Data Mining and Machine Learning

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 3rd International Conference on Data Mining & Machine Learning (DMML 2022)
- 11th International Conference on Software Engineering and Applications (SEAS 2022)
- 9th International Conference on Advanced Computing (ADCO 2022)
- 3rd International Conference on NLP & Information Retrieval (NLPI 2022)
- 8th International Conference on Signal Processing (SP 2022)
- 3rd International Conference on Big Data, Blockchain and Security (BDBS 2022)
- 11th International Conference on Control, Modelling, Computing and Applications (CMCA 2022)
- 8th International Conference on Computer Science, Information Technology (CSITEC 2022)

Published By

AIRCC Publishing Corporation

Volume Editors

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

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

ISSN: 2231 - 5403
ISBN: 978-1-925953-66-4
DOI: 10.5121/csit.2022.120701 - 10.5121/csit.2022.120717

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

Preface

3rd International Conference on Data Mining & Machine Learning (DMML 2022) April 23~24, 2022, Copenhagen, Denmark, 11th International Conference on Software Engineering and Applications (SEAS 2022), 9th International Conference on Advanced Computing (ADCO 2022), 3rd International Conference on NLP & Information Retrieval (NLPI 2022), 8th International Conference on Signal Processing (SP 2022), 3rd International Conference on Big Data, Blockchain and Security (BDBS 2022), 11th International Conference on Control, Modelling, Computing and Applications (CMCA 2022), 8th International Conference on Computer Science, Information Technology (CSITEC 2022) was collocated with 3rd International Conference on Data Mining & Machine Learning (DMML 2022). 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 DMML 2022, SEAS 2022, ADCO 2022, NLPI 2022, SP 2022, BDBS 2022, CMCA 2022 and CSITEC 2022. 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, DMML 2022, SEAS 2022, ADCO 2022, NLPI 2022, SP 2022, BDBS 2022, CMCA 2022 and CSITEC 2022 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 DMML 2022, SEAS 2022, ADCO 2022, NLPI 2022, SP 2022, BDBS 2022, CMCA 2022 and CSITEC 2022.

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

David C. Wyld,
Dhinaharan Nagamalai (Eds)

General Chair

Organization

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

Program Committee Members

Abdalhossein Rezai, University of Science and Culture, Iran
AbdelGhani, University of Tahri Mohamed of Bechar, Algeria
Abdelhadi Assir, Hassan 1st University, Morocco
Abderrahmane EZ-Zahout, Mohammed V University, Morocco
Abdullah Mohammed Alyateem, King Abdulaziz University, Saudi Arabia
Abdulraqeb Alhammadi, University of Technology Malaysia (UTM), Malaysia
Abhimanyu Kumar Patro, Manipal Institute of Technology, India
Abilash, Technopark, India
Addisson Salazar, Universitat Politecnica de Valencia, Spain
Adil Bashir, Islamic University of Science and Technology, India
Adnan Mohammed Hussein, Northern Technical University, Iraq
Adriana Carla Damasceno, Federal University of Paraíba, Brazil
Ahmed Farouk AbdelGawad, Zagazig University, Egypt
Ajay Anil Gurjar, Sipna College of Engineering and Technology, India
Ajit Singh, Patna University, India
Akhil Gupta, Lovely Professional University, India
Alexander Gelbukh, Instituto Politécnico Nacional, Mexico
Ali H. Wheeb, University of Baghdad, Iraq
Ali Rajabzadeh Ghatari, Tarbiat Modares University, Iran
Alireza Valipour Baboli, University Technical and Vocational, Iran
Allel Hadjali, LIAS/ENSMA, France
Amal Azeroual, Mohammed V University, Morocco
Amin Bazzazi, Islamic Azad University, Iran
Amina El murabet, Abdelmalek Essaadi University, Morocco
Ana Leal, University of Macau, China
Ana Luísa Varani Leal, University of Macau, China
Anirban Banik, National Institute of Technology Agartala, India
Anita Yadav, Harcourt Butler Technical University, India
António Abreu, Polytechnic Institute of Lisbon, Portugal
António Moreira, University of Aveiro, Portugal
Archit Yajnik, Sikkim Manipal University, India
Aridj Mohamed, University Chlef algeria, Algeria
Arvin Farid, Boise State University, Boise
Ashkan Ebadi, Concordia University, Canada
Ashraf Elnagar, University of Sharjah, UAE
Ashutosh Kumar Dubey, Chitkara University, India
Asif Irshad Khan, King Abdulaziz University, KSA
Asimi Ahmed, University Ibn Zohr, Morocco
Assem Abdel Hamied Moussa, Tech Support Systems for Egypt Air, India
Atanu Nag, IFTM University, India
Azeddine Chikh, University of Tlemcen, Algeria
Azeddine Wahbi, Hassan II University, Morocco
B Nandini, Telangana University, Nizamabad
Bandu B. Meshram, Veermata Jijabai Technological Institute (VJTI), India

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Artificial Intelligence Community (AIC)

Soft Computing Community (SCC)

Digital Signal & Image Processing Community (DSIPC)

3rd International Conference on Data Mining & Machine Learning (DMML 2022)

11th International Conference on Software Engineering and Applications (SEAS 2022)

9th International Conference on Advanced Computing (ADCO 2022)

3rd International Conference on NLP & Information Retrieval (NLPI 2022)

8th International Conference on Signal Processing (SP 2022)

3rd International Conference on Big Data, Blockchain and Security (BDBS 2022)

11th International Conference on Control, Modelling, Computing and Applications (CMCA 2022)

8th International Conference on Computer Science, Information Technology (CSITEC 2022)

TREATING CROWDSOURCING AS EXAMINATION: HOW TO SCORE TASKS AND ONLINE WORKERS?

Guangyang Han, Sufang Li, Runmin Wang and Chunming Wu

College of Computer and Information Sciences,
Southwest University, Chongqing, China

ABSTRACT

Crowdsourcing is an online outsourcing mode which can solve the machine learning algorithm's urge need for massive labeled data. How to model the interaction between workers and tasks is a hot spot. we try to model workers as four types based on their ability and divide tasks into hard, medium and easy according difficulty. We believe that even experts struggle with difficult tasks while sloppy workers can get easy tasks right. So, good examination tasks should have moderate degree of difficulty and discriminability to score workers more objectively. Thus, we first score workers' ability mainly on the medium difficult tasks. A probability graph model is adopted to simulate the task execution process, and an iterative method is adopted to calculate and update the ground truth, the ability of workers and the difficulty of the task. We verify the effectiveness of our algorithm both in simulated and real crowdsourcing scenes.

KEYWORDS

Crowdsourcing, Worker model, Task difficulty, Quality control, Data mining.

1. INTRODUCTION

With the rise of the Internet in the early 21st century, many traditional industries are undergoing changes, such is outsourcing. The first clear definition of crowdsourcing came in 2006, when Howe[1] defined it as: "Simple defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and Outsourcing it to an undefined (and generally large) network of people in the form of an open call."

Early crowdsourcing work mainly come from the field of database. Researchers package crowdsourcing platform and workers into a virtual database, and query some open or abstract questions through SQL-like query language. These queries are eventually transformed into crowdsourcing tasks, which are completed by a large number of Internet workers and fed back to the inquisitors. In recent years, with the rise of machine learning, especially deep neural network, researchers have an increasing demand for massive labeled data, and crowdsourcing has also attracted a lot of attention from data mining and machine learning field. The famous *ImageNet* dataset [2] was crowdsourced by Google through *Amazon Mechanical Turk* (*AMT*), a general crowdsourcing platform. There are other online platforms like *FigureEight* (formerly named as *CrowdFlower*), *InnoCentive, Upwork, etc.*

Crowdsourcing has various task patterns. For example, according to the task granularity, crowdsourcing can be divided into micro and macro task, such as annotating an image and

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 01-14, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120701

polishing a paper. Also, crowdsourcing could be classified into monetary, entertainment or voluntary services in terms of the incentive mechanism. *Foldit* [3] is an online game in which people match to optimize the 3d shape design of proteins, making people in the process of game also can contribute for biological sciences, and *Wikipedia*, for example, is enriched and maintained by volunteers. More, based on the purpose of crowdsourcing, individuals, collaborate and competition are three participation mode for workers in crowdsourcing. People solve independent sub-problems such as image annotation in parallel, collaborate on complex tasks such as writing a program system serially, and compete to win a creative design competition.

Crowdsourcing tasks that provide labeled data for machine learning include many forms, such as data collection, categorization, transcription, or relevance searching and sentiment analysis [4] , *etc.* Our main focus is the data annotation task, which categorizes and annotates unlabeled data. That is, tasks contain several options from which workers choose the most appropriate one. The general process for crowdsourcing can be roughly divided into three steps: requester designs and submits tasks to the crowdsourcing platform, the platform assigns tasks to appropriate workers, then collects and aggregates workers' answers and feeds back to the requester [5]. All three stages contribute to the quality of the final result. In a word, clear and intuitive task design, reasonable task assignment strategy and appropriate truth inference algorithm are the three hot spots of crowdsourcing research [6], furthermore, the last two stages all rely on pertinent worker and/or task model that can reveal the real situation.

We get the inspiration for modeling workers and tasks from school exams. There are two basic indicators in the evaluation of test questions, difficulty and discriminability. Difficulty corresponds to the overall scoring rate (correct rate) of the question. A good question should be able to distinguish between students of different levels of knowledge, that is, the high scoring rate of the good student and the low scoring rate of the student who does not work hard. The difference between the two evaluating indicator is the degree of discriminability. In general, the discriminability is highest when the problem is of moderate difficulty. In this way, we should pay more attention to medium difficult tasks when inferring workers' abilities from task results, since tasks that are too difficult or too easy contain little information about workers' abilities.

In general, we use two positive decimals to quantify task difficulty and worker ability, a probability graph model to simulate the annotating process, and comprehensively consider the impact of difficulty and ability on worker's performance in the deduction process. In particular, when inferring worker ability from the annotating results, we focus more on the performance on moderate difficulty tasks, reducing the impact of simple and difficult tasks, thus making our model of worker competence more accurate and robust. Our innovations are included as the following:

1. We draw inspiration from actual exam, using difficulty to measure tasks, and ability to measure workers. A probability graph model simulates the annotating process, which is simple and intuitive.
2. We *first* point out the effect of task discriminability on inferring workers' ability, and design a set of computational inference process to infer more accurate and robust worker model.
3. We designed simulation experiment and real experiment to verify the effectiveness and practicability of our algorithm.

2. RELATED WORK

Data mining in crowdsourcing focuses more on truth inference algorithms, and good raw data is indispensable. However, with limited budget, we need to select some from a large number of not so reliable online workers to complete our tasks, so task assignment has also attracted a lot of

research work. In order to better assign tasks or infer ground truth, we need to have a full understanding of workers and tasks, thus the modeling of workers and/or tasks is also inevitable. The modeling of workers mostly focuses on their ability level, the difference lies in different granularity, such as overall ability, or abilities in different domains, and even use matrix to express the cognitive preference of workers in various domains/options. For task modeling, besides focusing on the nature of the task itself, such as difficulty or domain, some researchers also measure the degree of completion of a task, such as uncertainty, information entropy, and so on.

2.1. Task Assignment

The core of task assignment is how to allocate tasks to appropriate workers so as to maximize the overall return [7]. Workers should have enough ability and even interest to complete the tasks assigned to them. According to whether it is assumed to know the global information of workers and tasks, such as skill level, difficulty, domain, price, *etc.*, task assignment algorithms can be divided into online and offline types. In the real crowdsourcing process, the information of workers and tasks is usually unknowable at the initial stage, so we mainly focus on online algorithms.

In the online algorithms, workers actively access the platform, and the platform has no prior information of the workers initially. The requester provides tasks to the platform, which allocates tasks when the workers arrive. Due to the lack of knowledge of the unknown workers' abilities, the assignment may be random at the beginning, but after workers returning their answers, platform can infer and learn stage by stage. With the increase of workers' answers, the platform can gradually grasp the characteristics of the workers and/or tasks, so that the subsequent task assignments can be targeted.

As far as we concern, there are the following typical work that focus on task assignment algorithms:

1. *DTA* [7]: It's a two-stage exploration-development assignment algorithm. In the exploratory stage, they used sampling methods to evaluate the ability of workers and the difficulty of the tasks. After obtaining enough information, the problem was transformed into the offline style, they further used the *Primal-Dual Formulation* to solve it.
2. *AskIt* [8]: A real-time application system for interactive crowd-data searching, which can effectively route questions to the due staff, and reduce the uncertainty of answers. The algorithm considers four constraints and two uncertainty measures of the workers and questions, and is used to solve the optimization problem. Entropy-like methods are used to quantify the uncertainty, while collaborative filtering is employed to predict the unseen answers.
3. *QASCA* [9]: It is an online task assignment framework that uses a confusion matrix to simulate workers' preferences when giving answers, and uses Accuracy and F-score to measure the quality of task completion. It gradually infers the truth value of the task and updates the worker confusion matrix, thereby dynamically assigning the task to the worker with the highest probability of giving the correct answer in the next assignment.

2.2. Truth Inference

Worker and task model are also crucial in the truth inference (answer aggregation) stages. Because building these models requires the ground truth of the tasks, which are unknown, many algorithms iteratively perform truth inference and model update calculation.

The truth inference algorithm can be divided according to whether it needs to be calculated iteratively. Non-iterative aggregation uses heuristics to calculate a single result for each task. Commonly used techniques include weighted majority voting (*WMV*) and filter-based techniques [10]. (Weighted) Majority voting gives all answers the same (different) weight, then counts, and the majority wins [11]. Filter-based methods rely on other techniques (such as qualification test) to filter out unreliable answers first, and then perform majority voting. Typical works are *HP* [12], *ELICE* [23], *BV* [5], *etc.*

The iterative aggregation algorithm performs a series of iterations and finally produces high-quality results. It usually consists of two stages: truth inference and model update. The most typical representative is the Expectation-Maximization (*EM*) algorithm [13]. In the E-stage, the truth value of the tasks is inferred based on the existing worker model, and in the M-stage, the model parameters are adjusted to maximize the occurrence probability of the current inferred result. The performance of the iterative algorithm is greatly affected by the initial parameters of the model, and is usually better than the non-iterative methods. However, the iterative process will cost a lot of time and computing resources. You can refer to the following works: *SLME* [14], *GLAD* [15], *ITER* [16], *etc.*

3. PROPOSED METHOD

We will elaborate on our proposed model and method in this section. First of all, let's introduce some common terms in crowdsourcing and the notations in this paper.

3.1. Terms and Notations

People with crowdsourcing needs will hire a group of *workers* to work for their *task* through a *website*. We call them *requester*, crowdsourcing *workers* $\mathcal{W} = \{w_1, w_2, \dots, w_m\}$, crowdsourcing *tasks* $\mathcal{T} = \{t^1, t^2, \dots, t^n\}$ and crowdsourcing *platform* respectively. The term task refers to different granularity according to the context, such as the entire task, a single sub-task or a batch of sub-task, so does the term worker. Each task t^i has a potential ground truth label l^i, and may receive some annotations $A^i = \{a_1^i, a_3^i, a_6^i, \dots\}$ from several workers $\{w_1, w_3, w_6, \dots\}$. At the same time, worker w_j may give out annotation set $A_j = \{a_j^2, a_j^4, a_j^9, \dots\}$ to tasks $\{t^2, t^4, t^9, \dots\}$. Please pay attention to the correspondence between serial numbers and subscripts in these two examples. In this paper, we prefer to use subscripts (letter j) to distinguish all m workers and superscripts (letter i) to mark total n tasks, a task may receive annotations from multiple (not all) workers, and a worker may also annotate multiple (not all) tasks. Due to the subjective or objective factors of tasks and workers, those annotations are not so reliable, we often need to annotate a task repeatedly to obtain more reliable results [17, 18]. We assume that each task gets r annotations on average, which is also the reason for the existence of truth inference algorithms. Truth inference algorithm is used to infer the ground truth set $\mathcal{L} = \{l^1, l^2, \dots, l^n\}$ from annotation set $\mathcal{A} = \sum_{i=1}^{n} A^i$ (or note as $\mathcal{A} = \sum_{j=1}^{m} A_j$).

As mentioned earlier, in data annotation task, workers need to choose one from several candidate labels which suits the task best. Intuitively, the more potential labels, the more difficult the problem will be. We assume the number of candidate labels is k. e_j is adopted to represent the potential ability of worker w_j, and d^i to represent the difficulty of task t^i. $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$ is chosen to denote the set of workers' abilities, and $\mathcal{D} = \{d^1, d^2, \dots, d^n\}$ to denote the set of tasks' difficulties. We use a probability graph model to simulate the generation process of a single annotation a_j^i, as shown in *Fig. 1.* In the following sections we will analyse in detail the reason

for the modeling of each step in *Fig. 1* and derive the calculation formulas which we rely on to establish our truth inference algorithm. According to *Fig. 1*, we can get the joint probability density formula as *Eq. 1*.

$$p(\mathcal{A}|\mathcal{D},\mathcal{E},\mathcal{L}) = p(\mathcal{P}|\mathcal{D},\mathcal{E})p(\mathcal{A}|\mathcal{P},\mathcal{L}) \qquad (1)$$

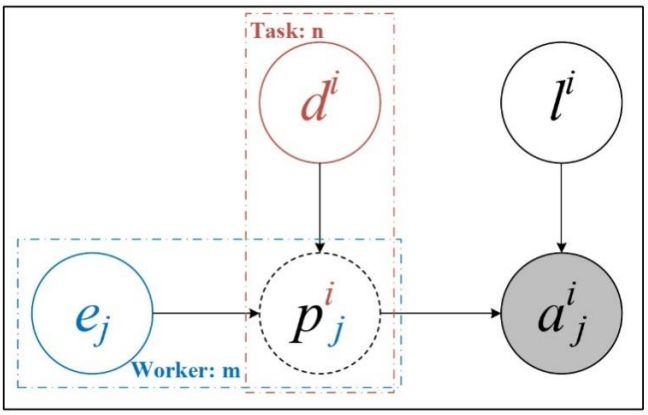


Figure 1. The probability graph model of proposed method. Above there are five elements: worker ability e_j, task difficult d^i, the probability p_j^i indicts the confidence of worker's annotation a_j^i being the same as ground truth l^i. The element in shaded circle (a_j^i) is the only variable we know, in dashed circle (p_j^i) is the intermediate variable, what we need to infer are the elements in the remaining three circles.

3.2. Worker-Task Interaction Model

Due to the open nature of the Internet, crowdsourcing workers may come from all corners around the world. They have different ethnicity, nationality, educational background, work skill, cognitive inclination, personal personality, *etc.* [19]. In order to highlight and study the different inherent performance of these workers during crowdsourcing, many worker modeling methods have been proposed, such as methods based on ability level (accuracy), methods based on ability domains, and methods based on confusion matrix, from simple to complex. [19] further classified workers to five classes based on their ability and behaviours.

Since the annotation matrix of a worker in real crowdsourcing scene is usually sparse, too complex models will suffer more from data lacking, so we employ the simplest and intuitive but practical ability model to represent the worker's ability. Each worker w_j is associated to a decimal $e_j \in [0,1]$ indicating the ability to get task right. With reference to the work of [19] we pre-established four categories of workers based on their abilities and behaviours: *expert*, *normal* worker, *sloppy* worker and *spammer*. Expert and normal worker will do their best to get things right while sloppy workers tend to give out answers imprudently and there may be very few workers deliberately give out wrong answers for various reasons, trying to sabotage the current task.

Supposing you are taking an exam, besides your own knowledge and ability, the factors that determine your final score also include the quality of test paper, such as the difficulty of the test questions, the scope of investigation, the quality of the question description, and the length of the test duration, *etc.* Reasonable task design means a lot to the quality of crowdsourcing [6], but this is beyond the scope of this paper. We synthetically use a decimal $d^i \in [0,1]$ to quantify the difficulty of task t^i, from easy to difficult.

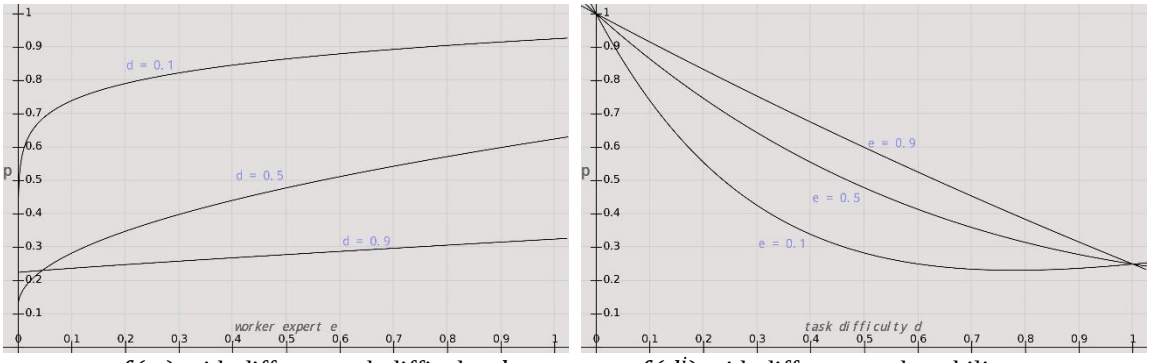
According to the experience in real life, a question may be too difficult for even experts to solve it well within limited time, identifying the authenticity of the porcelain in a picture, for example. On the other hand, even sloppy can easily answers the simplest problems, such as "*Which mountain is the highest in the world?*". Imagining what will happen if you encounter a question you don't know in an exam? picking an answer at random! The harder the question, the more difficult it is for workers to work out the correct answer, so the answers given tend to be chosen randomly. We use *Eq. 2* to express this trend:

$$p_j^i = f(d^i, e_j) = d^i \frac{1}{k} + (1 - d^i) {e_j}^{d^i} \tag{2}$$

here p_j^i is the intermediate variable showing the possibility that the worker w_j chooses what he thinks the correct answer is to task t^i with k choices. We choose *Eq. 2* because of the three important properties:

1. The function is concise, easy to calculate and understand. $f(d^i, e_j)$ can be seen as a competition between two factors, random selection and worker effort, difficulty (d^i, $(1 - d^i)$) act as weights.
2. Fixing one of the two variables, the function $f(d^i, e_j)$ basically changes monotonously with the other variable, which is in line with our intuition. Some special cases are shown in *Fig. 2*.
3. In two extreme cases, the function value meets our original assumptions. $f(e_j|d^i = 0) = 1$ and $f(e_j|d^i = 1) = \frac{1}{k}$ regardless of the value of e_j. Those mean that tasks that are too simple or too difficult are not enough to distinguish workers' ability difference.

