Furthermore, we construct a formal model of our design and prove its security in semi-honest model and finally we report on a prototype of implementation along with the performance study. Well established trusted computation and execution facilities will be leveraged to share, store and securely compute the key shares and reconstruction

The remainder of this paper is structured as follows. Section 2 presents preliminaries and basic design. Section 3 aims to present each protocol in the design, while section 4 presents the design implementation, security analysis and performances. Finally, we come-up with our conclusions and assumptions

2 preliminaries

2.1 Secret Sharing

The secret sharing theory is a very attractive research field. It has many applications, multiparty computation is by far the most relevant one. In this paper, we focus on particular Shamir based secret sharing schemes [10]. We assume that a dealer wants to share a secret s amongst n parties so that no less than $t + 1$ parties can recover the secret, whereas it can easily be recovered by any $t + 1$ or more parties. This is referred to as (t, n) secret sharing. Shamir based secret sharing scheme is built upon polynomials over finite field F , with $|F| > n$. For the sake of correctness and simplicity, we suppose that $F = F_p$ with $p > n$.

Whereas it can easily be reconstructed from any $t+1$ or more shares. Both of these facts are proved using Lagrange interpolation.

Shamir's early idea [10] of distributing shares of a secret as evaluations of a polynomial has become a standard building block in threshold cryptography. The scheme is based on polynomial interpolation. Given k couples (x_i, y_i) , with distinct x_i 's, there is one and only one polynomial $q(x)$ of degree $k - 1$ such that $q(x_i) = y_i$ for all i . This basic statement can be proved by using Lagrange interpolation. Without loss of generality, we can assume that the secret s is (or can be made) a number. To divide it into pieces $[s]_i$, we pick a random $k - 1$ degree polynomial $q(x) = a_0 + a_1x + \dots + a_{k-1}x^{k-1}$ in which $q(0) = s$, and evaluate:

$$s_1 = q(1) \cdots s_i = q(i) \cdots s_n = q(n).$$

Given any subset of $t + 1$ of these $[s]_i$ values (together with their identifying indices), we can find the coefficients L_i of $q(x)$ by interpolation, and then evaluate

$$s = q(0) = \sum_{i=1}^{i=t+1} L_i [s]_i, \text{ where } L_i = \prod_{j \neq i} \left(\frac{x_j}{x_j - x_i} \right)$$

The basic secret sharing scheme will have some flaws if some participants are dishonest [17]. For withstanding malicious participants, a new type of secret sharing scheme was proposed by Fieldman [18], called the verifiable secret sharing (VSS) scheme. The coefficients of this polynomial hidden in the exponent of the generator of a group in which the discrete-log assumption holds, are published. This allows that the participants can validate correctness only of their own shares distributed by the dealer in the distribution phase. In [19], Stadler introduced the publicly verifiable secret sharing (PVSS) scheme that allows that anyone can verify the validity of shares without revealing any secret information.

2.2 Model and assumptions

In this paper, we are dealing with scenarios where a software key management system Following the BYOK (Bring Your Own Keys) model is deployed in the cloud providers side as an ad-on facility to an existing on-premise key management system. The keys are managed on the on-premise side, following best practices, keys are exported, stored and handled in the cloud provider sides in a secure way. While they are in transit, conventional and well established techniques, including SSL, SSH DH key exchange are leveraged to achieve security. In the cloud providers side, keys are stored in a distributed way through n servers S_1, S_2, \dots, S_n . Latter, keys are reconstructed and their integrity is verified by using secure computation approaches. The proposed protocol can be modeled as follows:

- Secure storage. During the lifetime of the application, all servers possess some sensitive information to be stored in a secure and authenticated way. This could be shares of the keys or pieces of MAC's.
- Secure computation. The n servers S_1, S_2, \dots, S_n are involved in some secure computation phase taking place in the cloud providers side.
- Online and Offline phases. The protocol goes through periods where servers are active and other where they are idle. In the active periods, the servers are requested to send back their keys and MAC's shares to a dealer who conduct the secure computation for reconstructing and checking the validity of recovered keys. These periods, called on online phases, alternate with other where the servers are inactive. The latter periods are called Offline phases. The Offline and

Online phases must be synchronized in order to switch between these Offline and Online periods.

As for on-premise settings, the proposed key management system is implemented as a stand-alone application. The servers in cloud computing sides are fully autonomous, that is they can switch between offline and online phases without any interactions with outside the cloud instances. The servers can only communicate with each other in order to conduct the whole process. In other words, the only players involved are the servers themselves. This model comes with a limited level of confidentiality, availability that can be provided. This level is tightly related to the number of servers required for restoring the secret key without the leakage of any information about it.

The confidentiality threshold can be defined as the minimal number $Conf_{min}$ of server an adversary can break into to learn the secret key, whereas the availability threshold $Avail_{min}$ as the minimal number of uncorrupted server that should be available for restoring the key. In a fully autonomous scenario, the number of malicious servers $n-Avail_{min}$ must be at most $n/2$ for ensuring confidentiality and availability of the protocol. This limitation which is mainly due to the requirement that servers are not allowed to communicate with any instance from outside the cloud. The requirement about the number of malicious servers can be relaxed by limited interaction with on-premise key management system.

We make standard assumptions about the well established cryptographic techniques regarding the ability of an adversary to undermine their security. The techniques used for establishing secure and private channels or for authenticating parties, including SSL, SSH, DH key exchange are assumed secure in standard models.

2.3 Basic Design

We here give an informal description of our protocol that implement cloud key management system based on Robust Verifiable Secret Sharing with fully autonomous servers. The protocol consists in three main phases. A set up phase where the on-premise key management system computes the shares of the master key and the MAC and communicates them to the main instance (dealer) in cloud provider side through a secure channel. The second phase is where the servers enter into an offline period after receiving their shares and the final one consisting in conducting secure computation to reconstruct the secret key after they return to online period. The two subsequent phases are launched by the main instance.

Our protocol for key management in cloud computing, denoted $cloud_{KMS}$ consists of three main components, namely the key management system on-premise, an appliance acting as the dealer and n servers S_1, S_2, \dots, S_n . The appliance and servers, owned and managed by cloud users, are located in cloud side. We assume that the appliance is a trusted component. It is a semi-honest component (passive), which means that it behaves as prescribed by the protocol. This goal can

be achieved by issuing remote attestation for its software. The protection against passive attackers (an eavesdropper) is provided by leveraging a trusted execution mode such as SGX enclave and by using standard cryptographic tools like SSL, SSH.

The protocol $cloud_{KMS}$ assume that a key management system is already active on the user side (on premise). The KMS is responsible for generating the keys that will be used on the cloud computing side following the model BYOK. For the sake of simplicity, we assume that we are dealing with a single key, the master key K . The protocol $cloud_{KMS}$ can be conducted in three sub protocols. A Set up protocol generating the public parameters and computing the shares of the key k , denoted $[s]_i$. The sub protocol Sharing delivers the shares $[s]_i$ to the servers S_i in a confidential and authenticated way. The last sub protocol Reconstruction allows to get back the share and to conduct computations and reconstruct the master key. We now introduce the formal model of our protocol $cloud_{KMS}$ following the model BYOK (Bring Your Own Keys). As mentioned above, the protocol consists of three sub protocols.

- Protocol $Set_{up}(k)$: executed by the key management system on-premise, it takes the security parameter k . The protocol outputs the corresponding MAC of k denoted γ , the finite field F_p and the public shares $[s]_i$ of the master key along with the public shares of the MAC γ'_i for $i = 1, \dots, n$.
- Protocol $Sharing(\gamma, [s]_i, \gamma'_i)$: Executed by the trusted appliance, it takes shares of the master key and MAC along with value of the MAC. The protocol delivers/retrieves the shares $[s]_i, \gamma_j$, for $i, j \in \{1, \dots, n\}$ to/from the servers. The value of the MAC γ is stored by the trusted appliance.
- Protocol $Reconstruction([s]_i, \gamma'_i)$ takes as inputs the shares of the master key and the MAC. The protocol executed by the trusted appliance outputs the master key s . We note the protocol recover the MAC from its sharing and compares it with the public value γ stored by trusted appliance before reconstructing the master key

3 The proposed protocol

In this section we describe the main 3 sub protocols of our cloud key management system $cloud_{KMS}$, namely SetUp, Sharing and Reconstruction.

We denote the number of servers by n and the security parameter (master key) by k . We assume that there is one k from which we derived the specific shares. The k will be used during the life of the system. As mentioned before, the traffic between the components of the system is encrypted and authenticated using standard cryptographic tools.

3.1 SetUp protocol

To initiate the process, a user through the on-premise key management system denoted O_{KMS} generates the master key k . Executed by an application on behalf of the key management system on premise, it takes k . The protocol outputs the finite field F_p , the public shares $[s]_i$ of k along with its MAC (γ) and the public shares of the MAC γ'_i for $i = 1, \dots, n$. The algorithm 1 depicts the implementation of the protocol.

Algorithm 1 Sub protocol *Setup*

- 1: **function** SETUP(O_{KMS}, A)
 - 2: Compute the MAC $\gamma \leftarrow MACk$
 - 3: Sample random number $a_1, \dots, a_n \in Z_p$ and $b_1, \dots, b_n \in Z_p$
 - 4: Set $a_0 \leftarrow s$ and $b_0 \leftarrow \gamma$
 - 5: Set $q(x) \leftarrow \sum_{i=0}^{i=t} a_i x^i$ and $p(x) \leftarrow \sum_{i=0}^{i=t} b_i x^i$
 - 6: Compute $[s]_i \leftarrow p(i)$ and $\gamma'_i \leftarrow q(i)$ for $i = 1, \dots, n$
 - 7: Send to appliance A: γ and $([s]_i, \gamma'_i)$ for $i = 1, \dots, n$
-

3.2 Sharing

This protocol is executed with SetUp and Reconstruction protocols. It takes shares of the master key $[s_i]$ key and MAC γ along with value of the MAC γ'_i . The protocol delivers/retrieves the shares $[s]_i, \gamma_j$, for $i, j \in \{1, \dots, n\}$ to/from the servers. The value of the MAC γ is stored by the trusted appliance. (Algorithm 2).

Algorithm 2 Sub protocol *Sharing*

- 1: **function** SHARING(A, S)
 - 2: Store the MAC γ
 - 3: Send each servers $S_i: ([s]_i, \gamma'_i)$ for $i = 1, \dots, n$
-

3.3 Reconstruction

Takes as inputs the shares of the master key and the value of γ'_1 . The protocol executed by the trusted appliance outputs the master key s . We note the protocol recover the MAC from its sharing (γ'_i) and compares it with the public value γ stored by trusted appliance before reconstructing the master key.

Algorithm 3 Sub protocol *Reconstruction*

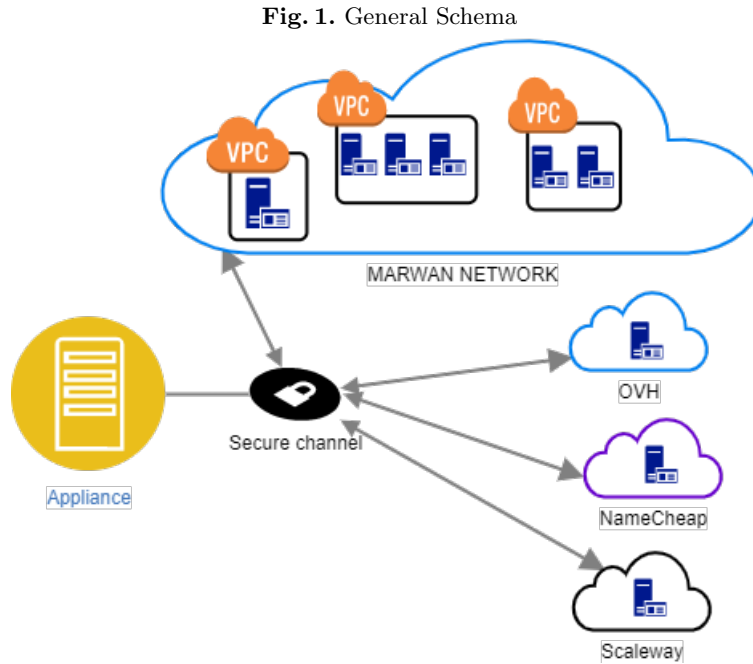
```

1: function POST(S, A)
2:   number_share Integer  $\leftarrow 1$ 
3:   Compute  $L_i \leftarrow \prod_{j \neq i} (\frac{x_j}{x_j - x_i})$ 
4:   Compute  $s \leftarrow q(0) \leftarrow \sum_{i=1}^{i=t+1} L_i[s]_i$ 
5:   Compute  $\gamma = p(0) \leftarrow \sum_{i=1}^{i=t+1} L_i[\gamma']_i$ 
6:   if  $\gamma == q(0)$  then
7:     compute  $k$ 
8:   else
9:     PickAnotherShare number_share  $\leftarrow 0$ 
10:  if number_share == 0 then
11:    Print : A share has been compromised

```

4 Security Analysis and Performance Evaluation

Figure 1 shows a presentation of how tests were conducted and where the shares were stored/retrieved.



We implemented our 3 protocols with Java 1.8 using a 64 bits Windows operating system with i7-8565 (1.8 GHz) processor and 16Go installed RAM. The Sharing protocol has been developed using JCraft library which is a pure Java implementation of SSH2 that is known to have more defensive mechanisms to avoid

vulnerabilities. Experiments are performed on different key size (AES-128, AES-192, AES-256, RSA-1024, RSA-2048 and RSA-4096) while the t and n of shamir secret sharing were $t = 3$ and $n = 9$.

We measured the performance of the proposed protocol using our developed prototype. We divided the overhead time of each measurement into :

- SetUp and Share : The split of secret into shares, MAC computation and the Upload time to Servers;
- Share and Reconstruction : The download time from Server, computation of MAC and reconstruction of secret

Tables 1 and 2 show the results of the running time for computation and file upload/download in seconds.

Table 1. SetUp and Share protocols performances

Key Size	SSS	Computation	Upload	Total	Latency	Upload
AES-128	0.005		18	18.005	± 0.8	
AES-192	0.005		18	18.005	± 0.8	
AES-256	0.005		18	18.005	± 0.8	
RSA-1024	0.03		18	18.03	± 0.8	
RSA-2048	0.03		18	18.03	± 0.8	
RSA-4096	0.03		18	18.03	± 0.8	

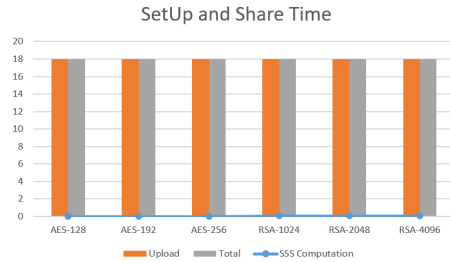
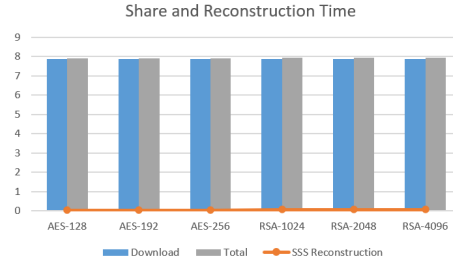
Table 2. Share and Reconstruction protocols performances

Key Size	SSS	Computation	Download	Total	Latency	Download
AES-128	0.04		7.86	7.9	± 0.43	
AES-192	0.04		7.86	7.9	± 0.43	
AES-256	0.04		7.86	7.9	± 0.43	
RSA-1024	0.07		7.86	7.93	± 0.43	
RSA-2048	0.07		7.86	7.93	± 0.43	
RSA-4096	0.07		7.86	7.93	± 0.43	

Analysis proves that data transmission is a dominant factor while shares computation do not heavily penalize the performance of the proposed design for different key size (Figures 2 and 3). The latency time depends mainly on network quality for file transfer and also the time taken by the appliance to authenticate to servers.

It is obvious that the proposed protocol adds a new security layer for the secret key confidentiality. Based on our proposed design security threats may be a :

- Malicious insider user who may attempt to gather information through side channel attacks

Fig. 2. SetUp and Share Time**Fig. 3.** Share and Reconstruction Time

- Malicious external attacker who may try to intercept communication and steal the shares.

The secret share addresses the specific need to enhance the security of the key during its lifetime. The use of secret share schema establishes a mechanism of sharing sensitive data securely amongst an untrusted network. Besides, our proposed design inherits some properties related to Shamir's (k, n) thresholds as:

- The system is *Information-theoretic security* meaning that an attacker with high computational power cannot break the secret without having minimum number of thresholds required to reconstruct the key.
- The system is extensible, where k_i could be dynamically added/removed without affecting other shares
- The size of each share does not exceed the size of the original data

However, during the upload/download of the shares it was noticed that the system freezes or takes longer than usual to upload or download the shares, this is usually due to the network connection and/or the interactions of the Appliance with the servers to authentication management. In addition, if a share is compromised, it will be difficult to know which part was affected.

The mere fact that the Appliance is hosted in on-premise does not mean that it is completely trusted. A malicious insider can still tamper with the application. This issue depends mainly on the organization hosting the KMS it self. Intel SGX may be leveraged to offer hardware-based memory encryption and isolates the running Appliance code and data in memory from processes running at a higher privilege level.

5 Conclusion and future work

In this paper, we proposed a secure cloud key management based on the robust secret share. The protocol is based on Shamir secret sharing that securely distribute fragments of secret key amongst a different distributed cloud server. We have also

implemented a prototype of our proposed prototype to demonstrate its practicality. The results are promising, the computation of shares and MAC are very negligible compared to data transfer. In the future, we plan to improve the proposed Appliance through the use of Intel SGX which will give it more protection from disclosure or modification. And implementing a sub-Appliance that will split the share of a server into other shares in order to improve the security of the secret.

References

1. IDG, "IDG Cloud Computing Survey", IDG (2020). <https://www.idg.com/tools-for-marketers/2016-idg-enterprise-cloud-computing-survey/>
2. P. J. Sun, "Security and privacy protection in cloud computing: Discussions and challenges", *Journal of Network and Computer Applications* 160 (2020) 102642. doi:10.1016/j.jnca.2020.102642.
3. A. Bentajer, M. Hedabou, K. Abouelmehdi, Z. Igarramen, S. El Fezazi, "An IBE-based design for assured deletion in cloud storage", *Cryptologia* 43 (3) (2019) 254-265. doi:10.1080/01611194.2018.1549123.
4. A. Bentajer, M. Hedabou, K. Abouelmehdi, S. Elfezazi, CS-IBE : AA data confidentiality system in public cloud storage system, in: *Procedia Computer Science*, Vol. 141, Elsevier B.V., 2018, pp. 559-564. doi:305 10.1016/j.procs.2018.10.126.
5. Z. Igarramen, M. Hedabou, FADETPM: Novel approach of file assured deletion based on trusted platform module, in: *Lecture Notes in Networks and Systems*, Vol. 49, Springer, 2019, pp. 49-59. doi:10.1007/978-3-319-97719-5_4.
6. J. Xiong, Y. Zhang, S. Tang, X. Liu, Z. Yao, Secure Encrypted Data with Authorized Deduplication in Cloud, *IEEE Access* 7 (2019) 75090-75104. doi:10.1109/ACCESS.2019.2920998.
7. R. Chandramouli, D. Pinhas, Security Guidelines for Storage Infrastructure, Tech. rep., National Institute of Standards and Technology, 315 Gaithersburg, MD (oct 2020). doi:10.6028/NIST.SP.800-209.
8. R. Chandramouli, M. Iorga, S. Chokhani, Cryptographic Key Management Issues and Challenges in Cloud Services, Tech. rep., National Institute of Standards and Technology, Gaithersburg, MD (sep 2013). doi:10.6028/NIST.IR.7956.
9. Bentajer A, Hedabou M, Chapter 6. Cryptographic Key Management Issues in Cloud Computing, in: Victoria M. Petrova (Ed.), *Advances in Engineering Research*, 34th Edition, Nova Science Publishers, Inc., 2020,
10. A. Shamir, How to share a secret, *Communications of the ACM* 22 (1979) 612-613. doi:10.1145/359168.359176.
11. W. Shi, T. Liu and M. Huang, "Design of File Multi-Cloud Secure Storage System Based on Web and Erasure Code," 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2020, pp. 208-211, doi: 10.1109/ICSESS49938.2020.9237703.
12. Katrina O., Saxena A. (2010) Secure Computation with Fixed-Point Numbers. In *Proceedings: Sion R. (eds) Financial Cryptography and Data Security. FC 2010. Lecture Notes in Computer Science*, vol 6052. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14577-3_6
13. M. Nassar, A. Erradi, F. Sabry and Q. M. Malluhi, "A Model Driven Framework for Secure Outsourcing of Computation to the Cloud," 2014 IEEE 7th International Conference on Cloud Computing, Anchorage, AK, USA, 2014, pp. 968-969, doi: 10.1109/CLOUD.2014.145.
14. Q. Wang, F. Zhou, C. Chen, P. Xuan and Q. Wu, "Secure Collaborative Publicly Verifiable Computation," in *IEEE Access*, vol. 5, pp. 2479-2488, 2017, doi: 10.1109/ACCESS.2017.2672866.

15. A. Bilakanti, Anjana N.B., Divya A., K. Divya, N. Chakraborty and G. K. Patra, "Secure computation over cloud using fully homomorphic encryption," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, 2016, pp. 633-636, doi: 10.1109/ICATCCT.2016.7912077.
16. X. Huang and R. Chen, "A Survey of Key Management Service in Cloud," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 916-919, doi: 10.1109/ICSESS.2018.8663805.
17. Schoenmakers B. (1999) A Simple Publicly Verifiable Secret Sharing Scheme and Its Application to Electronic Voting. In: Wiener M. (eds) Advances in Cryptology — CRYPTO' 99. CRYPTO 1999. Lecture Notes in Computer Science, vol 1666. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-48405-1_10
18. P. Feldman, "A practical scheme for non-interactive verifiable secret sharing," 28th Annual Symposium on Foundations of Computer Science (sfcs 1987), Los Angeles, CA, USA, 1987, pp. 427-438, doi: 10.1109/SFCS.1987.4.
19. Stadler M. (1996) Publicly Verifiable Secret Sharing. In: Maurer U. (eds) Advances in Cryptology — EUROCRYPT '96. EUROCRYPT 1996. Lecture Notes in Computer Science, vol 1070. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-68339-9_17
20. M. Hedabou and Y. S. Abdulsalam, "Efficient and Secure Implementation of BLS Multisignature Scheme on TPM," 2020 IEEE International Conference on Intelligence and Security Informatics (ISI), Arlington, VA, USA, 2020, pp. 1-6, doi: 10.1109/ISI49825.2020.9280511.

Authors

Ahmed Bentajer received his M.S. degree in National School of Applied Sciences from Cadi Ayyad University in 2012. In 2019, he received his Ph.D degree in computer science from EST of Safi from Cadi Ayyad University, Marrakech, Morocco. Currently he is a professor at ENSA of Tetouan. His area interest covers Information Security, Security architecture, Identity based cryptography and cloud computing.

Mustapha Hedabou received his M. Sc degree in Mathematics from the university of Paul Sabatier, Toulouse, France. In 2006, he received his Ph.D degree in computer science from INSA de Toulouse, France. He was a professor at ENSA of Safi, from Cadi Ayyad University Marrakech in Morocco, And now he is Associate Professor at Mohammed VI Polytechnic University, Benguerir, Morocco. His area interest covers Information Security, Public Key Cryptography based on Elliptic Curves, Identity based cryptography and cloud computing.

Sara Ennaama Sara ENNAAMA completed her Master's Degree in Business Intelligence and Big Data Analytics from Chouaib Doukhali University in 2020. Before that, she got her Bachelor's degree in Mathematical and Computer Sciences from

Cadi Ayyad University. She is currently pursuing a PhD in Computer Science at Abdelmalek Essaadi University and is passionate about cryptography, cloud storage, secure deletion and cloud computing.

Abderrahim Tahiri Obtained the Engineer degree in Computer Sciences in 2000 from Abdelmalek Essaâdi University (UAE) in Morocco and the Master degree in Telematics Engineering in 2007 from Polytechnic University of Cartagena (UPCT) in Spain, and the PhD degree in Computer Sciences from UAE and UPCT in 2009. Currently he is professor at the National School of Applied Sciences in the UAE, he is a full member of the Computer Sciences Engineering Department, specialized in Internet object models and applications and a full member of Information System and Software Engineering Laboratory of UAE. His research interests include software architecture integration and smart application models. He has cooperated in, and coordinated several projects on national level and on European level. His research output includes 35+ co-authored articles. He has been chair of multiple conference tracks related to Information System Engineering and Wireless Sensor Network.

Hiding Data in Plain Sight: Towards Provably Unbreakable Encryption with Short Secret Keys and One-Way Functions

Mircea-Adrian Digulescu^{1,2}

¹Individual Researcher, Russian Federation

²Formerly: Department of Computer Science, Faculty of Mathematics and Computer Science, University of Bucharest, Bucharest, Romania, Romania

Abstract

It has long been known that cryptographic schemes offering provably unbreakable security exist, namely the One Time Pad (OTP). The OTP, however, comes at the cost of a very long secret key - as long as the plain-text itself. In this paper we propose an encryption scheme which we (boldly) claim offers the same level of security as the OTP, while allowing for much shorter keys, of size polylogarithmic in the computing power available to the adversary. The Scheme requires a large sequence of *truly random* words, of length polynomial in the both plain-text size and the logarithm of the computing power the adversary has. We claim that it ensures such an attacker cannot discern the cipher output from random data, except with small probability. We also show how it can be adapted to allow for several plain-texts to be encrypted in the same cipher output, with almost independent keys. Also, we describe how it can be used in lieu of a One Way Function.

Keywords: Encryption, Provable Security, Chaos Machine, Truly Random, One Time Pad, One Way Function.

1 Introduction

Most existing encryption schemes work by identifying some suitable family of bijective functions, one for each potential secret key, from the universe $\{0, 1\}^n$ of plain-texts to the universe $\{0, 1\}^m$ of cipher outputs. In this paper, we, the author, take a different approach by supplementing the input with a large, *truly random* sequence. This sequence, with only a few minor, hopefully undetectable, changes is then outputted as the result of encryption. The core idea is the following: we will partially permute the words in the truly random sequence based on the plain-text and the short secret key. The main ingredient to security lies in using the elements of the random sequence itself to determine which positions to exchange. The assumption on which our security claims are based is that the partial permutation determined by our method will appear random to any adversary having less computing power than allowed. More precisely, no algorithm of running time less than exponential in some arbitrary chosen security parameter should be able to distinguish the cipher output from random, even when the plain text is known or can be chosen.

From the ancient polyalphabetic substitution cyphers, the idea of employing random decision making in the encryption method is not new. The security of a scheme depends, however, on how randomness is employed precisely. The scheme we propose paves the way to a plethora of similarly-built ciphers, all relying on the same core ideas we present in this paper. As such, it can be regarded as being the first specimen from a newly introduced class of encryption methods.

The proposed scheme is intended to be cryptographically secure. All of the author's knowledge of fields such as computability (e.g. Kolmogorov complexity), complexity (e.g. NP-Completeness), [CS]PRNGS ([Cryptographically Secure] Pseudorandom Number Generators) as well as knowledge of existent schemes such as AES [2] and RSA [3], together with the assumptions on which their security claims are based, played a role in developing the current proposal. The claim that the proposed scheme is cryptographically secure is thus intended to be taken as an educated statement, not a mere shot in the dark. Furthermore, the scheme can be adapted for use in lieu of a One Way Function (which is something actively sought by researchers, consisting of an easy to calculate function, whose inverse is computationally hard to determine - see [1]), for most scenarios.

1.1 Prior work

To the best of our knowledge there is no public body of literature pertaining to encryption schemes which involve making small changes to a large volume of random data, based on the secret key and the plain text as well as on said random data. Nevertheless, we, the author, strongly suspect that non-public research, by people such as Marius Zimand (see [4] and [5]), Leonid Levin (see [1]) and others who have a competent scientific interest in randomness, Kolmogorov complexity and the like, exists which includes ideas similar to those in this paper. Nevertheless, to the best of our knowledge, such research, if it exists, is not public. Chaos Theory has played a role in the development of some ideas in this paper. Its applications in encryption, in the form of Chaos Machines, are best described by Armour in [6]. Such machines can easily be employed to augment the security of the proposed encryption scheme.

1.2 Overview of this paper

The rest of the paper is organized as follows. In Section 2 we present the proposed Encryption Scheme, including an abridged version of its pseudocode. We include a brief natural language description of its steps and also very briefly discuss the theory behind it, before making bold claims regarding its cryptographic security. We provide no formal proofs, but, instead, claim that such exist. We conclude by analyzing the performance of the proposed algorithm. In Section 3 we present some important ideas on further increasing security, while in Section 4, we tackle the opposite tradeoff, by presenting a practically feasible, simplified version of the Scheme. Section 5 discusses performance considerations for the simplified scheme. In Section 6, we present an idea on how such ciphers can be modified to allow multiple, independent plain-texts to be encrypted within the same cipher output, using almost independent keys. In Section 7, we show how the scheme can be used in lieu of a one way function. Section 8 is dedicated to practical considerations. Conclusions are drawn in Sections 9 and Acknowledgments offered in Section 10.

For considerations of brevity, this paper includes an abridged version of the pseudocode of the proposed Scheme, sufficiently comprehensive nevertheless to illustrate it properly.

Appendix A contains proofs for theorems in this paper. The appendix is included at the end of this paper, with separate references.

2 The Encryption Scheme

In this section we proceed to describe the proposed encryption scheme. For lack of a better name, we shall call it Short Key Random Encryption Machine or SKREM in short.

2.1 The Encryption Scheme SKREM

The scheme takes the following inputs for encryption:

1. Three sequences $M1, M2$ and $M3$, of sufficiently large size, consisting of m independent, uniformly distributed, truly random w -bit words each. We call these **grand master tables**. When it is obvious to which we refer, we use the notation M . These are unique to a particular encryption and are never reused.
2. The **plain text** P consisting of n bits which need to be hidden. It can be arbitrary.
3. A sequence T of sufficiently large size, consisting of independent, uniformly distributed, truly random w -bit words, to be used for random decision making. This sequence is discarded after use (and never reused). We call it the **randomness well**.
4. A relatively short list of secret key elements, $K_small[]$, consisting of uniformly distributed, independent, truly random bits. This is the **secret key** which needs to be provided at decryption.
5. Two equally sized lists of secret key elements $K1_large[]$ and $K2_large[]$, consisting of uniformly distributed, independent, truly random bits. We call these, the **large secret keys**. These are to be discarded after use (and never reused).