When the difficulty of the task is 1 ($f(e_j|d^i = 1)$), *Eq. 2* degenerates into the probability of random selection, in another extreme cases ($f(e_j|d^i = 0)$), *Eq. 2* shows the all workers have full confidence to their answers. It can be seen from the *Fig. 2* that when the task difficulty is moderate (about 0.5), the distance between the confidence curves of the three types of workers is the largest, meaning that the task can distinguish workers of different abilities well at this time.



$f(e_j)$ with different task difficulty d $f(d^i)$ with different worker ability e

Figure 2. The figure on the left shows the change curve of the answer quality on three typical difficulty tasks with the ability of the workers, and the right one shows the change curve of the answer quality given by the three typical workers with the difficulty of the task. Note that we set $k = 4$ in above cases and the effective value of the horizontal coordinates is in the range of [0,1].

Intuition teaches us that the higher the credibility of the annotation given by the worker w_j, the more likely his annotation is the same as the potential ground truth of task t^i. Combining p_j^i and l^i, we can easily get the probability density function of the distribution of a_j^i as *Eq. 3*.

$$f(a_j^i = l^i) = p_j^i$$
$$f(a_j^i \neq l^i) = \frac{(1-p_j^i)}{(k-1)} \quad (3)$$

We simply assume that all wrong options are equally confusing, so we use a simple equalization process to represent the probability distribution density of the wrong answers. Although the use of more accurate models such as confusion matrix will be closer to the real situation, the improvement they bring is limited when the amount of data and information is relatively scarce [20].

$$f\left(a_j^i = l^i\right) = p_j^i = \frac{d^i}{k} + \left(1 - d^i\right){e_j}^{d^i} \approx \frac{1}{k} \quad \text{s.t. } d^i \approx 1 \quad (4)$$

3.3. Truth Inference Algorithm

In *Fig. 1*, there are four elements, of which only annotation a_j^i is what can be observed, how to infer the other three elements from the mere information becomes a problem. Generally speaking, when the value of a_j^i is given, l^i and p_j^i cannot be independent of each other. However, p_j^i is a hidden variable, and we have no idea about its value. In this case, e_j and d^i are independent of each other (see *Eq. 5, 6, 7*). These analyses are in line with our life experience. Therefore, when the value of a_j^i is observed, the three element we are concerned with: task label, worker ability, and task difficulty have some correlations. l^i depends on d^i and e_j both, while the later two are mutually independent.

Joint probability formula:

$$P(p_j^i, l^i, a_j^i) = P(p_j^i)P(l^i|p_j^i)P(a_j^i|p_j^i, l^i) \quad (5)$$

Bayesian hypothesis:

$$P(p_j^i, l^i, a_j^i) = P(p_j^i)P(l^i)P(a_j^i|p_j^i, l^i) \quad (6)$$

Compare above we have:

$$P(l^i|p_j^i) = P(l^i) \rightarrow P(l^i, p_j^i) = P(l^i)P(p_j^i) \quad (7)$$

Luckily, a task will usually be annotated by multiple workers and a worker will annotate a set of tasks. We assume that the ability of a worker is stable for a certain period of time, and the difficulty of the task is the attribute of the task itself, so we can transfer the ability of a worker on t^i task to t^j task, and the difficulty of a task inferred from the performance of worker w_p can also be used to predict the performance of worker w_q.

We propose an iterative calculation method based on the *EM* algorithm to update the values of $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$, $\mathcal{D} = \{d^1, d^2, \dots, d^n\}$ and $\mathcal{L} = \{l^1, l^2, \dots, l^n\}$. In each iteration, we use the current values of two of the three sets to update the other one, one iteration ends when the values of all three sets are updated once. In the related work section, we mentioned that the quality of the initial value is very important for the iterative algorithm. Therefore, in the following steps, we will explain in detail how to set the initial value at the start-up and how to update the above three sets.

Generally speaking, researchers will use a certain strategy to initialize the worker and task model, like random initialization or uniform initialization, and then derive the truth label set. Few people will assume the ground truths of the task to start the iteration, because this is contrary to our common sense. So, in each iteration we always use $\mathcal{E}$ and $\mathcal{D}$ to update $\mathcal{L}$ first, then $\mathcal{E}$ or $\mathcal{D}$.

3.3.1. Parameter Initialization

The work of [20] has summarized that when the amount of data is sufficient, simple models such as *MV* or *WMV* can also achieve good results, and have more stable performance and less calculation than complex models. In other words, simple algorithms such as *MV* or *WMV* tend not to be too bad and fluctuate less. So, we use simple *MV* as example to explain how to set the initial value of our model, it contains the two steps:

Count: First, for each task, we count all the annotations it receives, accumulate the number of counts for each option, and the option with the highest score is treats as the label of the task. Then we use the error rate as the initial difficulty of the task.

Weight: After obtaining the labels of all tasks, we in turn count all the annotations given by each worker, and use the correct rate as the initial value of the worker's ability.

$$d^i = 1 - \frac{\sum_{a_j^i \in A^i} \mathcal{I}(a_j^i = \hat{l}^i)}{|A^i|} \tag{8}$$

$$e_j = \frac{\sum_{a_j^i \in A_j} \mathcal{I}(a_j^i = \hat{l}^i)}{|A_j|} \tag{9}$$

In *Eq. 8&9*, $\mathcal{I}()$ is the indicator function, $\mathcal{I}(True) = 1, \mathcal{I}(False) = 0$, the hat of $\hat{l}^i$ denotes it is an approximate of the latent ground truth. If you want to get more accurate initialization parameters, or the number of annotations $r \leq k$, you can also use other iteration-based methods, such as *WMV* to initialize the parameters to avoid the embarrassing situations where there is no option to win. The main difference in *WMV* is that in the *Count* stage, when we count the score of each option, we will multiply the weight of each worker's annotation, that is, the accuracy of the worker in the current iteration. Then iteratively perform the *Count* and *Weight* steps until it stabilizes. The error rate is treated as task difficulty and correct rate as the worker ability.

In the following update process, different from the previous meaning, we use superscript on a set to mark the iteration rounds, $\mathcal{L}^{i+1} = f(\mathcal{D}^i, \mathcal{E}^i)$ means we use the task difficulty set $\mathcal{D}$ and worker ability set $\mathcal{E}$ obtained in round i to update the task label set in round $i+1$, for example.

3.3.2. Parameter Initialization

The term of the unobserved variable is called "latent variable". Let $\mathcal{X}$ denote the set of observed variables, $\mathcal{Z}$ denote the set of latent variables, Θ denote the model parameters, and $LL()$ denote the log-likelihood function. If we want to do the maximum likelihood estimation of Θ, we have *Eq. 10*. Since $\mathcal{Z}$ is unknown, we can only maximize the logarithmic marginal likelihood of the observed variables $\mathcal{X}$ by calculating the expectation of $\mathcal{Z}$, as shown in *Eq. 11*.

$$LL(\Theta|\mathcal{X}, \mathcal{Z}) = \ln P(\mathcal{X}, \mathcal{Z}|\Theta) \tag{10}$$

$$LL(\Theta|\mathcal{X}) = \ln P(\mathcal{X}|\Theta) = \ln \sum_{\mathcal{Z}} P(\mathcal{X}, \mathcal{Z}|\Theta) \tag{11}$$

In E-step, we need to infer the distribution of latent variables $\mathcal{Z}$ under current model parameters $P(\mathcal{Z}|\mathcal{X}, \Theta^i)$, and calculate the expectation of the log-likelihood function $LL(\Theta|\mathcal{X}, \mathcal{Z})$ for $\mathcal{Z}$, as in *Eq. 12*.

$$Q(\Theta|\Theta^i) = \mathbb{E}_{\mathcal{Z}|\mathcal{X},\Theta^i} LL(\Theta|\mathcal{X}, \mathcal{Z}) \tag{12}$$

In M-step, we can solve *Eq. 13* to update the model parameters Θ^{i+1} to maximize the expectation of the likelihood function $Q(\Theta|\Theta^i)$.

$$\Theta^{i+1} = \underset{\Theta}{max} Q(\Theta|\Theta^i) \tag{13}$$

Updating $\mathcal{L}$:

We first use $(\mathcal{D}^i, \mathcal{E}^i) \rightarrow \mathcal{P}^i$ to get $\mathcal{P}^i$, then take $\left(\mathcal{P}^i, \mathcal{A}\right) \rightarrow \mathcal{L}^{i+1}$ to get the distribution of $\mathcal{L}$ in $(i+1)$ turn. Those can be seen as the E-step.

Updating $\mathcal{D}$ & $\mathcal{E}$:

When we have the new distribution of $\mathcal{L}^{i+1}$, we can calculate the difficulty of all tasks ($\mathcal{D}^{i+1}$) by *Eq. 8*, showing as $(\mathcal{L}^{i+1}, \mathcal{A}) \rightarrow \mathcal{D}^{i+1}$. But how to update the worker ability is a bit different from *Eq. 9*. Remember the core idea of this paper, more than ten years of examination experience tells us that questions with different difficulty have different ability to reflect the ability of workers, we also designed the worker-task interaction function as *Eq. 2*, based on this concept. So in $(\mathcal{L}^{i+1}, \mathcal{D}^{i+1}) \rightarrow \mathcal{E}^{i+1}$, we modify *Eq. 9* to *Eq. 18*. It is not difficult to find that we also weighted the contribution of different tasks to the ability of a worker, just like weighting the contribution of different workers' annotations to a task label. *Eq. 14* shows the probability difference between two workers subject to task difficulty and *Eq. 15* is the derivative. From *Eq. 16 & 17* we can see that tasks with moderate difficulty ($d \approx 0.4$) have the highest distinguishing ability, so we should reduce the interference of extreme tasks on deriving the ability of workers. Finally, the weight is set as *Eq. 19*.

$$F(d) = f(e_1, d) - f(e_2, d) = (1-d)({e_1}^d - {e_2}^d) \;\; s.t. \;\; 1 > e_1 > e_2 > 0 \tag{14}$$

$$F'(d) = {e_2}^d - {e_1}^d + (1-x)({e_1}^d \ln e_1 - {e_2}^d \ln e_2) \tag{15}$$

$$\mathbb{E}(e_1 - e_2) = \mathbb{E}((e_2 + \tfrac{1-e_2}{2}) - e_2) = \mathbb{E}(\tfrac{1-e_2}{2}) = 0.25 \;\; s.t. \;\; 1 > e_1 > e_2 > 0 \tag{16}$$

$$F'(0.4) \approx 0 \; s.t. \; \mathbb{E}(e_1 - e_2) = 0.25 \tag{17}$$

$$e_j = \frac{\sum_{a_j^i \in A_j} \phi(d^i) \mathcal{I}(a_j^i = \hat{l}^i)}{\sum_{a_j^i \in A_j} \phi(d^i)} \tag{18}$$

$$\phi(d^i) = \begin{cases} 1 - \frac{d^i - 0.4}{0.6}, & d^i \geq 0.4 \\ 1 - \frac{0.4 - d^i}{0.4}, & d^i < 0.4 \end{cases} \tag{19}$$

We iteratively perform the above steps until the predetermined number of iteration is reached or the parameters no longer change. The final outputted $\mathcal{L}$ is the result of the tasks we inferred.

3.4. Task Assignment

Task assignment strategy is affected by many factors [5], such as platform task allocation method, modelling method, whether the budget is sufficient, the average number of workers' annotations and the average number of repeated annotations on a task, *etc*. Since our model needs to understand the task and worker information at the same time, this requires more information to build the model. At the same time, we have carried out a two-way correction on the mutual derivation between task difficulty and worker ability, which can reduce the impact of random assignment. Therefore, when the budget is not sufficient ($r < 3$), we recommend using random assignment. When the budget is relatively sufficient ($r > 5$), we can first use random assignment to build the task-worker model, then sort the workers and tasks separately, and then use the remaining budget for matching assignment.

Algorithm 1: Worker-Task Interaction Model

Input: task set $\mathcal{T}$, worker set $\mathcal{W}$, budget
Output: annotation set $\mathcal{A}$, task difficulty set $\mathcal{D}$, worker ability set $\mathcal{E}$, inferred label set $\hat{\mathcal{L}}$

while *($r < 3$)* **do**
  Random assign $\mathcal{T}$ to $\mathcal{W}$ to fill $\mathcal{A}$;
end
Initialize $\mathcal{D}^0$ and $\mathcal{E}_0$ using Eq. 8, 18, 19;
while *not converged* **do**
  Update $\mathcal{L}^{i+1}$: $(\mathcal{D}^i, \mathcal{E}_i) \rightarrow \mathcal{L}^{i+1}$ using Eq. 12;
  Update $\mathcal{D}^{i+1}$ and $\mathcal{E}^{i+1}$ using Eq. 13;
end
Sort $\mathcal{T}$ and $\mathcal{W}$ for remaining assignment;
while *($budget > 0$)* **do**
  Assign tasks in corresponding order to update $\mathcal{A}$;
end
Re-derive $\mathcal{D}, \mathcal{E}, \mathcal{L}$ and output the final $\mathcal{L}$;

4. EXPERIMENTS AND ANALYSIS

Since we have proposed a fresh worker-task interaction model (note as *WTIM*), we intend to prove the effectiveness and practicability of our algorithm in three steps. We first verify the effectiveness of our model under ideal conditions through a series of simulation experiments, then we run our algorithm on two real crowdsourcing data sets to test its practical performance. Finally, we hired colleagues around us to complete a special test to test the correctness of our hypothesis. In the real world, the actual model of crowdsourcing workers is affected by time, place, task, platform, workers' own situation and other random factors, and no worker model is confirmed to be universally correct [21]. Therefore, we believe that the method proposed by us is irreplaceable in some scenarios.

For single-choice tasks, the only evaluation indicator used is the correct rate or accuracy of the final answer. Since our truth inference method is based on an iterative calculation, in addition to the most common baseline algorithm *MV* and *WMV* in crowdsourcing, we also compared some classic iterative-based truth inference algorithms. Their brief introduction is as follows:

1. *MV*: *MV* directly uses majority vote to integrate annotations, without modeling tasks and workers.

2. *WMV*: *WMV* models each worker a weight when aggregating answers, that is, their accuracy. By iteratively updating task labels and worker ability, *WMV* can finally output a much more accurate result than *MV*, many other methods such as *ZC* [22] are based on *WMV*.
3. *ZC*: *ZC* [22] adopts a probabilistic graphical model to model the decision-making process of workers. In addition, its worker model uses the simplest accuracy and does not model the task. It also uses the *EM* algorithm to iteratively update the model parameters and task results.
4. *DS*: *DS* [13] is a classic worker model. It uses a confusion matrix to simulate the different tendencies of workers when choosing answers. The sum of each row or column in the matrix is 1, and the value of the i,j-th element represents the probability that the worker chooses item j when the true value is item i.
5. *GLAD*: *GLAD* [15] extend *WMV* in task model. Instead of treating tasks equally, *GLAD* gives each task a difficulty $d^i \in (0, +\infty)$ (the bigger, the easier), it further models workers by accuracy and the annotation distribution as $Pr(a^i = l^i | d^i, e_j) = 1/(1 + \exp^{(-d^i * e_j)})$, some way the same like us.

4.1. Simulation Experiment Setup

In the simulation experiment, we simulate the random process of a certain group of workers annotating a certain set of tasks according to certain rules. We first determine some parameters about the task, such as the number of options k, then we generate tasks following the setting in *Tab. 1*. For workers, we refer to the work of [19] and set up four groups of workers with different behaviour models. Their specific settings and ratios are shown in *Tab. 2*. We generated a total of n tasks and m workers, and their difficulties or abilities randomly fluctuate within the ability fluctuation range of their respective groups in two tables.

Table 1. Task classification and corresponding difficulty range and proportion of the simulated tasks. k is the number of options in the crowdsourcing task.

Worker Type	Difficulty Baseline	Fluctuation	Proportion
Hard	0.85	±5%	10%
Medium	(1 + 1/k)/2	±10%	60%
Easy	0.15	±5%	30%

Table 2. Worker classification and corresponding ability range and proportion of the simulated workers. k is the number of options in the crowdsourcing task.

Worker Type	Ability Baseline	Fluctuation	Proportion
Expert	0.85	±5%	30%
Normal	(1 + 1/k)/2	±10%	40%
Sloppy	1/k	+10%	20%
Spammer	0.15	±5%	10%

In crowdsourcing research, a lot of works focus on binary classification task, or True-False task, such assumptions are relatively simple but without loss of generality, they will also claim that multi-classification task can also be obtained by combining several binary classification tasks. An interesting property of the binary classification task is that the accuracy of random guessing will not deviate far from 0.5, so we can easily classify workers and correct the answers from

unqualified workers. However, in practice, many tasks have multiple alternative options, and the native binary classification algorithm is not easy to handle them well, therefore, we set up two kinds of simulation experiments, $k = 2$ and $k = 4$ respectively. According to *Tab. 1* and *Tab. 2*, we have simulated $n = 100$ tasks and $m = 10$ workers, and each task is annotated $r = 5$ times on average[1], which means that each worker will annotate 50 tasks on average.

4.2. Simulation Results & Analysis

The results of the simulation experiment are recorded in *Tab. 3*. We roughly recorded the running time of each algorithm and the accuracy of the results, and the best results are marked in bold. It can be seen from the table that the results of the methods (*WMV, ZC, DS, GLAD, WTIM*) with modeling of workers and/or tasks are far better than *MV*, and the differences between them are not obvious. Note that *WTIM* does not know the specific worker and task model in the truth inference stage. Because the model of our algorithm is the closest to the simulated worker-task model, we have achieved the best accuracy, which shows that our algorithm can indeed work well in certain scenarios. In addition, note that algorithms based on iteration and probabilistic graph models (*GLAD, WTIM*) often take a long time to stabilize, therefore, in actual crowdsourcing, quality, budget, and delay often need to be considered comprehensively [4].

Table 3. The performance of our method and various comparison methods on the simulated datasets. We calculate the algorithm accuracy and iteration time respectively. The result of the winning algorithm is highlighted in bold.

Dataset & KPI	**MV**		**WMV**		**ZC**		**DS**		**GLAD**		**WTIM**	
	ACC	Time	ACC	Time	ACC	Time	ACC	Time	ACC	Time	ACC	Time
Simulation data (k = 4)	52%	**<1s**	63%	<1s	62%	<1s	64%	<1s	66s	325s	**69%**	558s

Table 4. The performance of our method and various comparison methods on the two selected real world data sets. We calculate the algorithm accuracy and iteration time respectively. The result of the winning algorithm is highlighted in bold.

Dataset & KPI	**MV**		**WMV**		**ZC**		**DS**		**GLAD**		**WTIM**	
	ACC	Time	ACC	Time	ACC	Time	ACC	Time	ACC	Time	ACC	Time
BM(k = 2)	69.6%	**<1s**	69.6%	<1s	68.9%	<1s	69.5%	1s	**70.1%**	1154s	69.9%	2765s
AC(k = 4)	75.97%	**<1s**	76.27%	<1s	75.97%	2s	76.57%	3s	76.87%	3866s	**77.17%**	4529s

4.3. Dataset Experiment Setup

Thanks to the work of Sheng [17], Barzan Mozafari and Zheng [20], we have found some useful real world crowdsourcing dataset. Those datasets are available at here[2]. They are all about single-choice tasks, in which workers are asked to choose one label from four candidates. Brief introductions of the two data sets are as follows:

1. Barzan Mozafari's (note as *BM*) dataset: It is a binary classification task data set that contains a total of 1000 tasks to which AMT workers provide a total of 5000 annotations. All of them have ground truth labels and each task has 5 annotations on average.
2. Adult Content (note as *AC*) dataset: It contains about $100k$ annotations assigned by AMT workers to identify the adult level of the website, choosing from four classes, *G* (General Audience), *P* (Parental Guidance), *R* (Restricted) and *X* (Porn). Only 333 tasks annotated by

[1]As suggested in[15, 17, 20], the recommended value of r ranges in [3 ~ 7], and for single choice task with four options, the results keep stable for $r \geq 5$.
[2]https://github.com/ipeirotis/Get-Another-Label

workers have ground truth labels, so we only count the results of this part. Each task has around 10 annotations on average.

4.4. Dataset Results & Analysis

We run our algorithm and comparison methods on these two data sets until convergence, and their results are statistically in *Tab. 4*. As can be seen from the table, the difference between the various methods is not obvious for these two data sets, which may be explained by that the workers to complete the two set of tasks are all qualified and stable. For data set *AC*, our results and [20] 's are basically consistent, and the numerical differences may come from the inconsistency of the statistical standards on the number of annotated samples, we compare files containing gold truth and files containing labels, eventually only get 333 annotated tasks with ground truth label, the numbers of correct tasks in these true inference algorithms are all about 255. *WTIM* and *GLAD* are neck and neck in accuracy, but they both are the most time-consuming algorithms.

Another fact is that the results of *WMV* and *MV* are basically consistent on these two datasets. In other words, each task is well completed in the limit of workers' capabilities, therefore the improvement carried by various true inference algorithm is not obvious. On the other hand, this also shows that our method is feasible on real dataset, at least it does not deviate from the result of the mainstream algorithm too far.

4.5. Real World Test

We specially designed a quiz consisting of 20 four-option single-choice questions, including elementary mathematics content, computer science content and geography content, and their proportions are shown in *Tab. 1*. For students in the School of Computer Science, we think that the difficulty of these three parts is gradually increasing. We invite ten classmates to complete this quiz for free and give out scores and rankings. Interestingly, almost all the classmates successfully answered the first two parts of the quiz, and their scores on geography questions far exceeded the random choice $(12/(2 * 10) > 1/4)$. From this point of view, in real crowdsourcing applications, in addition to task assignment and truth inference algorithm, how to attract higher-quality workers and maintain their enthusiasm by good task design and incentive mechanisms are also important factors to improve crowdsourcing quality.

5. CONCLUSIONS

In this paper, we extracted the model of worker-task interaction in the process of crowdsourcing from actual exams, and transformed it into a probabilistic graphical model. We believe that the difficulty of the task and the ability of the workers will both affect the quality of the final task, and there is an interactive relationship between the two. Through simulation experiment and experiments on real datasets, we have proved that our algorithm is better than existing algorithms to a certain extent, and has irreplaceable advantages in some characteristic scenarios. Finally, experiments and quiz show that, in addition to the later data mining efforts, the early task design, incentive settings design and other humanistic efforts are equally important in crowdsourcing, in fact, good raw data can make truth inference algorithms less important.

REFERENCES

[1] Jeff, H., The Rise Of Crowdsourcing. Wired, 2006. 14(6): P. 1-4.
[2] J., D., et al. ImageNet: A large-scale hierarchical image database. in 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009.

[3] Seth, C., et al., Analysis of social gameplay macros in the Foldit cookbook. Foundations of Digital Games, 2011.
[4] Li, G., et al. Crowdsourced data management: Overview and challenges. 2017. Chicago, IL, United states: Association for Computing Machinery.
[5] Li, G., et al., Crowdsourced Data Management: A Survey. IEEE Transactions on Knowledge and Data Engineering, 2016. 28(9): p. 2296-2319.
[6] Shahzad, S.B., G. Xiaofeng and C. Guihai, General framework, opportunities and challenges for crowdsourcing techniques: A Comprehensive survey. The Journal of Systems & Software, 2020. 167(C).
[7] Ho, C. and J.W. Vaughan. Online task assignment in crowdsourcing markets. 2012. Toronto, ON, Canada: AI Access Foundation.
[8] Boim, R., et al. Asking the right questions in crowd data sourcing. 2012. Arlington, VA, United states: IEEE Computer Society.
[9] Zhengy, Y., et al. QASCA: A Quality-Aware task assignment system for crowdsourcing applications. 2015. Melbourne, VIC, Australia: Association for Computing Machinery.
[10] Luz, N., N. Silva and P. Novais, A survey of task-oriented crowdsourcing. Artificial Intelligence Review, 2015. 44(2): p. 187-213.
[11] Littlestone and Warmuth, The Weighted Majority Algorithm. Information and Computation, 1994. 108(2).
[12] Lee, K., J. Caverlee and S. Webb. The social honeypot project: Protecting online communities from spammers. 2010. Raleigh, NC, United states: Association for Computing Machinery.
[13] Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1979. 28(1).
[14] Raykar, V.C., et al. Supervised learning from multiple experts : Whom to trust when everyone lies a bit. 2009. Montreal, QC, Canada: Association for Computing Machinery (ACM).
[15] Whitehill, J., et al. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. 2009. Vancouver, BC, Canada: Curran Associates Inc.
[16] Karger, D.R., S. Oh and D. Shah. Iterative learning for reliable crowdsourcing systems. 2011. Granada, Spain: Curran Associates Inc.
[17] Sheng, V.S., F. Provost and P.G. Ipeirotis. Get another label? Improving data quality and data mining using multiple, noisy labelers. 2008. Las Vegas, NV, United states: Association for Computing Machinery.
[18] Chittilappilly, A.I., L. Chen and S. Amer-Yahia, A Survey of General-Purpose Crowdsourcing Techniques. IEEE Transactions on Knowledge and Data Engineering, 2016. 28(9): p. 2246-2266.
[19] Kazai, G., J. Kamps and N. Milic-Frayling. Worker types and personality traits in crowdsourcing relevance labels. 2011. Glasgow, United kingdom: Association for Computing Machinery.
[20] Zheng, Y., et al., Truth inference in crowdsourcing: is the problem solved? Proceedings of the VLDB Endowment, 2017. 10(5).
[21] Frénay, B. and M. Verleysen, Classification in the presence of label noise: a survey. IEEE transactions on neural networks and learning systems, 2014. 25(5).
[22] Demartini, G., D.E. Difallah and P. Cudre-Mauroux. ZenCrowd: Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. 2012. Lyon, France: Association for Computing Machinery.
[23] F. K. Khattak and A. Salleb-Aouissi, "Quality control of crowd labelling through expert evaluation," in Proceedings of the NIPS 2nd Workshop on Computational Social Science and the Wisdom of Crowds, vol. 2, 2011, p. 5.

MULTILINGUAL SPEECH RECOGNITION METHODS USING DEEP LEARNING AND COSINE SIMILARITY

P Deepak Reddy, Chirag Rudresh and Adithya A S

Department of Computer Science Engineering,
PES University, Bengaluru, Karnataka, India

ABSTRACT

The paper includes research on discovering new methods for multilingual speech recognition and comparing the effectiveness of the existing solutions with the proposed novelty approaches. The audio and textual multilingual dataset contains multilingual sentences where each sentence contains words from two different languages - English and Kannada.

Our proposed speech recognition process includes preprocessing and splitting each audio sentence based on words, which is then given as input to the DL translator (using MFCC features) along with next word predictions. The use of a Next Word Prediction model along with the DL translator to accurately identify the words and convert to text. Similarly the other approach proposed is the use of cosine similarity where the speech recognition is based on the similarity between word uttered and the generated training dataset. Our models were trained on an audio and textual dataset that were generated by the team members and the test accuracies were measured based on the same dataset.

The accuracy of our speech recognition model, using the novelty method, is 71%. This is a considerably good result compared to the existing multilingual translation solutions.

Communication gap has been a major issue for many natives and locals trying to learn or move ahead in this tech-savvy English-speaking world. To communicate effectively, it is not only essential to have a single language translator but also a tool that can help understand a mixture of different languages to bridge the gap of communication with the non-English speaking communities. Integrating a multilingual translator with the power of a smart phone voice assistant can help aid this process.

KEYWORDS

Natural Language Processing, Deep Learning, Multilingual Speech Recognition, Machine Learning, Speech to Text.

1. INTRODUCTION

In the developing country of India, it has become a common trait amongst the people, to speak in a mixture of their native language and English. The domain of recognising and translating multilingual speech is still a very less researched topic but there are a few top contenders who lead the market in translating and recognising monolingual sentences. Apart from the fact that these tools are not currently able to efficiently and accurately handle a mixture of multiple

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 15-24, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120702

languages in a single speech or text query, there are other drawbacks to the current models and tools.

In order to accurately solve the above issues and drawbacks, we have formulated a problem statement to tackle these limitations and solve the problem of multilingual speech recognition which is mentioned below.

- Recognizing the various languages used in a multilingual sentence and translating it into a single language sentence.
- Developing a model that comprehends multilingual voice navigation queries by recognizing the various languages used in a multilingual sentence and reply accordingly with a single language sentence.

India has 22 official languages and most of these are constrained to a specific state and are not spoken much outside these states. Hence, these languages are low-resourced and it is not easy to come across speech data for these languages. Adding to this, there does not exist any readily available multilingual dataset in combination with English for these languages.

One of the most popular translation and speech recognition tools currently available is the google api and its performance was tested on multilingual speech query containing Kannada and English. When the input language is chosen as English, the translator aims to translate the words using the English dictionary vocabulary. Thus, when another language (for example Kannada) is used in the sentence, that foreign word is mapped to the closest sounding English word irrespective of the meaning. For example: When the input speech is ‘How to go to Shaale’ all the English words are recognised correctly but the Kannada word ‘Shaale’ is mapped to the closest sounding English word ‘Charlotte’.

In our proposed approach, the model predicts each word based on the context and hence the entire sentence predicted so far is taken into consideration rather than recognising each word as a discrete component.

The challenges of multilingual speech recognition include the scarcity of data set for effective training and testing of the model. Limited existing literature for understanding the domain of multilingual speech recognition. One of the main obstacles is to convert the audio query to multilingual text.

2. LITERATURE REVIEW

[1]The paper is related to Speech Recognition using the LAS model, which uses the internal representations of the languages learnt by the model during training. The paper describes this method to perform better than existing single language translation and recognition models, as it combines the inferences drawn from training each language separately and then combining them to recognize monolingual sentences of various languages.

Although the paper is not directly related to our problem statement of multilingual speech recognition, the methodology used for combining multiple trained model pipeline gives us an idea of how to use DL models to train and test based on multiple language sentences. The paper has scope with respect to research on performance and working of existing speech recognition tools like Google API, Python libraries, etc and can further extend the use case towards solving the problem of multilingual translation and recognition using the same described model with a few tweaks. One major drawback of the paper’s described method, is that the models use the internal representations of each language to recognise the words spoken, whereas in reality the

languages vary in script and dialogue which are more practically applied for differentiating and recognizing the words.

[2]The paper is related to dynamic language identification and focused on the use case of a software that will help in the text-to-speech feature for applications that are developed for people who are visually challenged or have reading disabilities. This helped us in formulating the use case for our model which is regional voice assistant that can convert multilingual audio query to a single language query. In order to achieve this it was understood that language recognition is an important feature that is required for multilingual text-to-speech conversion. It is because the algorithms used in this process are different from those used in automatic language detection, since the recognition is done non synchronously on a continuous stream of texts. It mainly focused on the software component of multilingual text-to-speech. The results further indicated that for language detection algorithms, fragmentation of a piece of text is an important parameter. Tri grams provided better accuracy in language recognition as compared to single or bigram. But the limitation of this approach of changing language for another text is that since most of the users of this application are visually challenged, manually changing voice in the audio menu by following voice guidance was difficult and really time-consuming. So our proposed solution intends to build a single model that can understand the multilingual language queries.

[3]Paul Fogarassy and Costin Pribeanu in their paper 'Automatic Language Identification Using Deep Neural Networks, explored the performance of deep neural networks on the problem of Language identification. This deep neural network model works on the features extracted from short speech utterances. It was found that the proposed model using extracted form of short speech utterances outperforms the current state-of-the-art i-vector based acoustic model. From the research it was found out that when the data-set is large, the deep neural networks perform the language identification better.

The DNN outperforms the state-of-the-art models in most cases. This is when the training data for each language is more than 20 hours.

Similar approaches for our research problem may not work as desired as it is found that it is better to directly recognise the next word instead of trying to identify the language and then recognise the word.

3. DATA GENERATION

184 most common English queries were first generated but only 131 navigation related queries are chosen for this research purpose. For each of these English queries, all the possible multilingual sentences were framed resulting in a total of 412 multilingual sentences which were then POS tagged with 7 classes from English and 7 classes from Kannada. Most commonly used sentences were identified from this set for generating the speech data. A total of 64 words are picked and were then recorded by 3 different individuals and each word is recorded 10 times by each individual resulting in a total of 1920 recordings.

4. METHODOLOGY

The approach used for recognition and translation of multilingual audio query is as follows:

- The input audio wav file containing the query is split into individual wav files each containing the individual words of the query.

- These wav files are then passed to a predictive model that uses a deep learning model to map the audio to text and generate the text output for the corresponding words.
- The accuracy of the speech-to-text model is further increased using a next-word prediction model and a POS tag prediction model.
- Both these prediction models take in a sequence of words, tags respectively and use a RNN to generate the next possible 'n' words, tags that follow.
- These words, tags are then used to decrease the search space for the speech-to-text conversion model for better results.

The multilingual text query is passed through to Google Translation API to get the corresponding monolingual query which is then passed to the Search Engine to get the appropriate output. The entire process of recording the audio query to display the results is integrated into a user-friendly application.

The main constraint of this method is that the training audio and textual multilingual query dataset needed for the speech-to-text conversion model is very high (in terms of hundreds of thousands), which is not feasible with the team size and the time constraint. But this can be overcome, by expanding the dataset by generating recordings of different age groups, gender and language dialects.

This methodology also depends on the performance of the Google translation API and the Search Engine for the accuracy of the translation and output. The application depends on the storage constraints of the Database used as well as the limit on the amount of audio and textual queries that can be stored.

5. IMPLEMENTATION

5.1. Preprocessing

The wav files were converted to a numpy array using librosa where each value in the array represents the amplitude and the array was normalised between values -1 to 1. The silence factor that existed due to delay in pressing the record audio start/stop button in the beginning and end of the audio file was removed. The speed of audio files were changed by changing its frame rate such that the size of each file is 20000 numpy array length when read by librosa since each user can speak a particular word at different speeds.

5.2. Splitting of sentence

After careful analysis of a few recorded sentences, it was found that each individual word utterance was between 15,000 and 25,000 array length. It was observed that the amplitude is low between each utterance of words. Hence amplitude is used as a factor to split the audio file. Moving Average with a window size of 10,000 is used to smoothen the wav file and to clearly identify the minimas. Then, minimas were found in the smoothened signal at a window size of 15,000 array length. Thus when the original signal is sliced at these minimas, individual word utterances are obtained.

5.3. Word Predictor

[5] The Word Predictor model uses the concept of LSTM to take bags of words as input and predict the next possibly occurring words. LSTM uses the memory of previously occurring words and learns the weights of next occurring words, thus using this knowledge the model is able to

deduce the next possible words from the trained vocabulary. Sentences were tokenized and all n-gram (n=4) sequences were generated. The first three tokens were considered as features which were used to predict the fourth word. These sequences were passed to the LSTM model as input to generate the next top 'k' models. Since there was an ambiguity of prediction of first and second word, similar LSTM models with n=2 and n=3(bigram and trigram) predictors were also built.

5.4. Speech To Text

5.4.1. Methodology 1 : Deep Learning

This module receives a set of next possible words (classes) from the word predictor model and classifies the input chunk into one of these words. Pre-processed wav files of these classes were selected to train the model. The extracted mfcc (Mel Frequency Cepstral Coefficient) features were used for training the model. The neural network contains an input layer, two hidden layers and an output layer. The input layer contains 100 layers, the first hidden layer contains 200 neurons activated with ReLU, the second hidden layer contains 100 neurons activated with ReLU and the final output layer contains five neurons which is equal to the number of next possible words (given by the next word prediction module). Softmax activation function is used on the final layer.

5.4.2. Methodology 2 : Deep Learning

Cosine similarity is found between the input chunk and among all the recordings of next possible words and highest occurring class among the top 20 most similar recordings is then predicted as the next occurring word.

6. RESULTS AND DISCUSSIONS

6.1. Splitting of sentence

The splitting algorithm was tested on 30 sentences out of which 28 were correctly splitted into respective words as shown in the table. The incorrectly splitted sentences were re-recorded with sufficient gaps between words after which the splitting was done accurately.

6.2. Word Prediction

Word Predictor model gave 90% accuracy and when predicted top 5 possible next words for a current sentence, the desired word was present in this predicted possible words.

6.3. Speech to text

6.3.1. Methodology 1: Deep Learning

The average accuracy for each sentence was calculated by taking the accuracy of models while predicting each individual word of that corresponding sentence. Model Accuracy is the average accuracy of all the sentences tested. Prediction Accuracy is the number of words correctly predicted divided by the total number of words present.

6.3.2. Methodology 2 : Based on similarity of signal

Average similarity of each class is also found and the class with highest average is then predicted as the next occurring word. The second approach was to find similarity between the input chunk and among all the recordings of next possible words. Highest occurring class among the top 20 most similar recordings is then predicted as the next occurring word. The accuracy for both the methods is as shown in the table where it was calculated by taking the ratio of number of correctly predicted words by total number of words in the sentence.

The average accuracy of 0.59 was achieved when prediction was done using average similarity of each class and 0.64 was achieved when using the highest occurring class among the top 20 most similar recordings.

7. NOVELTY APPROACH

The concept of using a multilingual Next Word Prediction model in accordance with the DL translator is a novel approach used to tackle the problem of translation.

The input to the Word Predictor is a sequence of textual multilingual words, that is used to train the predictor to analyse and predict the next possible 5 words using the knowledge of prior occurrence of words in sentences. These 5 words are then provided as input to the DL method, which uses these 5 words to compare and figure out the word utterance rather than comparing it with the entire vocabulary.

This concept helps decrease the time for translation by reducing the DL translator search corpus and also increase the accuracy of the translation.

8. FIGURES AND TABLES

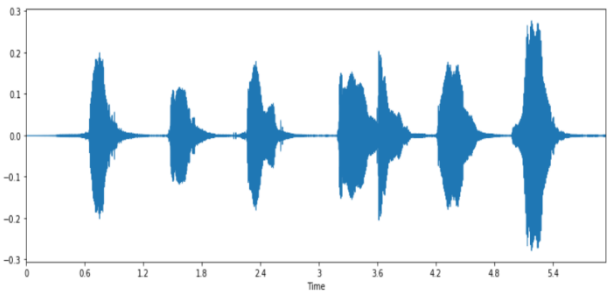


Figure 1. wav file of audio query recording

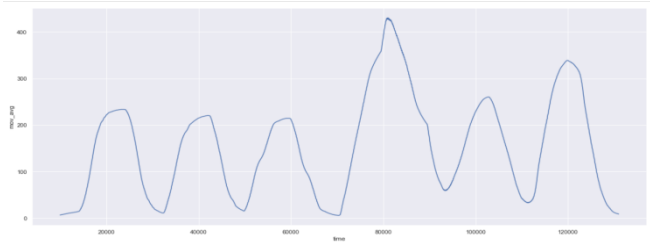


Figure 2. Wav file after smoothening which helps clearly identify individual words present in the audio query

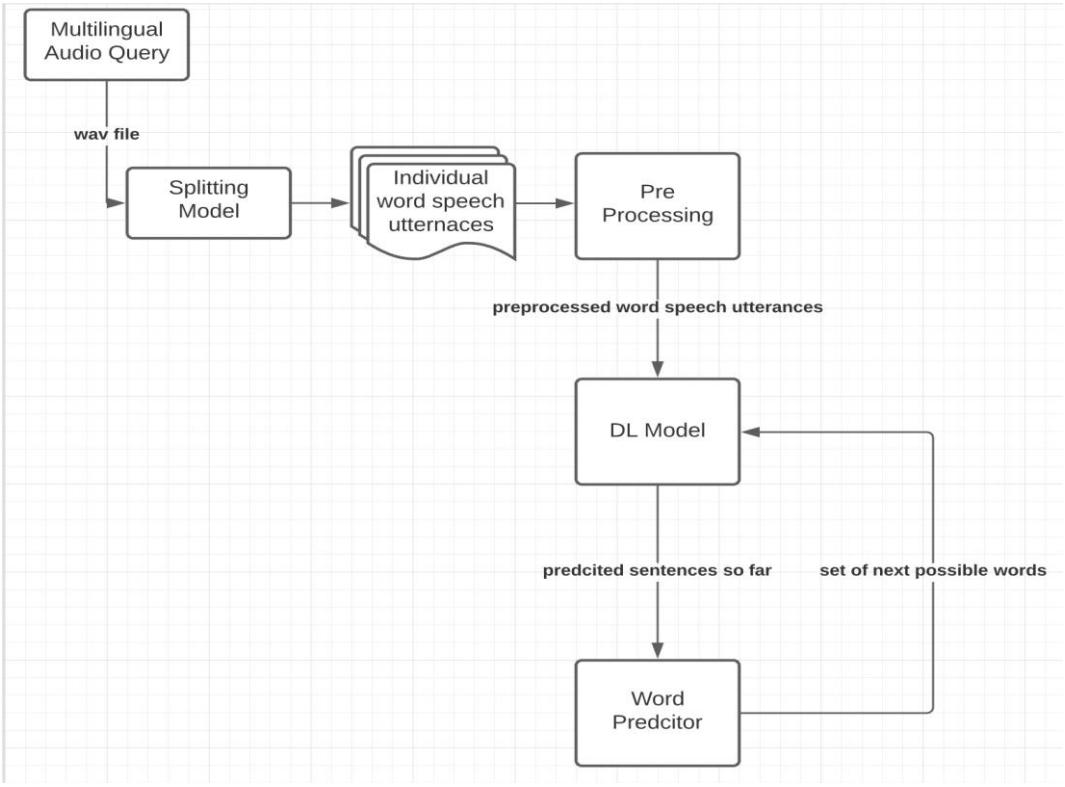


Figure 3. Flow chart of proposed methodology

Table 1. Test results of splitting module

Tested Sentence	Actual Number of Words	Predicted Number of Words
nanna next turn yenu	4	4
what is my next saradi	5	5
what is my mundina saradi	5	5
how is the datthane ahead	5	5
munde traffic hegidhe	3	3
how is the datthane munde	5	5
what are the nearby anila stations	6	6
what are the nearby gas kendragalu	6	6

Table 2. Test results of speech to text conversion using deep learning

Actual Sentence	Predicted Sentence	Average Accuracy
nanna next turn yenu	nanna next turn yenu	0.83
what is my next saradi	what is my next saradi	0.71

what is my mundina saradi	what aagothe any stalagalu saradi	0.68
how is the datthane ahead	how service the yaavaaga ahead	0.73
munde traffic hegidhe	munde traffic hegidhe	0.78
how is the datthane munde	how service the yaavaaga ahead	0.79
what are the nearby anila stations	what does the nearby anila stations	0.84
what are the nearby gas kendragalu	what aagothe the nearby gas kendragalu	0.84
what are the nearby anila kendragalu	what aagothe the nearby good kendragalu	0.85

Table 3. Test results of speech to text conversion using similarity of signal

Actual Sentence	**Similarity Predicted Sentence**	**Weights Predicted Sentence**	**Similarity Accuracy**	**Weights Accuracy**
nanna next turn yenu	nanna next turn yenu	nanna next turn yenu	1	1
what is my next saradi	what is any next yenu	how is my next saradi	0.6	0.8
what is my mundina saradi	are is any mundina yenu	what is my mundina yenu	0.4	0.8
how is the datthane ahead	what is some yaavaaga ahead	what aagothe my nearest ahead	0.4	0.2
munde traffic hegidhe	are traffic hegidhe	are traffic hegidhe	0.67	0.67
how is the datthane munde	are is my nearest munde	what is some nearest munde	0.4	0.4
what are the nearby anila stations	what is the nearby gas stations	how is some nearby gas kendragalu	0.67	0.17
what are the nearby gas kendragalu	what is some nearby gas kendragalu	what is some nearby gas kendragalu	0.67	0.67
what are the nearby anila kendragalu	what is the nearby gas kendragalu	what is the nearby gas kendragalu	0.67	0.67
hathira gas stations yavdu	what iro stations yavdu	what gas stations yavdu	0.5	0.75

Table 4. Accuracy of pre-existing models

Pre existing models	**Accuracy**
CMU Sphinx(HMM model trained and tested with our data)	57%
Similarity measure using Neural Network	64%
Google Translate(Kannada words recognition)	35.4%

Table 5. Accuracy of our deep learning and similarity models

Our Proposed Models	Accuracy
Deep learning model(using MFCC features)	71%
Deep learning model(using MFCC features, for kannada words recognition)	66.6%
Similarity model(using highest average similarity of each class)	0.59%
Similarity model(using the highest occurring class among the top 20 most similar recordings.)	0.64%

9. CONCLUSIONS

The methodology proposed in this paper is a completely novel approach which uses the self-learning abilities of a model to recognize and translate the multilingual queries to a monolingual query, as accurately as possible. The Deep Learning model uses the top predictions from the Word Predictor model to reduce the search space while identifying and translating each word input from the audio query. The new deep learning model, when tested on the generated multilingual dataset, gives an accuracy of 85%. And when it is tested live by the user, it gives an accuracy of 71%.For cosine similarity model The average accuracy of 0.59 was achieved when prediction was done using average similarity of each class and 0.64 was achieved when using the highest occurring class among the top 20 most similar recordings.