SKREM also incorporates a number of security parameters, grouped in the pseudocode in the *parameters_** structures which are an integral part of the scheme. Their values can be adjusted to obtain other SKREM-like schemes, however they are required to be identical at both encryption and decryption. The secret key and large secret keys can just as well be generated using the randomness well and provided as output, when no specific secret key generation method is required.

The first half of each grand master table is conceptually split into a number of smaller master tables, of equal size, to allow locations from it to be sampled using a lesser number of bits. The second half is used for replenishing values consumed from the first half. Each secret key element incorporates a few small numbers (called **key atoms**) which are used to generate a single location within a small master table.

The cipher output produced at encryption by the scheme, consists of the three, slightly modified, grand master tables $M1, M2$ and $M3$. The Decrypt routine, as expected, takes as input the output from the Encrypt and produces the original plain text.

Consider the following abridged version of the pseudocode for the SKREM encryption scheme.

Algorithm 1. *Abridged version of the pseudocode of the Encryption Scheme SKREM.*

- 1: **STRUCT** *params_normal*
- 2: $reqsec \leftarrow 256$ {Security strength parameter: Logarithm of computing power available to an adversary}
- 3: $dmod \leftarrow 0$ {Used to specify direct mode of key extension, using only one round}
- 4: $vrify \leftarrow 1$ {Specifies whether to return an error in case the security parameters do not offer the specified security strength}
- 5: $mtsize \leftarrow 85$ {Used to determine the size of one small master table}
- 6: $secrbase \leftarrow 4$ {Number of additional secret bits hidden in each key element to be used as base during key extension}
- 7: $secrtwo \leftarrow 1$ {Used to specify that an additional random exponent is to be sampled from the grand master table, during key extension}
- 8: $secrbpp \leftarrow 256$ {Length in bits of the offset used to select a random prime. Must be $\leq reqsec$.}

```

9:  $secrbpb \leftarrow 9$ {Number of bits, whose XOR is used to represent 1 emitted bit, during
   key extension}
10:  $secrbpn \leftarrow 9$ {Number of bits, whose XOR is used to represent 1 bit of plain text}
11:  $ppx \leftarrow 9$ {Number of key atoms per secret key element}
12:  $w \leftarrow 8$ {Number of bits in a word in  $M$  and  $T$ }
13:  $bopf \leftarrow 2$ {Multiplication factor for the number of bits in a number 1, necessary to
   sample it with (almost) uniform distribution from uniformly distributed individual
   bits}
14: END
15: METHOD Encrypt( $P, K1\_large[], K2\_large[], K\_small[], M1, M2, M3, T$ )
16: Use BasicEncryptDecrypt() to encrypt the two large keys,  $K1\_large[]$  and
    $K2\_large[]$  into  $M3$ , with secret key  $K\_small[]$ .
17: Use the randomness well  $T$  to generate an OTP for the plain text  $P$ .
18: Use BasicEncryptDecrypt() to encrypt the OTP into  $M1$ , with secret key  $K1\_large[]$ ,
   and its XOR with the plain text in  $M2$ , using secret key  $K2\_large[]$ .
19: return The modified grand master tables  $M1, M2$  and  $M3$  as the cipher output.
20: END
21: METHOD Decrypt( $M1, M2, M3, n, szLKey, K\_small$ )
22: Use BasicEncryptDecrypt() to retrieve the two large keys,  $K1\_large[]$  and
    $K2\_large[]$  from  $M3$ , using key  $K\_small[]$ .
23: Use BasicEncryptDecrypt() to retrieve the OTP from  $M1$ , using secret key
    $K1\_large[]$  and its XOR with the plain text from  $M2$ , using secret key  $K2\_large[]$ .
24: return The XOR of the two bit sequences obtained above, as the original plain text.
25: END
26: METHOD BasicEncryptDecrypt( $mode, n, P, K[], M, T, t, params$ )
27: {Performs encryption of P into M based on K[] or decryption from M into P based
   on K[], depending on mode}
28: {Initializations}
29: Initialize some constants based on the input and the security parameters. Of particular
   interest are the following. The values for when  $dmod = 1$  differ slightly, but their
   semantics remains the same.
30:  $k \leftarrow (secrbase * (1 + secrtwo) + (ppx) * mtsize * bopf + secrbpp) * bopf + secrbpp$  {Size
   of a key element in bits}
31:  $f \leftarrow 1 + ([1/((1 - 1/2^w))] - 1) * 10$  {Number of pairs of words sufficient to generate
   one random bit}
32:  $reqBitsPerKey \leftarrow k * 8 * (1 + secrtwo) * secrbpb$  {Number of bits required by a single
   key element, for extension}
33:  $reqWords \leftarrow (reqBitsPerKey * (secrbpn/ppx) * n/7 + secrbpn * n) * 2 * f$  {Total
   number of words consumed by the algorithm}
34:  $keyExtFactor \leftarrow 8$  {Number of new key elements to which a single old key element
   is extended}
35:  $mtSize \leftarrow Min(maxMTsize, NextPrime(2^{mtsize}))$  {Size of a small master table.
   Chosen to be a prime number, around  $2^{mtsize}$ }
36:  $noMTs \leftarrow Min(\lfloor \frac{m/2}{mtSize} \rfloor, 2 * f * reqBitsPerKey)$  {Total number of small master
   tables}
37: for  $i = 0$  to  $noMTs * mtSize - 1$  do
38:   Perm.Add(i) {Initialize Perm to be the identity permutation}
39: end for
40: {Security Parameters Validations}
41: if  $vr fy = 1$  then

```

```

42: Ensure that the following constrains are respected.
43:  $ppx * K[].Count * mtsize \geq reqsec + mtsize$ 
44:  $ppx * K[].Count * 4 \geq reqBitsPerKey/ppx$ 
45:  $mtsize + (mtsize - 1) * 2 + 3 \geq reqsec$ 
46:  $maxMTsize \geq NextPrime(2^{mtsize})$ 
47:  $noMTs \geq 2 * f * reqBitsPerKey$ 
48: end if
49: {Key Extension}
50: while There are  $\leq secrbpn * n$  key elements do
51:   for All old key elements and for increasing indexes of atoms within a key element
   and of small master tables do
52:     Use ExtractJthLocation() and GetLocation() to obtain two locations lp1 and lp2
     within the permutation list Perm, using a single atom of a single old key element.
53:     Use Perm to obtain two locations lm1 and lm2 in the grand master table M form
     lp1 and lp2.
54:     Use BurnLocation() to mark the locations lm1 and lm2 as used, and replenish
     Perm accordingly.
55:     Use GetBit() to obtain a random bit b2 from the values M[lm1] and M[lm2].
56:     Distribute b2 to the appropriate old key element, to be used for its extension.
57:   end for
58:   if A sufficient number of bits has been emitted to allow for extension of all old key
   elements then
59:     XOR every secrbpb bits from those distributed to each key element for extension,
     to obtain a usable bit.
60:     Extend each old key element into keyExtFactor new key elements, using
     ExtendKey(), thus concluding one key extension round.
61:   end if
62: end while
63: {Getting Locations For Encryption}
64: Use the last round key elements to emit a number of secrbpn * n bits from about twice
   as many locations in M, keeping a record of both.
65: {Encrypt / Decrypt}
66: Use the XOR of every secrbpn from the emitted bits above to represent a single bit
   of plain-text.
67: At encryption time, if this bit does not correspond to the desired one from P, swap
   the values at a single, randomly chosen, pair of locations in M, from the secrbpn used
   to generate it. Use the randomness well T to pick the exact pair.
68: {Return result}
69: return (P, M, t)
70: END
71: METHOD ExtendKey(K, ExtendBits, t, NewVals, k, count, params)
72: {Extends the key element K, by adding count new key elements to the NewVals list}
73: (_, KeyBits)  $\leftarrow$  ExpandKey(K, k, params)
74:  $lp \leftarrow 2^{k-secrbpp*(bopf+1)}$ 
75: for i = 0 to count - 1 do
76:    $x \leftarrow 17$ ,  $y \leftarrow 14$ ,  $v1 \leftarrow 17$ ,  $v2 \leftarrow 28$ 
77:    $q \leftarrow 0$ 
78:   (po, t)  $\leftarrow$  BuildValue(ExtendBits, t, secrbpp)
79:    $p \leftarrow NextPrime(lp + po)$ 
80:   (v1, q)  $\leftarrow$  BuildValue(KeyBits, q, secrbase)

```

```

81: if secrtwo > 0 then
82:    $(v2, q) \leftarrow \text{BuildValue}(\text{KeyBits}, q, \text{secrbase})$ 
83: end if
84:  $(x, t) \leftarrow \text{GenerateRandomFromBits}(\text{ExtendBits}, t, p, \text{params})$ 
85: if secrtwo > 0 then
86:    $(y, t) \leftarrow \text{GenerateRandomFromBits}(\text{ExtendBits}, t, p, \text{params})$ 
87: end if
88:  $\text{newK} \leftarrow (20 + [(v1 + 14)^{28+x} + (17 + v2)^{17+y} + 11431]^{-1}) \bmod p$ 
89:  $\text{newBits} \leftarrow \emptyset$ 
90:  $\text{GetBits}(po, \text{secrbpp}, \text{newBits})$ 
91:  $\text{GetBits}(\text{newK}, k - \text{secrbpp}, \text{newBits})$ 
92:  $(\text{genK}, \_) \leftarrow \text{BuildValue}(\text{newBits}, 0, k)$ 
93:  $\text{NewValus.Add}(\text{genK})$ 
94: end for
95: return  $t$ 
96: END
97: METHOD ExtractJthLocation $(K, k, j, \text{params})$ 
98: {Gets the value associated with the  $j$ -th atom used to index the small master tables,
    from key element  $K$ }
99: static  $\text{lastresult} \leftarrow 14$  {Retains value between method calls}
100:  $(\_, \text{KeyBits}) \leftarrow \text{ExpandKey}(K, k, \text{params})$ 
101:  $p \leftarrow \text{NextPrime}(2^{\text{mtsize}})$  {Gets the smallest prime larger than this value}
102:  $q \leftarrow \text{secrbase} * (1 + \text{secr\_two}) + j * \text{mtsize} * \text{bopf}$ 
103:  $(l, \_) \leftarrow \text{GenerateRandomFromBits}(\text{KeyBits}, q, p, \text{params})$ 
104: if  $\text{lastresult} < 2$  then
105:    $\text{lastresult} \leftarrow 28$ 
106: end if
107:  $l \leftarrow [17 + (\text{lastresult}^{l+28} + 14)^{-1}] \bmod p$ 
108:  $\text{lastresult} \leftarrow [17 + (\text{lastresult}^{l+20} + 11431)^{-1}] \bmod p$ 
109: return  $l$ 
110: END
111: METHOD BurnLocation $(\text{Perm}, l, m, \text{noMTs}, \text{mtSize})$ 
112: {Marks the location  $\text{Perm}[l]$  from the grand master table as used and performs some
    minor shuffling of  $\text{Perm}$ , based on  $l$ }
113:  $q \leftarrow \lfloor l / \text{mtSize} \rfloor$ 
114:  $j1 \leftarrow \lfloor \frac{q}{2} \rfloor$ 
115:  $j2 \leftarrow q + \lfloor \frac{\text{noMTs}-1-q}{2} \rfloor$ 
116:  $l1 \leftarrow j1 * \text{mtSize} + \lfloor \frac{l}{2} \rfloor$ 
117:  $l2 \leftarrow j2 * \text{mtSize} + l + \lfloor \frac{\text{mtSize}-l}{2} \rfloor$ 
118:  $\text{Perm}[l1] \leftrightarrow \text{Perm}[l2]$ 
119:  $\text{Perm}[l] \leftarrow m - 1$ 
120:  $m \leftarrow m - 1$ 
121: return  $m$ 
122: END
123: METHOD GetLocation $(l, x, \text{noMTs}, \text{orgNoMTs}, \text{mtSize}, \text{params})$ 
124: {Gets a location in  $\text{Perm}$  based on location index  $l$  in a small master table of size
 $\text{mtSize}$  and a value  $x$ , with  $0 \leq x < \text{orgNoMTs}$ }
125: END
126: METHOD ExpandKey $(K, k, \text{params})$ 
127: {Expands key element  $K$  into its constituent parts so they can be used directly. These

```

```

parts are the prime modulus used and the bits which represent all the atoms}
128: END
129: METHOD GetBit(w1, w2)
130: {Returns a uniformly distributed bit based on two random, but potentially not uni-
    formly distributed, words w1 and w2}
131: if w1 = w2 then
132:   return null
133: end if
134: if w1 < w2 then
135:   return 0
136: end if
137: return 1
138: END
139: METHOD GenerateRandomFromBits(Bits, t, l, params)
140: {Returns an almost uniformly distributed value between 0 and  $l - 1$  based on some
    uniformly distributed random bits found in the sequence Bits, starting at index t}
141: a ← GetReqGenerateBits(l, params)
142: x ← 0
143: for i = 0 to a - 1 do
144:   x ← x + Bits[t] *  $2^i$ 
145:   t ← t + 1
146: end for
147: step ←  $2^a/l$  {Noninteger value with double precision}
148: q ←  $\lfloor x/step \rfloor$ 
149: if r ≠ q and q < l and (q + 1) * step - x ≥ x - q * step then
150:   q ← q + 1
151: end if
152: return (q, t)
153: END
154: METHOD GenerateRandomBitsFromP(a, p, l)
155: {Returns the bits of an almost uniformly distributed value between 0 and  $2^a - 1$  based
    on some uniformly distributed value, between 0 and  $l - 1$ }
156: The method proceeds analogously to GenerateRandomFromBits.
157: END
158: METHOD GetReqGenerateBits(l, params)
159: {Returns the number of uniformly distributed random bits required to generate an
    almost uniformly distributed value between 0 and  $l - 1$ }
160: return bopf *  $\lceil \log(l) \rceil$ 
161: END
162: METHOD GetBits(val, noBits, Bits) {Adds noBits bits from val to a bit list}
163: METHOD BuildValue(Bits, t, noBits) {Builds a value from a bit list}
164: METHOD NextPrime(val) {Returns the smallest prime number ≥ val}

```

Discussion: The values 14, 28, 20, 11431, and 17 were chosen to be arbitrary beautiful constants ≥ 2 , which are used in such a way so as to have no impact on the security of the scheme. They can be replaced with anything else the reader finds more to his tastes, if desired.

SKREM proceeds as follows. It encrypts the two large keys $K1_large[]$ and $K2_large[]$ using the secret key $K_small[]$. This step is performed in order to allow for shorter keys in practice. SKREM then proceeds to encrypt the plain text, split in two via an OTP, into grand master tables $M1$ and $M2$. Each of these is, in effect, when taken

alone, truly random.

Actual encryption is performed in *BasicEncryptDecrypt()*. The security strength parameter - commonly identified with the key length in bits in most other ciphers - is *reqsec*. This must be taken such that 2^{reqsec} steps is beyond what is tractable by any adversary, within the intended lifetime of the cipher text. The security constraints put in place, and enforced via validation, are rather stringent. They are meant to be generously sufficient to allow us claim that SKREM can be proven to offer unbreakable security.

Encryption proceeds as follows. Firstly, a key extension stage takes place, where the number of available k -bit key elements is extended to $secrbpn * n$. Finally, each atom of the resulting last key elements is used to determine an unused location within the grandmaster table. The values at these locations are then altered to represent the plain text, encoded in the pair-wise relative order of consecutive such. A number of $secrbpn$ bits, represented by $secrbpn$ pairs of locations are XORed together to encode a single bit of P . When the existent bit does not correspond to the desired one, any of these pairs, randomly chosen, will have its values switched. Knowledge of the plain text could potentially be speculated only starting at the last stage, where actual encryption takes place. All transformations performed until then by SKREM are fully independent from it: the list of potential pairs of swappable locations from M is the same for any plain-text. As such, adaptive plain text attacks, as well as chosen cipher text attacks, should offer no noteworthy added benefit whatsoever, over simple known plain text attacks.

During the key extension round, the following occurs. Each old key element is processed by *ExtractJthLocation* (which uses some modular algebra to spice the result up a bit) to obtain a number of ppx locations between 0 and the size of a small master table. Each such location is used to reference $2 * f * reqBitsPerKey$ locations from the grand master table M , using *GetLocation* and the indirection vector *Perm*. Each successive pair of locations in M encode 1 (truly random) bit, with probability $(1 - 1/2^w)$, failing $1/2^w$ of the time, when the sampled values are equal. In order to ensure with reasonable probability that we obtain the required number of *reqBitsPerKey* from each key atom, the number of pairs of words is increased by a factor of f , which is taken to be close to the inverse of the former. The exact probability of failure for the entire process was not computed, but is instead left for further research.

Once *reqBitsPerKey* bits are generated for each of the new key elements (by all of the old key elements together), they are used to expand each such original key, into *keyExtFactor* (8) new ones. The bits used for extension, while generated by locations determined deterministically from original key K_small and the grand master table M , are understood to appear truly random and independent to the adversary: since M is truly random, unless we got astronomically unlucky for it to be something predictable (e.g. all 0s), the distribution of values across almost all of its permutations should be just as random. Do note, however, that the entropy (in terms of Shannon Entropy [7]) of the ensuing permutation of M will also never exceed the entropy of the secret key (which is large enough to exclude brute force guessing, fortunately).

The manner in which a single k -sized key element K is extended is as follows. At least 1 (potentially 2) k -sized, random values x, y , are generated based on the grand master table M , and on part of the key atoms from the other key elements, excluding itself. We expect each of these 8-16 values to be, effectively, indistinguishable from random by the adversary. The exact extension formula consists of the exponentiation of some primitive of $GF(p)$ to a random exponent based on the x, y values, thus producing a random result, regardless of the base used. The summing of two such exponentiations, as well as of the beautiful offset 11431 are meant to prevent a sequence of operations from simplifying to just one. Finally, the modular inverse operation was employed, given its apparent even-

today-valid strength with regard to cryptographic usages. It is known that the AES [2] scheme relies fundamentally on it. Given the OSINT available to the author, it seems the security of modular inversion holds in practice. We chose to work in $GF(p)$ (the Galois Field of order p) with varying, (almost) truly random primes p -s, of suitable length, rather than $GF(2^c)$ for some fixed c for three reasons. Firstly, we feel that the structure of $GF(2^c)$ might be the object of intense study and precomputations, particularly given its popular usage in cryptography. By choosing to work in a large number of simple fields, we preempt such possibilities. Secondly, the structure of $GF(p)$, simple as it may be, appeals more to us than the potential unknowns hidden by $GF(2^c)$. Thirdly, a primitive of $GF(p)$ is very be easy to find: any value ≥ 2 will do. The above formula produces a random result whenever x is random, regardless of the other values.

The method used for bit emission, by some key atom A , is designed to deny the attacker the possibility to guess more than a single one for any new key atoms, resulting from extension, by making some guesses about as many old key atoms as his computing power allows. We claim that no succinct characterization or useful property can be determined (except with astronomically low probability) by the adversary, with regard to the relationship between the set of new atoms and the set of old ones. This claim rests on the fact any such relationship will need to depend heavily on the existence of structure and order within M itself - which is, by the definition randomness involving constructive martingales, excluded (except with astronomically low probability).

Guessing all the *mtsize* bits of the location represented by an atom, could, under some pessimistic scenarios, potentially be used to determine one bit for all new atoms in an extension round. An attacker would then be able to take $\approx 2^{2*(mtsize-1)}$ random guesses for all the remaining bits (of which there are at least *mtsize* - 1 per atom - and this value could be increased as per the ideas in Section 3), in order to know with 1σ confidence ($\approx 68.5\%$) if he had guessed right (see [8]). Each time such an experiment is performed, the set of plausible initial guesses (2^{mtsize}) is reduced to the corresponding fraction (e.g. 0.3% for 3σ). We consider that the security check requiring *mtsize* to be no lesser than $\approx reqsec/3$ is sufficient from a theoretical perspective to allow for the desired level of voracity in our claims that SKREM offers security that can be formally proven to be unbreakable. Note that any speculated (guessed) property of any round's set of atoms will be impossible to verify before the very last stage, where actual plain text encryption occurs.

At the encryption stage, the attacker is faced with a set of $O(n)$ key elements, which, by the prior arguments, he should not be able to characterize in any useful manner. Each group of *secrbpn* key elements are used to encode a single bit of plain text. For *secrbpb* $\geq ppx$ (which is the case in SKREM), an intractable number of locations would need to be guessed, for the attacker to be able to constrain the last-round key elements universe to account for even just 1 bit of plain text. The attacker should, thus, be unable to deduce any useful properties of this set, allowing him to characterize the original key $K[]$ or discover some yet-unknown bit of plain text. In effect, the last round keys function like a practical OTP for the plain text (not an actual OTP since its entropy is limited to about *reqsec*). By introducing uncertainty about which pair of locations had its values changed in M , we feel we have made most attacks seeking a useful representation of the set of last round key elements, reduce to the counting problem #SKREM instead of merely the decision one. The counting problem is generally considered harder than the decision problem, as the difference in tractability between 2CNF-SAT and #2CNF-SAT illustrates [9].

Other, less crucial, decisions add some additional uncertainty over the entire execution of the algorithm. For example, the simple shuffling of just 2 locations used in

BurnLocation(), we expect to cause the indirection arrangement *Perm* to stray, before the start of the encryption stage, pretty significantly from anything which is very simple to describe. The usage of non-constant base in *ExtractJthLocation()*, based on effectively all the atoms encountered during the entire course of execution of the algorithm also complicates matters notably.

The method used by *GenerateRandomFromBits()* and *GenerateRandomBitsFromP()* to convert from one uniform probability distribution to another merits some attention. It involves dividing the source universe into even slots, each having an associated value from the target universe. This conceptual division entails non-integer, rational thresholds. Since each threshold is a rational number, it may fall between integers. Occasionally a value at a border between two slots is encountered. Since determining the true correspondent would require more precision than available, we simply employ some smart rounding. As such, each value in the target universe may have its distribution altered (increased or decreased) by inclusion / exclusion of a small part of it, situated around the two thresholds of the interval which represents it. Each such amounts to less than 0.5 (thanks to rounding), bringing the total to at most 1 value from the $\frac{2^{bopf * \log(l)}}{l} = 2^{(bopf-1) * \log(l)}$ representing it, thus making the suggested *bopf* = 2 overly generous. This method, along with the one in *GetBit()*, were discovered by the author in the context of this paper. Nevertheless, we strongly suspect the likes of Marius Zimand or Leonid Levin are also aware of them.

Decryption is identical to encryption, save the lack of need to switch values in the grand master tables. It warrants no separate discussion. Armed with the insights discussed above, we make the following claims.

2.2 Claims

Claim 1. *Any classical computer algorithm, using less running time than 2^{reqsec} , has a probability of less than a value below that corresponding to 1σ ($\approx 68.5\%$) of determining whether SKREM was used to encrypt some known, arbitrarily chosen plain text *P*.*

Discussion: Essentially, we claim that SKREM offers security *reqsec* against a classical computer. Note that this is a bit less than the size of the secret key, *K_small* (by no more than a polynomial factor). Also note, importantly, that the claim asserts an would-be attacker, not only is unable to characterize or determine the original key - or predict the unknown part (if any) of the plain text, but also that he is unable to discern the cipher text from random.

Claim 2. *Any quantum computer algorithm, using less running time than $2^{reqsec/2}$, has a probability of less than a value below that corresponding to 1σ ($\approx 68.5\%$) of determining whether SKREM was used to encrypt some known, arbitrarily chosen plain text *P*.*

Discussion: The reason for the reduction in the security strength against a quantum computer is obviously Grover's Algorithm [10]. If there was any doubt left, the discussion by Bernstein in [11] clarified this need. Although our key size is more than double *reqsec*, we choose to be conservative and consider the security strength of the secret to be just *reqsec*. Given Bennet et al [12], Grover's algorithm is asymptotically optimal for an arbitrary black box function - which is what we consider SKREM to be. Thus a reduction factor of 2 in the claimed strength should suffice. We, the author, consider humanity's current understanding of quantum physics incomplete and partially flawed. We believe it possible for physical phenomena to exist which allow computers, offering exponential speed-ups in finding feasible arguments to black boxes, to be built. This is adds on top of speed-ups possible under existing theory, such as taking refuge near the event horizon of a

black hole until a classical computer finishes breaking the encryption key, or sending the computer itself to a place where time flows relatively much faster. We believe that new encryption techniques will need to be developed, once this understanding is perfected. This highly speculative discussion is, however, outside the scope of the current paper.

In the “good old fashioned” practice of cryptology works, we provide no formal proofs of our claims. We do however claim that such formal proofs exist, which is, we believe, a far stronger assertion than simply saying they are true.

Claim 3. *There exists a formal proof of Claim 1.*

Claim 4. *There exists a formal proof of Claim 2.*

2.3 Performance Analysis of SKREM

Theorem 1. *The performance characteristics of SKREM are as follows:*

- Total number of random words required: $O(n * reqsec)$.
- Minimum size of secret key: $O(reqsec^2)$
- Space complexity: $O(n * reqsec)$.
- Running time complexity: $\tilde{O}(n * reqsec^7)$

Proof. The proof is included in Appendix A, as Lemma 1. □

Discussion: The running time is thus polynomial in the security strength parameter and the plain text size and linear in just the latter. The required minimum secret key size is only quadric in the security strength parameter, which is $O(1)$ with regard to plain text size. The extra space required, besides the output and randomness well, can be implemented to go as low as the $O(2^{mtsize} * 2 * f * reqBitsPerKey)$. This is no more than $2^{O(reqsec)}$, which is $O(1)$ with regard to the plain text size.

3 Further Strengthening Security

In this section, we present a few important ideas to further strengthen security. We consider these to hold significant value with regard to any SKREM-like cipher, including the simplified version presented in Section 4. Consider the following.

1. **Pepper the plain text with random changes, making use of error-correction codes.** First, transform the plain-text into a redundant form by encoding it using error-correction codes. We suggest using simple codes, like Reed-Solomon [13], and quadrupling the size of the plain text with error correction data. After this, the randomness well T can be used to randomly alter a large portion of the words from the transformed plain text. Using Reed-Solomon codes should allow for modifications in about 30% of the data and still ensure it can be deciphered correctly. Not knowing where exactly a modification has taken place complicates the ability of an adversary to exploit any potential known plaintext advantage.
2. **Permute the indirection list $Perm$ using a more sophisticated approach.** For example, a longer permutation, determined by the sequence of burned locations, could be applied instead one involving a single swap currently used. One idea for such is to choose a fixed permutation of large order and to raise it to a power determined by the burned location before every application to $Perm$.

3. **Perform Key Extensions / Refreshes between encryption of successive plain text bits.** Refreshing some key elements (having them generate 1 new element instead of 8) after encoding each bit of plain text could add further uncertainty.
4. **Encrypt the plain text and / or the master tables using an existing symmetric cipher like AES beforehand.** This is a relatively common idea to any cipher. In our context, it can be regarded as a mean to degrade the capacity of the adversary to know the actual plain text. Like with any new scheme, the security of SKREM and SKREM-like ciphers should not rely on the security of this additional cipher. AES encryption is regarded in this context as mere obfuscation. Nevertheless, the design of SKREM allows such additional ciphers to be applied to both the plain text and the master tables safely, without degradation of security.
5. **Use a Chaos Machine or a CSPRNG instead of, or along side, modular algebra.** Currently, the transformations performed by *ExtendKey()* are algebraically simple. However, SKREM could be altered to PUSH all emitted bits in a round into some Chaos Machine (see [6]). Whenever a certain value needs to be generated, its seed could simply be PULLED from this machine. If a full Chaos Machine is too slow, a different CSPRNG can be employed in lieu of it.
6. **Change the location sampling method to something more sophisticated.** Currently, once the first location has been extracted from an atom, a simple, predictable, linear probing method is employed. However, a more sophisticated approach could be taken in *GetLocation()* to translate l and x into an index in *Perm*. This could, for example, take the form of modular algebra again, combining the two in a manner similar to the expression in *ExtendKey()*. Alternatively, Chaos Machines, CSPRNGs or different one way function candidates (like the one in [1]) could be employed. Additionally, the size of one atom can be increased to something more suitable, irrespective of the size of a small master table, changing *ExtractJthLocation()* accordingly. The current approach was chosen firstly to prevent the arguments pertaining to the security of SKREM depend on the security of this additional enhancement, which we feel is besides the core of the proposed scheme. Secondly, we wanted a practically feasible version to be easy to describe and understand, thus prevent it from resorting to sophisticated hacks and optimizations.

Finally, the following salt generation idea presents significant value, especially when secure transmissions are envisioned (rather than mere secure storage or computations): Transmit some number of grand master tables M - say 2^8 . Among them, include the $M1$, $M2$ and $M3$ generated by *Encrypt()*. Each of these grand master tables will appear indistinguishable from truly random at any inspection (since indeed they are each actually truly random, taken apart) - for example when crossing a national border. The receiver will then try out all the $\binom{2^8}{3} \approx 2.7 * 10^6$ triplets and discover the correct plain-text, reasonably fast. For an attacker without knowledge of the secret key, discerning even if a single triplet is non-random is a very hard (claimed insurmountable) task. Performing this computationally expensive operation more than 2.7 million times should prove intractable.

4 Improving Performance

As powerful as SKREM is, its complexity (space in particular) - makes it impractical. In this section we present a simplified version, resulting from choosing non-default values for the security parameters, meant to drastically improve performance characteristics.

It comes at the cost of only a constant factor reduction in the voracity with which we assert our educated security claims 1 and 2 (although we expect finding a formal proof for them to be harder). For lack of better name, we call it SKREMS (Short Key Random Encryption Machine Simplified). It is intended to be just an example of how practically feasible SKREM-like ciphers can be derived. We aim to provide 100-year security against the top existing conventional supercomputer, as of 2019. Based on publicly available statistics, this constrains the value of the security strength parameter to about $\log(10^{30}) \approx 100$. Modifying SKREM to allow for defending against a quantum super-computer (thus requiring doubling this value) is possible, but it entails resorting to some refinements and hacks which are outside the scope of the current paper. Consider the following.