10. LIMITATIONS

The main drawback of our proposed method is that, the DL model runs every single time that a new word is predicted and given as input to the DL model, thus for recognising one sentence the model can be quite time consuming. Another limitation of our method is that it is heavily dependent on the performance and accuracy of the Word Predictor model, since the output of the next possible words is given as the input to the DL model. Since the predicted words do not have a probability attached to them, it is not possible to obtain a metric about the certainty of occurrence of the predicted words in the sentence.

11. FUTURE WORK

There are a few improvements that can be added which could further increase the accuracy of our proposed model :

- Increasing the data set, both audio and textual, with a variation in voice such as age, gender, noise level, etc.
- The similarity approach can be improved using similarity index comparison of the spectrograms of the words, instead of the previously mentioned method which compares the wav forms.
- The prediction of words and their parts of speech in a particular sentence by leveraging different and more useful language features

ACKNOWLEDGEMENTS

We would like to express our gratitude to PES University for providing us with continuous support and encouragement.

REFERENCES

[1] Multilingual Speech Recognition with a single end-to-end model - Shubham Toshniwal, 15th February 2018
[2] Multilingual Text-to-Speech Software Component for Dynamic Language Identification and Voice Switching ,September 2016 Paul Fogarassy,Costin Pribeanu
[3] Automatic Language Identification Using Deep Neural Networks,2016, Ignacio Lopez-Moreno, Javier Gonzalez , Dominguez, Oldrich Plcho.
[4] The research of feature extraction based on MFCC for speaker recognition,2014,Zhang Wanli,Li Guoxin
[5] LSTM Neural Networks for Language Modelling,2012,Martin Sundermeyer, Ralf Schlu ̈ter, and Hermann Ney

AUTHORS

P Deepak Reddy
Final Year Student Engineering studying at PES University, Bengaluru.

Chirag Rudresh
Final Year Student Engineering studying at PES University, Bengaluru.

Adithya A S
Final Year Student Engineering studying at PES University, Bengaluru.

GDGRU-DTA: PREDICTING DRUG-TARGET BINDING AFFINITY BASED ON GNN AND DOUBLE GRU

Lyu Zhijian, Jiang Shaohua, Liang Yigao and Gao Min

College of Information Science and Engineering,
Hunan Normal University, Chang Sha, China

ABSTRACT

The work for predicting drug and target affinity(DTA) is crucial for drug development and repurposing. In this work, we propose a novel method called GDGRU-DTA to predict the binding affinity between drugs and targets, which is based on GraphDTA, but we consider that protein sequences are long sequences, so simple CNN cannot capture the context dependencies in protein sequences well. Therefore, we improve it by interpreting the protein sequences as time series and extracting their features using Gate Recurrent Unit(GRU) and Bidirectional Gate Recurrent Unit(BiGRU). For the drug, our processing method is similar to that of GraphDTA, but uses two different graph convolution methods. Subsequently, the representation of drugs and proteins are concatenated for final prediction. We evaluate the proposed model on two benchmark datasets. Our model outperforms some state-of-the-art deep learning methods, and the results demonstrate the feasibility and excellent feature capture ability of our model.

KEYWORDS

Drug-Target Affinity, GRU, BiGRU, Graph Neural Network, Deep Learning.

1. INTRODUCTION

So far, due to the bottleneck of technological development, the development of new drugs is more difficult, and the exploration of new uses of developed drugs has become a new hot spot. Discovering new associations between drugs and targets is critical for drug development and repurposing, However, the traditional study of drug-protein relationships in the wet laboratory [1][2] is time-consuming and expensive due to the huge range of chemical spaces to be searched, to solve this problem, some virtual screening(VS) has been proposed to accelerate the experimental drug discovery and reposition studies in silico [3], some of the more commonly used VS methods, like structure-based VS, ligand-based VS and sequence-based VS have contributed to drug development to a large extent [4][5]. However, these VS methods have their own defects in application. For example, if the structural information of the protein is unknown, the structure-based approach cannot play its role. There is still a long way to go before accurately constructing the structure of proteins, to this end, some structure-free methods have sprung up.

In recent years, with the development and maturity of deep learning technology and its great breakthroughs in the field of computer vision(CV) and natural language processing(NLP) [6][7], many people in the field of drug research have begun to turn their attention to deep learning. Moreover, with the advent of more and more biological activity data, a great deal of work based on these data has been carried out to investigate the relationship between drugs and targets. These works are usually divided into two categories, one is a binary classification-based approach, that

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 25-37, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120703

is, to determine whether a drug and a target interact, and the other is a regression-based approach, which describes the relationship between the drug and the target by binding tightness. In binary classification-based drug-target (DT) prediction tasks, deep learning technologies seem to be used by more researches to deal with drug-target interactions (DTIs) problems. When doing DTIs prediction tasks in the past, compounds and proteins are represented using manually crafted descriptors and the final interaction prediction is made through several fully connected networks [8][9]. The problem with this approach is that the descriptors are designed from a specific perspective, that is, the design angle is too single, in addition, it remains fixed during the training process, so it cannot learn and adjust according to the results, and thus cannot extract task-related features. Therefore, some end-to-end models are proposed. Du *et al*. proposed a model called wide-and-deep to predict DTIs [10]. A generalized linear model and a deep feed-forward neural network are integrated to enhance the precise of DTIs prediction. Molecular structural information is also of great significance for feature extraction, to learn the mutual interaction features of atoms in a sequence, Shin *et al*. proposed a Transformer-based DTI model [11], which uses multi-layered bidirectional Transformer encoders [12] to learn the high-dimensional structure of a molecule from the Simplified Molecular Input Line Entry System (SMILES) string. Some researchers obtain structural information of compounds or proteins from another perspective, they represent the corresponding compounds or proteins as graphs and use graph neural networks to extract their spatial features, related work such as GraphCPI [13], Graph-CNN [14], etc.

However, the above methods have common defects, since it is a binary classification problem, the result is only yes or no, and so the distinction between compound-protein pairs is indistinguishable. In addition, many binary classification-based methods are based on setting a specific threshold as the basis of whether the drug and target interact or not. If the predictive value is higher than the threshold, it is considered interactive, otherwise it is not interactive. The deficiency of this design method is that the interaction information of many DT pairs is ignored and a proportion of these neglected information are actually significant for drug repurposing and discovery. In addition, the rationality of the threshold setting is also a factor that needs to be fully considered. Compared with the binary classification model, it seems more convincing to describe the relationship between drug and target through a regression task, the use of regression model can provide us with more information about the relationship between compounds and proteins, since continuous values can tell us how strongly the two are bound. What's more, the development of deep learning has also largely facilitated the affinity prediction of DT pairs. Related studies include KronRLS [15] and SimBoost [16], both of which based on regression and utilized the similarity information of drugs and targets to predict DTAs. DeepDTA [17] is the first framework for predicting drug and target affinity based on deep learning, which utilizes two CNN blocks to process SMILES strings of drugs and amino acid sequences of proteins, respectively. Works related to DeepDTA include WideDTA [18] and AttentionDTA [19]. The improvement of WideDTA over DeepDTA is that it combines several characters as words and proposes a word-based sequence representation method. The novelty of AttentionDTA compared to DeepDTA lies in that it proposes an attention mechanism for learning important parts of each other's sequences. In order to better capture the topological structure features of compounds, Nguyen et al. proposed GraphDTA [20] to predict drug and target affinity which utilizes RDKit technology to represent drug string sequences into graphs that could reflect its structural characteristics, and uses graph convolutional neural network to extract its spatial features. Furthermore, Lin proposed a similar approach called DeepGS [21], which uses advanced techniques to encode amino acid sequences and SMILES strings. DeepGS also combines a GAT model to capture the topological information of molecular graph and a BiGRU model to obtain the local chemical context of drug.

In this paper, we proposed a novel framework to predict DTAs. In most of the current DTA prediction research, the feature extraction of protein sequences is still dominated by CNN, this method considers the local correlation of sequences. However, most protein sequences are very long, so there are context dependences [22] in the sequence, and if we want to use CNN to capture these dependencies, then we need to use a large number of network layers. In contrast, GRU/BiGRU can capture the context dependencies of long sequences without using a large number of network layers due to its properties. Therefore, in the processing of protein representations, we interpret proteins as context-dependent time series and use GRU/BiGRU to capture the long-term dependencies of it. In the process of drug feature extraction, like GraphDTA, we still use graphs to represent drugs, and use two new graph convolution methods, namely GatedGraph and Transformer, to extract structural features of drugs. Of course, the four graph convolution methods mentioned in GraphDTA are also included for comparative experiments. Experimental results demonstrate that our model greatly improves the performance compared to previous models.

2. MATERIALS AND METHODS

2.1. Datasets

In our experimental evaluation, we used the two datasets most commonly used in DTAs prediction, namely Davis [23] and KIBA [24]. The Davis dataset contains 72 compounds and 442 proteins, along with their corresponding affinity values, where the affinity values are measured by K_d values (kinase dissociation constant) and the average length of SMILES strings for compounds is 64 and that of amino acid sequences is 788. There are a total of 30056 affinity values in Davis, and they range from 5.0 to 10.8. We convert K_d into the value of the corresponding logarithmic space, pK_d, as follows:

$$\mathrm{p}K_{\mathrm{d}} = -\log_{10}\left(\frac{K_{\mathrm{d}}}{10^{9}}\right) \tag{1}$$

The KIBA dataset contains 2116 compounds and 229 proteins, as well as 118,254 drug and target affinity values, where the affinity values range from 0.0 to 17.2. The average length of SMILES strings for compounds in KIBA is 58 and the average length of amino acid sequences is 728. The data information is summarized in Table 1.

Table 1. Summary of the benchmark datasets

Datasets	Compound	Protein	Affinity	AC	AP	DTAsRange
Davis	72	442	30056	64	788	(5.0,10.8)
KIBA	2116	229	118254	58	728	(0.0,17.2)

In Table 1, AC means the average length of compound strings, AP means the average length of protein amino acid sequences.

2.2. Overview of the proposed model

In this section, we will introduce an overview of our model. As mentioned earlier, GDGRU-DTA consists of three parts: GNN block, GRU/BiGRU block, and prediction block. After the SMILES strings of the drugs and the amino acid sequences of the proteins are given, these data are

preprocessed and converted into the corresponding graph representation and feature matrix. Subsequently, the GNN block is used to extract the features of the graph representation of the drug, and the GRU/BiGRU block is used to extract the feature matrix of the protein. Finally, the extracted features of drugs and proteins are concatenated and input to the prediction block for final prediction. The overall flow of GDGRU-DTA is depicted in Fig. 1.

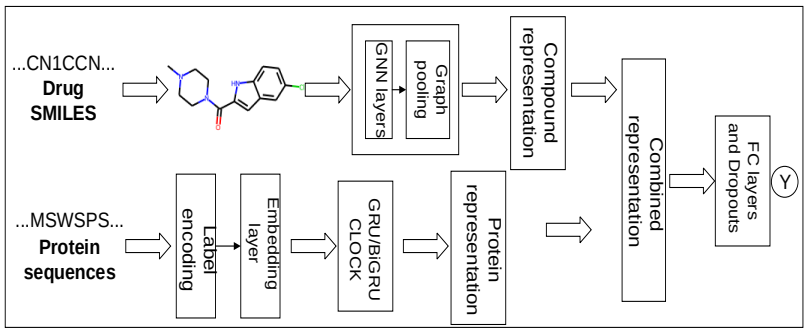


Fig. 1. Overall flow of GDGRU-DTA.

In Fig. 1, the drug and target are converted into corresponding feature representations, which are then input to the corresponding feature extraction model for feature extraction. Finally, the two extracted features are concatenated for final prediction.

2.2.1. Data Preprocessing

The feature extraction of drugs and targets are two independent input channels. Before drugs and targets are input into their respective feature extraction blocks, data preprocessing is required for drugs and targets, respectively. The implementation details are as follows.

2.2.1.1. Drug representation

For data preprocessing of drugs, we use the same method as GraphDTA, we use the open source technology RDKit to convert the SMILES strings of drugs into corresponding 2D molecule graphs. The molecule graph is denoted as $G = (V, E)$, and the vertexes V are represented as atoms and the edges E are represented as bonds, where $|V| = N$ is the number of nodes in the graph and $|E| = N^e$ is the number of edges. Each atom is embedded with 78-dimensional features such as the atom's type, degree, implied valence, aromaticity, and the number of hydrogen atoms attached to the atom. The feature of the node is encoded as a one-hot vector of shape $(N, 78)$. The chemical bonds index is encoded as $(2, E)$ vector, which is used to store the edges of the undirected graph.The schematic diagram of the SMILES string of a drug converted into a two-dimensional molecule map by rdkit technology is as follows:

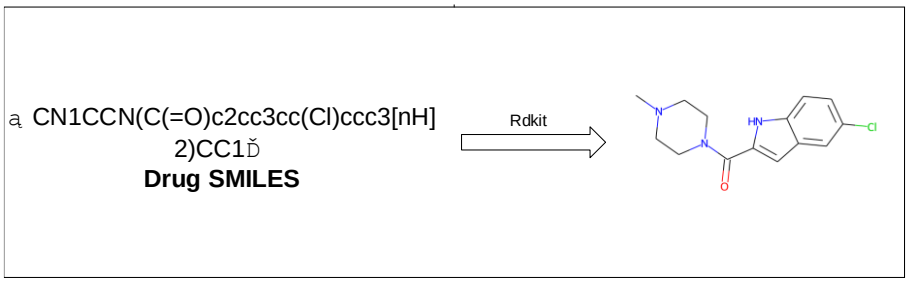


Fig. 2. Convert SMILES string to graph.

2.2.1.2. Target representation

The sequence length of each protein is different and varies greatly. For uniform feature representation, we fix the length of all protein sequences as 1000 according to the average length of protein sequences, if the sequence length of the protein exceeds 1000, the part more than 1000 will be cut off, and otherwise, the part less than 1000 will be padded with 0. In addition, since protein sequences are represented by different combinations of 25 amino acids, each represented by the one-letter code. We map each amino acid to an integer, and each integer is embedded as a 128-dimensional feature.

2.2.2. GNN Blocks

In the graph neural network block, we use two graph convolution algorithms to extract the 2D molecular graph features of drugs, namely GatedGraph and Transformer, and their details are as follows.

2.2.2.1. GatedGraph

GatedGraph [25] is a feature learning technique that studies graph-structured inputs, it modifies previous graph neural network work using gated recurrent units (GRU) and modern optimization techniques, and then extends to output sequences, so this method can make full use of long-distance information and fit well with our model of extracting protein features. In addition, GatedGraph has favorable inductive biases relative to purely sequence-based models when dealing with graph structure problems, and thus is a flexible and widely useful class of neural network models. The features of the node are updated as follows:

$$\mathrm{h}_{\mathrm{i}}^{(0)} = \mathrm{x}_{\mathrm{i}} \,||\, 0 \tag{2}$$

$$\mathrm{m}_{i}^{(l+1)} = \sum_{j \in N(i)} \mathrm{e}_{j,i} \cdot \Theta \cdot \mathrm{h}_{j}^{(l)} \tag{3}$$

$$h_{i}^{(l+1)} = \mathrm{GRU}\left(m_{i}^{(l+1)}, h_{i}^{(l)}\right) \tag{4}$$

Where in formula (2), $\mathrm{h}_{\mathrm{i}}^{(0)}$ is the input state, $\mathrm{x}_i \in \mathrm{R}^F$ is the feature of node i, $\mathrm{x}_i \,||\, 0$ represents padding 0 after feature x_i to the specified dimension. In formula (3), Θ is the parameter matrix to be learned, that is, the aggregation information of surrounding nodes. Formula (4) is to use a GRU unit to take the above two formulas as input and get an output, which can be functioned as a new feature of node i.

2.2.2.2. Transformer

Transformer [12] is a model proposed by Google researchers for seq2seq tasks, the special feature of Transformer is that it uses a lot of special layer such as self-attention in the model. Transformer breaks through the limitation that RNN models cannot be computed in parallel. Compared to CNN, Transformer does not grow with distance in the number of operations required to compute the association between two locations, and finally, self-attention can lead to more interpretable models. TransformerConv is a graph convolution method based on transformer idea [26], which takes into account the case of edge features by adopting Transformer's vanilla multi-head attention into graph learning and achieves ideal results. The feature extraction of the node is as follows:

$$x_i' = W_1 x_i + \sum_{j \in N(i)} \alpha_{i,j} W_2 x_j \quad (5)$$

Where the attention coefficients $\alpha_{i,j}$ are computed via mult-head dot product attention:

$$\alpha_{i,j} = \text{softmax}\left(\frac{(W_3 x_i)^T (W_4 x_j)}{\sqrt{d}}\right) \quad (6)$$

2.2.3. GRU/BiGRU Blocks

2.2.3.1. GRU block

When CNN is used to extract the context dependences of long sequences, the field of view is limited due to the influence of the size of convolution kernel, and multiple CNN layers need to be used, which makes the model bloated and complex. In order to overcome the inability of CNN and RNN to deal with long-distance dependence, LSTM (Long-Short Term Memory) [27] is proposed. GRU is a very successful variant of LSTM, both of them can capture the long-term dependencies of the sequence and have comparable performance on many tasks, but GRU has a simpler internal structure and fewer parameters than LSTM, so it is more efficient when dealing with the same task, therefore, using GRU to process the time series of proteins is an obvious choice. Compared to LSTM, GRU has only two gates, namely update gate and reset gate, so it is more efficient in handling the same task. The update gate is used to control the extent to which the state information of the previous moment is brought into the current state. The larger the value of the update gate is, the more state information of the previous moment is brought into the current state. Reset gate is used to control the degree of ignoring the state information of the previous moment. The smaller the value of reset gate is, the more state information is ignored. GRU is to make a prediction in the current time step by controlling the operation of these two gates and then realizing the selection of sequence context information. The update gate z_t and reset gate r_t in GRU can be expressed as follows:

$$z_t = \sigma(x_t U^z + h_{t-1} W^z) \quad (7)$$
$$r_t = \sigma(x_t U^r + h_{t-1} W^r) \quad (8)$$

Where σ is the sigmoid function, through which the data can be transformed into a value in the range of 0~1 to act as a gating signal. x_t is the input of the current node, h_{t-1} is the hidden state passed down by the previous node, and this hidden state contains the relevant information of the previous node. U and W are the corresponding weight matrices, respectively. When GRU is used to extract protein features, the feature extraction process of GRU/BiGRU blocks can be shown in Fig. 2.