Algorithm 2. *Changes to SKREM entailed by the Encryption Scheme SKREMS.*

- 1: **STRUCT params** normal
- 2: $reqsec \leftarrow 100$; $dmod \leftarrow 0$; $vrly \leftarrow 0$; $mtsize \leftarrow 33$; $secrbase \leftarrow 4$; $secrtwo \leftarrow 0$;
 $secrbpb \leftarrow 20$; $secrbpb \leftarrow 5$; $secrbpb \leftarrow 6$; $ppx \leftarrow 2$; $w \leftarrow 8$; $bopf \leftarrow 1.1$
- 3: **STRUCT params** short : **params** normal
- 4: $dmod \leftarrow 1$; $secrtwo \leftarrow 1$; $ppx \leftarrow 5$;

Theorem 2. *The performance characteristics of SKREMS are the following.*

- Total number of random words required for each of M1, M2 and M3: $9030 * n$, $9030 * n$ and 2^{34} respectively.
- Minimum size of secret key: 231 bits.
- Space complexity: $18060 * n + 2^{34}$.
- Running time complexity: $O(n)$, using about $n * 130/7$ CPU modular algebra operations over 127 bit numbers and about $18060 * n + 2^{34}$ disk and main memory operations.

Proof. The proof is included in Appendix A, as Lemma 2. □

Discussion: Given that n is actually the number of bits, not bytes of plain text, the actual byte of storage per byte of plaintext is in fact $\approx 9030 * 8 \approx 71GB/MB$. Note that the secret key is just a factor of 2.31x larger than the theoretical minimum for this security strength. Also, note that it can be coded using about 42 alphanumeric characters. With some special training and a bit of practice such a key can be stored in a human brain.

As per the detailed discussion in the section below, based on performance data from [14] and [15], using two, commercially available, portable SSD drives of 2 TB each, one could encrypt about 28 MB (which is sufficient to contain about one minute of video, not to mention the full text of this paper) in about 1 day, using high-end, but still commodity hardware. The CPU running time can be reduced by a factor of $\approx 10^7$, using super-computer grade hardware, allowing the entire process to complete in about 3 minutes.

The above illustrates that practically feasible SKREM-like schemes exist. Ideas to further significantly improve performance also exist. Two, immediately obvious ones, involve using a hardcoded, precomputed list of primes of all required sizes (e.g. the first prime above $2^k + 11431 * x$ for increasing values of x) or performing algebraic operations in $GF(p^x)$ for some fixed small prime p (which we strongly dislike).

Finally, note that splitting the plain-text into blocks - one of the worst practices ever adopted in cryptography - is not required by SKREM-like ciphers: their complexities are linear in the plain text size. As such, problems associated with chaining methods (such as XTS [16]) are fully avoided. We strongly dislike having block ciphers be used to encrypt large plain-texts, as we feel chaining methods conceptually open up the output to code book attacks of various sorts, like the watermarking attack for CBC [17].

5 Performance considerations of SKREMS using commodity hardware

Using commercially available storage hardware, consisting of 6 interconnected 120 TB hard-disks, providing 720 TB of volume, which, as of 2019, would cost around \$80,000, one could encrypt up to 5 GB of plain-text, with a careful implementation. This is the equivalent the capacity of about 1 DVD.

On the other hand, using two commercially available, portable SSD drives, of 2 TB each, one could encrypt about 28 MB, which is sufficient to contain about one minute of video, not to mention the full text of this paper.

Consider the information in [14] - stating that about $150 \cdot 10^3$ modular exponentiations can be performed per second, using 1024 bit numbers. Based on it, we expect at least $1500 \cdot 10^3$ modular operations per second to be achievable, for 127 bit numbers. The total delay thus introduced for the 28 MB example above, using two cores, will be $< 28 \cdot 2^{23} \cdot 130/7 / (1500 \cdot 10^3) \approx 2908$ seconds, or ≈ 49 minutes. The disk operations, involving processing $\approx 71 \cdot 28 + 16$ GB on two SSDs in parallel, could take up 1 year if random reads are allowed. However, a careful trick exists which allows for only sequential external reads to be performed. The total running time for disk operations thus becomes $(28 \cdot 71 + 16) \cdot 2^{30} \cdot 62 \cdot 10^{-12} \approx 134$ seconds (based on data from [15]), which is a bit over 2 minutes. Additional memory operations should take between $(28 \cdot 71 + 16) \cdot 2^{30} \cdot 100 \cdot 10^{-9} \approx 216 \cdot 10^3$ (about 60 hours) and $1/25x$ times this value (about 2.4 hours) over two cores of commodity hardware (based also on data from [15]). Overall, we expect encryption/decryption of the 28 MB to complete within about a day, using high-end, but still commodity hardware. According to public statistics, using super-computer grade hardware available as of 2019 will decrease CPU running time by a factor of 10^7 , making the entire process complete in less than 3 minutes (running time dominated by disk accesses).

6 Hiding Multiple Plain-Texts In The Same Cipher Output

While SKREM-like schemes, including SKREMS consume $O(n)$ space, the constant factors are rather large. However, of the total space used, only a small portion, wf , is ever changed, namely $2 \cdot f \cdot n$ words in total. For SKREMS $wf \approx 0.03\%$, with only $\approx 2.08 \cdot n$ words changing from the total of $9030 \cdot n$. The fact $wf \ll 1$ can be used to encrypt more than one plain text into the same master table. Consider q plain texts, each of length n , to be encrypted in the same master table M . Let m denote the minimum size of a grand master table required for encryption of a single one. Decryption of each plain text will succeed so long as all the locations which are “touched” at encryption are left unmodified by the encryptions of the others.

Consider the following approach for achieving this desiderate, making use of a Universal Perfect Hashing (see [18]) scheme to avoid writing to forbidden areas. Two hashtables $H1$ and $H2$ are used, with $|M|$ available slots each. They associate to each of their slots a distinct location in M . After the encryption of the i -th plain-text, for all locations $\{loc\}$ touched during encryption, the values $\{(j, loc) | i < j < q\}$ are removed from $H1$, resulting in some up to j slots becoming unavailable. Similarly, all write-zone locations touched, $\{locw\}$ have the corresponding values $\{(j, locw) | i < j < q\}$ removed from $H2$. During key extension of the i -th plain-text, whenever a reference to a location loc in M is made, it is interpreted to mean location $H2[(i, loc)]$ instead. Furthermore, (i, loc) is removed from both $H1$ and $H2$ in $BurnLocation()$. During the encryption stage, a reference to loca-

tion *locw* is interpreted to mean location $H1[(i, locw)]$. Thereafter, $(i, locw)$ is similarly removed from both $H1$ and $H2$. This way, all q plain texts will write to disjoint zones, and they will all read from unaltered locations. If removal from $H2$ fails, a different random seed is retried for both hash tables and the process starts over (from the first plain-text). If a collision occurs for $H1$, it can be ignored: the slots must have already been invalidated by some prior plain-text. It is crucial however, that there be no collision at all between the values $(i, 0) \dots (i, m)$ in either $H1$ or $H2$, for any i . This can be checked before the first encryption starts. The probability of collision for $H2$ should be rather small, since only $q * 2 * f * n = O(q * n)$ locations are removed. Having $|M| > O(m + wf * q * m)$ should suffice to get the probability of successful encryption for all q plain texts large enough. We leave calculation of the actual factor by which M needs to be increased to allow, with high probability, for precisely 0 collisions to occur in $H2$, for further research. Note that the hash tables $H1$ and $H2$ need to be common to all q plain-texts. The length of the random seed, needed for the theory of Universal Perfect Hashing to work, consists of just a few RAM words (over no less than $\log(|M|)$ bits). As such, it can be encrypted alongside each pair of large secret keys using the respective plain text's K_small . The secret key size thus increases with only $\log(q)$ bits, required to describe the index (or ID) i of the each plain text, from among the q possible.

We believe that a moderately ingenious application of the above could be used to modify SKREMS to allow it to encrypt 10-100x larger volumes of data using the same amount of space. We consider employing the above as a means to ensure “plausible deniability”, of little practical value, both as a legal defense and as a defense against a torturous interrogator, considering that the cipher output, as per Claims 1 and 2 will appear random to an adversary anyway. Potentially, having it represent more than one plain text could play a role in informative intoxication operations involving double agents.

7 Usages In Lieu Of A One Way Function

Consider the transformation, defined using any SKREM-like scheme, including SKREM and SKREMS themselves, mapping a secret key K_small to a plain text P resulting from decryption of a fixed, a priori chosen, cipher output M . For lack of better name, we shall call it SKREMOW (SKREM One Way). It is defined as follows: $SKREMOW_M : \{0, 1\}^k \rightarrow \{0, 1\}^n$, where $n = |P|$, an arbitrary chosen size for the plain text (small enough to be encryptable in M), k is the size of the secret key $|K_small|$ and let $m = |M|$ be the size of the three grand master tables $M1$, $M2$ and $M3$ combined (chosen to be of the minimum size, required for the specific plain text size n). Note that for a fixed key size k (chosen large enough to satisfy the security requirements for a specific *reqsec* security strength parameter) and a fixed plain text length n , SKREMOW actually defines a family of transformations: one for each of the 2^m possible cipher outputs. Reversing a function in this family is claimed to be, with high probability, provably impossible for any adversary having less computing power than 2^{reqsec} (2^{100} for SKREMS). SKREMOW thus fits the security requirements for almost all practical applications of one way functions.

There are two aspects preventing us from being able to formally consider $SKREMOW_M$ a one-way function. Firstly, it is one way only with high probability: in cases where M is chosen in an astronomically unlucky manner (for example it is all 0-s), it can be that $SKREMOW_M^{-1}$ could be easily computed. Discerning if a certain M is suitable (the quality of its randomness is sufficient) is a different, not at all trivial, task in itself. Secondly, the hardness of computing its inverse is only with regard to an adversary with a fixed, constant amount of computing power available (namely 2^{reqsec}). For

a sufficiently large constant (2^{reqsec} suffices), a simple $O(m)$ (linear in input) algorithm exists computing $SKREMOW_M^{-1}$ for any M . Although such an algorithm is intractable in practice, for a constant security strength parameter, $SKREMOW^{-1}$ does not meet the complexity class requirements of a one way function. This second aspect could be remedied by taking $reqsec$ to be a $\Omega(n)$. However, this would make the key size for SKREMOW too large for some practical applications.

Also note that SKREMOW is not bijective. When n is too small, the same plain text P could be decrypted using several secret keys - making SKREMOW non-injective. When n is large enough, there will exist bit sequences from $\{0, 1\}^n$ which do not admit any key to decrypt them from the chosen, fixed cipher output M , thus making SKREMOW non-surjective. Nevertheless, taking n to be large enough, and restricting the codomain of SKREMOW to its image $Im_{SKREMOW_M}$, makes it become, with high probability, bijective.

The above construction suffices for almost all scenarios where one way functions are required in practice, even though SKREMOW is not, per se, a one such itself.

8 Practical Considerations

We include some remarks, pertaining to practical usages of SKREM-like encryption schemes, such as SKREMS. Do note that implementation and actual usage pattern can make all the difference between perfect, unbreakable security and no security at all, for any scheme. The following is not intended to be even an exhaustive enumeration of potential pitfalls related to SKREM specifically. Nevertheless, we recommend paying particularly close attention to the following, before anything else.

Firstly, consider truly random numbers. As cryptologists, we love them, we hate them, we need them - all at the same time. The security of SKREM relies on the high quality of the randomness of the grand master table M and of the randomness well T . These should be harnessed from nature, rather than merely pseudo-randomly generated from a short seed. We can suggest the following as potential sources of entropy: (i) measuring the value of a single qubit passed through a single Hadamard gate on any quantum computer (and repeating the experiment for the number of bits required), (ii) sampling the phase of inbound solar radiation; (iii) measuring the interval between successive alpha-particle emissions during radioactive decay of certain atoms (if one such radioactive source is available to the user); (iv) sampling mouse gestures provided by the human user using a mouse; and (v) sampling a compressed, large patch of random text, provided by the human user using the keyboard. The output from several such sources should be combined using a Kolmogorov extractor (see [5]). Finally, automated statistical tests should be performed both on the originally sampled bits and on the cipher output, to detect cases of obvious deficiencies in the quality of the randomness.

Note that SKREM-like schemes can be used to create encrypted volumes - that is they allow for the plain text to be accessed randomly and even changed on the fly, given the secret encryption key. When used in this scenario, once any cipher output has been revealed to an attacker, the volume should be considered read-only. Generally speaking, the grand master tables and the randomness well should never be reused once a cipher output generated using them has been revealed, just as with an OTP.

All randomness used in the scheme, including the discarded randomness well T and the original grand master table M , must be kept secret. Revealing the grand master table to an adversary marks the moment he can start to preprocess it. A simple side by side comparison of the original master table and the cipher output will reveal the locations used to encrypt the plain text. Thus, the original sources of entropy used, as well as their outputs, must not become available to the attacker.

We are aware of several types of attack against some practical usage scenarios and implementations of SKREM-like ciphers, which do not break the encryption scheme itself. We, the author, choose not to share them as part of this paper, as being outside its scope.

Of particular concern can be that we, the author, suspect that it is possible for the laws of physics concerning quantum phenomena to allow for a particular kind of quantum computer to be build which allows for exponential speedups in search over a the universe given by the preimage of a function (thus being exponentially faster than Grover's algorithm). We fear that, an implementation of such, might, someday, exist, which would allow for an arbitrary circuit with imposed output to be modeled using a quantum computer. Such a theoretical computer would then be run to determine its input. While processing, we fear it might be possible for all non-feasible states to essentially "fade out of existence" (e.g. have cumulative probability $< 1/3$ of describing a particular evolution of the quantum system). Thus, such a theoretical computer would be able to provide, with high probability, a feasible input to the arbitrary circuit in a single run (although the universe of possible inputs can be exponential in the size of the circuit). While existing quantum computer models cannot, to the best of our knowledge, come even close to such a feat, physical phenomena might exists which allow it. The quantum phenomenon called "path-of-minimum-energy", believed to be involved in photosynthesis, and harnessed by some for optimal route computation, is a particular inspiration for this concern of ours.

Finally, no encryption scheme, no matter how secure - even provably unbreakable - can protect against spyware installed on the devices where the encryption key is entered, or on the machines which perform actual encryption / decryption. Given that the we, the author, have used a personal, commercially available, not particularly secured, laptop to write this paper, we expects that there is a high probability some foreign intelligence services had already gained access to this research, by the time it was released by us.

9 Conclusions And Further Research

We have proposed two encryption schemes, which we claim offer the same level of security as the OTP in their strength parameter, while allowing for keys of constant size with regard to the plain text. For one of them, we further claim that formal proofs for these claims exist. Both schemes are claimed to make encrypted data indiscernible from random to any attacker having less computing power than 2^{reqsec} for some fixed, arbitrary parameter *reqsec*, thus achieving the main desiderate of encryption.

In the "good old fashioned" tradition of cryptography works, we offer \$17 to the first 9 people who present an argument falsifying Claims 3 or 4 and an additional \$14 to the first 28 people who present an attack falsifying Claims 1 or 2.

10 Acknowledgments

Warm thanks for their existence and patient understanding to the four beautiful persons who inspired the author to strive to make the world a better place for them, and furthermore served as an inspiration for some of the constants used in the algorithms. This paper would never have existed without them. These persons are: Anca Pană (now Anca Baliatti), Cătălina Ghiorghiuță (now), Tatyana Nevidima and Diana Paiu.

11 Authors

Mircea Digulescu is an independent computer science researcher, software engineer, entrepreneur, military and intelligence enthusiast and amateur writer.

He has a PhD (ABD) at the University of Bucharest, Department of Computer Science of Faculty of Mathematics and Computer Science, from which he also obtained both a Bachelors and a Master's degree in Computer Science. He gained recognition by being awarded medals and prizes at international contests and Olympiads in Computer Science, including 1 Bronze Medal at CEOI and 2 prizes at ACM SEERC. He is a Div1 coder on Codeforces.com. He also has over 15 years of experience in the software industry, creating systems that currently run in production and scale to billions of transactions.

His research interests include Cryptography, Game Theory, Complexity Theory as well as Algorithms and Data Structures.



A Proofs of the theorems

This appendix includes proofs of the theorems stated throughout the paper.

Lemma 1. *The performance characteristics of SKREM are as follows:*

- Total number of random words required: $O(n * reqsec)$.
- Minimum size of secret key: $O(reqsec^2)$
- Space complexity: $O(n * reqsec)$.
- Running time complexity: $\tilde{O}(n * reqsec^7)$

Proof. Consider the encryption using large keys first. Each key element has at most $k = (secrbase * (1 + secrtwo) + (ppx) * mtsize * bopf + secrbpp) * bopf + secrbpp$ bits. During key extension, precisely this many bits are required to be emitted for the each of the $8x$ new key elements, times $(1 + secrtwo)$ - the number of secret exponents per new key element. Finally, for each useful bit required, $scrbbpb$ bits need to be emitted. This brings the total to $[(secrbase * (1 + secrtwo) + (ppx) * mtsize * bopf + secrbpp) * bopf + secrbpp] * 8 * (1 + secrtwo) * scrbbpb$ required bits to be emitted by each old key element. For each such required bit, $2 * f$ words are consumed from M .

The total number of key expansions occurring is no more than the sum

$$(1 + 8 + 16 + \dots) = \sum_{x=1}^{\log(secrbpn*n)/\log(8)-1} 8^x$$

. The upper bound is given by the need to have $secrbpn * n$ last-round key elements, for representing the n bits of plain-text. This resolves to $(secrbpn * n)/7$.

During encryption, no key extensions occur. However, an extra $secrbpn * n$ bits are emitted. For each such bit, $2 * f$ locations in M are used. Finally, one time more the full

number of words consumed need to remain untouched at the end, doubling this value. The grand total number of locations used by a large key is thus: $[(\text{secrbase} * (1 + \text{secrtwo}) + (\text{ppx}) * \text{mtsize} * \text{bopf} + \text{secrbpp}) * \text{bopf} + \text{secrbpp}] * 8 * (1 + \text{secrtwo}) * \text{secrbpb} * (\text{secrbpn}) / 7 + (\text{secrbpn})] * 2 * f * n * 2$. All the values in this rather long expression are constants, except mtsize which is constrained by the security validations to be at $O(\text{reqsec})$. In terms of this parameter and n , the expression becomes $O([(O(1) + O(\text{reqsec}) + O(1)) * O(1) + O(1)] * O(1) + O(1)) * O(n)$. This resolves to $O(n * \text{reqsec})$, representing the total for the grand master table.

The required size of the large key is at most $O(\text{reqsec})/O(1)$ times k , which is $O(k) * O(\text{reqsec}) = O(\text{reqsec}) * O(\text{reqsec}) = O(\text{reqsec}^2)$. We can assume $n > \text{reqsec}^2$, since this is trivially true in practice. The length of the plain text representing the two large keys, encrypted by Encrypt is no more than $O(\text{reqsec}^2 * \text{reqsec}) = O(\text{reqsec}^3)$, when using params_short no more stringent than params_normal , as is the case for the default values in SKREM.

The total complexity in the number of words consumed is by SKREM for the master tables is thus $O(n * \text{reqsec}) + 2 * O(\text{reqsec}^3) = O(n * \text{reqsec}) + O(n * \text{reqsec}) = O(n * \text{reqsec})$.

The randomness well is polled to determine the OTP used for P , consuming $O(n)$ words. Another up to $2 * O(n) + 2 * O(\text{reqsec}^2)$ words are used to determine a value up to secrbpn which describes which locations in the grand master table to switch, per bit of plain-text encoded. The total required number of words for the randomness well is thus $O(n) + O(n) + O(\text{reqsec}^2) = O(n)$.

The size of the secret key needs to be no larger than the size of a large key, which was computed above to be $O(\text{reqsec}^2)$.

The operations performed in SKREM are either direct reads or writes from or to one of M or T , temporary storage of bits emitted for either key extension or expansion, some modular algebra on $O(k) = O(\text{reqsec})$ bit numbers and accessing the Perm indirection vector. Since no super linear data structures are employed, the total complexity is dominated by the amount of random words consumed, which is $O(n * \text{reqsec})$.

Time complexity is as follows.

SKREM essentially consists of key expansions and extensions, plus a number of grand master table accesses, only a constant factor times the total number of words consumed. The latter is $O(n * \text{reqsec})$. The total number of key extensions, as computed above is $O(n)$. The number of key expansions is less than the total number of words consumed, thus only $O(n * \text{reqsec})$. For all word accesses, at most a constant number of algebraic operations are performed, with numbers of at most $O(k) = O(\text{reqsec})$ bits. These are additions, subtractions, multiplications and division, modular exponentiation and modular inverse. Furthermore, for key extension, sampling a prime less than some value is performed. The maximum total number of sampled values is secrbpp which is less than reqsec , thus being $O(\text{reqsec})$. Additions and subtractions take $O(k)$, multiplications takes $O(k * \log(k))$ using Fast Fourier Transformation (FFT), while divisions take $O(k * \log^2(k))$ using recursive division due to Brunikel and Ziegler [19]. Fast exponentiation involves at most $O(\log(k))$ multiplications and divisions, taking $O(\log(k) * k * \log^2(k)) = O(k * \log^3(k))$. Modular inverse can be computed in $O(\log(k))$ using the Extended Euclid algorithm. Generating the first prime beyond or below some value can be achieved by brute force trial and error, given the high density of prime numbers (established thanks to the Prime Number Theorem). The number of trials required is logarithmic in the upper bound. The probabilistic Miller-Rabin primality testing algorithm [20] can be used to obtain a probable prime which can then be verified using AKS algorithm [21]. Because AKS has rather large - even if still polynomial - complexity, it is likely in practice this verification step will be skipped. Generation of a prime less than $2^{O(\text{reqsec})}$ bits takes no

more than $O(\log(2^{O(reqsec)})) = O(reqsec)$ applications of a prime testing algorithm on $O(k) = O(reqsec)$ bit numbers. If Rabin-Miller [20] is used with $O(reqsec)$ rounds, this adds up to $\tilde{O}(reqsec * reqsec * reqsec^2) = \tilde{O}(reqsec^4)$, when using FFT for multiplications. There are no more than $O(reqsec)$ rounds in total, thus a single prime number generation takes no more than $O(reqsec) * \tilde{O}(reqsec^4) = \tilde{O}(reqsec^5)$ in total. If AKS is used instead, this becomes $\tilde{O}(reqsec^7)$. The total running time complexity is thus no more than $O(n) * \tilde{O}(reqsec^7) + O(n * reqsec) * O(reqsec * \log(reqsec)^3) = \tilde{O}(n * reqsec^7) + O(n * reqsec^2 * \log^3(reqsec)) = \tilde{O}(n * reqsec^7)$. \square

Lemma 2. *The performance characteristics of SKREMS are the following.*

- Total number of random words required for each of M1, M2 and M3: $9030 * n$, $9030 * n$ and 2^{34} respectively.
- Minimum size of secret key: 231 bits.
- Space complexity: $18060 * n + 2^{34}$.
- Running time complexity: $O(n)$, using about $n * 130/7$ CPU modular algebra operations over 127 bit numbers and about $18060 * n + 2^{34}$ disk and main memory operations.

Proof. The total number of words used, expressed precisely in terms of the security parameters is given in the proof of Lemma 1.

Substituting the effective values used for the security parameters, results in getting $k < 127$ and the total number of words consumed for each of the grand master tables M1 and M2 to be $< 9030 * n$.

The minimum number of elements in each of the large keys $K1_Large$ and $K2_large$, required for them to satisfy security constraints is $k * 8 * (1 + secrtwo) * secrbpb = 127 * 8 * (1 + 0) * 5 = 5080$.

Thus, the total number of bits for a single large key is $5080 * k = 5080 * 127 = 645160$.

It is thus straightforward to note that the number of words complexity of encrypting the two of large keys, the using $dmod = 1$ is: $[(secrbase * (1 + secrtwo) + (ppx) * mtsize * bopf + secrbpb) * bopf + secrbpb] * [secrbpn] / ppx * (1 + two) * secrbpb + secrbpb * 2 * f * (2 * 655360) * 2 = [(4 * (1 + 1) + 5 * 33 * 1.1 + 20) * 1.1 + 20] * 6/5 * (1 + 1) * 5 + 6 * 2 * 1.04 * 2 * 645160 * 2$. This is $< 2^{34}$.

Each element of the $K_small[]$ private key will have $(secrbase * (1 + secrtwo) + (ppx) * mtsize * bopf + secrbpb) * bopf + secrbpb$ bits, which is $(4 * (1 + 1) + 5 * 33 * 1.1 + 20) * 1.1 + 20 = 231$ bits, with ceiling. Each such element represents 5 secret locations, over 33 bits each, bringing the secret size beyond $reqsec$. Thus, we can allow $K_small[]$ to contain only one element. Its total size is thus just 231 bits.

As with SKREM, SKREMS uses no data structures of super-linear size. The space consumed is thus expected to be only slightly above the $18060 * n + 2^{34}$ words consumed for the three master tables combined.

In terms of time complexity, SKREMS can be optimized to perform only a constant number of modular algebraic operations per key, by using memoization. Doing so, will bring the number of modular arithmetic operations to just a small constant factor (which we estimate to be 10) times the total number of keys ever in existence. The number of key expansions, was computed in the proof of Lemma 1 as $(secrbpn * n) / 7$, which, resolves to $n * 6/7$. An additional $n + 1$ key elements account for the initial and final stages, bringing the total to about $n * 13/7$, or $n * 130/7$ considering the chosen constant factor. The hardest operation performed is prime number identification, taking $O(k * \log^3(k))$

using Rabin-Miller [20]. However, given the small seed of only 20 bits, these could be all precomputed. The outstanding operations take at most $O(k * \log^2(k))$ each. The time consumed for performing modular algebra is added on top of the linear running time (with a constant factor close to 1) in the total number of touched words (which is close to the $18060 * n + 2^{34}$ computed above), consisting of both disk and memory operations. The total running time is thus $O(n)$. \square

References

- [1] Levin, L.A., "The tale of one-way functions", *Problems of Information Transmission*, **39(1)** (2003), 92-103.
- [2] Daemen, J. and Rijmen, V., "The block cipher Rijndael", *In International Conference on Smart Card Research and Advanced Applications*, **Springer** (1998), 277-284.
- [3] Rivest, R.L., Shamir, A. and Adleman, L., "A method for obtaining digital signatures and public-key cryptosystems", *Communications of the ACM*, **21(2)** (1978), 120-126.
- [4] Zimand, M., "On the topological size of sets of random strings", *Mathematical Logic Quarterly*, **32(6)** (1986), 81-88.
- [5] *Extracting the Kolmogorov complexity of strings and sequences from sources with limited independence*. Zimand, M. *arXiv preprint arXiv:0902.2141*, 2009.
- [6] Czyzewski, M.A., "Chaos Machine: Different Approach to the Application and Significance of Numbers", *IACR Cryptol. ePrint Arch.*, 2016, 468.
- [7] Shannon, C.E., "Prediction and entropy of printed English", *Bell system technical journal*, **30(1)** (1951), 50-64.
- [8] *Chebyshev's & Empirical rules*. CSUS. <https://www.csus.edu/indiv/s/seria/lecturenotes/chebyshev.htm..>
- [9] Valiant, L.G., "The complexity of enumeration and reliability problems", *SIAM Journal on Computing*, **8(3)** (1979), 410-421.
- [10] Grover, L.K., "A fast quantum mechanical algorithm for database search", *Proceedings, 28th Annual ACM Symposium on the Theory of Computing*, 1996, 212.
- [11] Bernstein, D.J., "May. Grover vs. mceliece", *International Workshop on Post-Quantum Cryptography*, **Springer** (2010), 73-80.
- [12] Bennett, C.H., Bernstein, E., Brassard, G. and Vazirani, U., "Strengths and weaknesses of quantum computing", *SIAM Journal on Computing*, **26(5)** (1997), 1510-1523.
- [13] Reed, I.S. and Solomon, G., "Polynomial codes over certain finite fields", *Journal of the society for industrial and applied mathematics*, **8(2)** (1960), 300-304.
- [14] Emmart, N., Zhengt, F. and Weems, C., "Faster Modular Exponentiation Using Double Precision Floating Point Arithmetic on the GPU", *IEEE 25th Symposium on Computer Arithmetic (ARITH)*, 2018, 130-137.
- [15] *Interactive Latency - Latency Numbers Every Programmer Should Know (for 2019)*. Berkley., 2019.
- [16] Martin, L., "XTS: A mode of AES for encrypting hard disks", *IEEE Security & Privacy*, **8(3)** (2010), 68-69.
- [17] *Redundancy, the Watermarking Attack and its Countermeasures*. Markus Gattol. https://www.markus-gattol.name/ws/dm-crypt_luks.html, 2015.
- [18] Dietzfelbinger, M., Karlin, A., Mehlhorn, K., Meyer Auf Der Heide, F., Rohnert, H. and Tarjan, R.E., "Dynamic perfect hashing: Upper and lower bounds", *SIAM Journal on Computing*, **23(4)** (1994), 738-761.
- [19] Hasselström, K., "Fast division of large integers", *Department of Numerical Analysis and Computer Science, Royal Institute of Technology*, 2003.
- [20] Rabin, M.O., "Probabilistic algorithm for testing primality", *Journal of number theory*, **12(1)** (1980), 128-138.
- [21] Lenstra Jr, H.W. and Pomerance, C., "Primality testing with Gaussian periods", *FSTTCS*, 2002, 1.