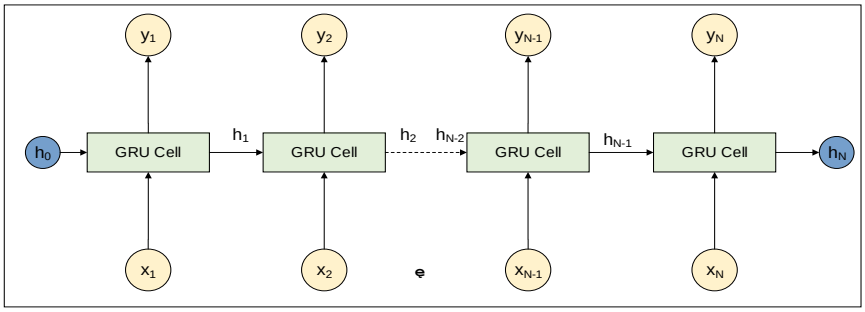


Fig. 3. GRU structure diagram.

In Fig. 2, the output of each stage is jointly determined by the hidden state of its previous stage and the input of the current stage.

2.2.3.2. BiGRU block

For some specific tasks, the information at a certain moment is not only related to the previous state, but also has some connection with the later state. When dealing with such problems, the traditional unidirectional GRU is obviously not competent, therefore, the bidirectional GRU is introduced. For protein sequences, we consider that the features of a certain part of the protein sequence are not only related to the previous part, but also related to the later part. Therefore, we also use a bidirectional GRU to extract the amino acid sequence features of the protein. BiGRU is composed of two unidirectional GRUs with opposite directions. At each moment, the input will fuse the outputs of the two opposite GRUs at the same time, and the output is jointly determined by these two unidirectional GRUs. The feature extraction process of BiGRU is as follows:

$$\overrightarrow{h_t} = \mathrm{GRU}\left(x_t, \overrightarrow{h_{t-1}}\right) \quad (9)$$

$$\overleftarrow{h_t} = \mathrm{GRU}(x_t, \overleftarrow{h_{t-1}}) \quad (10)$$

$$h_t = w_t\overrightarrow{h_t} + v_t\overleftarrow{h_t} + b_t \quad (11)$$

Where the function GRU () represents converting the corresponding input to its hidden layer state. $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ represent the hidden layer state in the corresponding direction, respectively, w_t and v_t represent the weights corresponding to the forward hidden layer state $\overrightarrow{h_t}$ and reverse hidden layer state $\overleftarrow{h_t}$ of the bidirectional GRU at time t, respectively. b_t represents the bias corresponding to the hidden layer state at time t. When BiGRU is used to extract protein features, the feature extraction process of GRU/BiGRU blocks can be shown in Fig. 3.

2.2.4. Prediction block

The features of the drug and the features of the protein are concatenated after being extracted and then fed into the prediction block. The prediction block consists of two fully connected layers, each of which is followed by a Dropout of rate 0.5 to prevent over fitting. The activation function of fully connected layer is the Rectified Linear Unit (ReLU). The output of the last layer identifies the final predicted affinity value for the drug and protein.

2.3. Implementation

GDGRU-DTA is implemented in Pytorch. We use the Adam optimizer with the default learning rate of 2e-4. The SMILES string for each drug is converted into 2-dimensional molecular graph where each node of the molecular graph is embedded with 78-dimensional features. GNN block consists of three stacked GNN layers with 78, 156 and 312 output features, respectively, which followed by a global max pooling layer to get the most striking features. The protein input embedding is of size 128, which means that we represent each character in amino acid sequence with a 128-dimensional dense vector. The GRU block is made up of 2 GRU layers, the first of which is followed by a Dropout of rate 0.2 and the output dimension of each layer is 8. For the BiGRU block, the number of layers of GRU is set to 1, and the output dimension is also 8. The prediction block is made up of three fully connected layers, in which the numbers of neurons are 1024, 512 and 1, respectively. The dropout rate is set to 0.5 and for the Davis dataset, the batch size is set to 128, while for the KIBA dataset the batch size is set to 512 because it is much larger than the Davis dataset, about four times larger than the Davis. Each drug and protein are converted into a 128-dimensional vector after their respective feature extraction, and are concatenated into a 256-dimensional vector for the final prediction. In this experiment, we

divided the dataset into five equal parts, four of which were used as training set and one was used as test set, we deal with over fitting problem by setting up cross-validation. The number of training epochs is set to 1000. Our experiments are run on Windows 10 professional with Intel(R) Core(TM) i5-10400F CPU @ 2.90GHz and GeForce GTX 1660Ti(6GB).

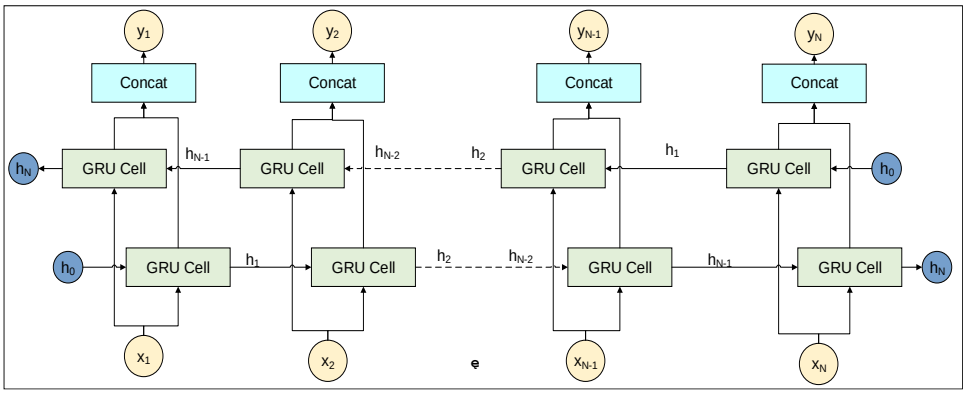


Fig. 4. BiGRU structure diagram.

In Fig. 3, the output of each stage is jointly determined by the hidden states of its previous and subsequent stages and the input of the current stage.

3. EXPERIMENTS AND RESULTS

3.1. Evaluation Metrics

MSE (Mean Squared Error), CI (Concordance Index) and r_m^2 (Regression toward the mean) are the most commonly used evaluation metrics in regression tasks to study drug-target interactions [15-21]. Since our research is also in this field, we continue to use these evaluation metrics, the details of each metric are as follows:

MSE is the mean square error, which is used to measure the gap between the predicted value of the model and the actual label value. The smaller the gap is, the better the performance of the model is; otherwise, the worse the performance of the model is.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(P_i - Y_i)^2 \tag{12}$$

Where P_i is the prediction value, Y_i corresponds to the label value and n is the total number of samples.

CI is the Concordance Index, which is a measure of whether the order of predicted binding affinity values for two random drug-target pairs is consistent with their true values, which value exceeds 0.8 indicates a strong model.

$$\mathrm{CI} = \frac{1}{Z}\sum_{y_i > y_j} h(p_i - p_j) \tag{13}$$

$$h(x) = \begin{cases} 1, x > 0 \\ 0.5, x = 0 \\ 0, x < 0 \end{cases} \tag{14}$$

In (13), sample *i* has a bigger label value than sample *j*.

R_m^2 index is the regression toward the mean, which is used to evaluate the external predictive performance. The metric can be described as follows:

$$r_m^2 = r^2 * \left(1 - \sqrt{r^2 - r_0^2}\right) (15)$$

Where r^2 and r_0^2 are the squared correlation coefficients with and without intercept, respectively. A value of r_m^2 above 0.5 is considered an ideal model.

3.2. Results

3.2.1. Performance comparison of GRU/BiGRU and CNN

To demonstrate that the GRU model we use is more efficient than the CNN model in extracting protein sequence features, in this section, we compare the above two. Our experiments were carried out on the Davis database and modified on GraphDTA. We changed the way of extracting proteins from the four models of GraphDTA from CNN to GRU and BiGRU to observe their experimental results. The results are shown in Table 2, from which we can conclude that GRU/BiGRU is facilitating to capture context dependencies in sequence.

Table 2. Performances of GRU and BiGRU compared to CNN on Davis dataset

	CI			**MSE**			**r_m^2**	
Method	GRU	BiGRU	CNN	GRU	BiGRU	CNN	GRU	BiGRU
GCN	0.899	0.896	0.880	0.211	0.220	0.254	0.712	0.705
GAT	0.902	0.903	0.892	0.218	0.220	0.232	0.715	0.706
GCN-GAT	0.895	0.897	0.881	0.223	0.232	0.245	0.697	0.699
GIN	0.901	0.896	0.893	0.214	0.218	0.229	0.726	0.714

As can be seen from the table, after the extraction method of protein is changed from CNN to GRU and BiGRU, both MSE, CI and r_m^2 are improved to varying degrees. For using the GRU model, the CI of GCN, GAT, GCN-GAT and GIN increases by 2.2%, 1.1%, 1.6%, and 0.9%, respectively, and the MSE decreases by 16.9%, 6.0%, 9.0%, and 6.6%, respectively. For using BiGRU model, the CI of GCN, GAT, GCN-GAT and GIN increases by 1.8%, 1.2%, 1.8%, and 0.3%, respectively, and the MSE decreases by 13.9%, 5.2%, 5.3%, and 4.8%, respectively. Besides, except for GCN-GAT, the r_m^2 values of the other methods using these two models all exceed 0.7, which indicate its excellent linear correlation and acceptability.

3.2.2. Comparison with other models

The GDGRU-DTA model combines GNN and RNN, and we conduct experiments on two different datasets, Davis and KIBA. The experimental results demonstrate that compared with other DTA methods, GDGRU-DTA has a huge improvement in performance. For each DTA model, we use its optimal data for comparison, the results on Davis and KIBA dataset are shown in Table 3 and Table 4 respectively.

Table 3. Results of various DTA prediction models on the Davis dataset

Method	Protein	Compound	CI	MSE	r_m^2
Baseline models					
KronRLS[15]	S-W	Pubchem	0.871	0.379	0.407
SimBoost[16]	S-W	Pubchem	0.872	0.282	0.644
WideDTA[18]	CNN	CNN	0.886	0.262	—
DeepDTA[17]	CNN	CNN	0.878	0.261	0.630
DeepGS[21]	GAT+Smi2Vec	CNN(Prot2Vec)	0.882	0.252	*0.686*
GraphDTA[20]	CNN	GNN	*0.893*	0.229	—
AttentionDTA[19]	CNN	CNN	*0.893*	*0.216*	0.677
Proposed model – GDGRU-DTA					
Transformer-BiGRU	BiGRU	GNN	**0.902**	**0.214**	**0.697**
GatedGraph-BiGRU	BiGRU	GNN	**0.904**	**0.214**	**0.708**
Transformer-GRU	GRU	GNN	**0.903**	**0.212**	**0.730**
GatedGraph-GRU	GRU	GNN	**0.906**	**0.207**	**0.711**

Table 4. Results of various DTA prediction models on the KIBA dataset

Method	Protein	Compound	CI	MSE	r_m^2
Baseline models					
KronRLS[15]	S-W	Pubchem	0.782	0.411	0.342
SimBoost[16]	S-W	Pubchem	0.836	0.222	0.629
DeepDTA[17]	CNN	CNN	0.863	0.194	0.673
DeepGS[21]	GAT+Smi2Vec	CNN(Prot2Vec)	0.860	0.193	0.684
WideDTA[18]	CNN	CNN	0.875	0.179	—
AttentionDTA[19]	CNN	CNN	0.882	0.155	0.755
GraphDTA[20]	CNN	GNN	*0.891*	*0.139*	—
Proposed model – GDGRU-DTA					
GatedGraph-BiGRU	BiGRU	GNN	**0.892**	**0.137**	**0.775**
GatedGraph-GRU	GRU	GNN	**0.894**	**0.136**	**0.781**
Transformer-BiGRU	BiGRU	GNN	**0.894**	**0.134**	**0.780**
Transformer-GRU	GRU	GNN	**0.895**	**0.132**	**0.785**

The models in the above two tables are arranged in descending order of MSE. The data for the baseline model is obtained from [15-21]. For the proposed model, two methods of drug feature extraction and two methods of protein feature extraction are randomly combined. It is not difficult to conclude from the table that the four methods of the proposed model outperform some current DTA methods to varying degrees in three indicators. In the table above, italics represent the best data of the baseline model, and bold represent the data that is better than the baseline model. In the above baseline method, KronRLS and SimBoost are traditional machine learning

methods which based on similarity. DeepDTA and WideDTA are sequence-based feature extraction methods, and AttentionDTA introduces an Attention block based on it to learn mutual features. DeepGS and GraphDTA are novel in that they both use graph structure and graph convolution network to extract features.

In the analysis based on Table 3, the four approaches of the proposed model outperform the baseline model on all data, with the lowest MSE of 0.207, a 4.2% reduction compared to the lowest baseline method, and the highest CI of 0.906, compared to the highest baseline method improves by 1.5%, and the highest r_m^2 is 0.730, which is 6.4% higher than the highest baseline method. In addition, it can be seen from table 3 that the CI values of four methods of the proposed model all exceed 0.9, which proves that they have strong consistency, moreover, the r_m^2 values are all over or close to 0.7, indicating that they have strong external prediction performance. To sum up, among the above four methods of the GDGRU-DTA, the combined method of GatedGraph and GRU shows the best performance in comprehensive consideration of MSE, CI and r_m^2, while the combination of Transformer and BiGRU is relatively poor. The data in Table 4 shows that the performance improvement of our model on large data sets is not so obvious compared to small data sets, which indicates that our model is insufficient in some aspects, and this is a problem that we need to consider and solve.

Combining the above results of Table 3 and Table 4, we can conclude that our model has better performance than some other DTA models and has great significance for the research of DTA, and thus will greatly promote its development.

4. CONCLUSION

In this paper, we describe our model in detail earlier, which is an end-to-end bio-inspired deep learning-based model for DTA prediction. In this work, Since the graph structure of the drug can better represent the structural features of the drug, we represent the SMILES string of the drug as its graph structure, and use two graph convolution methods different from those used by GraphDTA, these two new graph convolution methods exhibit excellent performance on the one hand, and also demonstrate the generalization ability of the GRU/BiGRU model on the other hand. To address the feature extraction problem for long amino acid sequences, we use GRU and BiGRU to capture the long-term dependencies, in order to confirm that the model is better in protein feature extraction, we change the protein extraction method of the four models in GraphDTA to the method we used, and the results of the four models have been improved to varying degrees. GRU and BiGRU also show excellent performance when combined with our two new graph convolution methods, which demonstrate their excellent generalization ability. Finally, we combine the two newly proposed graph convolution methods and two GRU models, and compare them with the previous DTA methods and some state-of-the-art DTA methods, and the results show that our method outperforms the previous methods. Our model can greatly facilitate the affinity prediction of drugs and targets, and provide a good reference for future research.

However, there is still room for improvement in our work. For example, for the feature extraction model of drugs, our structural innovation of the model is not very large. In addition, the attention mechanism is currently widely used in the model of drug and target interaction prediction. Therefore, our next work is to investigate how to improve the structure of drug feature extraction and add attention to the proposed model to better improve its performance.

REFERENCES

[1] A. Ezzat, M. Wu, X. L. Li, and C. K. Kwoh, "Computational prediction of drug-target interactions using chemogenomic approaches: An empirical survey," *Brief. Bioinform.*, vol. 20, no. 4, pp. 1337–1357, 2018, doi: 10.1093/bib/bby002.

[2] X. Chen *et al.*, "Drug-target interaction prediction: Databases, web servers and computational models," *Brief. Bioinform.*, vol. 17, no. 4, pp. 696–712, 2016, doi: 10.1093/bib/bbv066.

[3] A. S. Rifaioglu, H. Atas, M. J. Martin, R. Cetin-Atalay, V. Atalay, and T. Doğan, "Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases," *Brief. Bioinform.*, vol. 20, no. 5, pp. 1878–1912, 2019, doi: 10.1093/bib/bby061.

[4] E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, and A. G. Taranto, "Structure-Based Virtual Screening: From Classical to Artificial Intelligence," *Front. Chem.*, vol. 8, no. April, 2020, doi: 10.3389/fchem.2020.00343.

[5] M. Himmat, N. Salim, M. M. Al-Dabbagh, F. Saeed, and A. Ahmed, "Adapting document similarity measures for ligand-based virtual screening," *Molecules*, vol. 21, no. 4, pp. 1–13, 2016, doi: 10.3390/molecules21040476.

[6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[7] A. Hazra, P. Choudhary, and M. Sheetal Singh, *Recent advances in deep learning techniques and its applications: An overview*. Springer Singapore, 2021.

[8] M. Wen *et al.*, "Deep-Learning-Based Drug-Target Interaction Prediction," *J. Proteome Res.*, vol. 16, no. 4, pp. 1401–1409, 2017, doi: 10.1021/acs.jproteome.6b00618.

[9] K. Tian, M. Shao, S. Zhou, and J. Guan, "Boosting compound-protein interaction prediction by deep learning," *Proc. - 2015 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2015*, pp. 29–34, 2015, doi: 10.1109/BIBM.2015.7359651.

[10] Y. Du, J. Wang, X. Wang, J. Chen, and H. Chang, "Predicting drug-target interaction via wide and deep learning," *ACM Int. Conf. Proceeding Ser.*, pp. 128–132, 2018, doi: 10.1145/3194480.3194491.

[11] B. Shin, S. Park, K. Kang, and J. C. Ho, "Self-Attention Based Molecule Representation for Predicting Drug-Target Interaction," pp. 1–18, 2019, [Online]. Available: http://arxiv.org/abs/1908.06760.

[12] A. Vaswani *et al.*, "Attention Is All You Need," *CoRR*, vol. abs/1706.03762, 2017, [Online]. Available: http://arxiv.org/abs/1706.03762.

[13] Z. Quan, Y. Guo, X. Lin, Z. J. Wang, and X. Zeng, "GraphCPI: Graph Neural Representation Learning for Compound-Protein Interaction," *Proc. - 2019 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2019*, pp. 717–722, 2019, doi: 10.1109/BIBM47256.2019.8983267.

[14] W. Torng and R. B. Altman, "Graph Convolutional Neural Networks for Predicting Drug-Target Interactions," *J. Chem. Inf. Model.*, 2019, doi: 10.1021/acs.jcim.9b00628.

[15] T. Pahikkala *et al.*, "Toward more realistic drug-target interaction predictions," *Brief. Bioinform.*, vol. 16, no. 2, pp. 325–337, 2015, doi: 10.1093/bib/bbu010.

[16] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines," *J. Cheminform.*, vol. 9, no. 1, pp. 1–14, 2017, doi: 10.1186/s13321-017-0209-z.

[17] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: Deep drug-target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018, doi: 10.1093/bioinformatics/bty593.

[18] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity," 2019, [Online]. Available: http://arxiv.org/abs/1902.04166.

[19] Q. Zhao, F. Xiao, M. Yang, Y. Li, and J. Wang, "AttentionDTA: Prediction of drug-target binding affinity using attention model," in *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, Nov. 2019, pp. 64–69, doi: 10.1109/BIBM47256.2019.8983125.

[20] T. Nguyen, H. Le, and S. Venkatesh, "GraphDTA: prediction of drug–target binding affinity using graph convolutional networks," *BioRxiv*, p. 684662, 2019, doi: 10.1101/684662.

[21] X. Lin, K. Zhao, T. Xiao, Z. Quan, Z. J. Wang, and P. S. Yu, "Deepgs: Deep representation learning of graphs and sequences for drug-target binding affinity prediction," *Front. Artif. Intell. Appl.*, vol. 325, no. i, pp. 1301–1308, 2020, doi: 10.3233/FAIA200232.

[22] S. Hu, R. Ma, and H. Wang, "An improved deep learning method for predicting DNA-binding proteins based on contextual features in amino acid sequences," *PLoS One*, vol. 14, no. 11, pp. 1–21, 2019, doi: 10.1371/journal.pone.0225317.

[23] M. I. Davis *et al.*, "Comprehensive analysis of kinase inhibitor selectivity," *Nat. Biotechnol.*, vol. 29, no. 11, pp. 1046–1051, 2011, doi: 10.1038/nbt.1990.

[24] J. Tang *et al.*, "Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis," *J. Chem. Inf. Model.*, vol. 54, no. 3, pp. 735–743, 2014, doi: 10.1021/ci400709d.

[25] Y. Li, R. Zemel, M. Brockschmidt, and D. Tarlow, "Gated graph sequence neural networks," *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, no. 1, pp. 1–20, 2016.

[26] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun, "Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification," pp. 1548–1554, 2021, doi: 10.24963/ijcai.2021/214.

[27] S. Hochreiter and J. Schmidhuber., "Long Short-Term Memory," *Neural computation 9(8)*,vol. 1780, pp. 1735–1780, 1997.

THE IMPACT OF USING A CONTRACT-DRIVEN, TEST-INTERCEPTOR BASED SOFTWARE DEVELOPMENT APPROACH

Justus Posthuma[1], Fritz Solms[2] and Bruce W. Watson[3]

[1]Centre for AI Research, School for Data-Science & Computational Thinking,
Stellenbosch University, Stellenbosch, South Africa
[2]Department of Information Science,
Stellenbosch University, Stellenbosch, South Africa
[3]Centre for AI Research, School for Data-Science & Computational Thinking,
Stellenbosch University, Stellenbosch, South Africa

ABSTRACT

Contract Driven Development formalizes functional requirements within component contracts. The process aims to produce higher quality software, reduce quality assurance costs and improve reusability. However, the perceived complexity and cost of requirements formalization has limited the adoption of this approach in industry.

In this article, we consider the extent to which the overheads of requirements formalization can be netted off against the reduced quality assurance costs arising from being able to auto-generate functional test interceptors from component contracts. Test-interceptors are used during testing to verify that component contracts are satisfied. In particular, we investigate the impact of contract-driven development on both the quality attributes of the software development process and the quality of the software produced by the process.

Empirical data obtained from an actual software project using contract-driven development with test interceptor generation is compared to that obtained from similar projects that used a traditional software development process with informal requirements and manually written functional tests.

KEYWORDS

Software and its engineering, Software development methods, Software implementation.