ELECTROMAGNETIC ANALYSIS OF AN ULTRA-LIGHTWEIGHT CIPHER: PRESENT

Nilupulee A. Gunathilake¹, Ahmed Al-Dubai²,
William J. Buchanan¹, Owen Lo¹

¹Blockpass ID Lab, School of Computing, Edinburgh Napier University, UK

²School of Computing, Edinburgh Napier University, UK

ABSTRACT

Side-channel attacks are an unpredictable risk factor in cryptography. Therefore, continuous observations of physical leakages are essential to minimise vulnerabilities associated with cryptographic functions. Lightweight cryptography is a novel approach in progress towards internet-of-things (IoT) security. Thus, it would provide sufficient data and privacy protection in such a constrained ecosystem. IoT devices are resource-limited in terms of data rates (in kbps), power maintainability (battery) as well as hardware and software footprints (physical size, internal memory, RAM/ROM). Due to the difficulty in handling conventional cryptographic algorithms, lightweight ciphers consist of small key sizes, block sizes and few operational rounds. Unlike in the past, affordability to perform side-channel attacks using inexpensive electronic circuitries is becoming a reality. Hence, cryptanalysis of physical leakage in these emerging ciphers is crucial. Among existing studies, power analysis seems to have enough attention in research, whereas other aspects such as electromagnetic, timing, cache and optical attacks continue to be appropriately evaluated to play a role in forensic analysis.

As a result, we started analysing electromagnetic emission leakage of an ultra-lightweight block cipher, PRESENT. According to the literature, PRESENT promises to be adequate for IoT devices, and there still seems not to exist any work regarding correlation electromagnetic analysis (CEMA) of it. Firstly, we conducted simple electromagnetic analysis in both time and frequency domains and then proceeded towards CEMA attack modelling. This paper provides a summary of the related literature (IoT, lightweight cryptography, side-channel attacks and EMA), our methodology, current outcomes and future plans for the optimised results.

KEYWORDS

Side-channel attacks, electromagnetic analysis, lightweight cryptography, PRESENT and IoT.

1. INTRODUCTION

Internet of things (IoT) is a wide-spread infrastructure that consists of millions of connected devices. The main purpose of the IoT is to produce useful insights through data analytics. This causes collection, process and distribution of information continuously privately and publicly. Because of that, avoidance of data breaches is challenging. A summary of overall IoT communication strategies, technologies and challenges is accessible in [1] and [2]. IoT devices are constrained in terms of physical size, memory allocation and data rates (kbps) because they target sensor-based applications, *i.e., wearables to monitor health features, fault detection in factory automation, vehicular communications with real-time insights, etc.* Therefore, conventional cryptographic algorithms such as Advanced Encryption Standard (AES) used in general computing devices are impractical to function well on the IoT platform, due to long block

sizes and key lengths that would require high-end processors. As a solution, a concept towards light versions of the ciphers named lightweight cryptography was introduced [3 - 4].

One of our recent publications [5] covers a complete survey of lightweight cryptography. It includes history, categories of all existing lightweight ciphers (symmetric, asymmetric and hash) up to today, along with their parametric values. Furthermore, it critically analyses cryptologic as well as cryptanalysis studies available in the literature. Among hundreds of recommended lightweight cryptographic algorithms and their versions, some of the well-known ciphers in the field are included in Table 1 compared to the AES.

Table 1. Lightweight ciphers versus AES

Cipher	Architecture	Block Size (b)	Key Size (b)	Rounds	S-box (b)
AES	SPN	128	128/192/256	10/12/14	8
KLEIN	SPN	64	64/80/96	12/16/20	4
PRESENT	SPN	64	80/128	31	4
Fantomas	LS-design	128	128	12	5
LBlock	Feistel Network	64	80	32	4
Piccolo	Feistel Network	64	80/128	25/31	4
PRINCE	FX construction	64	128	12	4
Simon	Feistel Network	64	96/128	42	-
Speck	Feistel Network	64	96	26	-
LED	SPN	64	64/128	32/48	4
SPN: Substitution Permutation Network, LS: Linear diffusion and non-linear S-box					

Side-channel attacks are external phenomena that create an environment for unauthorised third parties to investigate sensitive information leakage through physical parameters such as power consumption, thermal radiation, time duration, cache files and optical changes when cryptographic functions are running on hardware. An attacker can approach the tasks via invasive and non-invasive ways. Invasive methods involve de-packaging the chip to get the connectivity to its inside elements, *i.e.*, *access to a data bus to monitor data transfers, etc.*, whereas non-invasive trials investigate externally available details only, *i.e.*, *energy drainage, electromagnetic (EM) emission, etc.* The major areas in side-channel attacks are, but not limited to:

- **Probing attack:** Direct observations of the internal parameters of the device.
- **Cache attack:** Cache access is monitored in a shared physical system [6, 7], *i.e.*, *virtual machines, cloud, etc.*
- **Data remanence attack:** Sensitive data is recycled after deletion, *i.e.*, *coldboot attack* [8].
- **Timing attack:** Running time is monitored.
- **Acoustic attack:** Sound generated is concerned [9].
- **Optical attack:** The surrounding of the device is visualised to see any indication using high resolution cameras [10].
- **Fault analysis attack:** Clock, temperature, voltage, radiation, light and Eddy current¹ are measured.
- **Power analysis attack:** Simple, differential and correlation power analysis (SPA, DPA, CPA) [11 - 12].
- **EM analysis attack:** Measures of EM field around the device.

¹ Loops of electrical current induced within conductors by a changing magnetic field in the conductor according to Faraday's law of induction

Preliminary analysis of those scenarios is essential to prevent physical leakages for guaranteed security of a cipher. General countermeasures that could be taken to mitigate weaknesses in physical security are:

- Cryptographic operation obfuscating enabled firmware [13].
- Randomisation of the operation sequences and or lookup tables in ciphers.
- Access of critical data using pointers instead of values when data structures are chosen.
- Asynchronism of the clock in the chip with respect to the cryptographic functions.
- Design of mathematical models in a manner that the leakage is misleading.
- Application of masking technique where appropriate (research [14] and [15] prove that masking would be insufficient against EM attacks).
- EM shielding via suitable materials when the chip/item is manufactured, *i.e.*, inclusion of Faraday cages, etc.
- Excessive noise addition to hiding leakages in EM radiation.

Since lightweight cryptography is still emerging in academia, the main focus can be seen in algorithm optimisation in cryptanalysis. Among the few studies, power analysis takes the highest percentage, *i.e.*, correlation power analysis (CPA) of PRESENT [12], Fantomas, LBlock, Piccolo, PRINCE, Simon and Speck [16], differential power analysis (DPA) of PRESENT [17], Simon and LED [18]. In EM analysis (EMA), the study [19] shows their results for a differential EMA (DEMA) of PRESENT, the studies [20] and [21] of a correlation EMA (CEMA) of PRINCE and Twine respectively. Yet any research outcome regarding CEMA of PRESENT seems to be unavailable in the literature. Other types of attacks on these novel ciphers remain to be considered in academia as well.

1.1. Our Contribution

In comparison, CEMA tends to offer more efficacy than DEMA theoretically. However, results may vary practically. Due to the fact that the unavailability of a study on CEMA of PRESENT, which is the most promising lightweight block cipher to include in IoT in the near future, we thought of filling a research gap by conducting a CEMA of PRESENT against firmware robustness. Here, performances from simple analysis in both time (SEMA) and frequency (SFEMA) domains to correlation white-box attack modelling were analysed. The work is still ongoing for improved results and evaluation.

The major contents of this paper are:

- A comprehensive overview about EMA including types of attack models and related mathematical equations.
- In-depth details about our attack model including code snippets, so it would help freshers in the field to begin with their own experimental setups.
- Current results of our SEMA, SFEMA and CEMA attack models.
- Discussion over the observations and real-world scenarios.
- Up-coming plans for optimising the attack model.

2. ELECTROMAGNETIC ANALYSIS OF PRESENT

2.1. Electromagnetic Analysis

Although EMA seemed to be conducted in the time domain in the past, an increased interest can be seen in the frequency domain recently. However, methods need to be well designed accordingly as just domain conversion via fast Fourier transform (FFT) would not create an accurate attack model. Our literature findings conclude an existing model classification as in Fig.1. There are two mathematical approaches used in these analyses based on Hamming calculations:

- **Hamming distance (HD) method:** XOR operation between two binary values. In a binary number, bit by bit operation is reflected here.

$$W = HD(D, R) = a.HW(D \oplus R) + b \tag{1}$$

where,

W - Hypothesised value

D - Intermediate value

R - Reference state value

a - Gain

b - noise

- **Hamming weight (HW) method:** Non-zero elements of a binary number where the reference state is zero. For example, if a number is 10010101, then the Hamming weight would be 4.

$$W = a.HW(D) + b \tag{2}$$

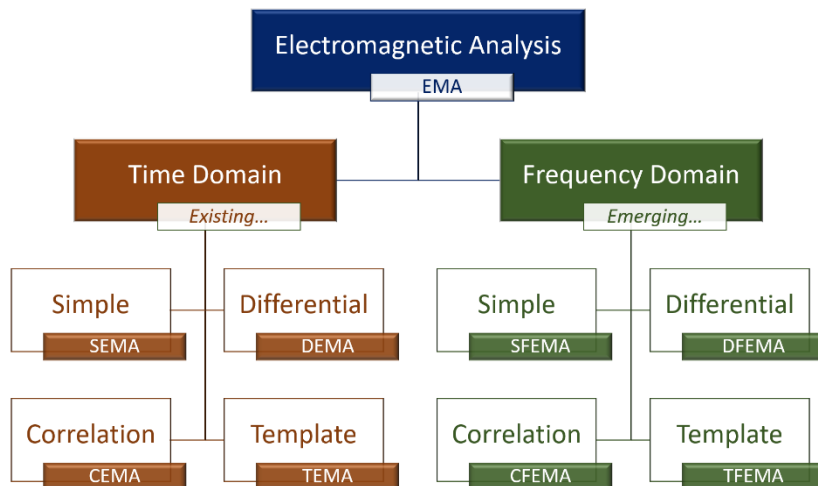


Figure 1. General EMA classification

SEMA and SFEMA are visual inspections of EM traces recorded using an oscilloscope or a software-defined radio (SDR). It helps identify where the leakage occurs, but it does not involve extracting secret information such as the encryption key to breaking into the secret data. On the other hand, locating the exact operation of the cipher and guessing the key via clock information followed by brute-force analysis are possible sometimes, if thorough knowledge of the device is there [22].

DEMA and DFEMA [19, 23] are complex statistical models which do not require much information about the device. These derived from the differential power analysis (DPA) method introduced by [24]. The analysis continues with guesses while corresponding traces are divided into two groups based on the guess for a bit to be 0 or 1. Then all traces of each group added together and averaged. Finally, the difference between the averages of group 1 and group 0 is calculated. If the trace alignment is correct, a considerable amount of spikes will illustrate supportive information for the derivation of the key.

CEMA and CFEMA [20, 21, 25] are efficient versions of DEMA and DFEMA respectively where grouping is not required. This also does not need knowledge of the device. The model focuses on several bits at a time. The analysis is based on the correlation of either the HD or HW method. It offers a hypothetical intermediate value that would indicate the possibility of the attack. The correlation between the EM emanation and hypothesised intermediate value could be calculated using the equation 3.

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}} \quad (3)$$

where,

$$Cov(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$Var(Y) = E(Y^2) - (E(Y))^2$$

$E(X)$ - Mean of X

$E(Y)$ - Mean of Y

TEMA and TFEMA [15] require a complete copy of the device with full control. Then via pre-processing using a large number of EM traces, a template is created. It further needs capturing a small number of traces from the victim's side to complete the attack.

2.2. PRESENT Block Cipher

PRESENT is a lightweight cryptographic algorithm that is suitable in ultra-lightweight² conditions too. It is a block cipher designed by the authors of [26] in 2007. It has been approved by several international standardisation authorities, the International Organisation for Standardisation (ISO), the International Electrotechnical Commission (IEC) and the National Institute of Standards and Technology (NIST).

It is based on a SPN architecture with 64b block, 80b key and 31 rounds. It operates as in Fig. 2, offering moderate security. It targets hardware optimisation by having a tiny footprint of 1570 gate equivalent (GE) and a low power consumption of 5 μ W over 32 clock cycles. It has the smallest substitution box (S-box), which is 4b-4b equivalent to 28 GE as well as the simplest permutation (pLayer) resulting in 0 GE. The S-box mapping as in Table 2. Moreover, this cipher has another version with a 128b key that would require 1886 GE. The authors mention that the PRESENT is more prone to side-channel and invasive hardware attacks.

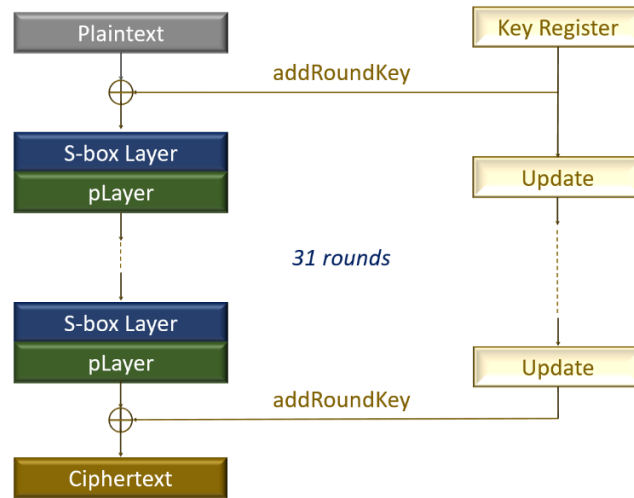


Figure 2. Operational process of PRESENT cipher

Table 2. S-box mapping of PRESENT (in hexadecimal)

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$S(x)$	C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

In S-box, the output of the addRound key step, which is a 64b block is split into sixteen 4b blocks. Then each is fetched through the S-box layer simultaneously to look up against the mapping values. After, the block value is replaced by the resultant values.

²Correspondence only to specific areas of the algorithm, *i.e.*, selective micro-controllers (μC), selected cipher sections, etc.

2.3. Methodology

Near field (NF) EM compatibility (EMC) probes are used to capture EM radiation around an embedded device. According to Faraday's law, changes of the magnetic flux in a magnetic field generate a voltage in the probe's loop (equation 4). Induced voltage fluctuations in the electric signal would give an idea about the EM emanation of a device. Besides, probes with smaller loops have higher frequency resolution, but less sensitive in the acquisition. The waveforms can be analysed using an oscilloscope or a software-defined radio (SDR). Here, the 80b key version of PRESENT was chosen due to the expectation of application in IoT devices. The resources used for the experimental setup are in Table 3.

(4)

$$V = 2\pi BA$$

where,

V - Voltage

π - Pi constant, 22/7

B - Average magnetic field

A - Area perpendicular to the magnetic field

Table 3. Resources used for the testbed

Hardware	
Oscilloscope	KEYSIGHT InfiniiVision MSOX4101A (1GHz, 5GSa/s)
EM emanation acquisition	TekBox EMC NF probe set – H20, H10, H5, E5 (9kHz – 6GHz)
Trigger connectivity	KEYSIGHT passive probe N2894A 700MHz
Pre-amplifier	TekBox 40dB wide-band amplifier
Encryption device	Arduino UNO R3
Instrument control	Microsoft Surface Laptop with an i5 processor
Software	
Encryption	Arduino IDE
Data processing	Matlab R2020b

The hardware connectivity of the testbed is as in Fig.3. The probe was initially placed on the top of the chip in a manner that the angle becomes 90^0 to the chip, because a study [25] mentions that it is the optimum position to acquire the highest EM radiation. However, different positions and angles are expected to take into consideration in the future. A trigger signal was used to locate the S-box function (Fig.4). The trigger function was monitored using a separate channel of the oscilloscope. Consequently, the maximum possible sampling rate became 2.5GSa/s for EM waveforms. The LED of the Arduino UNO board (pin 13) was connected to the trigger channel, and the encryption code was made in a way that the pin gets high (LED on) when the operation starts and then gets low (LED off) after the completion. In addition, primary level precautions were taken to reduce ambient and system noises such as:

- Waveforms were captured in average mode by using five encryption cycles per trace.
- Aluminium foil was used to cover the setup as an EM shield.
- The computer was operated in flight mode.

The accuracy of trace capturing and reconstruction using Matlab was verified by known test data values. The Arduino code for encryption was derived from [27] and had been previously validated using test vectors for [12].

2.3.1. Attack Modelling

Our model targets the S-box implementation because a successful attempt would cause key extraction straightaway. Due to its non-linear behaviour, significant differences may appear in correlation calculations. The encryption key used was **AC DE FB 21 F9 23 43 75 C0 E6**. The code snippet for the output implementation of the S-box in Arduino is shown below.

```

for (int i = 0; i < 8; i++)
{
state [i] = sBox [state [i] >> 4] << 4 | sBox [state [i] & 0xF];
}

```

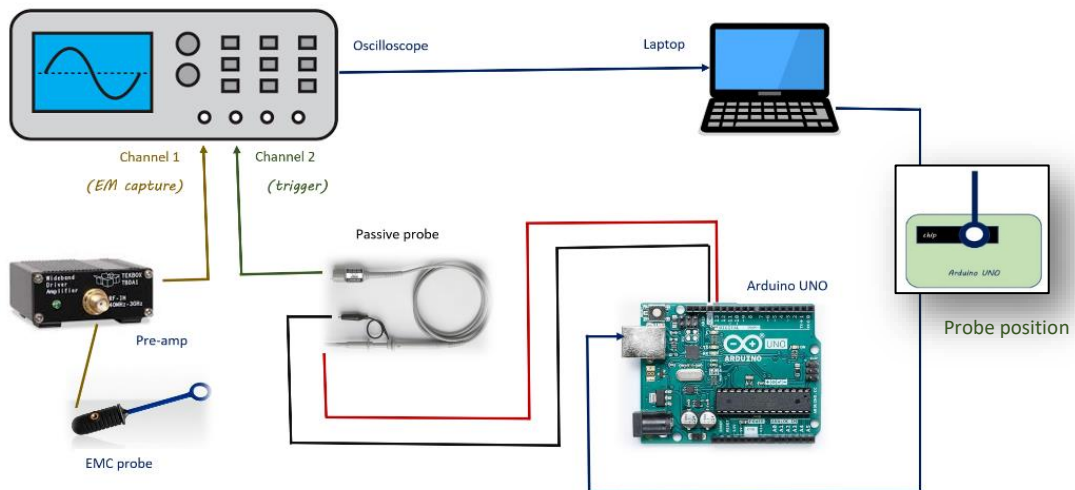


Figure 3. Hardware connectivity of the experimental setup

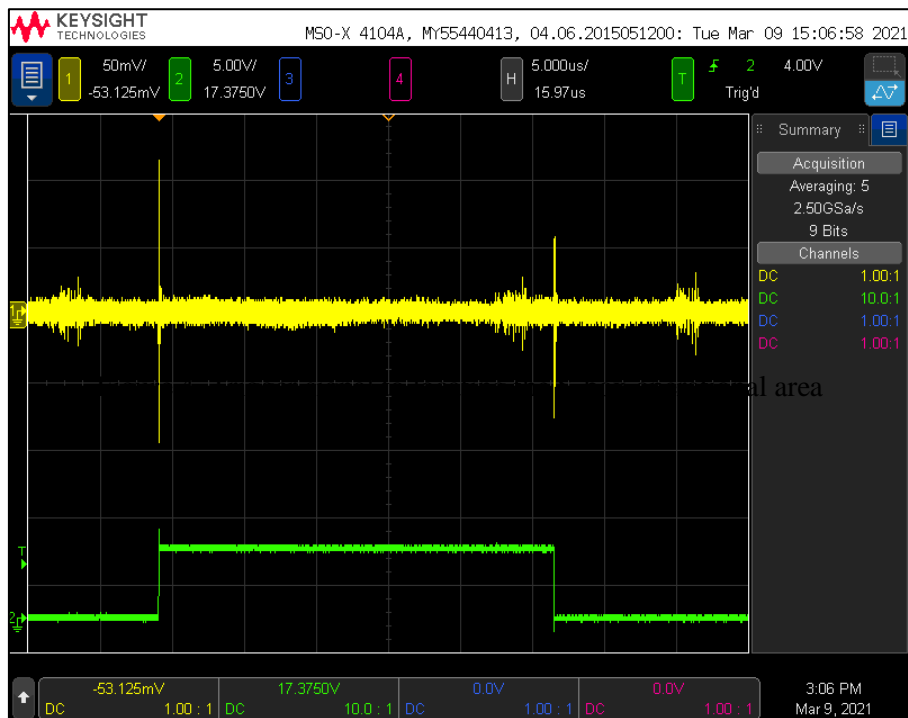


Figure 4. Trigger signal to locating the S-box operational area

The first part of the state $[i]$, $sBox [state [i] \gg 4] \ll 4$ looks up against the first four most significant bits (MSB) where $sBox [state [i] \& 0xF]$ does against the least significant bits (LSB). Next, the bitwise OR logic operator, $|$ combines the two 4b numbers into a single 1B value. The states from 0-7 are the outputs after each S-box substitution.

The plaintext values start to increment from 0-255 in each bit (00 00 00 00 00 00 00 00 – FF FF FF FF FF FF FF FF) once a command is received via Arduino UNO's serial port from a Matlab script. Then, each waveform for each plaintext was captured and saved for post-processing. Thus, the total number of traces collected was 256, and each trace originally contained 16000 data

points. When the signal was trimmed later on Matlab by extracting related values into a new matrix to filter the S-box functional area, the useable number of data points were reduced to 8800.

In CEMA modelling, the choice of method was HW because of its efficacy. The accuracy of the model was validated using the test results for [8]. The steps of the procedure:

- Firstly, hypothesised values considering both key and plaintext values from 0-255 were calculated using HW.
- Meanwhile, the actual EM values that had been recorded in 256 rows (0-256 plaintexts) for each data point (total of 8800) were compared to its correspondent hypothesised array of 256 items for each key to get the correlation coefficient values (ρ) using the equation 3.
- Finally, the maximum ρ value was taken as the result for each data point.

The highest value in ρ throughout all data points should indicate possible leakages of the key Bytes. The pseudo-code for the model is shown below.

```

for k = 0:255 (Key values),
    for p = 0:255 (Plaintext values),
        set output of AddRound key step as the input to the S-box
        lookup MSB of the input (4b)
        lookup LSB of the input (4b)
        combination of the MSB and LSB (1B)
        HW calculation and saving
    For x = 1:length of data (8800 data points for 256 waveforms),
         $\rho$  calculation between arrays of actual values and HW array
        if empty,
            save key value
            save  $\rho$  value
        else if  $\rho >$  previous one,
            save key value
            update  $\rho$  value
    plot data points vs.  $\rho$ 

```

3. RESULTS

Currently, the results are available only for the H20 and H10 probes' measurements.

3.1. SEMA

Fig.5 and Fig.6 show EM emission in the time domain (time versus voltage) for the S-box function for the H20 and H10 probes, respectively. Fig.7 and Fig.8 show data distribution over the operation in histogram plots for each case. Subplots in each figure illustrate the difference between encryption and non-encryption scenarios.

3.2. SFEMA

Fig.9 and Fig.10 show EM emission in the frequency domain (frequency versus amplitude) for the S-box function for the H20 and H10 probes, respectively. Fig.11 and Fig.12 show changes in

frequency strengths over the operation in spectrogram plots for each case. Subplots in each figure visualise the difference between encryption and non-encryption scenarios.

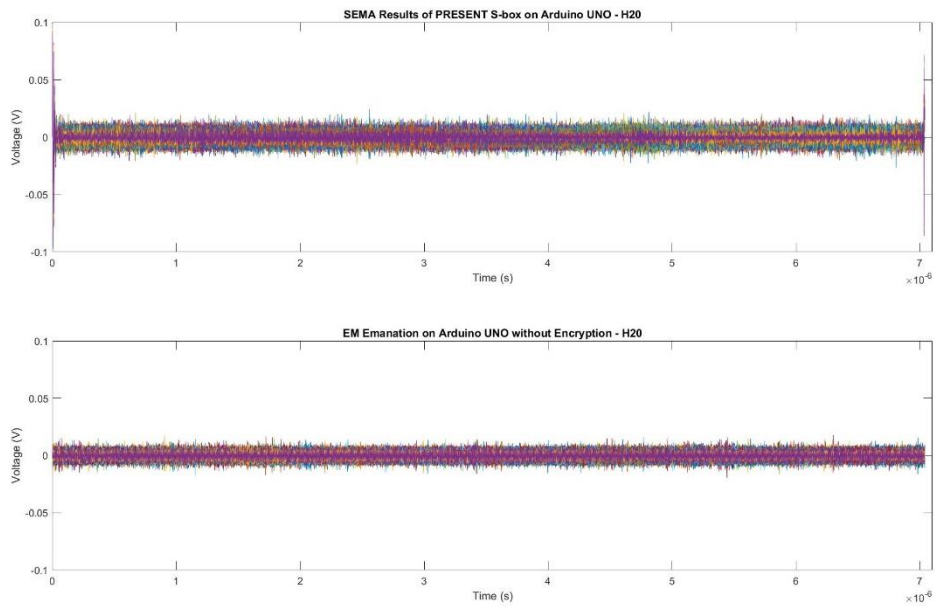


Figure 5. EM emanation in the time domain - H20 probe

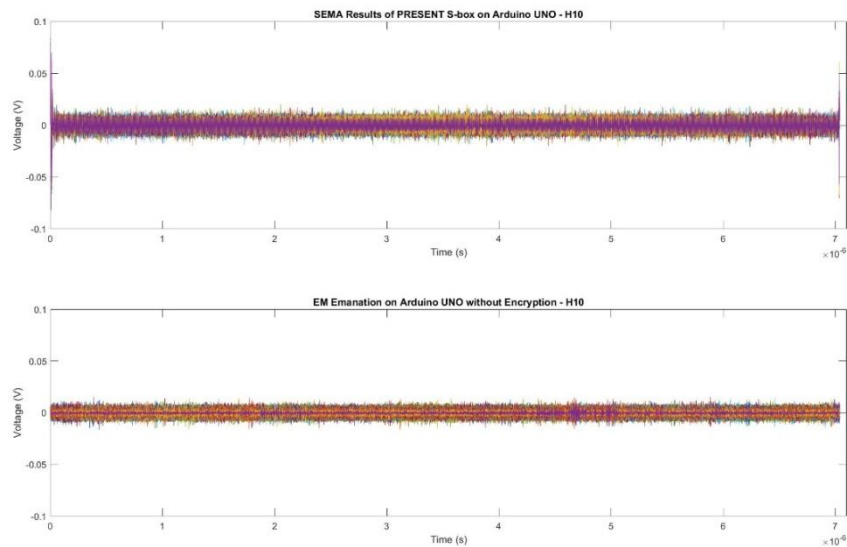


Figure 6. EM emanation in the time domain - H10 probe

3.3. CEMA

Fig.13 and Fig.14 illustrate the outcomes of the CEMA attacks performed using the H20 and H10 probes, respectively. Guesses for key Bytes were made by checking on notable peaks and troughs

of the graph. Correspondent results for each setup are summarised in Table 4 and Table 5. Due to the repetition values in the troughs in both cases, non-encryption data values were fetched through the model to check on possible false-positive appearances because of the system noise. The results are as in Table 6 and Fig.15.

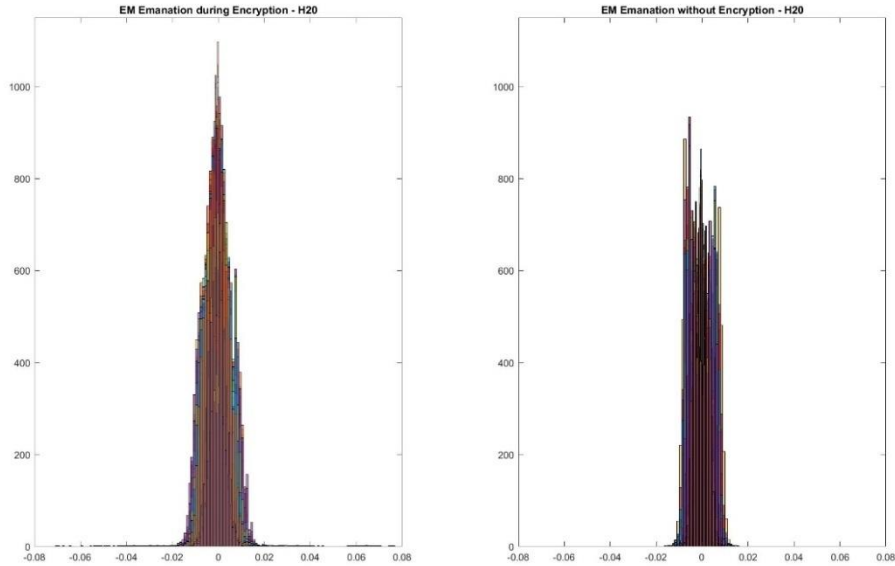


Figure 7. Histogram of the H20 probe data for the encryption and non-encryption states

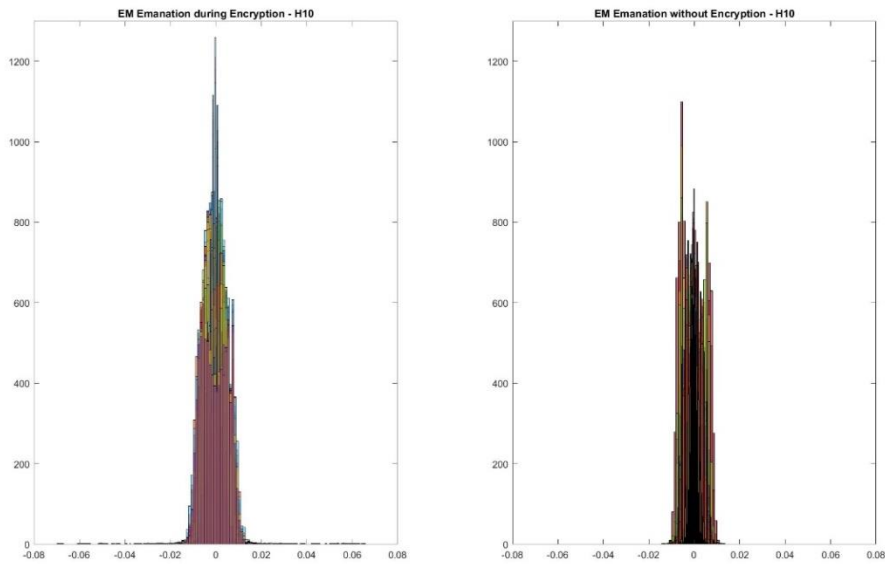


Figure 8. Histogram of the H10 probe data for the encryption and non-encryption states

Table 4. Key guess for CEMA - H20 probe

Type	No.	Key Byte Guesses												
Peaks	12	C0	B2	9E	40	1A	02	90	4F	D2	10	4D	D6	-
Troughs	13	90	99	9C	90	90	99	99	90	90	90	99	91	72

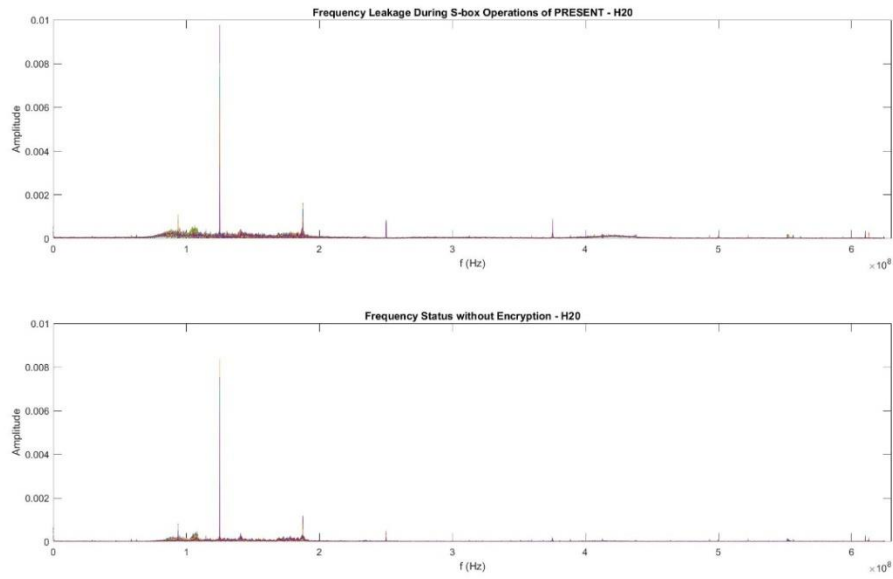


Figure 9. EM emanation in the frequency domain - H20 probe

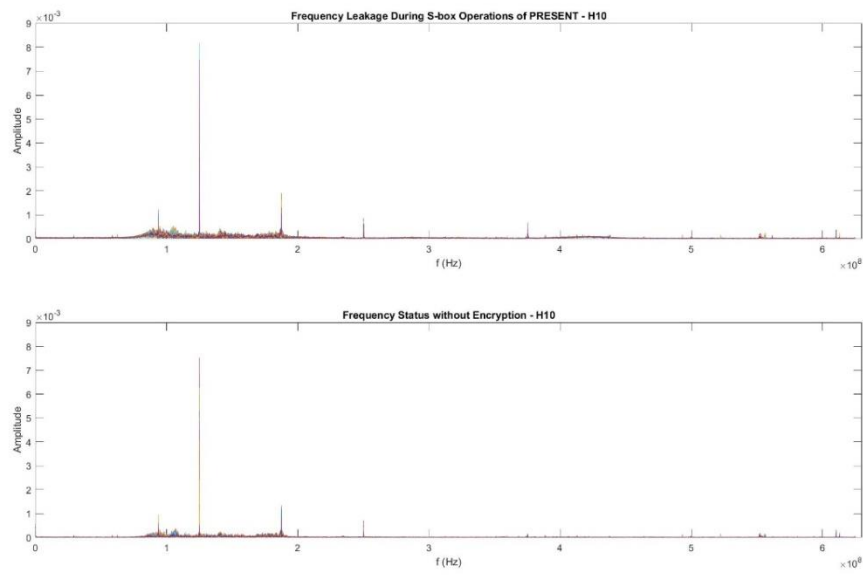


Figure 10. EM emanation in the frequency domain - H10 probe

Table 5. Key guess for CEMA – H10 probe

Type	No.	Key Byte Guesses										
Peaks	10	E4	DF	88	48	C4	D8	E6	49	54	44	-
Troughs	11	56	56	56	E5	56	35	56	56	F6	51	56

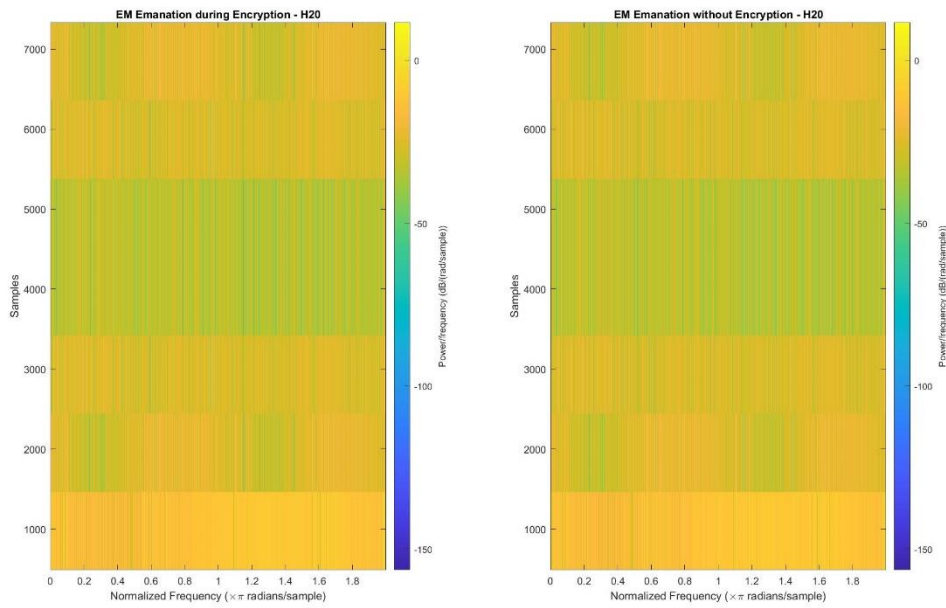


Figure 11. Spectrogram of the H20 probe data for the encryption and non-encryption states

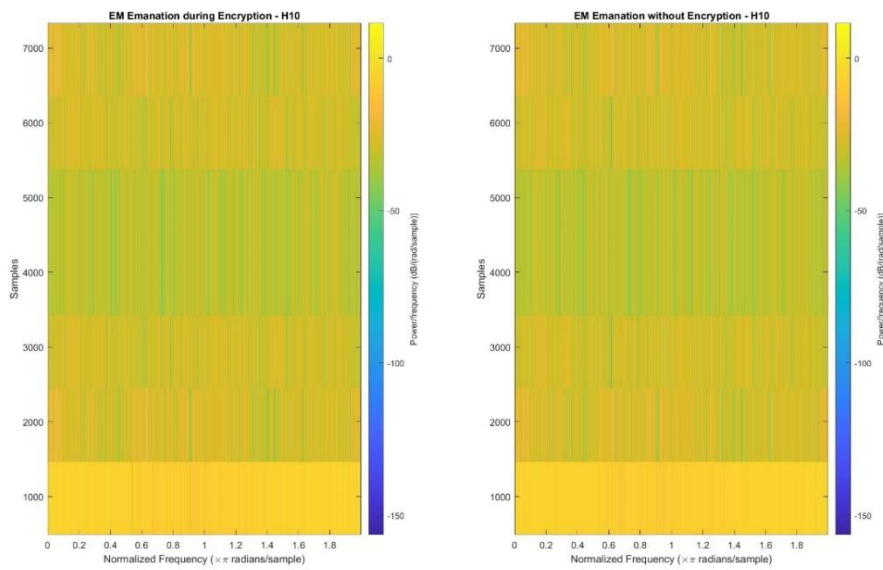


Figure 12. Spectrogram of the H20 probe data for the encryption and non-encryption states

Table 6. False positive appearance check

Type	No.	Pattern of Repetition											
H20													
Peaks	-	-	-	-	-	-	-	-	-	-	-	-	-
Troughs	12	C1	C9	CD	CD	CD	C3	C3	CD	C3	C3	C3	CD
H10													
Peaks	7	1A	2D	2B	CA	99	2B	49	-	-	-	-	-
Troughs	12	11	1F	C5	3A	C6	58	3A	AE	3D	31	C9	C9

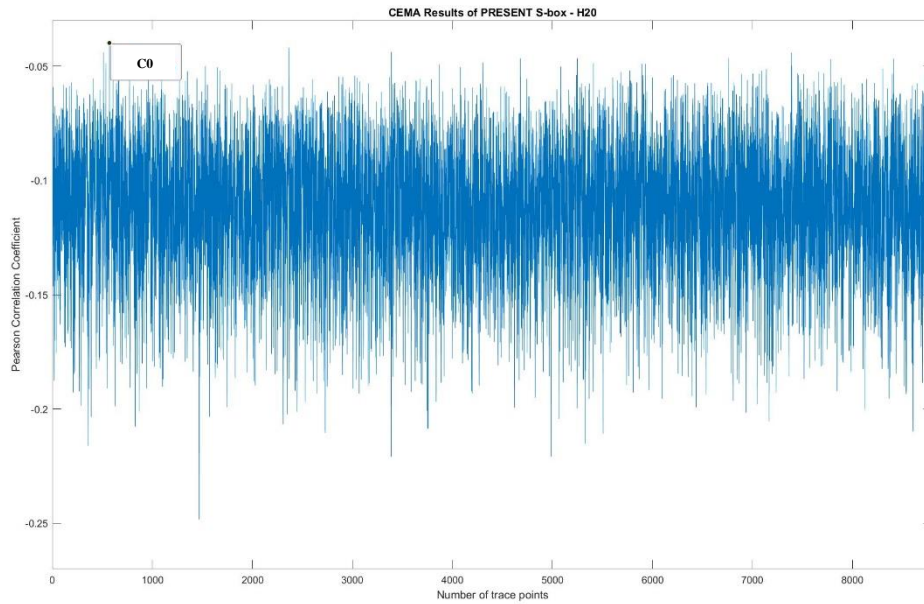


Figure 13. CEMA results for the H20 probe

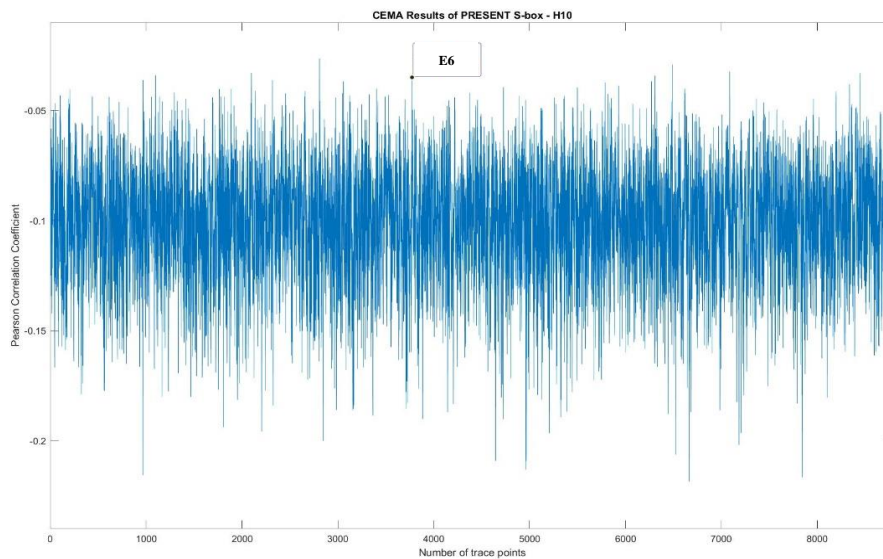


Figure 14. CEMA results for the H10 probe

4. DISCUSSION

Even though visible changes do not appear in the time domain graphs, slight differences can be seen in frequencies between encryption and non-encryption states. The observations of the SEMA and SFEMA are as follows:

- According to Fig.9 and Fig.10, higher density can be seen from 75MHz – 200MHz during encryption. Furthermore, a set of new frequency components seem to appear from

375MHz – 450MHz. Increased amplitude is seen in already existing components in 93MHz, 187MHz, 250MHz and 375MHz itself.

- Regarding data distribution (Fig.7 and Fig.8), values tend to shrink sharply towards zero with one peak head during encryption in both setups. Nevertheless, non-encryption state shows flattened distribution with more than one high peak.
- The outputs show similar results for both H20 and H10 probes in comparison of frequencies as well as of histograms.
- However, spectrogram diagrams do not show any significant difference between encryption and non-encryption states, but they do between the probes where the H20 probe's data has a higher density than the H10 probe's.

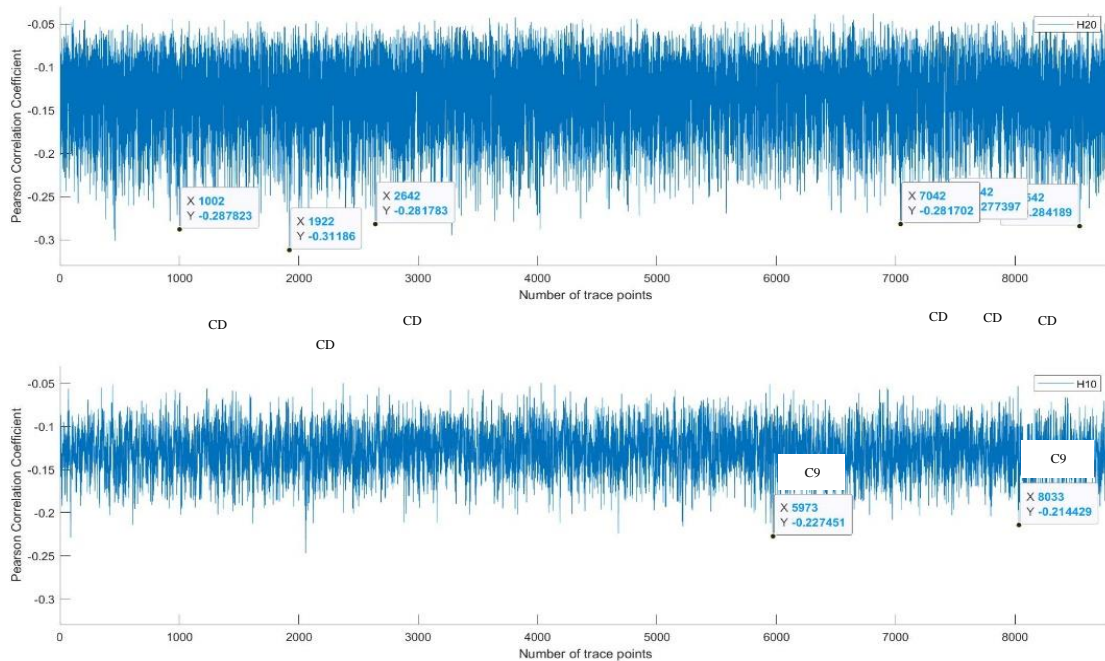


Fig.15: CEMA on non-encryption data

In spite of these differences, the induced values seem to vary on a very small scale, usually from (-20) mV to +20mV. The largest value can be seen when the trigger rises and falls, which is around 200mV.

In CEMA, the general outcomes are extremely noisy and do not appear sharp in comparison to CPA results of the study [8]. Therefore, it is quite disappointing from the attacker's side. Despite this, the H10 probe seems to offer slightly clearer results. In any case, EM emission is highly interfered with by radio waves, UHF/VHF signals, Wi-Fi transmission, Bluetooth communications and any type of EM radiations emitted by electronic devices. That is why EMA has become quite challenging and requires careful attention to perform successfully. On the other hand, it can be seen that the emission dramatically increases when the LED goes to the on position from off as well as the off position from on. Thus, it proves the probes are picking up the emanation created by the circuit board properly, but the radiation created in response to cryptographic functions may not be significant. If that is really the case, lightweight ciphers that come with such shielding would be less prone to EM side-channel attacks. Nevertheless, further analyses are needed with many more sample data to confirm the fact.

In the results, the H20 probe was able to give the 9th Byte of the encryption key in one of its peaks. Moreover, the H10 probe gave the 10th Byte correctly. For both cases, the peaks of the graph caused the correct guess rather than troughs. Anyway, verification is expected to be obtained via further tests because there is a high probability of false-positives due to the excessive noise factors mentioned in the previous paragraph. The issue is further confirmed by the repeated values in several troughs, both during the encryption and non-encryption states. This may occur due to the clock cycles running inside the chip after power-up. In contrast, repetition did not occur in the peaks. Hence, there is a greater probability of key leakages in peaks than in troughs in this model.

What is more, in the work [13], [14] and [15], more than a thousand traces have been used for their EMA research in lightweight ciphers, but in our model, only 256 waveforms for 256 plaintexts were used. Therefore, we assume that a larger number of EM waveforms would increase the probability of attacking in comparison to the quantity of power traces. The main disadvantage was the decrease in the number data points once the signal was trimmed to filter the S-box functional area. It reduced 7200 points, thus it could have been a negative impact on the outcome. To mitigate the identified weaknesses and enhance the performance of the testbed, we expect to be concern with:

- Increase in the number of waveform collections beyond a thousand for increased probability in successful attacks.
- Confirmation of correct waveform alignment to the S-box starting point.
- Bandpass filter application for identified frequency leakage ranges to reduce system noise impacts.
- Analysis of the correlation coefficient results in the frequency domain (CFEMA).
- Checking the behaviour of EM fluctuations as an electric current response rather than the voltage in the time domain.
- Finding the optimised position for probe placement on the chip.
- Choice of the ideal probe for the testbed among the five probes.
- Application of a Faraday cage using Faraday fabrics (low-cost approach compared to expensive manufactured cages) to reduce ambient EM noise.
- Comparison of the performance using different hardware, *i.e.*, *FPGA* (this would take our research limitation beyond the firmware consideration).

While side-channel attack research is encouraged in practical cryptanalysis, it is also important to contemplate real-world scenarios. Usually, side-channel analyses are based on a particular operation of the cipher for a chosen round. In reality, encryption runs through all rounds in pipeline processing. Regarding this project's target of the S-box of PRESENT, AddRound key and pLayer operate before and after in all 31 rounds. Therefore, it would be problematic to split waveforms having proper alignment for the correct function and round. Also, the most probable attack type will be a black-box one where the attacker has no access to the model's parameters. The problem could be worse for the attacker if the manufacturer has properly taken countermeasures against possible EM leakages. Furthermore, complete noise reduction of either ambient or system is an unavoidable task.

5. CONCLUSIONS

In this 5G era, connected devices are increasing massively. The IoT has become a widespread infrastructure in communications and analytics of data. However, provision of sufficient security is challenging because of the resource-constrained environment in the IoT platform. The devices neither contain high processing power nor large onboard memory capacities. In contrast, the data rates are in the least range, in kbps to be specific, and battery power is expected to be maintained

for several years. Apart from the intended purpose of producing insights using sensor data, security must be validated in this enormous data distribution in real-time. Traditional cryptographic algorithms are computationally demanding for this ecosystem. Consequently, lightweight cipher inventions continue to be introduced in academia in recent years. However, confirmation of sufficient security is still open to question.

Cryptanalysis is as essential as cryptography in performance validation of a cipher in cryptology. In that context, side-channel attack resilience plays a huge role in practical cryptanalysis. Therefore, thorough attention must be kept continuously on these emerging ciphers as the matter is quite critical with short encryption key sizes. In addition, side-channel vulnerabilities are very diverse from invasive to non-invasive and white-box to black-box attack types, *i.e.*, *energy drainage, thermal radiation, optical, EM emission, fault injection, etc.* Yet the highest contribution is seen in power analysis. There are few studies available for EMA as mentioned in section 1. As a result, our analysis covers firmware robustness of PRESENT block cipher against CEMA. No CEMA study regarding PRESENT seems yet available in the literature. SEMA and SFEMA were also covered prior to correlation attack modelling.

Our attack model:

- Uses the maximum number of 256 EM waveforms for 256 different plaintexts.
- Performs a white-box non-invasive attack.
- Reduces noise inference by taking averaged waveforms.
- Produced noisy results.
- Was able to guess one Byte of the encryption key correctly in a random position.

What is more, the identified limitations of the model are:

- A total number of 256 traces may not be enough to cause a sharp output.
- Ambient noise was not affected by aluminium foil coverage.
- Reduced trace data points when the waveform was trimmed to locating the S-box functional area may have reduced the accuracy of the final outcome.

In conclusion, the current analysis outcomes indicate that:

- There are no significant fluctuations in the EM emanation of the Arduino UNO in accordance with the cryptographic operations of PRESENT.
- Encryption key leakage tends to occur in peaks rather than troughs in the resultant correlation graphs.

This work is still on-going, and we expect to optimise our model by:

- Increasing the number of EM trace collections by more than a thousand to enhance accuracy.
- Finding ways to increase the number of data points in the trimmed waveforms.
- Using a Faraday box made of Faraday fabrics for the Arduino UNO placement to reduce ambient noise.
- Applying bandpass filters for the identified frequency leakage range.
- Considering both voltage as well as electric current response in data analytics.
- Comparing the performance on various hardware types if possible, *i.e.*, *FPGA*

Finally, this study further discusses the practical possibilities in successful EM attacks once the encryption runs over all 31 rounds in a commercially-ready manufactured item where accessibility to internal parameters is unavailable.

ACKNOWLEDGEMENTS

We would like to thank Dr. Peiter Robyns, the presenter of [25] for sharing his experience in electromagnetic attack modelling with us throughout our journey.

REFERENCES

- [1] Gunathilake, N. A., Buchanan, W. J. and Asif, R. (2019) “Next Generation Lightweight Cryptography for Smart IoT Devices: Implementation, Challenges and Applications”, in *proc. IEEE 5th world forum on Internet of Things (WF-IoT)*, Limerick, Republic of Ireland, pp707-710, Apr. 15, doi: 10.1109/WF-IoT.2019.8767250
- [2] Gunathilake, N. A., Al-Dubai, A. and Buchanan, W. J. (2020) “Internet of Things: Concept, Implementation and Challenges”, in *proc. international conference on IoT and its Applications (ICIA)*, Jamshedpur, India, Dec. 26, available at https://www.researchgate.net/publication/347987943_Internet_of_Things_Concept_Implementation_and_Challenges
- [3] NIST Interagency Report 8114 (2017) “Report on Lightweight Cryptography”, *National Institute of Standards and Technology (NIST)*, Mar., available at https://csrc.nist.gov/CSRC/media/Publications/nistir/8114/draft/documents/nistir_8114_draft.pdf (accessed: 17 October 2018)
- [4] Dinu, D., Corre, Y., Khovratovich, D., Perrin, L., Großschädl, J. and Biryukov, A. (2019) “Triathlon of lightweight block ciphers for the Internet of things”, in *Journal of Cryptographic Engineering*, vol.9, pp283-302, doi: 10.1007/s13389-018-0193-x
- [5] Gunathilake, N. A., Al-Dubai, A. and Buchanan, W. J. (2020) “Recent Advances and Trends in Lightweight Cryptography for IoT Security”, in *proc. 16th international conference on Network and Service Management (CNSM)*, Izmir, Turkey, pp1-5, Nov. 2, doi:10.23919/CNSM50824.2020.9269083
- [6] Canteaut, A., Lauradoux, C. and Seznec, A. (2006) “Understanding Cache Attacks”. *Research Report RR-5881, inria-00051387, INRIA, Le Chesnay Cedex, France*, available at <https://hal.inria.fr/inria-00071387> (accessed: 23 February 2019)
- [7] Raïkin, S., Gueron, S. and Sheaffer, G. (2013) “Protecting Private Data from Cache Attacks”, in *pat. United States Patent – US 8516201 B2*, Aug. 20, available at <https://patentimages.storage.googleapis.com/bf/ba/9c/93aa33d508fed0/US8516201.pdf> (accessed: 24 February 2019)
- [8] Villanueva-Polanco, R. (2019) “Cold Boot Attacks on Bliss”, in *Thèriault N. (eds) Progress in Cryptology – LATINCRYPT 2019*, Lecture Notes in Computer Science, vol.11774, pp40-61, Springer, Cham, Sept. 9, doi: 10.1007/978-3-030-30530-7_3
- [9] Toreini, E., Randell, B. and Hao, F (2015) “An Acoustic Side Channel Attack on Enigma”, in *Computing Science Technical Report series No. CS-TR-1455, Newcastle University, United Kingdom*, available at https://eprint.ncl.ac.uk/file_store/production/211469/2196D5C7-7CC2-494C-A004-E2E5B9288B74.pdf (accessed: 02 March 2019)
- [10] Wang, H. S., Ji, D. G., Zhang, Y., Chen, K. Y., Chen, J. G. and Wang, Y. Z. (2015) “Optical Side Channel Attacks on Single-chip”, in *proc. international conference on Industrial Technology and Management Science, Tianjin, China*, pp364-369, Nov. 27, doi: 10.2991/itms-15.2015.87
- [11] Lo, O., Buchanan, W. J. and Carson, D. (2017) “Power Analysis Attacks on the AES-128 S-box using Differential Power Analysis (DPA) and Correlation Power Analysis (CPA)”. in *Journal of Cyber Security Technology*, vol.1.2, pp88-107, doi: 10.1080/23742917.2016.1231523
- [12] Lo, O., Buchanan, W. J. and Carson, D. (2018) “Correlation Power Analysis on the PRESENT Block Cipher on an Embedded Device”, in *proc. 13th international conference on Availability, Reliability and Security (ICPS)*, Hamburg, Germany, pp1-6, Aug., doi: 10.1145/3230833.3232801
- [13] Sayakkara, A., Le-Khac, N. and Scanlon, M. (2019) “A Survey of Electromagnetic Side Channel Attacks and Discussion on their Case-Progressing Potential for Digital Forensics”, in *Digital Investigation*, vol.29, pp43-54, doi: 10.1016/j.diin.2019.03.002

- [14] Kim, C., Schl affer, M. and Moon, S. (2008) ‘‘Differential Side Channel Analysis Attacks on FPGA Implementations of ARIA, in *ETRI journal*, vol.30, No.2, pp315-325, Apr. 1, doi: 10.4218/etrij.08.0107.0167
- [15] Chari, S., Rao, J. R. and Rohatgi, P. (2003) ‘‘Template Attacks’’, in *Kaliski B.S, Ko, K., Paar C. (eds) Cryptographic Hardware and Embedded Systems (CHES) 2002, Lecture Notes in Computer Science*, vol.2523, Springer, Berlin, Heidelberg, doi: 10.1007/3-540-36400-5_3
- [16] Biryukov, A., Dinu, D. and Grosch adl (2016) ‘‘Correlation Power Analysis of Lightweight Block Ciphers: From Theory to Practice’’, in *Manulis M., Sadeghi A.R., Schneider S. (eds) Applied Cryptography and Network Security (ACNS) 2016, Lecture Notes in Computer Science*, vol.9696, Springer, Cham, doi: 10/1007/978-3-319-39555-5_29
- [17] Duan, X., Cui, Q., Wang, S., Fang, H. and She, G. (2016) ‘‘Differential Power Analysis Attack and Efficient Countermeasures on PRESENT’’, in *proc. 8th IEEE international conference on Communication Software and Networks (ICCSN), Beijing, China*, pp8-12, doi: 10.1109/ICCSN.2016.7586627
- [18] Shanmugam, D., Selvam, R. and Annadurai, S. (2014) ‘‘Differential Power Analysis Attack on SIMON and LED Block Ciphers’’, in *Chakraborty R.S., Matyas V., Schaumont P. (eds) Security, Privacy and Applied Cryptography Engineering (SPACE) 2014, Lecture Notes in Computer Science*, vol.8804, Springer, Cham, doi: 10.1007/978-3-319-12060-7_8
- [19] Nozaki, Y., Iwase, T., Ikezaki, Y. and Yoshikawa, M. (2017) ‘‘Differential Electromagnetic Analysis for PRESENT and its Evaluation with Several Selection Functions’’, in *Journals of International Council on Electrical Engineering*, vol.7, No.1, pp137-141, doi: 10.1080/22348972.2017.1344014
- [20] Yoshikawa, M. and Nozaki, Y. (2016) ‘‘Electromagnetic Analysis Attack for a Lightweight Cipher PRINCE’’, in *proc. IEEE international conference on Cybercrime and Computer Forensic (ICCCF), Vancouver, BC, Canada*, pp1-6, doi: 10.1109/ICCCF.2016.7740423
- [21] Yoshikawa, M., Nozaki, Y. and Asahi, K. (2016) ‘‘Electromagnetic Analysis Attack for a Lightweight Block Cipher TWINE’’, in *proc. IEEE international conference on Wireless Information Technology and Systems (ICWITS) and Applied Computational Electromagnetics (ACES), Honolulu, HI, USA*, pp1-2, doi: 10.1109/ROPACES.2016.7465354
- [22] Lakshminarasimhan, A. (2011) ‘‘Electromagnetic Side-Channel Analysis for Hardware and Software Watermarking’’, *thesis presented to the University of Massachusetts Amherst, Amherst, Massachusetts*, pp39, available at <https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1822&context=theses> (accessed: 21 March 2020)
- [23] Tiu, C. C. (2005) ‘‘A New Frequency-based Side Channel Attack for Embedded Systems’’, *thesis presented to the University of Waterloo, Ontario, Canada*, doi: 10.1.1.69.1441
- [24] Kocher, P., Jaffe, J. Jun, B. (1999) ‘‘Differential Power Analysis’’, in *Wiener M. (eds) Advances in Cryptology (CRYPTO) 1999, Lecture NOTES IN Computer Science*, vol.1666, Springer, Berlin, Heidelberg, doi: 10.1007/3-540-38305-1_25
- [25] Robyns, P. (2019) ‘‘Performing Low-cost Electromagnetic Side-Channel Attacks using RTL-SDR and Neural Networks’’, in *FOSDEM, Brussels, Belgium*, Feb.2, available at <https://youtu.be/cs08QSIbp-A> (accessed: 10 April 2020)
- [26] Bogdanov, A., Knudsen, L., Leander, G., Paar, C., Poschmann, A., Robshaw, M., Seurin, Y. and Vikkelsoe, C. (2007) ‘‘PRESENT: An Ultra-lightweight Block Cipher’’, in *Paillier P., Verbauwhede I. (eds) Cryptographic Hardware and Embedded Systems (CHES) 2007, Lecture Notes in Computer Science*, vol.4727, Springer, Berlin, Heidelberg, doi: 10.1007/978-3-540-74735-2_31
- [27] lightweightcrypto.org by Ruhr-Universit at Bochum (2007), ‘‘C PRESENT Implementation (8bit)’’, in *Bochum, Germany*, available at http://www.lightweightcrypto.org/downloads/implementations/PRESENT%208-bit_implementation.rar(accessed: 05 September 2020)

AUTHORS

Nilupulee A. Gunathilake is a doctoral student of Blockpass ID Lab in the School of Computing at Edinburgh Napier University, UK. She is a multi-disciplinary researcher who is currently engaged in lightweight cryptography for IoT devices. Her mainstream towards the PhD completion targets side-channel analysis of lightweight ciphers. Nilupulee had previously involved in research in communication engineering related areas; free-space optics (FSO), vehicular communication and microwave propagation. She graduated in MSc (Smart Networks) at the University of the West of Scotland, UK in 2017. Nilupulee obtained a PgDip (Telecommunication and Electronic Engineering), Sheffield Hallam University, UK in 2016 as well as her first degree, BEng (Hons) in Electronics and Communications Engineering, University of Wolverhampton, UK in 2013. She also holds professional qualifications in project management and international spoken English language ESOL. Nilupulee has been involved in academia for over five years and is the first author of five international conference papers. She was a Lecturer in the Faculty of Engineering at Sri Lanka Institute of Information Technology (SLIIT) in 2018 and an Assistant Lecturer (Grade 1) at International College of Business and Technology (ICBT), Sri Lanka in 2015. She worked as a Research Assistant in the Department of Electronic and Telecommunication Engineering at the University of Moratuwa, Sri Lanka in 2016. Nilupulee was offered a PhD scholarship by Edinburgh Napier University in 2018, the Dean's Scholarship by the University of the West of Scotland in 2016 and research grants by the University of Moratuwa in 2015. She was awarded "the University Court Medal" for the best academic excellence of the MSc program.