1. INTRODUCTION

Contract driven development (CDD), also known as Design by Contract (DbC), Contract Programming and Programming by Contract, is a development methodology that aims to improve the quality of the produced software by formalizing functional requirements in the form of component contracts [10]. The benefits of CDD have been demonstrated [6], yet the methodology has never really gained significant traction. One of the main reasons for this is the perception that the added effort and complexity of formalizing component requirements in component contracts is tedious and does not deliver "enough bang for your buck" [14, 17].

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 39-58, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120704

There are a number of modern software development methodologies that are used by companies in their efforts to produce quality software. The two most popular methodologies today are:

- Test Driven Development (TDD) which requires software requirements to be converted into test cases before development starts, and then testing the developed software repeatedly against these test cases as it is being developed. TDD is primarily used to test units of code (like methods and classes) and aims to ensure a set of operations is performed correctly.

- Another popular methodology is Behaviour Driven Development (BDD) which is similar to TDD in that it requires tests to be written first, but in the case of BDD, the tests describe software behaviour by using a domain specific language (DSL). The DSL uses natural-language constructs that describe the required behaviour and expected outcomes, and is converted into executable tests by specialised software. BDD is also sometimes referred to as Story Test Driven Development (STDD), Acceptance Test Driven Development (ATTD) or Example Driven Development (EDD). [19]

The main disadvantage of TDD and BDD is that it requires a great amount of extra effort from developers to write tests. In the case of BDD, developers additionally have to learn and use the DSL and specialised software, resulting in extra time and costs on a project. Another major disadvantage is that the tests need to be updated and maintained when the code is updated.

In this study we investigate whether the cost benefits of generating functional test logic from component contracts is sufficient to nett off the cost of requirements formalization, whilst preserving the benefit of improving the quality of the produced software. In particular, we consider an approach where the process of formalizing requirements is simplified by encoding component contracts within the programming language used for the project (in this study we used the *Java* programming language) and generate test interceptors, which monitor contract compliance of wrapped components. These test interceptors are used for unit testing, integration testing and operational testing.

Empirical data is obtained from an actual software development project by using a contract-driven development methodology to specify component contracts. We empirically assess the impact on both quality attributes of the software development process and quality attributes of the software produced by the process. Process qualities include *development productivity* (the rate/cost at which software is produced), *development reliability* (the ability of the process to reliably produce software at a given rate and quality), *process usability* (the ease with which CDD can be introduced) and *process scalability* (the ability of introducing CDD to large development teams).

We also assess the impact of contract-driven development on selected quality attributes of the produced software. In particular the impact on correctness (the software meets the requirements), re-usability (whether contract-driven development leads to components that are more reusable), and simplicity are considered.

Section 2 examines related work and other frameworks that utilizes CDD. In Section 3 we discuss the details of our solution and exactly how test interceptors are implemented and how they function. Section 4 describes the results we obtained by using our solution in a real-world software developments project, and in Section 5 we discuss the conclusions we have drawn from the results. Finally, in Section 6 we look at future work.

2. RELATED WORK

Nebut et al describe a requirement-based testing technique that leverage use-cases in order to create functional and robustness test objectives [11]. Requirements are specified as UML use cases, which are annotated with contract specifications attached as notes. The contracts are specified using a custom language to construct predicates for pre- and post-conditions from state assertions (e.g., open(door-1)) and standard logical operators. In addition, they support two use case relationships: includes and generalization, which they effectively use for requirements aggregation and refinement. In order to facilitate automated test generation, the authors narrow down the differentiation between use cases and services by supporting parametrized use cases and assuming that there is a mapping of use cases onto system services.

The authors then use pre- and post-conditions of use cases to construct a graph of all possible transitions between use cases. Any system process can then be related to a path through the graph. Test processes (called test objectives) are constructed as specific finite paths through the Use Case Transition System (UCTS) graph to satisfy certain test criteria. The authors found that selecting a set of test paths (test objectives) that ensures that each use case is instantiated by at least one path and that each scenario for which all pre-conditions for a use case are satisfied, is included on one of the test paths.

Finally, test scenarios are constructed by setting the system into specific states and providing particular user inputs, i.e., each test scenario is a combination of system state and system input.
Strengths of the approach is that both test processes (test functions) and test scenarios (test data) are generated and that the tests are generated from a semi-formal requirements specification, without taking design considerations into account. But even though the proposed framework provides automated requirements-based functional and robustness testing from contract-annotated requirements models, it has not been widely adopted for automated requirements testing. Possible reasons for this failure include:

- Lack of Scalability: The framework was demonstrated in a very simple academic project. The approach does not test across levels of granularity. Instead, the entire requirements specification for the system is flattened into a single UCTS graph across which finite test processes are found using a breadth-first search algorithm. For industrial systems, this approach is unlikely to be feasible as the search space will explode with increasing system size and complexity. The approach investigated in this paper aims to manage scalability through automated contract-based test function generation for services across levels of granularity.

- Incompleteness: The authors propose the use of a custom language for the specification of pre- and post-conditions. These are attached via notes to UML Model elements. The expressiveness of the language is very limited and does not support the specification of non-trivial constraints. For specifying constraints against an object graph, UML has included the Object Constraint Language (OCL) as part of the standardization of the UML. The OCL enables one to specify UML model constraints using first-order predicate logic with quantifiers on finite domains. Furthermore, although the ability to generate test scenarios is one of the strengths of this approach, the authors have shown the generated test scenarios to be incomplete. If testing completeness is required (e.g. for certifiable projects) a manual process needs to identify scenarios that need to be added to achieve completeness.

- Limited usability: The use of a custom language for constraint specification requires new tools for validating constraint syntax and correctness. These tools are not available. Furthermore, the proposed approach is only usable in the context of model-driven engineering. Only a very small fraction of industrial projects is based on model-driven engineering. The approach analysed in this study uses code annotations of contract specifications in the programming language (e.g. against Java interfaces), using the programming language itself to specify the pre-condition, post-condition and invariant constraint predicates. Users thus need not learn a new language and the compiler tools can themselves be used to verify the constraint specification. Furthermore, the approach can be used for non-model-driven engineering projects.

Evolving service providers and consumers present a number of challenges, particularly when they change their document schemas (contracts), Robinson has found. He identifies two strategies for mitigating such issues: performing "just enough" validation of received messages and adding schema extension points [12]. Robinson notes that both of these strategies help to protect consumers when the provider changes the contract, but they are of little help to the provider to know how the contract is being used and what obligations it must preserve as it evolves. Robinson continues to discuss the "Consumer-Driven Contract" pattern which addresses this shortcoming by drawing on the assertion-based language of the "just enough" validation strategy, which imbues the provider with insight into its obligations towards consumers.

By observing contracts between providers and consumers, Robinson expresses the following insights: The business function capabilities of a service provider are expressed by a provider contract in terms of the collection of exportable elements that are necessary to support that functionality. The main characteristics of provider contracts are:

- Closed and complete: Provider contracts show the business function capabilities of a service as a complete set of exportable elements available to consumers. As such they are complete and closed regarding the functionality of the system.

- Singular and authoritative: Provider contracts are singular and authoritative in their expression of the business functionality of the system.

- Bounded stability and immutability: Provider contracts are stable and immutable for a bounded period and/or locale. They typically use a form of versioning to differentiate differently bounded instances of a contract.

Consumer contracts on the other hand, are entered into when a provider accepts and adopts the reasonable expectations expressed by a consumer. The characteristics of such a contract are:

- Incomplete and open: A consumer contract is incomplete and open regarding the business functionality of the system. It shows a subset of the capabilities of the system in terms of the expectations that the consumer has of the contract of the provider.

- Non-authoritative and multiple: Each consumer contract is non-authoritative regarding the total set of contractual obligations placed on the provider, and they are multiple in relation to the number of service consumers.

- Bounded stability and immutability: Similar to provider contracts, consumer contracts are valid for a particular location and/or period of time.

Robinson concludes that consumer contracts, by showing and asserting expectations of a provider contract, allow us to know precisely which parts of the provider contract presently support business value by the system, and which parts do not. He further suggests that services might benefit from being specified in terms of consumer contracts from the start. In this way, provider contracts emerge to meet the demands and expectations of consumers - consumer-driven contracts or derived contracts.

The characteristics of Consumer-driven contracts are:

- Complete and closed: A consumer-driven contract is complete and closed with respect to the complete set of functionalities required of it by its existing consumers. The mandatory set of exportable elements is represented by the contract, which is required to support consumer expectations for the period in which those expectations are required by their parent applications.

- Non-authoritative and singular: Provider contracts are non-authoritative because they are derived from the union of existing consumer expectations, and singular in their expression of the business functionality available to the system.

- Bounded stability and immutability: In respect of a particular set of consumer contracts, a consumer-driven contract is stable and immutable. That means the validity of a consumer-driven contract can be determined according to a specified set of consumer contracts, thereby binding the backwards- and forwards-compatible nature of the contract in space and time. The compatibility of a contract remains immutable and stable for a specific set of consumer contracts and expectations, but it is subject to change as expectations change.

Robinson identifies two significant benefits of consumer-driven contracts in terms of evolving services. Firstly, the focus is on the delivery and specification of functionality around key business value drivers - the value if a service is determined by the extent to which it is consumed. And secondly, consumer-driven contracts provide the fine-grained insight and rapid feedback needed to plan changes and assess their impact on applications currently in production. However, there are certain liabilities: in the context of a closed community or a single enterprise (an environment in which providers can exert influence over how consumers establish contracts with them), the consumer-driven contract pattern is applicable. Consumers and providers must accept, adopt and know about an agreed-upon set of channels and conventions. This adds a layer of protocol dependence and complexity on an already complex service infrastructure.

Although the pattern allows for better management of breaking changes to contracts, it is not a cure. At best, it provides better insight into what constitutes a breaking change, and as such may serve as the foundation of a versioning strategy.

Consumer-driven contracts do not necessarily reduce the coupling between services, but it does identify "hidden" couplings, which allow providers and consumers to better manage them. Finally, there is a risk that when the specification of a service provider is driven by consumer contracts, the conceptual integrity of that service provider could be compromised. Services are identifiable, discreet and reusable business functions whose integrity should not be undermined by demands that fall outside their mandate.

Pact [13] and Spring Cloud Contract [16] are popular testing frameworks that utilise Consumer Driven Contracts.

Belhaouari et al created an experimental platform known as Tamago-Test for software analysis and automated testing [2]. They focus on the automation of test-case generation from specifications written as Design by Contract and rely on First-order logic assertions to express contracts between components in the generation of test-cases. The pre-conditions are used to infer the relevant values to provide as input parameters to methods, and the post-conditions are then used as natural oracles (an oracle determines whether the test results are correct [3]). The creation of values for method input parameters which are correct with regards to the specification, corresponds to a constraint-satisfaction problem (CSP) [8]. Initially, the variable domain is defined by the type, and the CSP reduces the range by the various constraints represented by each atomic term extracted from the contract. This term is obtained in the disjunctive normal form [4] of assertions and is included in the pre-conditions. When the CSP must instantiate a variable, it draws a random value from its reduced domain. Finally, the CSP architecture the authors propose does not restrict the constraint language to finite-domain and predefined types - they came up with a "type builder" that enables the framework to be extended in a flexible way.

The most important characteristics of the Tamago platform created by the authors, are the provision of a specification language, a runtime, and a set of tools for analysis and support. Their approach also emphasizes the Separation of Concerns principle: The client provides the specifications for components, and the providers implement those specifications. The runtime and tools are responsible to realise the contract between these two parties.

The specification language is similar to the grammar of assertions in Java Modelling Language (JML) and the authors use only a subset of the features currently available. The resulting language is abstract, and based on a model of co-algebraic observable properties with first-order logic assertions (pre-conditions, post-conditions and invariants). It also includes descriptions of service behaviours based on finite-state automata with conditional transitions.

The authors believe the contract language that they have designed is a good compromise between expressiveness and tractability. For tractability, they have developed a set of tools to analyse the contract specification at design time. The first tool performs structural analysis by doing type-checking and using finite-state automata techniques to detect unreachable states or unused functionalities in the contract. The second tool attempts to uncover inconsistencies in the dynamics of the contract by utilizing a symbolic interpreter to generate a set of scenarios from the service behaviour. Using the effective pre-conditions, invariants and the guard of the predecessor transition in the behaviour automaton, the constraints to be enforced are partially evaluated. With this conjunction of assertions, a CSP-based minimization algorithm is used to narrow the domains of observable properties. Complementary to the conjuncts of the effective post-conditions, the invariants and the guards of the next transition (if any) is able to expand/reduce the domain of properties. If a domain becomes empty, there is no more solution and another branch is inspected. This analysis uses various fixed point detection heuristics to ensure the termination of the analysis.

Leitner et al recognizes that unit testing is a resource-intensive and time-consuming activity [9]. They introduce Contract Driven Development as a method to solve this problem by extracting the contracts that are present in code, and use those contracts to generate test cases. This is achieved by taking advantage of the activities that developers normally perform during the development process: when a new feature is being developed, the developer will run the application in such a way so that the new feature is used. By placing assertions along the way, the developer verifies that the new feature works as expected. This is done by triggering the new feature with the correct input, and also with incorrect input (by changing parts of the application) to ensure that error-handling is done correctly.

These implicit human-generated tests are easy to create and run, and they don't require any maintenance since they are not permanent. Other advantages over automated tools are that the automated testing strategies cannot make up for the insights that a human tester has into the semantics of the software and the relationships between different components. Automated tools also cannot distinguish between meaningful and meaningless input data, and although the quality of the automated tests can be estimated using certain measures or combinations of measures (such as code coverage, number of bugs found, mutation testing, proportion of fault-finding tests out of total tests, etc.), the characteristics of the project under test make it difficult if not impossible for a tool to determine automatically which measures to use.

The drawback of the implicit human-generated tests is that they only exist for one or very few runs and are not kept for later execution. This is because the developer manually provided specific inputs, and if the application was changed to force a certain path to be tested, that change was undone after the tests.

Leitner proposes a method to capture these implicit tests and to make them explicit and persistent. They created a tool called Cdd (which expands to Contract Driven Development), which targets Eiffel code since Eiffel natively supports contracts, and is installed into the Eiffel Studio IDE. Cdd monitors program executions and, when a failure occurs, Cdd detects the last safe state and takes a snapshot of this state. It then recreates this snapshot, which serves as the starting point of the extracted test case. Cdd determines the time to take this snapshot so that it is early enough for the state not to be infected but also late enough to reduce execution time. Furthermore, in order to make the test case more robust relating to system change, the snapshot excludes that part of the state that is not relevant for reproducing the failure.

Since Cdd employs continuous testing, these test cases will be evaluated on every run. The IDE will show the tests as failed, until the developer has implemented the relevant methods and they work correctly, in which case the IDE will show that the tests have passed.

Kramer created iContract, a freely available source-code pre-processor that instruments Java source-code with checks for class invariants, as well as pre- and post-conditions that may be associated with methods in classes and interfaces [7]. His inspiration came from the Eiffel language, which has native support for design by contract. In iContract, special comment tags (@pre, @post, @invariant, added to the Javadoc of the classes and methods) are interpreted and converted into assertion check code that is inserted into the source code. It also caters for the four hierarchical type relations in Java: class extension, interface implementation, interface extension and inner classes (collectively referred to as "type extension").

To use iContract, Java source code is annotated with formalized functional requirements in the form of three specific comment paragraph tags:

- @invariant, to specify class- and interface-invariants;
- @pre, to specify pre-conditions on methods of classes and interfaces;
- @post, to specify post-conditions on methods of classes and interfaces.

iContract is run as a pre-processor over annotated source code files. It instruments the methods in these files with assertion check code that enforces the specified functional requirements. In the background, a repository of contract related information about the classes, interfaces and methods is built up. This information is used to support subcontracting, which propagates functional requirements along interface-implementation, multiple interface-extension, class-extension, and

inner class relations. The instrumented files are compiled instead of the originals, resulting in the same classes, interfaces and methods except that the functional requirements are being enforced by means of automatically generated specification checks. Despite being annotated with functional requirements, the original files remain fully compliant to standard Java due to the annotations being a part of the (optional) comment paragraphs.

It is not clear when work on iContract stopped, but an attempt was made to resurrect it in the form of Java Contract Suite (JContractS) which is available on SourceForge (https://sourceforge.net/projects/jcontracts/). However, at the time of writing, the last update to this project was made in April, 2013.

3. IMPLEMENTATION

A common approach to requirements analysis and design is incremental refinement of requirements and design across levels of refinement or granularity [15]. At each level of refinement, one identifies the services required to assess the pre-conditions and address the post-conditions, and categorize them to responsibilities, which, in turn are assigned to components representing responsibility domains.

We represent component contracts in Java as Java interfaces, which are annotated with the component's functional requirements, i.e., the pre- and post-conditions and, in the case of stateful components, invariance constraints specifying enforced symmetries. A class implementing the Java interface is required to meet the component contracts and will be tested against the functional test logic generated from the component contract.
From component contracts we generate test interceptors, which are interface compatible with the component and which, in the context of service provision, verify that the wrapped component fulfils the functional requirements specified in the component contract. In particular, test-interceptors intercept service requests in order to verify that:

1. if all pre-conditions are met, the service is provided (no exception is raised) and all post-conditions hold after service provision;

2. if one or more pre-conditions is not met, that an exception specified for one of the pre-conditions that are not met is raised;

3. that all invariant constraints hold when the service is requested (if not, then the system was in a broken state already); and

4. that all invariance constraints hold after service provision.

In order to achieve this, the generated test interceptors perform the following steps:

1. assess any invariance constraints and raise an invalid state exception (which is not a component error) when any invariance constraint is violated;
2. assess and store the truth value of each pre-condition on service request;

3. delegate, within a try block, the request to the underlying component for processing;

4. upon catching an exception:

    a. verify that the pre-condition associated in the component contract with the caught exception did not hold upon service request;

5. if no exception has been caught, verify that:
    a. all pre-conditions were met, and that
    b. all post-conditions hold.
6. in either case that all invariance constraints hold.

Note that invariance constraints fall away if one follows a service-oriented analysis and design method [15]. In this case components are stateless and serve solely to package services within responsibility domains.

Some programming languages like Eiffel natively support the concepts required for contract specification, i.e., pre- and post-conditions and invariance constraints. Other languages like Java, C# and Python natively support language extensions via mechanisms like annotations and attributes. For programming languages that do not provide a syntax for extending the language one may need to either use an external framework for this purpose, or embed contract annotations within comments. In Java, we define the concepts and syntax for specifying functional requirements as pre-condition, post-condition, and invariance annotations. These annotations are processed using custom annotation processors that are bound into the compilation step.

The following annotated pseudo-code illustrates a Java interface with a method that is annotated with pre-conditions and post-conditions:

```
@Invariant(constraint="invariant boolean expr")
public interface the_interface
{
  @Precondition(constraint = "boolean expr //pre-assessment //",  raises =
Exception.class)
  ... repeat for multiple pre-conditions
  @Postcondition(constraint = "boolean expr"}
  ... repeat for multiple post-conditions
  public returntype methodName(param1, param2, ...) throws Exception;
}
```

The pseudo-code of the interceptor class that is generated from the annotations, look as follows:

```
public class className_TestInterceptor implements the_interface {
  private the_interface counterpart = null;

  public className_TestInterceptor(the_interface counterpart){
    this.counterpart = counterpart;
  }

  /*
  Test the invariants
  */
  private validateConsistency() throws ExceptionRaisedDuringInvariantCheck,
InvariantViolatedError {

    boolean invariantHolds = false;
    try {
```

```
            invariantHolds = invariant boolean expr;
    }
    catch (Exception e){
            throw ExceptionRaisedDuringInvariantCheck("Message","<invariant boolean
expr>");
    }
    if (!invariantHolds) {
            throw InvariantViolatedError("Message","<invariant boolean expr>");
    }

  }

  @Override
  public returntype methodName(param1, param2, ...) throws Exception {
    /*
    Evaluate the pre-conditions
    */
    boolean _pre_1 = boolean expr == true;
    ... repeat for every pre-condition

    /*
    Evaluate the pre-assessments
    */
    int _preAs_1 = pre-assessment expression;
    ... repeat for every pre-assessment

    /*
    Validate the invariants before the method call
    */
    validateConsistency();

    returnValue = 0;
    try {
      /*
      Call the wrapped method of the counterpart component to be tested
      */
      returnValue = counterpart.methodName(param1, param2, ...);

      /*
      If method was implemented correctly, it would have thrown exceptions if any of the
pre-conditions were false.
      */
      if (!_pre_1){
        /*
        The method above did not throw an exception but pre-condition 1 is false.  This
means there is a problem with the implementation of the method.
        */
        throw new PreconditionNotEnforcedException("boolean expr", "methodName");
      }
      ... Repeat check for all pre-conditions.

      /*
```

```
    Evaluate post conditions and their pre-assessments
    */
    if (!(post-condition boolean expr)){
      throw new PostconditionNotMetException("post-condition boolean expr",
"methodName");
    }

    if (!(post-condition expr == _preAs_1)){
      throw new PostconditionNotMetException("post-condition expr == //pre-assessment
expr//", "methodName");
    }

  /*
  Validate the invariants after the method call
  */
  validateConsistency();
  }
  /*
  The wrapped method threw an exception, catch different types of exceptions
  */
  catch (InvalidArgumentException iae) {
    if (!_pre_1) {
      /*
      Pre-condition 1 is false, so the method was supposed to throw this exception.  This
means it implemented the check for this pre-condition correctly.  Rethrow the valid
exception.
      */
      throw iae;
    }
    ... repeat for all preconditions with this type of exception

    /*
    If this point is reached, it means that the pre-conditions are valid, but the method still
threw an exception.  So there is a problem with the implementation of the method.
    */
    throw new
PreconditionsHoldButServiceRefusedException("InvalidArgumentException",
"methodName");
  }
  catch (InvalidStateException ise){
    if (!_pre_3){
      /*
      Pre-condition 3 is false, so the method was supposed to throw this exception.  This
means it implemented the check for this pre-condition correctly.  Rethrow the valid
exception.
      */
      throw ise;
    }
          ... repeat for all preconditions with this type of exception
    /*
    If this point is reached, it means that the pre-conditions are valid, but the method still
threw an exception.  So there is a problem with the implementation of the method.
```

```
      */
     throw new PreconditionsHoldButServiceRefusedException("InvalidStateException",
"methodName");
    }
    /*
    All pre-conditions and post-conditions are true.
    */
    return returnValue;
  }
}
```

A working example is available on GitHub: https://github.com/JustusPosthuma/example

The test-interceptor wrapped components are used in the following scenarios:

Unit testing:

In the case of unit testing, the component is meant to be tested in isolation, assuming that any components it depends on perform their tasks correctly. To this end, one substitutes any dependencies of the component under test with mock objects for these dependencies, which (under the scenarios for the given test data) behave as would be expected from the components they are mocking. Dependency injection [18] can be used to inject mock objects instead of actual components for the component dependencies.