Ahmed Al-Dubai is currently Professor in the School of Computing, Edinburgh Napier University, where he leads the IoT and Networked Systems Research Group. He is also the Director of the Postgraduate Research Degrees Programme. Ahmed was awarded his PhD from the Department of Computing Science, University of Glasgow in 2004. He leads interdisciplinary research and initiatives on the area of Group Communications, High-performance Networks, Internet of Things, Future Networks, E-Health, Smart Cities and Security. His research has been supported by the EU, Universities UK and the Royal Society, Carnegie Trust, EPSRC and Scottish Funding Council. His findings have been published widely in top tier journals including different IEEE Transactions journals, and in prestigious international conferences including IEEE IPDPS, IEEE ICC, IEEE GLOBECOM, IEEE WCNC, ICPP and IEEE IPCCC. Ahmed has been the recipient of several international academic awards and recognition including Best Papers Awards at IEEE IUCC 2015, ACM MoMM 2013 and ACM SAC 2002. He has been regularly invited to give keynote and plenary speeches at several international conferences. He has been a member of several editorial boards of scholarly journals. Ahmed been the Guest Editor for over 25 special issues in scholarly journals and the Chair/Co-Chair of over 30 International Conferences and workshops. He is a Fellow of the British Higher Education Academy and a Senior Member of the IEEE.

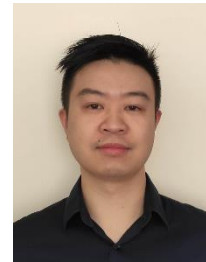


William (Bill) J. Buchanan OBE is a Professor in the School of Computing at Edinburgh Napier University, and a Fellow of the BCS and Principal Fellow of the HEA. He was appointed an Officer of the Order of the British Empire (OBE) in the 2017 Birthday Honours for services to cybersecurity. Bill lives and works in Edinburgh, and is a believer in fairness, justice, and freedom. His social media tagline reflects his strong belief in changing the world for the better: "A Serial Innovator. An Old World Breaker. A New World Creator". He also has a strong belief in the power of education, and in supporting innovation from every angle. Bill currently leads the Blockpass ID Lab and the Centre for Cybersystems and Cryptography. He works in the areas of blockchain, cryptography, trust and digital identity. He has one of the most extensive cryptography sites in the World (asecuritysite.com), and is involved in many areas of novel research and teaching. Bill has published over 30 academic books, and over 300 academic research papers. Along with this, Bill's work has led to many areas of impact, including three highly successful spin-out companies (Zonefox, Symphonic Software and Cyan Forensics), along with awards for excellence in knowledge transfer, and for teaching. Bill recently received an "Outstanding



Contribution to Knowledge Exchange” award, and was included in the FutureScot “Top 50 Scottish Tech People Who Are Changing The World”.

Owen Lo is a Research Fellow (PhD) at Edinburgh Napier University. He obtained a BEng (Hons) in Computer Networks and Distributed Systems at Edinburgh Napier University in 2010 before continuing to complete a PhD on the topic of e-Health in 2015 at the same institute. Awards received by Owen include: Young Software Engineer of the Year Award (Lumison Prize) in 2010, Team Prize Raytheon Cyber Challenge Award in 2011 and Student of the Year ENU Award in 2012. During his PhD, Owen contributed to the Data Capture and Auto Identification Reference (DACAR) Project – a project funded in part by EPSRC and TSB – which aimed to create a secure cloud-based information sharing platform for patient data in healthcare environments. During his time as a researcher at Edinburgh Napier University, Owen has also contributed to the success of two spin-out of two companies: Symphonic Software and Cyan Forensics. For Symphonic Software, Owen helped develop an information-sharing engine used for the secure and trusted sharing of information between different sectors including finance, healthcare, social care and law enforcement. His work on Cyan Forensics included the development of a fully-featured contraband detection software used to rapidly determine if an individual is suspected of storing illegal data on a computer. The contraband detection software was designed specifically to be used by digital forensics experts within the law enforcement sector. Owen's research interests include side-channel analysis, cryptography and computer security. Currently, he is working on a third university spin-out, MemCrypt, which uses novel techniques to combat ransomware.



© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.

WEB SCRAPER UTILIZES GOOGLE STREET VIEW IMAGES TO POWER A UNIVERSITY TOUR

Peiyuan Sun¹ and Yu Sun²

¹Webb School of California, Claremont, CA 91711

²California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

Due to the outbreak of the Covid-19 pandemic, college tours are no longer available, so many students have lost the opportunity to see their dream school's campus. To solve this problem, we developed a product called "Virtourgo," a university virtual tour website that uses Google Street View images gathered from a web scraper allowing students to see what college campuses are like even when tours are unavailable during the pandemic. The project consists of 3/4 parts: the web scraper script, the GitHub server, the Google Domains DNS Server, and the HTML files. Some challenges we met include scraping repeated pictures and letting the HTML dropdown menu jump to the correct location. We solved these by implementing Python and Javascript functions that specifically target such challenges. Finally, after experimenting with all the functions of the web scraper and website, we confirmed that it works as expected and can scrape and deliver tours of any university campus or public buildings we want.

KEYWORDS

Web scraping, virtual tour, cloud computing.

1. INTRODUCTION

Due to the Covid-19 pandemic, in-person campus tours are no longer possible. This means that a lot of seniors in high school will miss the opportunity to see their dream schools' campuses, which is significant in helping them form an idea of what college is like. Currently, there lacks a persuasive virtual tour platform on the internet that can deliver to students and families a persuasive overview of what school campuses really look like. In recent years, more than two million American high school seniors that graduate each year enroll in a college or university, which is about 66% of each class. Although it is hard to predict the percentage of these two million students that tried to get a virtual tour, it is reasonable that many of them wanted one as a guide to know what kind of school they might be attending. In the future, the number will only rise, as both the total number of high school students and high school graduation rates are projected to increase. Even after the peak of the pandemic, when colleges and universities reopen their campuses to the public, there will still be a need for virtual tours, as not every applying senior will be able to take a tour at every school they are interested in. This leaves a strong demand for a platform hosting virtual tours for universities. The same applies to other public organizations or buildings. Not only do schools need virtual tours, public buildings as big as sports stadiums and airports, or as small as restaurants and stores might also find themselves one day in need of developing a virtual tour to increase publicity or improve customer experience.

Some existing methods have already been used to address this problem. Some universities and public places have had outside companies develop virtual tours in the form of 360-degree images, which allows the user to switch between images and navigate around the place [11, 12, 13].

David C. Wyld et al. (Eds): ITCSE, ICDIPV, NC, CBIoT, CAIML, CRYPIS, ICAIT, NLCA - 2021

pp. 207-219, 2021. CS & IT - CSCP 2021

DOI: 10.5121/csit.2021.110916

However, this requires people to go around the place to take pictures, which is inefficient and often takes a lot of time and effort to complete. Also, it is hard to get permissions for such tours, especially during the pandemic when schools are extra cautious about the potential spread of the virus. Therefore, if images were not made before, tours using this method would be extremely hard to develop during the pandemic.

In fact, immersive tours in the form of 360-degree pictures or, if possible, virtual reality will produce the best results because they will let people have better long-term memory results [14]. One important goal of virtual tours is to make people remember the university, so immersive tours are better than 2D pictures at letting people experience the school visually, and are therefore more preferable. However, because the technology of virtual reality is still not mature, there are obstacles in solely using virtual reality for campus tours [15]. A major challenge is that virtual reality is not popular enough so that everyone has a pair of goggles, which are often expensive. Also, people will likely not buy a pair of goggles just to go on a school virtual tour. As a result, because 2D pictures cannot provide an immersive experience, and virtual reality has not been popularized enough for everyone, using 360-degree images seems the best approach. It is also better to develop a website than a phone application, because computer screens are larger and will give a better view, and people don't have to download an application in order to access the virtual tour.

In this paper, we follow the same line of research to provide immersive experiences in the form of 360 images. Our goal is to create simple tours that can generate tours within a short period of time that do not require people to travel to the campus during the pandemic. Our method is inspired by web scraping and Google Street View [1, 2]. Google Street View is a function developed by the Google Earth [3, 4] team to provide users with the opportunity to explore the world in the form of 360-degree images, which can be taken with special panoramic cameras that can capture the surrounding environment such as Ricoh Theta [9,10]. First, we developed a Python scraping script that finds 60 locations near a given university and an available 360-degree picture around each location [5, 6]. Web scraper is a popular method that gathers information from the internet for analytical or personal use. Second, after deleting the repeated pictures, we output the scraped information into a json file. Finally, we developed a website that pulls the information from the json file and shows the tour. Google has many users; some of them own 360 cameras and have taken and uploaded the pictures of a school campus before. This effectively solves the problem of needing people to go to the school and take pictures, as we can find those that are already available on the internet through a web scraper.

In two application scenarios, we not only demonstrated the effectiveness of our scraping script, but also proved the usefulness of our website in an actual environment. First, we generated a list of the Top 100 schools in the United States, as well as every registered university or college in California. We then ran our python scraping script and waited for it to iterate through every school on the list. When each school was completed, the terminal would print the name of the school and its number on the list (the first school is 1, the second is 2, etc.) After the scraping part was done, two processing functions went through the results, checking for repeated pictures and changing the descriptions. We successfully acquired a file named "data.json" that contains all the scraped information. The scraping script worked just the way we expected it to. Secondly, we copied the data.json file and put in the same directory as our website html file [7]. We pushed the information to GitHub, our website's server [8], and after a while the data showed up on our domain "https://www.virtourgo.com." After we typed in the name of a random school in our list of schools and clicked the button to start the virtual tour, we were led to another tour page, where we could not only see the street view panorama up and working, but also a places list dropdown menu that shows all the available buildings at the school, two descriptions that tell people the

name of each building and the next one, and a button to go to the next building on the list. The website also worked just as we intended.

The rest of this paper is organized thus: Section 2 provides insight on the challenges we experienced during the development process of the web scraper and website; Section 3 focuses on the details and gives a guide for how we designed the different components, as well as how we solved the problems mentioned in Section 2; Section 4 gives an overall evaluation of our final product; Section 5 presents the related works done in this field. Finally, Section 6 allows for concluding remarks and future possibilities for the project.

2. VOICE USER INTERFACE

In order to build a university virtual tour website that uses Google Street View images gathered from a web scraper, a few challenges have been identified as follows

2.1. Challenge 1: The Scraping Script Does Not Tell What the Next Place Is

The scraping script generates a json file, which consists of two descriptions for each location. Description1 is a sentence that states the name of the current location, while description2 is a sentence that either tells the viewer what the next location is, or concludes the tour if it is the last location. However, when the python script tries to find the available 360 pictures, it finds the locations near the targeted school one by one, which means that it cannot get the name of the next location and put it in “description2.” In order to fix the problem, we implemented a function called “processing(inp)” (see Figure 1) for when the scraping script finishes running. The processing function will iterate through every location in the data dictionary, changing the “description2” of each location with the name of the next location before outputting the data as a json file.

```
# Add the descriptions
def processing(inp):
    data = inp
    for school in data["data"]:
        for i in range(len(school["buildings"])):
            if i+1 == len(school["buildings"]):
                school["buildings"][i]['description2'] = "You have reached the last stop of the tour. Hope you enjoyed. Have a wonderful day!"
            else:
                school["buildings"][i]['description2'] = "Your next stop is " + school["buildings"][i+1]['title'] + '.'
```

Figure 1. “processing(inp)”

2.2. Challenge 2: The Scraping Script will always Produce Duplicate Locations

Our scraping script is designed to find available 360 pictures around several locations in each school. However, because there might not be that many 360 pictures on Google Street View, every school will more or less have locations with the same pictures, depending on the school and its available pictures. If a school has abundant images (meaning that people in the past have taken a lot of 360 pictures and uploaded them), then it will have fewer places with the same images. Obviously, users will not want to see the same image twice for a school, so in order to avoid this problem we implemented another processing function called “check_repeat(inp)” (see Figure 2). The function will iterate through every location with every school scraped, truncate the latitude and longitude to four decimal places (this is because although many locations show different latitude and longitudes in the later decimal places, they still represent the same image), and delete the location if the image already appears before the data gets outputted as a json file. Meanwhile, the function will also print out the name and the number of locations deleted in the terminal.

Although this will not eradicate the problem entirely since sometimes the latitude and longitude will still show the same image, it will effectively solve the problem in 90% of the instances.

```
def check_repeat(inp):
    data = inp
    for school in data["data"]:
        li = []
        index_factor = 0
        for i in range(len(school["buildings"])-1):
            lat = float('%%.4f'%school['buildings'][i-index_factor]['lat'])
            lng = float('%%.4f'%school['buildings'][i-index_factor]['lng'])
            school['buildings'][i-index_factor]['lat'] = lat
            school['buildings'][i-index_factor]['lng'] = lng
            if {lat:lng} not in li:
                li.append({lat:lng})
            else:
                school['buildings'].pop(i-index_factor)
                index_factor += 1
        print(school['school_name'], index_factor)
    return data
```

Figure 2. “check_repeat(inp)”

2.3. Challenge 3: The “List of Places” Menu Cannot Jump to the Correct Location

In the touring page, each school will have its own “List of Places” menu in the navigation bar that shows a list of all the places available on the site. This is intended for users to see all the locations available and jump to the one they want. However, since we did not make a new html page for every location, it is difficult to let them jump to the correct location of the correct school, especially when there are a lot of locations stored in the database. Therefore, we developed a function, “goToNextPlaceWithIndex(gotoIndex)” (see Figure 3). The function will take the index of the location that appears on the List of Places menu, find the data stored with the index, and start a street view panorama with that location.

```
function goToNextPlaceWithIndex(gotoIndex) {
    if (index < allData[schoolIndex]["buildings"].length) {
        index = gotoIndex;
        $("#title").text(allData[schoolIndex]["buildings"][index].title);
        $("#description1").text(allData[schoolIndex]["buildings"][index].description1);
        $("#description2").text(allData[schoolIndex]["buildings"][index].description2);
        $("#heading").text('Virtourgo | ' + allData[schoolIndex]["buildings"][index].title);
        var panorama = new google.maps.StreetViewPanorama(
            document.getElementById('pano'), {
                position: {lat: allData[schoolIndex]["buildings"][index].lat, lng: allData[schoolIndex]["buildings"][index].lng},
                pov: {
                    heading: allData[schoolIndex]["buildings"][index].heading,
                    pitch: allData[schoolIndex]["buildings"][index].pitch,
                }
            });
        map.setStreetView(panorama);
    }
}
```

Figure 3. “goToNextPlaceWithIndex(gotoIndex)”

3. SOLUTION

Virtourgo is a website that displays virtual tours for well-known universities in the world. Virtourgo implements a web scraper to gather information from the internet and compile it into a virtual tour. The scraping process is done in the background, so individual users will not need to

see or change anything to access the tour. The scraper uses Selenium in Python to search 1) buildings on the campus of the desired school, 2) available Google Street View pictures around each building, and 3) each picture's latitude and longitude. The information scraped is then re-evaluated by another script, which goes through each location to make sure they are not repeated. To conclude the scraping process, we stored the information scraped in a json file within a GitHub repository. Like any other websites, users can access the Virtourgo by entering the URL, which in this case is virtourgo.com, in a web browser. The DNS Server we chose is Google Domains, and the server is GitHub. The users will automatically be transported to a page named "index.html," the main page of the website. On the homepage, there is a search box with an auto-correct function. Users will enter the name of their desired school and click the "Start Virtual Tour" button to go to the tour page that shows the scraped locations. On the tour page, most of the screen will be the Google Street View images embedded on the page, and below those will be some descriptions about the place, as well as what the next stop will be. When the tour is concluded, users will be prompted to return to the home page, where they can feel free to start another tour. In sum, there are three main components of our system (see boxes in Figure 4):

- a Python Web Scraper that gathers information about all the locations
- a GitHub repository that acts as the server, hosting the website and storing the scraped information
- a DNS Server by Google Domains that processes the request

The following sections will describe the implementation process for each of the three components, as well as the website itself, in detail.

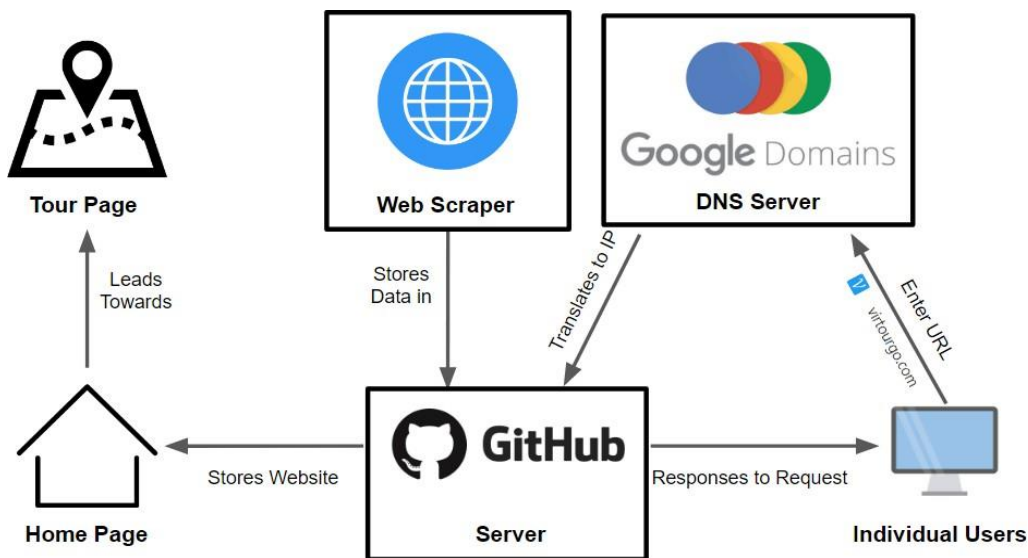


Figure 4. Three main components of our system

3.1. Web Scraper

For the Web Scraper part, our goal for this project is to find the latitude and longitude of places of interest at and around the given school. First, we need to make sure that the location we want is the same one that appears on Google Maps. We can do this by finding the place with the

closest name to the one we inputted. Then, we can start searching for the actual locations around the place we want. For each place, we will find the closest 60 locations around it; for each location, we will gather the name of the place and the latitude and longitude of the closest available images. To be able to search many schools at a time, we implemented another

function, which simply iterates through a list of different schools to find the information above for each one. The code to find places around each school is shown in Figure 5. Finally, we will output the data we scraped into a json file, which we will use in the actual website.

```
def findPlaces(loc=(location["lat"],location["lng"]), pagetoken = None):
    lat, lng = loc
    url = "https://maps.googleapis.com/maps/api/place/nearbysearch/json?location={lat},{lng}&rankby=distance&key={APIKEY}{pagetoken}"
    ".format(lat = lat, lng = lng,APIKEY = APIKEY, pagetoken = "&pagetoken="+pagetoken if pagetoken else "")
    response = requests.get(url)
    res = json.loads(response.text)

    for result in res["results"]:
        building_name = result["name"]
        lat = result["geometry"]["location"]["lat"]
        lng = result["geometry"]["location"]["lng"]
        buildings.append(
            {
                "title": building_name ,
                "description1": "This is the " + building_name + '.',
                "description2": "",
                "lat": lat,
                "lng": lng,
                "heading": 34,
                "pitch": 10,
            })
    pagetoken = res.get("next_page_token",None)

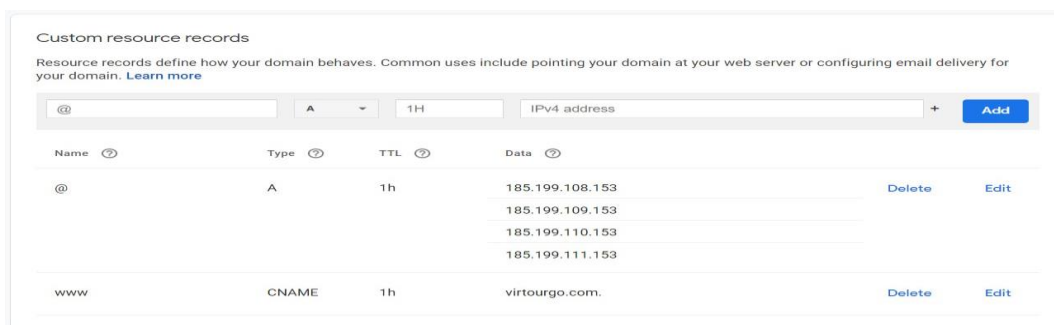
    return pagetoken
```

Figure 5. Code to find places around each school

Of course, because not every location has its own 360 picture on Google Street View available, many locations will show the same picture, so we have to use another function to rule out locations with the same images.

3.2. Google Domains DNS Server

The DNS Server used for the website is Google Domains, which hosts our domain. First, we entered the webpage to build a domain through “https://domains.google.com/registrar/search.” Then, we entered the name that we wanted the domain to be: “virtourgo.” Next, we chose the ending of my domain, which is “.com” in our case. Finally, we needed to pay an annual fee of \$12 for Google to host our domain. After we paid for the domain and logged in to the google account that bought it, we navigated to “My Domains,” and selected the domain we just registered, and clicked on “DNS.” There are a couple of things we need to put in before it starts working (see Figure 6).



Custom resource records

Resource records define how your domain behaves. Common uses include pointing your domain at your web server or configuring email delivery for your domain. [Learn more](#)

Form: @, A, 1H, IPv4 address, Add

Name	Type	TTL	Data		
@	A	1h	185.199.108.153 185.199.109.153 185.199.110.153 185.199.111.153	Delete	Edit
www	CNAME	1h	virtourgo.com.	Delete	Edit

Figure 6. DNS records

In the box above, we entered “185.199.108.153” into the “IPv4 address” section, and clicked the + button to the right of it. Then we entered the rest three of the IPv4 addresses, “185.199.109.153,” “185.199.110.153,” and “185.199.111.153.” These are the IP addresses that belong to GitHub. Then we clicked “Add” after finishing. Finally, in the dropbox that says “A,” we clicked to find a button called “CNAME,” and in the “Domain name” section, we put in the domain we just registered.

3.3. GitHub Server

All the html programs of the website, as well as the scraped information, are stored in a GitHub repository, which also serves as the server of the website. To do this, we first created a GitHub repository, and pushed the html and json program files on the repository. Then, we went to the “settings” page of the repository, which can be found at the top of the repository menu, and found a section called “GitHub Pages” (see Figure 7). In this section, we entered the domain registered before in the “custom domain” part and clicked “save.” Finally, we checked the box that says “enforce HTTPS,” which stands for Hypertext Transfer Protocol Secure. This gave our website secure communication by encrypting the data requests automatically.

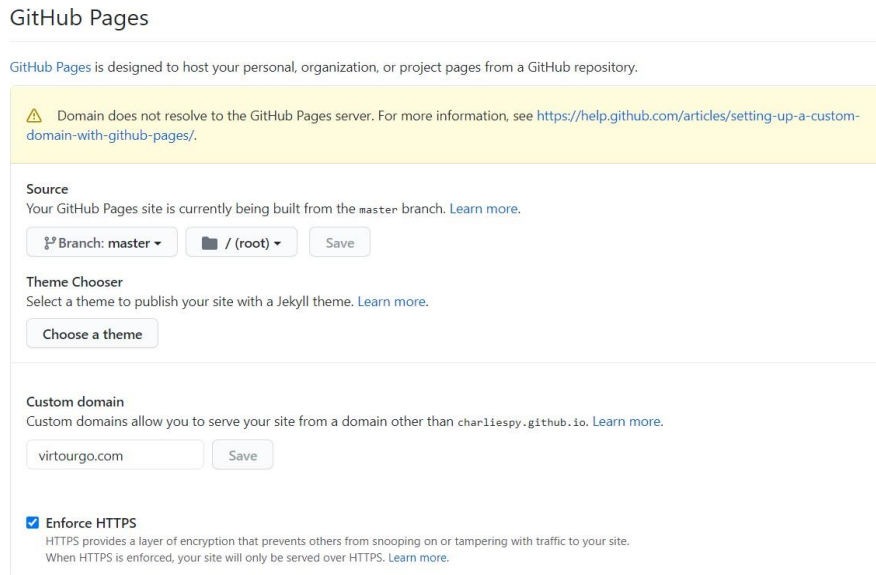


Figure 7. “GitHub Pages”

3.4. Website Development

There are three main technical parts of the website development part. The first one is the search box, the second one is the function to find the desired school, and the third one is the embedding of the Google Maps Street View window into our website. For the search box, we want to add an auto-complete function for it, so users will not need to enter the precise name that appears on our website. We implemented the autocomplete function using JQuery. We created a “#schoolInput” tag (see Figure 8), and used the tag when defining our search box by saying (id=”schoolInput”). The source “schools” is a var that we created with all the names of our school list, and the delay is the time in milliseconds the search box will wait before giving its suggestions.

```
$( "#schoolInput" ).autocomplete({
  source: schools,
  delay: 600
});
```

Figure 8. Autocomplete function using JQuery

For our search box to be able to jump to the correct school, we created another function called “#startVirtualTour” (see Figure 9). In this function, we call and retrieve the name of the school from the previous function “#schoolInput,” and find the index of that school’s name. Then, we navigate to the index of that specific school with another function “goToNewPage,” which leads us to the actual tour page.

```
$("#startVirtualTour").click(function(){
  var school = $("#schoolInput").val();
  var indexForTheSchool = school_mapping[school];
  dataToParse = allData[indexForTheSchool];

  function goToNewPage() {
    url = 'new_view.html?index=' + indexForTheSchool;
    document.location.href = url;
  }

  goToNewPage();
});
```

Figure 9. “#startVirtualTour”

For the Google Maps Street View function, there are a couple of stats we have to enter in order for the street view panorama to initiate (see Figure 10): the position, which includes the latitude and longitude of the location, and the pov, which includes the heading and the pitch. The latitude and longitude determine the exact position the place is located on the map, so that Google Street View can find it in its system. The heading signifies the position the users are looking at (e.g., whether north or south), and the pitch determines the angle of the pov (whether looking straight ahead or pointing at the sky). Therefore, we accessed the information which we scraped in the previous section in a function by id “pano,” and entered those numbers. Notably, besides linking the javascript file in our html file as usual, we also inserted the following script with our google api in the html file:

```
<script async defer
src="https://maps.googleapis.com/maps/api/js?key=AIzaSyDXUNSnZVRHG9S0aS1SUr_TNz
5iQHwgKJo&callback=initialize"></script>
```

```
var panorama = new google.maps.StreetViewPanorama(
  document.getElementById('pano'), {
    position: {lat: allData[schoolIndex]["buildings"][index].lat, lng: allData[schoolIndex]["buildings"][index].lng},
    pov: {
      heading: allData[schoolIndex]["buildings"][index].heading,
      pitch: allData[schoolIndex]["buildings"][index].pitch,
    }
  });
map.setStreetView(panorama);
```

Figure 10. Stats to enter for street view panorama

4. EXPERIMENT

To evaluate the major components described above, we designed two experiments to see if they work as expected. The first experiment aimed to test the autocomplete search box and the jump to school function. On the website, we typed in the first few letters of a school scrape. If our targeted school appeared in the autocomplete list, we clicked the “Start Virtual Tour” button to see if we are navigated to the correct school. We then picked 20 schools randomly from the list, and if all of the schools passed the aforementioned test, the whole system would be considered to fulfill our expectations.

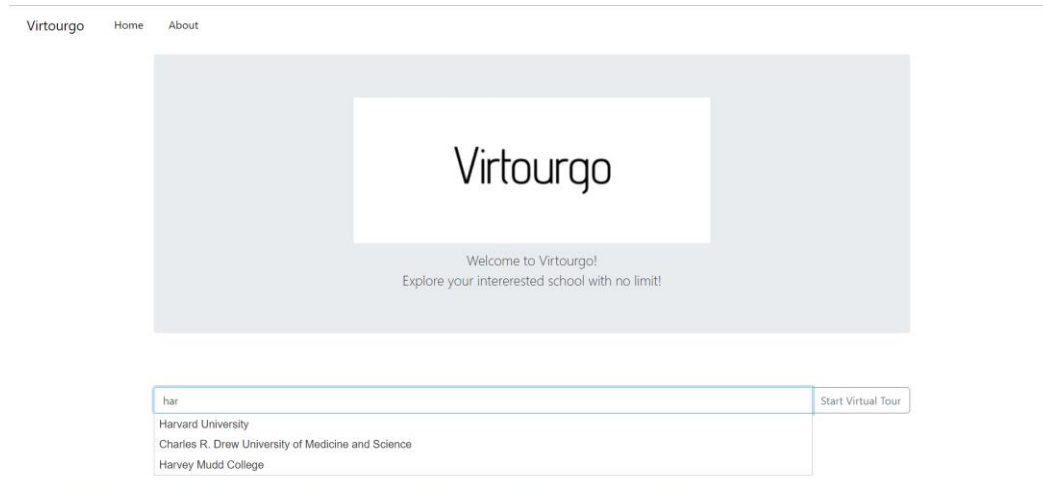


Figure 11. Typing fist three letters, “Har...”

We tried the test with 20 schools randomly selected from the school list, one of which is Harvard University. We typed in the first three letters of the name “Harvard” (see Figure 11), and the autocomplete function successfully provides us with all the schools whose names include “har.” In our database now, there are three schools with “har” in its name, “Harvard University,” “Charles R. Drew University of Medicine and Science,” and “Harvey Mudd College.” After we clicked and selected our targeted school Harvard University, we clicked the “Start Virtual Tour” button at the right of the search box. The page then changed to the main tour page (see Figure 12), which shows the first stop of Harvard University.

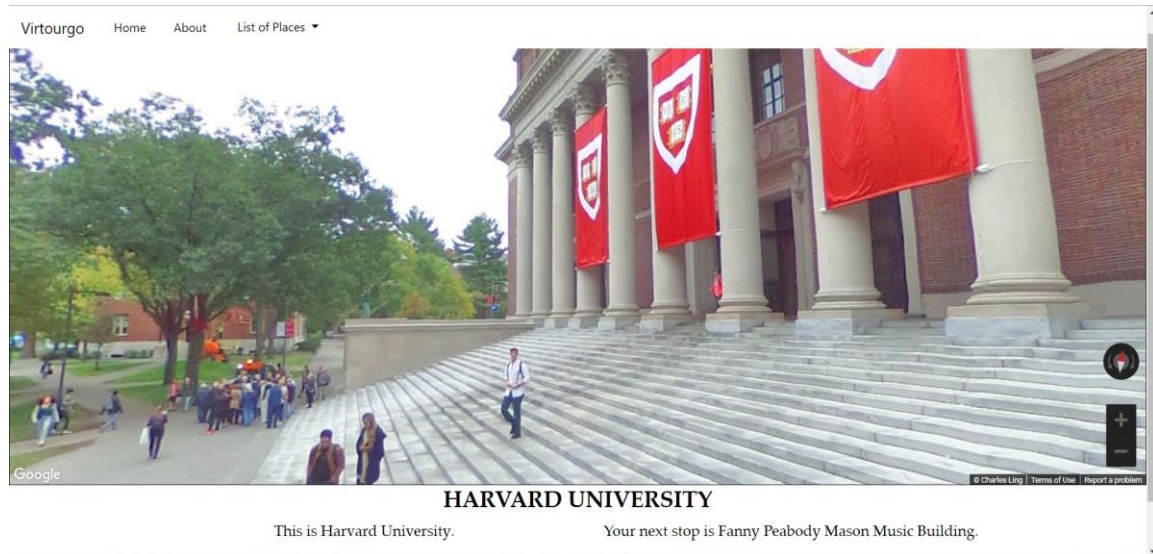


Figure 12. The main tour page for Harvard, which shows the first stop of Harvard University

We repeated the process for another 19 randomly-selected schools and all got the results we expected - the autocomplete function showed us the school's name, and the start tour button led us to the tour page with the correct school's images.

Experiment 2

The second experiment was designed to test the scraping function. We randomly generated 20 schools and put them in a list named "schools." Then we called for the scraping function and the processing function (see Figure 13).

```
schools = [
    'Andreas University',
    'University of Maryland, College Park',
    'University of West Los Angeles',
    'California State University San Marcos',
    'Star King School for Ministry',
    'University of Pittsburgh',
    'Claremont Graduate University',
    'Vanderbilt University',
    'University of Illinois, Chicago (UIC)',
    'Providence Christian College',
    'San Jose State University',
    'California State University, Los Angeles',
    'Golden Gate University',
    'Epic Bible College',
    'Fielding Graduate University',
    'Biola University',
    'University of Southern California',
    'Alliant International University',
    'Illinois Institute of Technology',
    'The Ohio State University'
]

# Add the descriptions
def processing(inp):
    data = inp
    for school in data["data"]:
        for i in range(len(school["buildings"])):
            if i+1 == len(school["buildings"]):
                school["buildings"][i]["description2"] = "You have reached the last stop of the tour. Hope you enjoyed. Have a wonderful day!"
            else:
                school["buildings"][i]["description2"] = "Your next stop is " + school["buildings"][i+1]["title"] + "."
    return data

result = generateSchoolList(schools)
final = processing(result)
with open('data.json', 'w') as outfile:
    json.dump(final, outfile)
```

Figure 13. Scraping and processing function

The processing function's purpose is to fix the problem with the school's names. During the scraping process, we could not get the name of the next location, so we cannot tell users where the next stop is. Therefore, the processing function aims to go into every location and change the part where it says what the next stop is, which is in a key called "description2." After the scraping is done, we then uploaded it into a json file called "data.json."

After the scraping is done (after each school's scraping process is completed, the terminal will print the index and the name of the school), we can open the output file "data.json" and see the results (see Figure 14).



```

data.json
1  {
2    "timestamp": 1617508974.146392,
3    "data": [
4      {
5        "school_name": "Andrews University",
6        "address": "8975 Old 31, Berrien Springs, MI 49104, United States",
7        "location": {
8          "lat": 41.9646127,
9          "lng": -86.35946609999999
10       },
11       "buildings": [
12         {
13           "title": "Andrews University",
14           "description1": "This is the Andrews University.",
15           "description2": "Your next stop is AUSA.",
16           "lat": 41.9646127,
17           "lng": -86.35946609999999,
18           "heading": 34,
19           "pitch": 10
20         },
21         {
22           "title": "AUSA",
23           "description1": "This is the AUSA.",
24           "description2": "Your next stop is Andrews University Dining Services.",
25           "lat": 41.96432169999999,
26           "lng": -86.3600211,
27           "heading": 34,
28           "pitch": 10
29         }
30       ]
31     }
32   ]
33 }

```

Figure 14. Open output file "data.json"

The file is constructed with many dictionaries in json format. In the main dictionary, there is a timestamp, which represents the time the scraping script was performed, as well as a list called "data." The list is compiled of 20 dictionaries that each point to their respective school. In each

school dictionary, there includes a data point for the school's name, the school's address, the school's location in terms of latitude and longitude, and a list of buildings. In the list of buildings, there are 30 different dictionaries that each contain the information for a particular building. The information includes: the name of the building marked as "title," a sentence describing the name of this building called "description1," a sentence that tells the users where the next stop is stored in "description2," the latitude and longitude of the 360 picture scraped, and the heading and pitch (also called "pov" in the html file of the website) of the location.

The experiment results show that the website brings up the correct school and the scraping script met our expectations.

5. RELATED WORK

Andri, C., et al. created a virtual tour for Management and Science University (MSU) with augmented reality and virtual reality in 2019 and recorded that the majority of the users were happy with the product [16]. Their work is more detailed and entails more time with each specific information. Our work is a more general script that can cover the tours of many schools at the same time. Perdana, D., et al. organized a virtual tour with 360-degree pictures using 3DVista for 3 buildings of Telkom University (Tel-U) in Indonesia [17]. Their 360 pictures were made using stitched-together 2D images, while the source of our 360-degree pictures were scraped from Google Street View. Similar to the augmented 360-degree panoramic safety training done by Eiris, R. et al., their works are more interactive than ours, but require more precise planning of the tour route [18]. Li, X. et al. developed a program to use 360-degree panoramic pictures from Google Street View to map urban landscapes and quantify environmental features [19]. Both of our works use Google Street View as the source of 360 pictures, but instead of analyzing them, our project tries to develop a tour with the pictures, while they try to analyze the city.

Thennakoon, M. et al. advanced a tour mobile application, using web scraping to compile information from Wikipedia for tourists [20]. Both our applications are web-based products, but their work requests information from Wikipedia, while ours is from Google Street View.

6. CONCLUSION AND FUTURE WORK

In this project, we successfully developed a python web scraper that scrapes images from Google Street View to compile a tour for a given list of universities. We also created and launched a website that makes use of the data scraped from a json file and can show the school the user wants based on a search using a search box.

The entire project can be split up into 3/4 parts: the scraping script, the GitHub server, the Google Domains DNS Server, and the HTML website program. We discussed the implementation process of all of them in Section 3. In section 4, we did a series of experiments and proved that each part can function and the website can work as planned and deliver a user's tour smoothly. The experiments also showed that we have solved the challenges effectively.

During the Covid-19 pandemic, many students are locked in their homes, unable to attend an actual university tour. The purpose of this project was to provide a solution to this problem by developing a virtual tour using 360-degree images. Through the experiments, we have proven that users can achieve this goal effectively on our website, and that the scraping script can create tours for many schools efficiently. In fact, the tour can not only be used for universities, but for other public buildings or gathering places. It is even possible to scrape other information, such as detailed descriptions, people's comments, or 2D images of each school.

However, there are still limitations to the system. For instance, although the function that checks repeats works in the majority of cases, it is still unable to detect every repeated location with just latitude and longitude data. This requires human effort to physically go through the scraping

script and delete repeated locations. Also, the quality of the 360-degree images cannot be ensured. Everybody can upload 360 pictures to the internet, and since Google Street View has been around for more than 10 years, some pictures were taken many years ago and therefore have lower image resolution and quality.

In the future, we can try to find another way to analyze and delete the repeated images, and also find ways to analyze the resolution of the images in order to maximize the tour quality. Improving the look of the website UI can also give users a better experience with the tour.

REFERENCES

- [1] Anguelov, Dragomir, et al. "Google street view: Capturing the world at street level." *Computer* 43.6 (2010): 32-38.
- [2] Rundle, Andrew G., et al. "Using Google Street View to audit neighborhood environments." *American journal of preventive medicine* 40.1 (2011): 94-100.
- [3] Gorelick, Noel, et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote sensing of Environment* 202 (2017): 18-27.
- [4] Patterson, Todd C. "Google Earth as a (not just) geography education tool." *Journal of Geography* 106.4 (2007): 145-152.
- [5] Mitchell, Ryan. *Web scraping with Python: Collecting more data from the modern web.* " O'Reilly Media, Inc.", 2018.
- [6] Lawson, Richard. *Web scraping with Python.* Packt Publishing Ltd, 2015.
- [7] Marrs, Tom. *JSON at work: practical data integration for the web.* " O'Reilly Media, Inc.", 2017.

- [8] Dabbish, Laura, et al. "Social coding in GitHub: transparency and collaboration in an open software repository." *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 2012.
- [9] Aghayari, S., et al. "Geometric calibration of full spherical panoramic Ricoh-Theta camera." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-1/W1* (2017) 4 (2017): 237-245.
- [10] Kavanagh, Sam, et al. "Creating 360 educational video: A case study." *Proceedings of the 28th Australian conference on computer-human interaction*. 2016.
- [11] Napolitano, Rebecca K., George Scherer, and Branko Glisic. "Virtual tours and informational modeling for conservation of cultural heritage sites." *Journal of Cultural Heritage* 29 (2018): 123-129.
- [12] Ashmore, Beth, and Jill E. Grogg. "Library virtual tours: A case study." *Research Strategies* 20.1-2 (2004): 77-88.
- [13] Kabassi, Katerina, et al. "Evaluating museum virtual tours: the case study of Italy." *Information* 10.11 (2019): 351.
- [14] Parsons, Thomas D., and Albert A. Rizzo. "Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test." *CyberPsychology & Behavior* 11.1 (2008): 17-25.
- [15] Burdea, Grigore, and Philippe Coiffet. "Virtual reality technology." (2003): 663-664.
- [16] Andri, Chairil, Mohammed Hazim Alkawaz, and A. Bibo Sallow. "Adoption of mobile augmented reality as a campus tour application." *Int. J. Eng. Technol.* 7 (2018): 64-69.
- [17] Perdana, Doan, Arif Indra Irawan, and Rendy Munadi. "Implementation of a web based campus virtual tour for introducing Telkom university building." *International Journal of Simulation—Systems, Science & Technology* 20.1 (2019): 1-6.
- [18] Eiris, Ricardo, Masoud Gheisari, and Behzad Esmaeili. "PARS: Using augmented 360-degree panoramas of reality for construction safety training." *International journal of environmental research and public health* 15.11 (2018): 2452.
- [19] Li, Xiaojiang, Carlo Ratti, and Ian Seiferling. "Mapping urban landscapes along streets using google street view." *International cartographic conference*. Springer, Cham, 2017.
- [20] Thennakoon, M. S. B. W. T. M. P. S. B., et al. "TOURGURU: tour guide mobile application for tourists." *2019 International Conference on Advancements in Computing (ICAC)*. IEEE, 2019.

A VIDEO NOTE TAKING SYSTEM TO MAKE ONLINE VIDEO LEARNING EASIER

Haochen Han¹ and Yu Sun²

¹Haochen Han, Troy High School, Fullerton, CA 92831

²California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

Recent coronavirus lockdowns have had a significant impact on how students study. As states shut down schools, millions of students are now required to study at home with pre-recorded videos. This, however, proves challenging, as teachers have no way of knowing whether or not students are paying attention to the videos, and students may be easily distracted from important parts of the videos. Currently, there is virtually no research and development of applications revolving specifically around the subject of effectively taking digital notes from videos. This paper introduces the web application we developed for streamlined, video-focused auto-schematic note-taking. We applied our application to school-related video lectures and conducted a qualitative evaluation of the approach. The results show that the tools increase productivity when taking notes from a video, and are more effective and informational than conventional paper notes.

KEYWORDS

Web Service, Note Taking, React JS.

1. INTRODUCTION

Note-taking is one of the most important factors in facilitating students' learning [1, 2, 3]. To be competent learners, students must be able to capture key information within fast-moving lectures [4, 5]. Traditionally, this was done by pen and paper. The recent popularization of online courses and video learning has presented a new challenge to learners worldwide, namely, how to take effective notes in a fast manner. There are existing note taking apps, but most of them are not taking advantage of the digital nature of online learning, and are merely an online version of pen and paper notes (where students type info into a digital document). ***There are two missed opportunities for existing note taking apps. First***, online videos are a brand-new learning format compared to traditional lectures in terms of their recordability and repeatability. Thus, a good note-taking app should be able to take advantage of online videos. However, this feature is not present in most modern note-taking apps. Parrotnote, however, does take advantage of the video format of online courses and streamlines the note-taking experience for online learning. ***Second***, traditional note taking applications require users to painstakingly format their notes, potentially taking away valuable learning time and distracting them from the content of their lessons. This manual formatting can also be chaotic and messy, as users might not understand what is the best format for taking notes. Thus, one may consider auto-formatting a natural solution for such issues. Taking away users' freedom to edit in exchange for organization, effectiveness, and a more traditional formatting is what sets Parrotnote apart from other note-taking applications. We believe this is the right design choice because a learner's priority is to have a format that allows written information to be easily memorized without wasting time struggling with a custom editor

(e.g., Evernote). The goal of Parrotnote is to improve the speed of note-taking and thus make video learning more effective, accessible, and convenient.

One example of a digital note taking system that has been proposed to help improve learning effectiveness is Evernote [6, 7, 8]. However, Evernote does not take full advantage of the digital nature of content and is still confined by some of the limitations of traditional documents, such as requiring students to painstakingly format their notes and constantly worry about the readability and structure of their information. This slows the learning progress, and even worse, many students may not know what note-taking format is best for them, thus creating messy notes in which information gets scrambled and becomes hard to recognize [9, 10]. Parrotnote recognizes such issues and incorporates a new approach to note-taking in response to them—full automation with limited customization. Instead of presenting the user with a blank document sheet, our application provides a clear, concise, and effective user interface that lets the user enter info. On the other hand, a more traditional approach to taking video notes would be writing down info on either a google doc or a piece of paper. These methods, however, have their own downsides beyond the formatting issues mentioned previously. Taking notes on a google doc or paper requires constantly closing or pausing the video, switching tabs, and typing new info. This can easily interrupt the concentration of the user, and potentially distract them while switching tabs [11]. Also, such methods do not take advantage of the digital nature of video. Parrotnote, however, does not require either pausing video or switching tabs. Our user simply needs to enter a link to their video. Parrotnote will then automatically fetch the video source from the web and display it inside the app, where the student can then take notes. This eliminates the chore of tab switching and helps the user stay focused. One of the most prominent developments in recent years on the subject of note-taking is the Cornell Note Format [12, 13, 14, 15]. Parrotnote's auto-format generation system is essentially designed to follow such a format to produce maximum clarity and readability.

Our tool to solve the problem of online digital note taking is a website that allows users to simply type notes while watching videos without worrying about formatting. The main advantage of this method is that students can have a fluid experience without being overly distracted from the video. Also, students will not have to worry about the formatting of their notes, which makes our tool convenient and simple to use. The core feature behind our tool is the ability to generate a digital, Cornell note formatted pdf based on the information the user types while watching the video. Users simply have to enter the video lecture link on the website for an embedded video to show up. Beside the video panel is the note panel, in which the user can type individual pieces of information. Each piece of information is indexed and time stamped, and later on is used to generate a PDF note in Cornell Note Format.

The effectiveness of our web note taking application is much greater than that of the traditional switching tabs approach or the traditional pencil and paper approach, so long as the user is familiar with how to use a computer and a website. As our first step, we conducted research on user experience, asking several users to test the effectiveness of the website. The results were mostly positive, as many users reported that they enjoyed the convenience of the note taking process, as facilitated by the website. Several other users found the different slots for different types of notes quite useful (our website has two tabs in the note taking section, one for information and one for questions). This shows that our website is user friendly. Secondly, we studied the behavior of less than a dozen users. We found that users were less likely to be worried about the format of their notes while using our website (because our website could format notes automatically) and could instead focus on grasping the key concepts of the lecture videos. Because they could focus more on the video and less on the actual formatting of the notes, their note taking speed and effectiveness greatly improved.

The rest of this paper is organized into different sections. First, Section 2 will list three challenges that we faced while implementing our ideas. The section following will explain the core features and implementation details.

2. CHALLENGES

In order to create a video note taking system to make online video learning easier, a few challenges were identified, as follows.

2.1. Challenge 1: Generation of the PDF Note

One challenge that we faced is the actual generation of the PDF note. The core idea of Parrotnote is to create a user-friendly experience of taking notes from a lecture video, as video lectures are likely to be prevalent during quarantine. We adopted a two-step solution and implemented a backend server. First, the frontend website will send all the note data to our backend server through a HTTP GET request. The server uses mustache.js, a templating language, to render a HTML string from the data. Afterward, the HTML string is sent to the frontend web as the response and is generated into a file through Javascript BLOB object.

2.2. Challenge 2: Embedding Videos into Parrotnote

Another major feature that posed a great challenge was embedding videos into Parrotnote. Parrotnote allows users to grab their favorite lecture videos from YouTube, then play them on the Parrotnote webpage while allowing notetaking through the notetaking features. To implement browser video, we had three options: 1. YouTube Embedded Video (Iframe tag in HTML), 2. saving the video on our server first, then streaming it to the user through a custom video streaming, and 3. Letting the user upload videos to the webpage. We decided to use YouTube Embedded Video, or the Iframe tag, simply because it would be easier for users to pick videos from YouTube rather than uploading them from files. There is an enter field in the application prompting users to enter their video links. After getting the video links, Parrotnote automatically generates YouTube Iframe Embedded links by grabbing the unique YouTube video IDs from the videos. We use a Regular Expression statement to grab the IDs.

2.3. Challenge 3: Implementing A Secure, Reliable Server and Providing the Note Generation Feature

There was also the challenge of implementing a secure, reliable server to actually host our website and provide the note generation feature mentioned in the first challenge. We chose not to do server side rendering, and instead used React.js to build a single page application. This is because server side rendering usually results in more overhead on the server, which is something we did not want. By reducing the overhead of the server side rendering, we cut the work of the server down to essentially two functions: serving the website, and providing the formatting note GET route. Another step we took to make our server more robust was to cloud host it on an AWS server, which provides faster internet speed.

3. SOLUTION

Parrotnote is an online note taking web application that formats note segments such that users can enter notes digitally while watching lecture videos on the same tab. While using the website, users will first be prompted to enter a YouTube Link to the lecture video they want to study. Afterward, an embedded video will appear on the right side of the screen, which users will watch.

On the left side are the note sections. Users can enter pieces of information or questions there, and the note sections will automatically be numbered. When users finish taking notes they can click on the PDF button to get their notes generated in Cornell format, and then get the html page for their notes. There are three main subsystems. The first two are at the frontend, and they comprise the system for grabbing YouTube video embedded codes from YouTube links and the system for saving note data from user input. These were coded in react.js. Finally, there is a backend server, which will open an API GET route, which receives users' note data, and generates html page strings to send to the frontend.

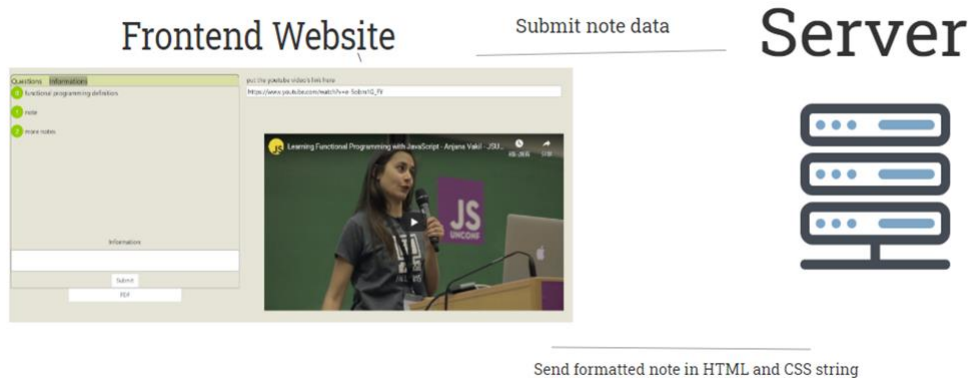


Figure 1. Parrotnote system

Backend

We used node.js as the primary language for the backend. For routing and regular items, we simply used express to handle them. The core feature of the backend is the ability to turn user submitted note data into pdf notes, save them, and serve them to users. This feature was achieved with templating language and css styling. In essence, we used html and css to first create the style for the pdf, then used Mustache.js templating language to fill data in its place.

```

{{#sections}}
<h1>Video Note</h1>
<div class="container">
  <div class="questions">
    Questions:
    <ul>
      {{#questions}}
      <br>
      <span class="time-stamp">--{{timeStamp}}</span>
      <li>
        {{text}}</li>
      {{/questions}}
    </ul>
  </div>

  <div class="main">
    Informations:
    <ol>
      {{#informations}}

```

Figure 2. Html formatting

However, one tricky issue we faced with html formatting was actually turning this into a pdf, as users may not need html files. There are many techniques to achieve this. We used a headless browser to pre-convert html to pdf before sending to the user.

```

1  const puppeteer = require('puppeteer')
2
3  async function printPDF(html) {
4      const browser = await puppeteer.launch({ headless: true });
5      const page = await browser.newPage();
6      await page.setContent(html);
7      const pdf = await page.pdf({ format: 'A4' });
8      await browser.close();
9
10     return pdf;
11 }
12 module.exports=printPDF;
13 process.on('exit', function(){
14
15 });

```

Figure 3. Code

```

app.post('/api/datatopdf',function(req,res){
  console.log("recieved data");
  console.log(req.body);
  var filePath = '';
  // const Filename = './pdfs'+req.connection.remoteAddress.toString().replace+'.pdf';
  //console.log(req.connection.remoteAddress.match(/[0-9]+\.[0-9]+\.[0-9]+\.[0-9]+/)[0].replace(/\./g, '_'));
  if(ValidateIPAddress(req.connection.remoteAddress)){
    filePath= '/pdfs/'+req.connection.remoteAddress.replace('.', '_')+'.html';}
  else{
    filePath = '/pdfs/HahaYourIPisBroken.html';
  }
  console.log(req.connection.remoteAddress);

  fs.writeFile('.'+filePath,template(req.body),//Use template to get the domstring
  function(err) {
    if(err){
      console.log(err,"error");
    }

    console.log("wrotefile");//respond and the client will send another request to the file
    res.sendFile(__dirname+filePath)
  });

```

Figure 4. Here we generate the file with a unique URL based on the user IP. Afterward, the user is redirected to the file and they can download it from their browser.

Frontend

The most important feature of our application is the system that allows users submit note data to our server and receive a HTML and CSS string for formatted notes, which the frontend eventually converts to BLOB, and downloads. The technologies we used include frontend javascript fetch and BLOB. In the backend, where the server renders a CSS formatted HTML page, we used Mustache.js as the rendering engine.

The second most important feature, note recording and data saving, was accomplished through javascript and react.js. We simply made a styled list. The program then objectifies each piece of note, and displays its data through the styled list.

To enable users to watch lecture videos on our website, we used Iframe technology and YouTube embedded videos. First, users will input the link to their video. Then, our program will use a Regular Expression to extract the unique YouTube identifier from each video (contained in query string). Lastly, our program will convert that unique identifier to an HTML iframe tag, based on YouTube video embedding iframe format.

4. EXPERIMENT

To test the effectiveness of Parrotnote, we conducted several user studies involving taking video notes. After dividing our testers into experimental and control groups, we randomly selected a YouTube video on programming in C# (<https://youtu.be/5MmhSfyO3kg>), and asked them to take notes using different techniques. Specifically, the experimental group was asked to use the Parrotnote app player to play a video and take notes using the Parrotnote editor, while the control group was asked to open a blank google doc and take notes using it. We summarized the resulting data derived from their notes and took the mean to create the following chart:

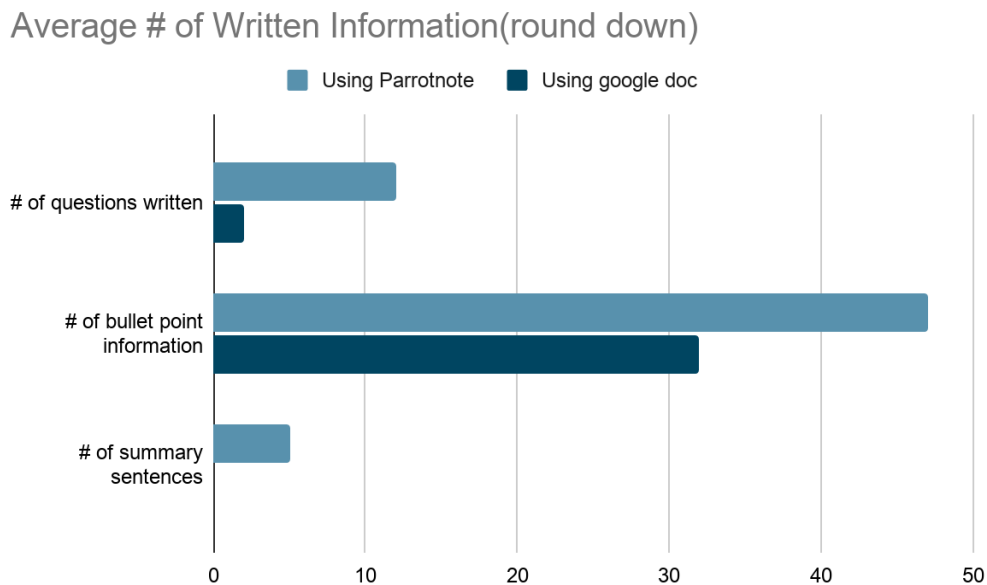


Figure 5. Chart of note-taking data

When it came to bullet point information, the experimental group (using Parrotnote) had a slightly higher number of items written down. This could likely be due to the convenient, embedded in-app video player Parrotnote offers, as it allows users to type information while watching videos without the need of pausing or switching tabs. More noticeably, Parrotnote users had a significantly higher number of both questions and summaries written down in their notes. This is likely due to the fact that Parrotnote's note editor provides a clear place for summaries and questions, thus reminding users to keep track of such information. It is also likely that when manually formatting notes using google docs, the control group users could simply get lazy and skip questions and summaries, whereas the experimental group users wrote them down because it was easier and faster to record them with Parrotnote (due to its auto-formatting feature). As the

Cornell note study points out (see also related work), summaries and questions are crucial components to help students learn.

5. RELATED WORK

Jenni Donohoo at Cornell University proposed the technique known as Cornell note [16]. It was devised to help students organize their lecture notes. This paper introduces a method by which teachers can teach students about how to make good Cornell notes. One key aspect of it is a gradual process of teaching students how to write down important information and delete irrelevant information. The strategy we employed is different. Since Parrotnote focuses on speed, students are able to write down multiple bits of information to use as review materials.

Korzaan and Lawrence have written on how Evernote pioneered the field of online, document-based note taking and cloud storage [17]. Evernote uses a document interface that allows users to format notes themselves. Parrotnote differs from Evernote in that Parrotnote automatically formats notes that users input into Cornell Note Format, instead of letting users format their notes manually.

Leann Mischel has written on the efficiency of EDpuzzle within online learning situations involving videos [18]. Our work differs from EDpuzzle in terms of method. EDpuzzle is an application that enforces video completion by locking down the browser, which may be suitable for a classroom environment. Parrotnote, on the other hand, is an application that boosts note-taking speed for self-paced learners.

6. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new way of taking notes online while watching video lectures. While there are many tools such as EDpuzzle that are suitable for classroom use, Parrotnote is suitable for self-paced learners. We have proposed an auto-formatting approach to online note taking, which will greatly boost the speed of users when combined with embedded video players and potential speech-to-text notes. Furthermore, we have conducted experiments on our auto-formatting approach, comparing people who manually format with people who use Parrotnote. The result shows that students who use Parrotnote to help them perform auto-formatting are more likely to write down key information, such as questions and summaries, while students who type on a blank document will likely only write down plain information. This shows that our auto-formatting approach helps students gain a better and a more critical understanding of their videos. The increased speed and amount of information written down also suggests that Parrotnote can be a great tool to create review sheets or study guides. In essence, Parrotnote is a tool suitable for individual learners who seek to boost their note-taking speed and general understanding of key concepts while watching lecture videos.