Integration testing:

In the case of integration testing the component is tested within its actual environment. In this case, one injects real components for component dependencies.

Operational testing:

For operational testing or testing on the live system, one can dependency-inject test-interceptor wrapped components for those aspects of the system one wants to monitor for contract compliance.

Our annotation processor is invoked as a pre-compilation step by the Java compiler. Apart from the initial set-up and configuration of the project to use the annotation processor (in the project build-scripts or in the Integrated Development Environment (IDE) that is used), there are no extra steps required. This results in interceptor classes being automatically generated each time the project is built.

For unit and integration testing the test-interceptor wrapped component is called with test data sets, which are systematically created from an analysis equivalence partitions for the problem. Future work will look at augmenting the CDD tool suite with tools that automatically generate test data sets [5]. It is the clean separation of test logic and test data which facilitates the reuse of test logic which also applies for operational testing.

Generating the interceptor classes is trivial in terms of speed and resources, and even in extreme cases where large numbers of annotations are processed, the build process is not noticeably affected. An interceptor can process any contract as long as it is expressed as a valid Boolean expression.

4. RESULTS

In January 2021, work started on a Covid-19 vaccination project at one of our clients. The goal of the project was the creation of software components that enable the client to:

- gauge the interest of students and staff to get the vaccine via a short survey;
- enable students and staff to make appointments to receive the vaccine;
- sign digital waivers;
- push-notifications with reminders for vaccine appointments and other important information;
- report symptoms and side-effects; and to
- provide digital certificates as proof of vaccination.

The system also provides additional admin functionality for officials to create appointment slots, record a medical history for each person, and generate statistics.

The system was developed using an Amazon Web Services (AWS) “serverless” infrastructure for the back-end. This infrastructure hosted a MySQL relational database and provided AWS Lambdas, (event-driven computing services that run code in response to events from clients that connect to it via a Representational State Transfer (REST) protocol). The types of clients that connect to the back-end include Android mobile applications, iOS mobile applications, and web-applications (see Figure 1).

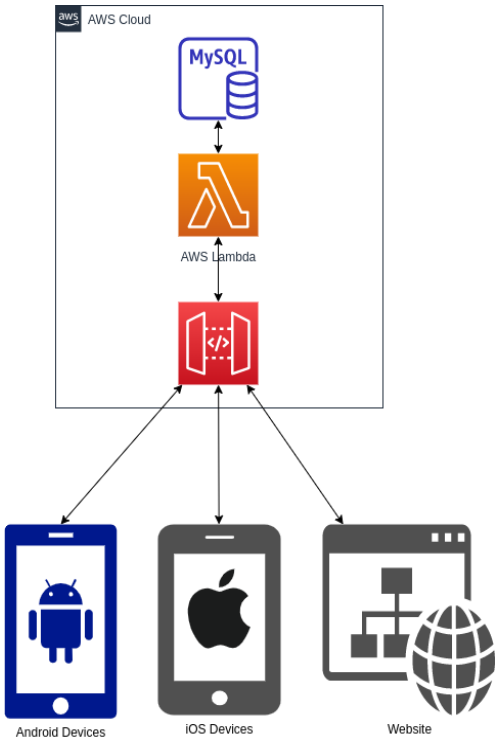


Figure 1. System Architecture

Different teams were responsible for each of the different components: One team focused on the back-end, another on the web front-end, and two other teams focused on the iOS and Android applications respectively. An informal Agile development process with the following elements was followed:

- Continuous software delivery through the use of sprints. At the end of each sprint, workable components were delivered for testing;
- Changing requirements were accommodated;
- Frequent meetings with stakeholders were organised;
- Online face-to-face interactions were common;
- The teams were self-organizing;
- After each milestone was delivered, reflections (retro-actives) were held to discuss what worked, what didn't work and how things could be done better.

The back-end team was responsible for setting up and maintaining the AWS environment, creating and maintaining the database and creating and maintaining the REST endpoints via AWS Lambdas. The web front-end team created web-based applications that run inside a browser, and the iOS team and Android team implemented the same functionality for the respective mobile platforms. Contracts were informally discussed between the back-end team and the other teams to determine what data the REST endpoints expected and what they provided.

The Android application was created with Domain Driven Design (DDD) in mind. The "domain" in this context is the sphere of knowledge and activity around which the application logic revolves [1]; in this case, all the activities around the Covid-19 vaccine roll-out. And keeping to the principles of DDD, the application was architected using four layers:

1. A **User Interface layer**, which contains both, the user views (forms) as well as the user work-flows around assembling the information required for service requests and the information provided with responses.

2. An **Application adapter layer**, which maps user requests from the user interface layer onto the infrastructure layer and service responses back onto response domain objects provided to the user interface layer.

3. An **Infrastructure layer**, which implements the application services API (also implemented on the server side) and maps the Java requests onto REST requests using data transfer objects (DTOs) and responses back onto Java responses. This layer is responsible for communication with external systems and persistent storage.

4. A **Domain layer**, containing objects from the user domain as objects encapsulating the request and response data. This layer does not have technical details like database connections and should be understandable to those who do not have technical knowledge.

Normally, contract validation should happen on the back-end (in this case, the AWS Lambdas), since it is the service provider for three different types of clients. But due to the following factors this was not possible:

- The AWS Lambdas are written in TypeScript. Our interceptor-generator is written in Java;

- The author was a member of the team that focused on Android.

However, implementing contract validation via interceptors in the infrastructure layer of the Android application that communicates with the AWS Lambdas, does in theory, have, from the perspective of the Android application, the same effect as if it were implemented on the back-end itself because the contracts are the same. The only difference is that the validation happens before the data leaves the client instead of when it arrives on the server.

To determine the impact of our approach, we compare the Covid-19 project (with contracts and interceptors) to a different project called Travel (with no contracts or interceptors). The Travel project allows users to book personal or business trips, upload digital copies of their passports and visas, notify the user if the destination is high-risk (Covid-19 infections or other reasons like unrest) and provide digital waivers for business trips. Although the Travel project is functionally completely different, it was developed on the exact same architecture as well as the same four layers relating to DDD. We analysed the requirements specification, UI interfaces, number of REST calls and project plans to ensure that the projects are as similar as possible in scope:

Table 1. Project properties

	Covid-19	**Travel**
Activities (screens)	11	8
REST API calls	30	26
Planned scope (in days)	30	30

The following measures were used to quantify the impact:

- *Number of man-hours to develop*. This includes requirements specification, requirements refinement and design, test and bug fixing.
- *Correctness of the produced software*. This is the rate at which the produced software meets the client requirements. It entails producing correct results and handling exceptions properly, and can be measured by counting defects over a period of time with bug-tracking software.
- *Bug Density*. This measure is defined as defects per thousand lines of code (KLOC) and measured by dividing the size of the module by the number of confirmed defects.

Number of man-hours to develop

The number of hours were very similar between the project with contracts and interceptors, versus the project without contracts and interceptors. This result shows that the introduction of contracts does not significantly impact the duration of a project (Figure 2).

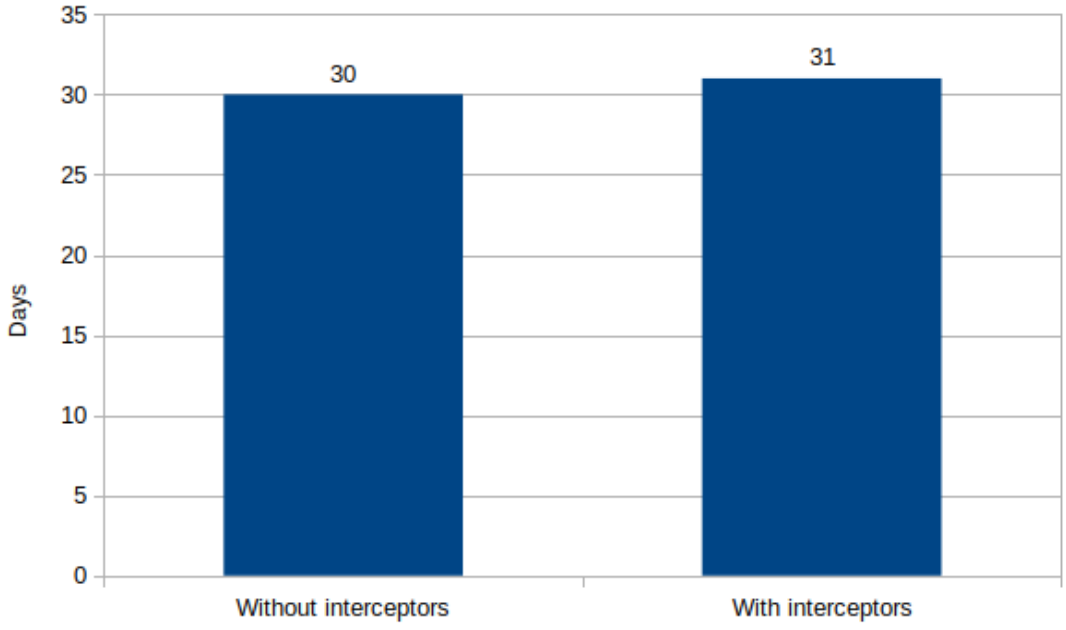


Figure 2. Man-hours to develop

Correctness of the produced software

The average bugs per day was significantly less in our project with contracts and interceptors versus a project without contracts and interceptors. We did not include cosmetic bugs in this measurement (Figure 3).

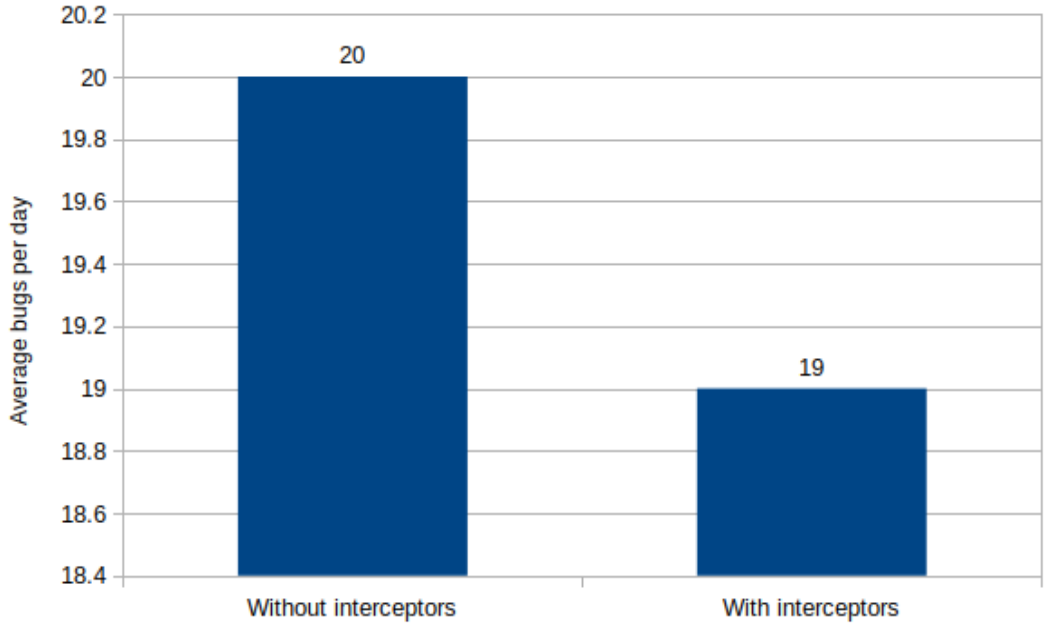


Figure 3. Correctness of the produced software

Bug density

There was a significant improvement in the bug density of the project that utilized contracts and interceptors, compared to the project without them. This was especially true for integration and logic bugs. Cosmetic bugs are high in both projects due to the fact that the client is notorious for always requesting layout, label and graphical changes, and these changes are logged as bugs (Figure 4).

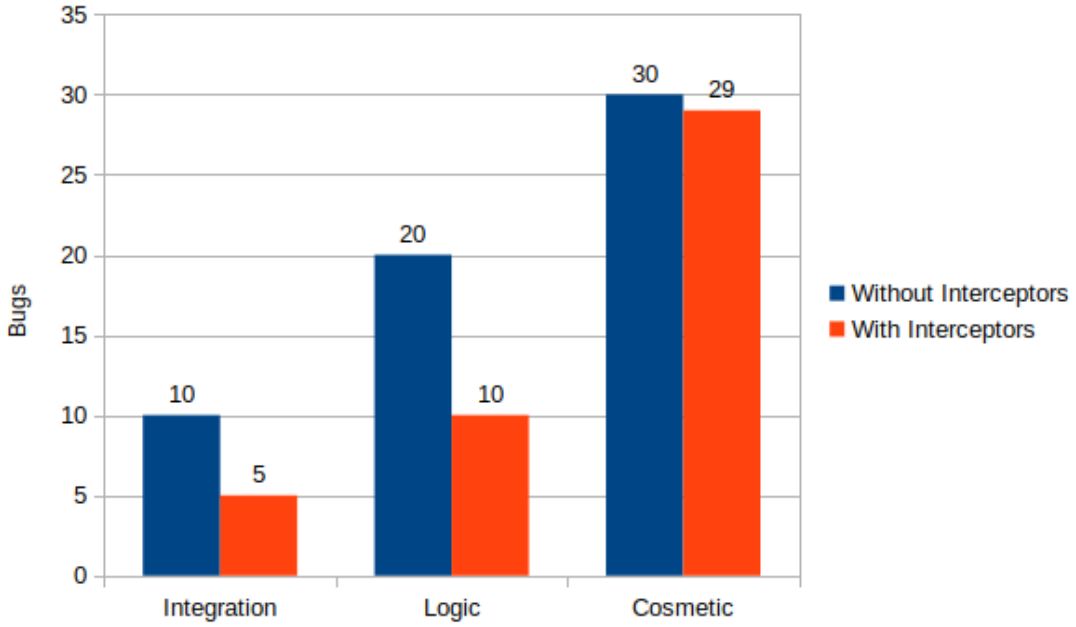


Figure 4. Bug density

4.1. Limitations and Deficiencies

Contracts are defined as Boolean expressions in annotations. These Boolean expressions are written as strings in the annotations, for example:

@Precondition(constraint = "**itemPrice > 0**", raises = InvalidStateException.class)

Since the expressions are strings, the compiler will not evaluate the syntax of the Boolean expression and return an error if it is incorrect – the compiler will simply evaluate the string as any other string. Only once the interceptor has been generated and the Boolean expression has been converted into code, will any syntax errors be discovered by the compiler, and will an error be reported in the interceptor class. Unfortunately, due to how Java annotations and Reflection work, the Boolean expressions (contracts) have to be written as strings.

Modern Integrated Development Environments (IDE) show syntax errors in code as you are typing, and it would have been ideal if any syntax errors related to the Boolean expressions (contracts) could have been identified similarly.

An important function that the interceptors cannot do (yet) is to test for errors or exceptions as part of a contract - a good example of this is date or time formats: it is common practice in Java to use a date parser (for example *SimpleDateFormat*) which throws an exception when a date or time is not in the expected format. It would be a very powerful feature if date and time formats can be specified in contracts.

5. CONCLUSION

We have found the Java programming language to be very suitable for contract driven development - language features like annotations and reflection were extensively used to define

contracts and generate interceptors. The fact that interceptors are generated as a pre-compilation step also makes the process easy and seamless from the perspective of a developer.

Developers were able to easily implement contracts since the learning curve was not steep at all – everyone was familiar with annotations already (since it is an existing Java language feature) and everything else (interceptor generation, evaluation of contracts) happens automatically. It was easy to convince the team to try our implementation of CDD since it had so little impact on the normal development routine.

Our empirical measurements show that providing a tool suite for CDD enabling developers to specify syntax-checked component contracts and generate test interceptors used for unit, integration and operational testing, significantly improved the quality of the software produced by the software development process without increasing the cost significantly. In particular, logic and integration errors were reduced by approximately 50%, whilst the reduction of cosmetic errors and the increase in number of man-hours were both less than 5%. Further efficiency benefits can be obtained by automating the generation of test data for unit testing.

We expect that having components that perform rigorously specified functionality (in the form of component contracts) would improve re-usability. The scope of the current study is, however, too small to assess this quality attribute and this is left for future work.

Based on the results obtained so far in this study, we are confident that we are on the right track to rekindle a renewed interest in CDD in industry and we feel that this study also demonstrates that a developer-centric approach and tool suite for CDD does deliver "enough bang for your buck!"

6. FUTURE WORK

Further efficiency benefits can be obtained by automating the generation of test data from the contracts for unit testing, in particular generating unit test templates which require only the population of data structures with test data.

We also feel that these results warrant a CDD approach in other programming languages, especially Javascript which is very popular on the server side. One would have to scrutinize the specific language features to decide how best to implement contracts, because not all languages have built-in annotations. But in the case of Javascript, early investigations show that comment blocks can be used to define pre-conditions and postconditions for methods, and interceptors can be generated by launching an external process that parses the comment blocks and generate interceptors.

REFERENCES

[1] airbrake.io. 2017. Domain-Driven Design: What is it and how do you use it? https://airbrake.io/blog/software-design/domaindriven-design

[2] Hakim Belhaouari and Frederic Peschanski. 2008. Automated Generation of Test Cases from Contract-Oriented Specifications: A CSP-Based Approach. In HASE '08: Proceedings of the 2008 11th IEEE High Assurance Systems Engineering Symposium. IEEE Computer Society, Washington, DC, USA, 219–228. https://doi.org/10.1109/HASE.2008.15

[3] Yoonsik Cheon and Gary T. Leavens. 2002. A simple and practical approach to unit testing: The jml and junit way. (2002).

[4] M. Donat. 1997. Automating Formal Specification Based Testing.(1997).

[5] Stefan J. Galler, Martin Weiglhofer, and Franz Wotawa. 2010. Synthesize It: From Design by Contract to Meaningful Test Input Data. In 2010 8th IEEE International Conference on Software Engineering and Formal Methods. 286–295. https://doi.org/10.1109/SEFM.2010.33

[6] J.-M. Jazequel and B. Meyer. 1997. Design by contract: the lessons of Ariane. Computer 30, 1 (Jan. 1997), 129–130. https://doi.org/10.1109/2.562936

[7] R. Kramer. 1998. iContract - The Java(Tm) Design by Contract(Tm) Tool. In Proceedings of the Technology of ObjectOriented Languages and Systems (TOOLS '98). IEEE Computer Society, Washington, DC, USA, 295–. http://dl.acm.org/citation.cfm?id=832254.832856

[8] Vipin Kumar. 1992. Algorithms for constraint-satisfaction problems: A survey. (1992).

[9] Andreas Leitner, Ilinca Ciupa, Manuel Oriol, Bertrand Meyer, and Arno Fiva. 2007. Contract Driven Development = Test Driven Development - Writing Test Cases. In Proceedings of the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering (Dubrovnik, Croatia) (ESEC-FSE '07). ACM, New York, NY, USA, 425–434. https://doi.org/10.1145/1287624.1287685

[10] Bertrand Meyer. 1992. Applying "Design by Contract". Interactive Software Engineering (1992).

[11] Clementine Nebut, Franck Fleurey, Yves Le Traon, and Jean-Marc Jezequel. 2003. Requirements by Contracts allow Automated System Testing. In ISSRE '03: Proceedings of the 14th International Symposium on Software Reliability Engineering. IEEE Computer Society, Washington, DC, USA, 85.

[12] Ian Robinson. 2006. Consumer-Driven Contracts: A Service Evolution Pattern. https://www.martinfowler.com/articles/consumerDrivenContracts.html

[13] Beth Skurrie. 2020. Pact: Getting Started. https://docs.pact.io/

[14] slashdot.org. 2007. Why is design by contract not more popular. https://ask.slashdot.org/story/07/03/10/009237/why-isdesign-by-contract-not-more-popular

[15] Fritz Solms. 2018. URDAD for System Design.