In this paper, we also introduced the techniques we used, including aspects from both backend and frontend. In the backend, the key feature is the pdf note generation from data, in which we apply templating language and a headless browser. In the frontend, we described the way we present the information, and how we attempted to make it intuitive. We also introduced our embedded video player feature.

Parrotnote can currently record text-based notes. It provides a division of question notes and informational notes. There are still many limitations to the current Parrotnote application. For example, the formatting does not allow customization. It is also currently limited to text-based notes.

In future, we would devise a better system that allows for greater customization of the existing note pieces. Such a system would allow for switching orders and changing font size, etc. Also, we are planning to include the styling of users' note pieces so they can choose to highlight, bold, or italicize text.

REFERENCES

- [1] Boch, Francoise, and Annie Piolat. "Note taking and learning a summary of research." *Writing*. 2005.
- [2] Barnett, Jerrold E., Francis J. Di Vesta, and James T. Rogozinski. "What is learned in note taking?" *Journal of Educational Psychology* 73.2 (1981): 181.
- [3] Di Vesta, Francis J., and G. Susan Gray. "Listening and note taking." *Journal of educational psychology* 63.1 (1972): 8.
- [4] Howe, Michael JA. "The utility of taking notes as an aid to learning." *Educational Research* 16.3 (1974): 222-227.
- [5] Horney, Mark A., et al. "Exploring the effects of digital note taking on student comprehension of science texts." *Journal of Special Education Technology* 24.3 (2009): 45-61.
- [6] Zhumabekova, G. B., and A. T. Kenzhebekova. "THE EFFECTIVENESS OF SMART TECHNOLOGY AS "EVERNOTE" IN FORMATION OF WRITTEN COMMUNICATION." *Абылай хан атындағы ҚазХҚЖӘТУ*: 58.
- [7] Ka, Jay. "Using the Mobile Application Evernote for Diagnostic Assessment to Enhance Foreign Language Proficiency." (2016).
- [8] Ansari, Mohd Shoaib, and Aditya Tripathi. "An investigation of effectiveness of mobile learning apps in higher education in India." *International Journal of Information studies and libraries* 2.1 (2017): 33-41.
- [9] Schuh, Allen J. "Effects of an early interruption and note taking on listening accuracy and decision making in the interview." *Bulletin of the Psychonomic Society* 12.3 (1978): 242-244.
- [10] Piolat, Annie, Thierry Olive, and Ronald T. Kellogg. "Cognitive effort during note taking." *Applied cognitive psychology* 19.3 (2005): 291-312.
- [11] Zureick, Andrew H., et al. "The interrupted learner: How distractions during live and video lectures influence learning outcomes." *Anatomical sciences education* 11.4 (2018): 366-376.
- [12] Donohoo, Jenni. "Learning how to learn: Cornell notes as an example." *Journal of Adolescent & Adult Literacy* 54.3 (2010): 224-227.
- [13] Quintus, Lori, et al. "The impact of the Cornell note-taking method on students' performance in a high school family and consumer sciences class." *Journal of Family & Consumer Sciences* 30 (2012): 1.
- [14] Evans, Bradley P., and Chris T. Shively. "Using the Cornell Note-Taking System Can Help Eighth Grade Students Alleviate the Impact of Interruptions While Reading at Home." *Journal of Inquiry and Action in Education* 10.1 (2019): 1-35.
- [15] Susanti, Luh Eka. "CORNELL NOTE-TAKING: EFFECTIVE WAY TO ACTIVATE STUDENTS' AUTONOMOUS LEARNING." *Journal on Studies in English Language Teaching (JOSELT)* 1.2 (2020): 43-55.
- [16] Donohoo, Jenni. "Learning how to learn: Cornell notes as an example." *Journal of Adolescent & Adult Literacy* 54.3 (2010): 224-227.
- [17] Korzaan, Melinda, and Cameron Lawrence. "Advancing student productivity: an introduction to Evernote." *Information Systems Education Journal* 14.2 (2016): 19.
- [18] Mischel, Leann J. "Watch and learn? Using EDpuzzle to enhance the use of online videos." *Management Teaching Review* 4.3 (2019): 283-289.

RATIONAL MOBILE APPLICATION TO DETECT LANGUAGE AND COMPOSE ANNOTATIONS: NOTESPEAK APP

Yingzhi Ma¹ and Yu Sun²

¹Crean Lutheran High School, Irvine, CA 92618

²California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

Students in international classroom settings face difficulties comprehending and writing down data shared with them, which causes unnecessary frustration and misunderstanding. However, utilizing digital aids to record and store data can alleviate these issues and ensure comprehension by providing other means of studying/reinforcement. This paper presents an application to actively listen and write down notes for students as teachers instruct class. We applied our application to multiple class settings and company meetings, and conducted a qualitative evaluation of the approach.

KEYWORDS

Digital learning aids, digital note-taking, note-taking mobile applications.

1. INTRODUCTION

The term “language barrier” refers to the limitations imposed between two or more people when trying to communicate using different accents, wording, or language. This is a common problem that applies to everyday situations, and also school or educational settings. In the modern world, people move or travel across continents every day, and with more than 6,500 languages worldwide, there is bound to be miscommunication [1]. Resolving these issues with an application capable of language detection and translation is an effective and popular solution.

Some of the benefits of utilizing such an application include: having indefinite access to the written information, being able to translate the content to other languages, and being able to store the audio/video for future analysis. This technology is being improved over time, but is not yet fully developed. Unfortunately, the solutions previously mentioned still have barriers to overcome before they can reach maximum efficiency. For instance, the transcribing and translation of audio may not be accurate. In other words, the devices or applications used are not yet perfected, so there exists the possibility that information or context may be lost. Nonetheless, the ultimate goal of this application is to facilitate learning for students or other interested parties by allowing them to digest information at their own pace.

Some of the app technologies that have been developed so far that are available on either Google Play or the Apple Store include: One Note, Nono Notes, Microsoft Notes, Ulysses, and Noted. However, these apps assume that the user is only focused on making and sharing the notes they create. The Notespeak App, on the other hand, is capable of doing this while also providing instant word recognition and image services as well.

Many of the aforesaid apps are actually less efficient than people may assume. In some cases, the cost of the application is greater than the standard of the service it provides. For example, Evernote does not utilize the full potential of dictation, which is usually the case for note apps. Their implementations are also limited in scale, and keyword detection with image examples are not offered either [10]. Other techniques, such as instant sharing through the platform, are not offered by apps such as Ulysses. This simplified approach does not satisfy the need of students and businesspeople who need to take notes in the moment. The methods and algorithms such companies employ are not equipped for such fast-paced environments, which is one of the major reasons the Notespeak app was developed. A second practical problem is that note app services are often not user friendly or intuitive. They have lots of information or features that complicate navigation. For example, Bearn Note looks aesthetically pleasing, but has various visual distractions.

In this paper, we trace the development of our own mobile application—Notespeak—that offers a specific set of services to attain better results than those already mentioned. Our main goal is to provide users with an audio transcription service with multiple storing and editing options, which may be expanded upon in future updates. This was inspired by a team member's desire to perform better within classes taught in languages other than his native language.

The first and most prominent feature of our app is the audio detector. It actively listens for recognizable speech patterns, identifies the language spoken, and initiates note-taking. This feature is provided through Google Audio services and has an error rate of approximately 5% [2]. The second feature compiles and organizes the data written by Google Audio service and adds images to the data collected for identified keywords. These keywords are detected using natural language processing techniques and the images are provided by Bing. The images provide context and visual aid to make the notes easier for users to comprehend. The third and last feature offered is saving or sending the recorded information to other electronic devices for storage.

The application was tested multiple times to gather concrete results on whether the included features worked properly or not. In three application scenarios, we examined how the three features mentioned previously work with different volume levels and speech patterns. First, we tested the audio detector on a recorded lecture given by a Harvard professor. It detected over 95% of the audio and provided multiple pictures to let people understand the context of the lecture much better. This is especially helpful for users whose understanding of English is not strong, since it offers the opportunity to learn from images as well. In the second scenario, the focus was on keyword detection to decide which words would have images linked to them. This required testing on diverse lectures using complex words, such as engineering and science. Adding a mode where users can select important words for consideration to improve the service in future updates is also under consideration. In the third scenario, the saving of data was tested using a couple of scenarios: saving data after users are done, and also while transcribing. As further updates are made, we would like to be able to offer temporary data storage in case of disconnection, as well as audio recording while lectures are transcribed.

The rest of this paper is organized into different sections. First, Section 2 will list three challenges that we faced while implementing our ideas. The section following will explain the core features and implementation details.

2. CHALLENGES

In order to build an application to actively listen and record notes for students' lectures, a few challenges have been identified as follows.

2.1. Challenge 1: Flutter Coding Language

One of the major challenges in creating the Notespeak App was using the Flutter coding language. Chosen originally for its compatibility with both Android and IOS devices, it was complex to advance consistently, since it does not have as much information online as other coding languages do. There were also more than a handful of bugs that had to be dealt with to get the different services Notespeak App offers up and running. After looking up documentation on various files, employing API's, optimizing code, and a couple of months of dedicated work, most of the issues were solved. Going through this experience should diminish further challenges and make future updates and development changes easier.

2.2. Challenge 2: Acquiring a Good Transcription Service

Acquiring a good transcription service was another challenge. Various services had to be researched and tested, and many unfortunately had issues. The majority had high error rates, higher than 10% in most cases, and some of them could not detect words at all. There were payment barriers too, but fortunately many of them had free testing for a limited number of words per day. Some of the API's were not fast enough and the process of evaluating API's took longer than other aspects because each test required looking up the documentation for every API along with the JSON response format. Implementing it correctly into the code took multiple tries, and on some occasions the API's were outdated or even completely dysfunctional. The last step was considerably time consuming, but Google translation services was ultimately chosen for its good audio detection rate, easy to read documentation, and fast and free transcription access.

2.3. Challenge 3: Keyword Detection Services

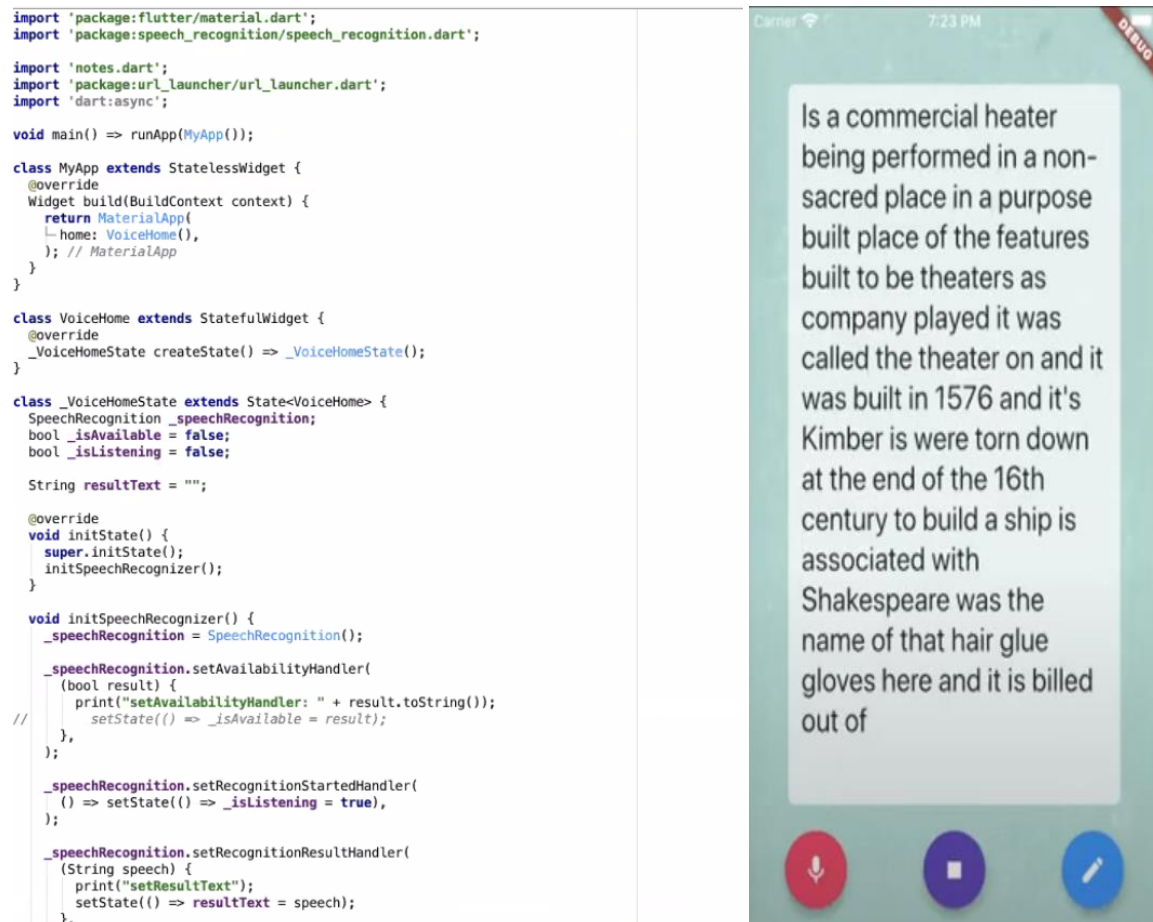
The last challenge was the keyword detection services. The keywords were sent to an image searcher. The first test for selecting keywords consisted of making a stopword list, which is a document including an alphabetized list of common and filler words. The stopword list was then used to compare it to the transcribed text and delete any words both shared. The use of API's was also considered to achieve the same effect, and much like with the transcription service, a few had to be tested and compared along with the stopword list method. In the end, the most reliable and useful was Amazon Comprehend. It uses machine learning to identify and extract key phrases from the given context of the English transcription. Since we plan to expand the language transcription features and have the app become more useful in international situations, we selected this service [8].

3. SOLUTION

Notespeak is a mobile application focused on transcription and note-taking services primarily for educational environments. The application starts when the user prompts it by clicking on a button. It then actively listens for surrounding voice input. The system then transcribes all detectable language spoken in real-time until the user indicates it should stop. The system then proceeds to analyse, detect keywords, and organize data into an organized note file that may be stored and shared. The main functions of our application include the use of proper language detection, transcription services, recording information into a text document, and identifying keywords within the recorded material that may connect to and find relevant images. The notes can be titled and stored within the user's phone with the intention of studying or referencing the material at a later time. Since Notespeak integrates the use of multiple API's offering high-quality services, it can be utilized in similar, functional contexts. Besides a classroom setting, Notespeak is perfect for conference meetings, speeches, and day-to-day conversations [9].

The main technical challenges of the system are proper information organization and display readability. We look forward to improving the current services and implementing more features in the future. Within the application scope, we look forward to adding translation, manual editing, and audio storage components.

Notespeak, as previously mentioned, was coded using the Flutter language. The figures below detail the components utilized to make it work. Figure 1 depicts the speech recognition function. It uses the dart speech recognition library to process audio input. When initialized, it actively listens through the use of the device's microphone. The data captured is then sent to a speech recognition function provided by Google [2]. It then converts this data into a string of text, which is displayed on the user's screen as seen in Figure 2.

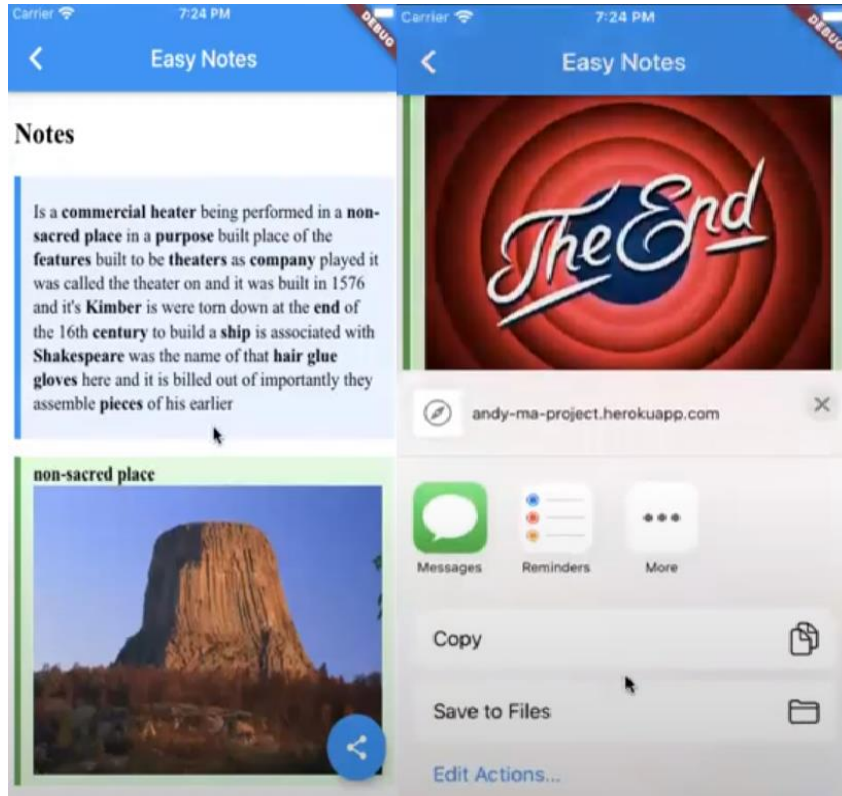


Figures 1-2. Flutter speech recognition code (left); Notespeak App UI transcription (right)

Notespeak image display is used to depict visuals for keywords. These become available once the user selects the “convert to notes” blue button, shown in the lower right of figure 2. The transcribed text is tokenized and filtered to only keep the keywords. Keywords are then sent to an image search function, where an image is rendered using Bing. A portion of code showing this utilization is shown in Figure 3. An example of how the UI displays to the user is shown in Figure 4.

```
img = search_image(name)
if img != None:
    innerHtml += '<div class="success"><center><p class = "title">' + (name.upper()) + '</p></center><br><img src = "' + img + '" /></div>'
    #innerHtml += '<div class="success"><strong>' + name + '</strong><br><img src = "' + img + '" /></div>'
```

Figure 3. Flutter image display / selection code



Figures 4-5. Notespeak App UI images of keywords (left); Notespeak App share button options (right)

A sharing function is available by clicking the share button displayed on the note transcription screen, as shown in figure 4. When the button is pressed, a sub-screen will be displayed at the bottom that provides multiple options. The copy option saves the entire text and image data to the device’s clipboard. The save to files option stores the data on the phone’s local storage, if available. The send-to-application option directly transfers data from Notespeak to a selected secondary application. The airdrop feature can send the data to devices nearby if the phone allows this. Figure 6 shows the code used to make the data display and share button features.

```

launchURL(url) async {
  // const url = 'https://flutter.io';
  if (await canLaunch(url)) {
    await launch(url);
  } else {
    throw 'Could not launch $url';
  }
}

@override
Widget build(BuildContext context) {
  return Scaffold(
    body: Container(
      child: Column(
        mainAxisAlignment: MainAxisAlignment.center,
        crossAxisAlignment: CrossAxisAlignment.center,
        children: <Widget>[
          new Flexible(
            flex: 1,
            child: Row(
              mainAxisAlignment: MainAxisAlignment.center,
              children: <Widget>[
                FloatingActionButton(
                  heroTag: "btn2",
                  child: Icon(Icons.cancel),
                  mini: true,
                  backgroundColor: Colors.deepOrange,
                  onPressed: () {
                    if (_isListening)
                      _speechRecognition.cancel().then(
                        (result) => setState(() {
                          _isListening = result;
                          resultText = "";
                        })),
                  );
                }, // FloatingActionButton
                FloatingActionButton(
                  heroTag: "btn1",
                  child: Icon(Icons.mic),
                  onPressed: () {
                    if (!_isAvailable && !_isListening)
                      _speechRecognition
                        .listen(locale: "en_US")
                        .then((result) => print('$result'));
                  },
                  backgroundColor: Colors.pink,
                ), // FloatingActionButton
              ],
            ),
          ],
        ),
      ),
    ),
  );
}

```

Figure 6. Code to create the data display and share button features

4. EXPERIMENT

Notespeak should be able to aid people in accomplishing all their note-taking goals. However, we need the features to provide efficient results, including transcription word accuracy. To evaluate transcription accuracy, we tested the detection of technical terms spoken in real time to assure that they were being transcribed appropriately within their given context. The experiment was applied to a series of files that contained audio from different settings involving individuals giving speeches in English. We transcribed three speeches by ear using a team of two people. Each excerpt had an exact duration of five minutes. The first of these was a business meeting with a single person taking the lead describing a product. The second consisted of a mathematics class lecture. The last audio was a debate between two opposing presidential candidates. The transcription done by humans was then compared to the one made by Notespeak. All missing, additional, and incorrect words were subtracted from the grand total identified by humans. The results are shown in Table 1.

Table 1. Word correctness for the three audio files

Audios	Fractional Result	Correct %
Audio #1	453/520	87.12%
Audio #2	528/566	93.29%
Audio #3	499/601	83.03%

Table 1 shows a high success rate for all three videos, averaging 87.81% overall. This indicates that Google speech recognition services are very accurate [13, 14]. When compared to studies conducted by Emil Protalinski, there was some discrepancy [2]. He indicated a 4.9% error rate

as of 2017 versus our 12.19%. This may suggest that our experiment was not executed under the best conditions, but it did approximate the expected result.

Images displayed for the notes sometimes differed from the given context. For example, the word “bat” could refer to either a flying mammal or a piece of baseball equipment. We identified the quantity of correct images given the context of the notes taken. To accomplish this, we conducted a new test utilizing the same three audio files from our first experiment. The images displayed were carefully analysed by two testers who labelled them as correct or incorrect within their given contexts. The results are shown in Table 2.

Table 2. Image correctness for the same audio files

Notes	Fractional Results	Correct %
Note #1	27/32	84.38%
Note #2	35/40	87.50%
Note #3	29/34	85.29%

The rate of overall correctness averages 85.72%. This suggests that Notespeak is reliable in producing images that match the context of the notes, which are helpful in understanding and visualizing the notes [12]. The images were selected by a search algorithm that identifies select keywords in the notes. This algorithm could possibly be further improved by utilizing some of the search strategies employed by Bing.

The experiment results demonstrate that Notespeak’s main features, transcription and image creation, are reliable but can be improved. Speech recognition services provided improved results in environments with low background/white noise. This was especially observed with our target audiences within classroom and business settings. Other scenarios also produced good results, however (see Tables 1 and 2). This correlates with the outcomes achieved by Bokhove, Christian, and Christopher, who deem a good digital transcription as achieving within the 90-percentile range in terms of accuracy [6]. This is approximately 2.2% off from our average.

Correct image selection based on keywords averaged 85.72%, when we expected it to be closer to 70% [7]. This is a good rate, although image filtering would likely further improve the selection and relation of the images to the context of the notes. We also have access to pseudocode that could improve our current algorithm in future updates.

5. RELATED WORK

Yu Fu, et al. demonstrated how Mobile Application UI is perceived by the public and designers alike, with a focused analysis on users’ preferences: “Selective user involvement which treats users mainly as information sources is adopted to efficiently incorporate users’ insights in practical UI designs” [3]. They found that UI is more impactful as it relates to certain functions, such as Multi-icon or activity, and that color patterns also have impact. This article explores multiple apps in controlled environments and compares them to one another with detailed results. Since Notespeak focuses on providing a good user interface, UI research is important. Although Notespeak is different in comparison to the apps studied by Yu Fu, et al., it does contain some of the same features [3]. Notespeak also caters to the note-taking needs of students and businesspeople.

Jolanda-Pieta van Arnhem demonstrated how Evernote, a note-taking app, offers various useful services, which include sketching, multiplatform access, and text/image/audio integration [4]. Evernote offers more services than Notespeak, such as sketching restaurant information, Notebook services, etc. [11]. Notespeak, on the other hand, focuses on the simplicity and notational aspects of quick and easy note-taking. Overall, both applications have their strengths and weaknesses. Whereas Notespeak is better suited for people who only need accessible notes, Evernote offers extra features that may be needed by others willing to pay more for them [15].

R. Ranchal, et al. studied the benefits and constraints of speech recognition between real-time captioning (RTC) and post lecture transcription (PLT) for classroom settings [5]. Note taking in PLT was executed using a video recorder, while RTC employed a note taking application similar to Notespeak. Their investigation found that PLT is better than RTC by a considerable margin for various functions, including word error rate and recognition accuracy, 22% and 78% respectively. Notespeak provided even better results, but we used updated speech recognition services available seven years after this paper was published. Notespeak's average audio recognition accuracy was found to be 87.81%. The PLT gave 85% accuracy, which is lower than Notespeak's. According to Ranchal, et al., "students felt that RTC improved teaching and learning in class as long as word recognition was greater than 85 percent and the transcription and display lag was negligible" [5].

6. CONCLUSION AND FUTURE WORK

Notespeak is a phone application used to take down live notes and save them for users to use. We experimented with users' input, errors, and feedback based on usage of the app. The results indicate that Notespeak can collect audio data and transcribe it efficiently with an average success rate of 87.81%, provided the user's phone has the minimal requirements to run the program.

The application is currently limited by the fact that its services can be impacted by outside sources. The audio recognition, for example, is heavily dependent on each phone's microphone quality, so the transcription accuracy may be impacted by this. The app is practical, but can also be affected by background noise; further testing of and search for optimal audio service continues. Optimization of data organization and image filters continues as well, since some images may still not be appropriate given the audio context.

To solve these issues, we plan to implement an error report feature to collect phone data. This data will be used to identify which microphones don't work properly and allow the app to send messages to users alerting them that the minimal technical requirements are not being met. Using an improved image search and filtering algorithm will also allow the app to select images that better fit the context of the audio transcription.

REFERENCES

- [1] Klappenbach, Anna. "Most Spoken Languages in the World 2020." Busuu Blog, 20 Dec. 2019, blog.busuu.com/most-spoken-languages-in-the-world/.
- [2] Protalinski, Emil. "ProBeat: Has Google's Word Error Rate Progress Stalled?" VentureBeat, VentureBeat, 10 May 2019, venturebeat.com/2019/05/10/probeat-has-googles-word-error-rate-progress-stalled/.
- [3] Fu, Yu, et al. "Comparison of perceptual differences between users and designers in mobile shopping app interface design: Implications for evaluation practice." *IEEE Access* 7 (2019): 23459-23470.
- [4] Van Arnhem, Jolanda-Pieta. "Unpacking Evernote: Apps for note-taking and a repository for note-keeping." *The Charleston Advisor* 15.1 (2013): 55-57.
- [5] R. Ranchal et al., "Using speech recognition for real-time captioning and lecture transcription in the classroom," in *IEEE Transactions on Learning Technologies*, vol. 6, no. 4, pp. 299-311, Oct.-Dec. 2013, doi: 10.1109/TLT.2013.21.
- [6] Bokhove, Christian, and Christopher Downey. "Automated generation of 'good enough transcripts as a first step to transcription of audio-recorded data." *Methodological innovations* 11.2 (2018): 2059799118790743.
- [7] K. Wnuk and S. Soatto, "Filtering Internet image search results towards keyword based category recognition," 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-8, doi: 10.1109/CVPR.2008.4587621.
- [8] Saindon, Richard, and Stephen Brand. "Systems and methods for automated audio transcription, translation, and transfer." U.S. Patent Application No. 11/410,380.
- [9] Cloran, Michael Eric, et al. "Real-time transcription system utilizing divided audio chunks." U.S. Patent No. 9,710,819. 18 Jul. 2017.
- [10] Viitaniemi, Ville, and Jorma Laaksonen. "Keyword-detection approach to automatic image annotation." (2005): 15-22.
- [11] Walsh, Emily, and Ilseung Cho. "Using Evernote as an electronic lab notebook in a translational science laboratory." *Journal of laboratory automation* 18.3 (2013): 229-234.
- [12] Keegan, Shobana Nair. "Importance of visual images in lectures: case study on tourism management students." *Journal of hospitality, leisure, sport and tourism education* 6.1 (2007): 58-65.
- [13] Kępuska, Veton, and Gamal Bohouta. "Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx)." *Int. J. Eng. Res. Appl* 7.03 (2017): 20-24.
- [14] Assefi, Mehdi, et al. "An experimental evaluation of apple siri and google speech recognition." *Proceedings of the 2015 ISCA SEDE* 118 (2015).
- [15] Van Arnhem, Jolanda-Pieta. "Unpacking Evernote: Apps for note-taking and a repository for note-keeping." *The Charleston Advisor* 15.1 (2013): 55-57.

AUTHOR INDEX

<i>Abderrahim Tahiri</i>	149
<i>Ahmed Al-Dubai</i>	185
<i>Ahmed Bentajer</i>	149
<i>Cong zhong Wu</i>	23
<i>Francis Lugayizi</i>	47
<i>Han tong Jiang</i>	23
<i>Hao Dong</i>	23
<i>Haochen Han</i>	221
<i>Hiroshi Unno</i>	13
<i>Hu Shaolin</i>	61
<i>Ishmael Rico</i>	69
<i>José Carlos Franceli</i>	81
<i>Kazutake Uehira</i>	13
<i>Li quan Wang</i>	23
<i>Li Xiwu</i>	61
<i>Maphuthego Etu Maditsi</i>	47
<i>Michael Esiefarienrhe</i>	47
<i>Mircea-Adrian Digulescu</i>	131, 163
<i>Mohamed N. Sweilam</i>	105
<i>Mustapha Hedabou</i>	149
<i>Nathan Ji</i>	97
<i>Nikolay Tolstokulakov</i>	105
<i>Nilupulee A. Gunathilake</i>	185
<i>Owen Lo</i>	185
<i>Peiyuan Sun</i>	207
<i>Rohan de Silva</i>	37
<i>Sara Ennaama</i>	149
<i>Silvia Novaes Zilber Turri</i>	81
<i>Su Naiquan</i>	61
<i>Sunil Seneviratne</i>	37
<i>Thomas Xiao</i>	117
<i>Thulani Phakathi</i>	47
<i>Vladimir Tregubov</i>	01
<i>Wei kai Shi</i>	23
<i>Wenhua Liang</i>	69
<i>William J. Buchanan</i>	185
<i>Xin zhi Liu</i>	23
<i>Xuan jie Lin</i>	23
<i>Yingzhi Ma</i>	229
<i>Yu Sun</i>	69, 97, 117, 207, 221, 229
<i>Zhang Qinghua</i>	61