[16] spring.io. 2020. Spring Cloud Contract. https://spring.io/projects/spring-cloud-contract

[17] Stackoverflow.com. 2018. Why is design-by-contract not so popular compared to test-driven development? https://stackoverflow.com/questions/481312/why-is-design-bycontract-not-so-popular-compared-to-test-driven-development

[18] Hong Yul Yang, E. Tempero, and H. Melton. [n.d.]. An Empirical Study into Use of Dependency Injection in Java. In Software Engineering, 2008. ASWEC 2008. 19th Australian Conference on (2008-03). 239–247. https://doi.org/10.1109/ASWEC.2008.4483212

[19] agilealliance.org. Acceptance test driven development (atdd), 2016

A DATA-DRIVEN REAL-TIME ANALYTICAL FRAMEWORK WITH IMPROVED GRANULARITY USING MACHINE LEARNING AND BIG DATA ANALYSIS

Yubo Zhang[1] and Yu Sun[2]

[1]Shenzhen College of International Education,
3 Antuoshan 6th Rd, Futian District, Shenzhen, China, 518043
[2]California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620, USA

ABSTRACT

During daily studying and working, people have to research massive amounts of information on the internet and download numerous files. Some of these files can be easily categorized to relevant files. However, there are always some files left unorganized due to their difficulty in categorization [1]. Such files pile up in the download folder as time passes, making the folder extremely messy. Many people do not have the motive to clean up the folder as it requires a lot of energy and time. Based on this common problem, my group developed an app that can clean the messy folder up. After applying our program which is based on machine learning, the files will firstly be divided into five general parts – document, video, music, photos and package. Then the files will be further categorized based on the contents they present. For example, photos are divided into animals, families and so on. In order to achieve the content-categorizing function, several powerful apis were introduced in our program, and they helped us to achieve the optimal results.

KEYWORDS

Data Processing, Deep Learning, Machine learning.

1. INTRODUCTION

I am a high school student who was previously annoyed by the messy download files on my computer. Everyday I download many files from websites, but I do not have enough time and energy to clean up these files one by one -- I just leave them there. Files gradually piled up and messed up my computer, so I wanted to find a way to clean up the files [8]. My first action was to find a suitable app in the app store. However, many of them were way too expensive and they were not personalized enough to achieve my wish – categorizing the files perfectly as if they were done by myself. Thus, I decided to create a personalized program that can help me clean up the files automatically. The first version of my program could only fit the environment on my computer as it was based on my personal preferences. It was indeed powerful when I tested it on my laptop – every file went to the most desirable place as I wished. However, it was not general enough to cater for others' needs to organize their files [3]. Thus, I made several amendments on the original code to reach the second version, which had a larger range of applications and could be used for the public. As a result, others encountering the same problem can also apply my program and then clean their folders up effectively.

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 59-66, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120705

There are numerous related methods on this topic. One example is the system developed decades ago to classify emails by scanning over their bodies, senders, receivers, and the titles. The results come from the comparisons between the preset keywords and the features extracted from the emails. Therefore, such classification can be unreliable in many cases. For example, if none of the preset keywords matches perfectly with the features, then the system will not be able to classify the documents, inhibiting its capability. Another example is using machine learning to detect the contents of images [9]. Such techniques have become very mature. By feeding the computer with numerous sample images, the computer can learn the features of nearly every commonly-used item such as cars and desserts, and its accuracy is quite high – higher than 85% in most cases. Thus, I will apply such techniques to achieve parts of my project. Another example is the concept of a new file organization system. Different with the traditional system that follows a tree structure, such systems import 'concepts' to replace the folders. One single file can be placed under numerous concepts so that the paths to the file vary. However, such a new idea has not been widely adopted yet, and the majority of the operation systems are still using the traditional organization methods. Thus, the new concept, though providing some novel approach to organization, is yet to be widely used.

In this paper, we follow the same line of research by adopting existing mature techniques into our system and develop the rest parts by using our own logic and algorithms. Our goal is to develop a system that can automatically clean up the messy download folder and then transfer each file to the proper place of the computer. Some of our methods are inspired by the well-researched methods such as the use of machine learning on classifying the pictures. Based on the results given by the system developed by such a classifying technique, our system is able to place each image into a proper place under the image folder, depending on the content of the image. Another strength of the program is its capability to categorize the files missing key information. For example, if there is an untitled audio under the download folder, the program is able to detect the content of the audio and then transfer it to the right place. It may find out that the audio is a song by a famous singer. As a result, the audio will be placed under the singer's folder, which is rational and effective.

In order to prove my results, I first operated the program on my laptop. The results were satisfying: all the documents and files went to the ideal places. My first experiment proved the effectiveness of the program on my laptop. However, when I ran the program on some of my friends' laptops, the results were not very good initially. One main reason was that they had different organizing habits with me: my categorizing methods did not match up with theirs. As a result, the program simply created several new folders in their laptops and directly transferred the files from the Download folder to the new ones. This was really bad as it actually further messed up their laptop. My friends still had to do the categorization work on their own, and they had to drag each file from the new folders rather than the download folder to the optimal ones, which was more tiring [2]. Despite detecting this problem, I was not able to solve it due to the fact that every person has unique habits. Fully personalizing the program requires techniques that are far beyond my abilities. Thus, I could just make some amendments on the program to make the logic of my categorization more reasonable at the current stage.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Choosing an appropriate categorization method

The first challenge that I needed to confront is choosing an appropriate categorization method. This problem was crucial and was required to be tackled before entering the code developing stage: it directly determined the structure of the program [10]. Numerous aspects had to be considered carefully in order to deal with this challenge perfectly. Firstly, the first layer of the categorization system was to be set up, dividing the files into several general parts, for instance, videos, pictures and documents. After that, a second layer was to be built, further categorizing each part in reasonable ways. For instance, the picture part could contain family, pet, travel and other sections. The categorization work was somehow difficult as the system was supposed to be reasonable for every user. Therefore, many considerations had to be taken into account to develop a general system that could cater for everyone's basic needs.

2.2. How to realize the categorization system designed in the last stage

The second challenge was how to realize the categorization system designed in the last stage. Firstly, I tried to develop the system based on the names of the file. However, I soon found that I could only complete the first layer of the system. This was because even though I could have a general categorization of the files by detecting the differences between their extensions, I could not further divide them up in most cases as the majority of their names did not contain useful information. Therefore, in order to make further categorization possible, I needed to enable the program to scan over the contents of files and make categorizing decisions based on the results. This was difficult to me as it required the application of deep learning, which I hardly knew anything about [11][12]. Thus, in order to achieve this, I searched numerous apis on the internet and fitted the most proper and powerful ones into my project, maximizing the rationality of my categorization.

2.3. Compatibility

Compatibility was the third challenge after resolving the last problem. For example, one particular api used for detecting the contents of pictures could work pretty well on my computer, whereas it could not run properly on my laptop as my laptop did not have a discrete graphics card. This problem could happen on many other laptops, weakening the power of my program. Thus, I needed to find apis that were not only powerful but also could be applied by the majority of the computers. This was somehow difficult. Even though some apis were powerful and were supposed to be compatible with every computer, running them successfully could use a lot of memory, leading to crashes on weak computers [13].

3. SOLUTION

In order to categorize the files, X Cleaner will firstly examine all files in the download folder. Based on the extensions of their names, the files will then be divided up to five general parts – Documents, Musics, Images, Videos, and others. If the file name contains enough information for the program to complete the categorization (it may include the author's name, subject name, etc.), then the program will directly use the existing information and transfer the files to proper places. Otherwise, the program will use additional tools to make up key information. For the music, documents, and images, the program will upload the files to the internet and use existing

algorithms to recognize related information. For example, information can cover the name (if missed in the file names), the album and the artists of different musical files. The existing online algorithm can also recognize the general contents of images and put them into various groups based on the results. Additionally, we have set numerous key words for each commonly used subject. Thus, by comparing the highest frequently used words in the documents and the existing key-word base, the program can make a precise judgment on the related subjects of documents, finally transferring the files to the right places. However, a different categorizing method is applied on the videos. Since examining the contents of videos is too difficult at the current stage, videos lacking crucial information will be directly transferred to a folder named 'other videos' under the video folder. Finally, the rest of the files like zip files will be moved to the other folder.

```
all_file_path = get_all_file_path('%s/Downloads' % (get_my_user_folder_path()))
i = 0
while i < len(all_file_path):
    try:
        source_file_path = all_file_path[i]
        split_filename = os.path.splitext(source_file_path)
        file_ext = str.lower(split_filename[1])
        if file_ext in music_ext:
            if music(source_file_path):
                num_organized += 1
            else:
                num_not_organized += 1
        elif file_ext in video_ext:
            if video(source_file_path):
                num_organized += 1
            else:
                num_not_organized += 1
```

Figure 1. Screenshot of code 1

In order to achieve the first step, firstly five general parts are given with frequently used extensions respectively. For example, extensions like 'jpg', 'jpes', and 'png' are under the image part. After assigning the extensions, the program will then loop through the whole download folder, as shown in the picture abode, and finish the first step up with all files flowing into five general sections. Further categorizing work will then take place.

```
def images(source_file_path: str) -> bool:
    try:
        api_key = 'acc_b580a33df0d99bb'
        api_secret = '508be5bf54a6c634ee73057797de606f'
        upload_info = requests.post(
            'https://api.imagga.com/v2/uploads',
            auth=(api_key, api_secret),
            files={'image': open(source_file_path, 'rb')},
        )
```

Figure 2. Screenshot of code 2

A crucial part of the second step is the disposal of images. The code abode is used to upload the files assigned to the image section to the existing online api. Then, the well-trained online program will go through each file and compare its features with the past information, determining

the type of its content. After that, the result will be sent back to the program. Based on the result, the program will transfer each image file to the appropriate place under the image folder.

```
try:
    audio_file = eyed3.load(source_file_path)
    if audio_file.tag == None:
        pass
    else:
        if audio_file.tag.album_artist == None:
            audio_file.tag.album_artist = 'unknown_artist'
        else:
            audio_file.tag.album_artist = change_forbidden_filename(audio_file.tag.album_artist)
        if audio_file.tag.album == None:
            audio_file.tag.album = 'unknown_album'
```

Figure 3. Screenshot of code 3

Another crucial part is processing the music files. The logic behind this part is very similar to the one of image section. Online api is applied again to get the necessary information. However, the result will be more precise and detailed this time. By examining the features of the audios and making comparisons to the audios around the world, the api can not only send back information about the general type of the music but also the name, the album, and even the author of the music. This is very beneficial for further categorization. If an artist has never appeared on the laptop, the program will create a folder using his name and then move the file to a minor folder named by the album under the artist name. Otherwise, if the artist folder exists but the album folder is absent, the program will create the minor folder again. Such a categorizing method is very detailed and effective.

The program will categorize each file one by one, according to their sequence in the download folder. After the last file entering the right place, the program will print out a line, notifying that the categorizing work has been finished. The notice will also include the number of files categorized. The whole process follows a simple but rigorous logic: recognizing the type of the file and then further categorizing it by examining its features and content [4]. In order for the program to run successfully, having a reliable network connection is indispensable. This is because the online apis take an important role in examining the contents of the files. Otherwise, every laptop using the program has to be trained properly long before starting the program to acquire the ability of recognizing contents offline.

4. EXPERIMENT

4.1. Experiment 1

In order to test the usefulness of our program, I borrowed five laptops, which showed some variety between the samples, from my friends and then ran the program on them. The process of the first experiment is very simple. Firstly, the program will be run on a laptop with multiple documents in its download folder. After that, we will check if the program crashes during the previous process. If the program does not crash, we will check if all files are cleaned up from the download folder and moved to the proper places. The time taken to clean up the files will also be measured.

The experiment went through quite well. All files from five laptops were cleaned up and were moved to ideal places. The average time taken to clean up each file was estimated to be around 2.5 seconds – – this was not very fast but still acceptable. Thus, a simple conclusion could be drawn: the program was powerful enough to cope with daily problems for a high-school student.

However, one thing worth mentioning was that 4 laptops, taking up 80% of the samples, were macbooks. Therefore, even though the program ran smoothly on the only laptop with a windows system, the program had yet been proven to be strong and stable enough in such a system.

4.2. Experiment 2

In order to test the effectiveness of the results, I designed a survey for 20 of my friends and asked them about the feeling of running my program. The survey had two parts: the first part was about the attitude towards the general categorization idea behind the program and the second part was their feeling about the final results.

The results showed that 17 people, taking up 85% of the samples, deemed the logic behind the program was reasonable. However, only 10% of them, which were 2 people, were satisfied with the final categorization results. The results of the survey were reasonable. Firstly, the general logic of categorizing matched up with the general needs and habits of the majorities – putting files of the same type together. However, as the categorizing work moved into more detailed parts such as the separations between document files, the effectiveness of the program was significantly limited.Since everyone has unique detailed habits and requirements, the unsatisfactory results were reasonable and predictable.

The results matched up with previous challenges to a large extent. The first experiment shows that the program could work smoothly in the majority of the environments. This means that the third challenge has been tackled successfully with the environment becoming compatible with numerous devices. In addition, the second experiment shows that the general logic behind the categorizing method is rational while it will encounter problems when dealing with more detailed parts. This matched up with the first and the second challenge. The first challenge has been successfully solved – we have created a categorizing system that could properly function. The difficulty of solving the second challenge is further addressed. Even though we could use the program to clean up the download folder, the result is still not personalized enough to cater to the customer's specific needs and habits.

5. RELATED WORK

In the first reference work, Kendrick the author has created a system to categorize the emails by using a nearest-neighbor classifier [5]. Such a classifier will compare the features extracted from the emails and the given target message, and then move the emails to the most matching group based on the result. Such a structure is similar to my approach to categorizing the documents: both having preset key words for classification. Kendrick's system is more powerful than mine as it will consider more features of the emails. While my work only extracts the features from the body of a document, Kendrick's system also takes the author and other parts into account. This can further increase the reliability of the classifier with more information provided.

In the second reference work, a new file management method, called CMF, is introduced [6]. Unlike traditional hierarchical file systems that follow a 'tree structure', CMF contains two essential parts, concepts and containment, which are converted from a folder. Like being assigned to a folder in a traditional system, a file is assigned with concepts in CMF. However, rather than being limited in a single place, a file can be under multiple concepts in such a system. The concepts all contain paths to the file, and users can access to the file through various paths. CMF is very different from my approach as my system follows the traditional way. The biggest advantage of CMF is that it can create quick access to files and maximize personalization. Once I

introduce such a system effectively to my existing system, the problem of lack of personalization can be tackled.

In the third reference work, the method of analyzing the content of images is introduced [7]. The model is trained numerous times to acquire the features of various groups of images. The api used in our program for image categorization shares the same concept with the reference work – training the model with a large number of preset images. Unlike the reference work in which the model is trained directly in the computer, my program does not download the model. When there is a need for image categorization, my program will send the image to the online api and retrieve the result. The advantage is that the users will not have to download the picture-categorizing system that can take up much space in their laptops.

6. CONCLUSIONS

In conclusion, we developed a system that can categorize different types of files and make further classification within the same type [14]. Machine learning is applied in the program, enabling it to identify the contents of pictures and audios. In addition, in order to categorize documents lacking key information, the program can extract features from their bodies and make comparisons with preset key words, reaching final results. Based on the classification results, the program is able to create branches under each type and then transfer the files to proper places. Two experiments have been conducted to test the effectiveness of the program. The first experiment showed that the program can run quite steadily on different devices. And the results matched perfectly with the assumptions. Thus, the first and second challenges were solved. A rational categorization system was built, and the program perfectly realized such a system. The second experiment further proved the rationality of the system — the majority of the users agreed with the logic behind the categorization system. However, the second experiment also revealed the fact that the third challenge is yet to be solved. Almost all the users deemed that the system is not personalized enough. Though the overall structure is reasonable, the system cannot be altered by users to satisfy their special needs and habits. Thus, the system is not very effective when it comes to details, and more personalized parts need to be created.

The accuracy of my method is proved to be very high – almost all the files are transferred to the most ideal places on the laptops. However, its practicability is relatively low. As mentioned before, the method does not contain choices for personalization. Therefore, all users have to follow the exact categorization structure designed by me. This is not practical enough as every person has different habits and needs. The users may find their personal needs unsatisfied in places where my habits conflict with theirs. In short, the program still has room for further optimization.

In order to improve the practicability of the program, I will add parts for personalization to the program [15]. As a result, the users can enter their special needs and habits for the program to follow. In addition, I will improve its capability of classifying the contents of pictures. Currently, the program is able to classify the pictures into over twenty types based on their contents. A refined training can help the program make further categorization within existing types, maximizing its effectiveness.

REFERENCES

[1] Rosch, Eleanor, and Barbara Bloom Lloyd, eds. "Cognition and categorization." (1978).

[2] Medin, Douglas L., and Evan Heit. "Categorization." Cognitive science. Academic Press, 1999. 99-143.

[3] Mock, Kenricj. "An experimental framework for email categorization and management." Proceedings of the 24th annual international ACM Sigir conference on research and development in information retrieval. 2001.

[4] Badashian A S, Mahdavi M, Afzali S H, et al. Supporting Multiple Categorization using Conceptual File Management[J]. American Journal of Scientific Research, ISSN, 2011: 129-136.

[5] Chen Y, Wang J Z. Image categorization by learning and reasoning with regions[J]. The Journal of Machine Learning Research, 2004, 5: 913-939.

[6] Truong B T, Dorai C. Automatic genre identification for content-based video categorization[C]//Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. IEEE, 2000, 4: 230-233.

[7] Lin C C, Chen S H, Truong T K, et al. Audio classification and categorization based on wavelets and support vector machine[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 644-651.

[8] Tuemmler, Brian. "Network shared drives: How to clean up files for better information management." Information Management 46.1 (2012): 26.

[9] El Naqa, Issam, and Martin J. Murphy. "What is machine learning?." machine learning in radiation oncology. Springer, Cham, 2015. 3-11.

[10] Rist, Robert S. "Program structure and design." Cognitive science 19.4 (1995): 507-562.

[11] Deng, Li, and Dong Yu. "Deep learning: methods and applications." Foundations and trends in signal processing 7.3–4 (2014): 197-387.

[12] Hao, Xing, Guigang Zhang, and Shang Ma. "Deep learning." International Journal of Semantic Computing 10.03 (2016): 417-439.

[13] Gu, Xiaodong, et al. "Deep API learning." Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering. 2016.

[14] Cormack, Richard M. "A review of classification." Journal of the Royal Statistical Society: Series A (General) 134.3 (1971): 321-353.

[15] Vesanen, Jari. "What is personalization? A conceptual framework." European Journal of Marketing (2007).

COST-EFFICIENT DATA PRIVACY PROTECTION IN MULTI CLOUD STORAGE

Artem Matveev

Buryat Institute of Info communication (branch of) Siberian State University of Telecommunication and Information Science, Ulan-Ude,
Republic of Buryatia, Russia

ABSTRACT

Data privacy in the cloud is a big concern for all of its users, especially for public clouds. Modern trends in studies utilise multiple clouds to achieve data privacy protection. Most of the present studies focus on business-oriented solutions, but current study aims to create a solution for individual users which would not increase the cost of ownership, and provide enough flexibility and privacy protection by combining password protection, key-derivation, multilayer encryption and key distribution across multiple clouds. New design allows to use single cloud to store protected user data, meanwhile use free plans on other clouds to store key information on others and thereby does not rise a cost of the solution. As a result, proposed design gives multiple layers of protection of Data Privacy while having a low cost of use. With some further adaptation it could be proposed as a business solution.

KEYWORDS

Multi Cloud Storage, Privacy Protection, Password, Key Distribution, Cloud Data Security.

1. INTRODUCTION

Increasing demands in processing data, quick service deployments with high-availability and at low costs resulted in forming a cloud computing model, also known as clouds, where separate users are sharing common resources, maintained by the cloud provider. Cloud storage is one of the services provided by cloud providers which allows users to store data on the provider's servers.

As the amount of data increases [1], [2], more and more data ends up uploaded to the cloud storage. Despite cloud providers does not provide direct statistic of data volume inside cloud storages, trend could be seen through growth of both revenue and user base, like in Drop box reports: [3], [4], [5], [6], [7], or can be found in analytics reports, such as [8].In many cases, especially in public clouds, the cloud provider is an independent organisation or person which means that data uploaded to the cloud is maintained and accessible not only by the end user itself, but by the cloud provider and its affiliated organisations or persons. These raise concerns over data privacy in the clouds. Past occurrences [9], [10], [11], [12], [13], [14], [15] have shown that this concern is not groundless, and some incidents with data leakage, business espionage and even government spying over the data confirms that.

The most effective solution to maintain data privacy is local side encryption performed shortly before data would be send into cloud storage. Meanwhile, modern times have shown that data located in cloud storage is required to be accessible at any time from any device. This is

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 67-81, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120706

especially stimulated by the Bring Your Own Device (BYOD) trend in businesses where employees can access data from their own mobile devices. On the other hand, individuals or small business users can have only a single device to access the data and still want to maintain their data privacy. It also needs to be considered that this device can be stolen, lost or severely broken (further referred to as lost).However, encryption requires an encryption key. As a result, the questions are raised: where to locate a key; how to keep it secret and perform its safe sharing and recovery in cases of loss.

The first simple solution is to use a password-based key generation, for example, described in [16]. This approach solves the issues where data would be inaccessible from another device or in case of device loss. But from the other point of view it needs to be considered that this way opens attack on brute forcing passwords (including usage of password dictionaries) and entire crypto strength fully relies on the password. Furthermore, it needs to considered that typically users are not using strong passwords or reusing their passwords so that making such attacks are a real issue. Finally, it also needs not to be forgotten about possible social engineering attacks on the end user. Altogether, these facts making this approach vulnerable to the end user behaviour.

Another solution is to save the encryption key on some end user's device such as a computer or smart-card. This way increases the cost of infrastructure and limits the list of devices from where access is available. Also, in case of device loss the data becomes unrecoverable. On the other hand, lack of proper infrastructure or improper actions with keys can lead to encryption key leakage [17]. As a result, this way is more oriented for some business applications rather than for individual users.

Next generation solutions no longer rely on single cloud provider and use the "divide and conquer" idea, but applied to data privacy security. The basic concept is to slice data into separate chunks and store them in multiple cloud storages from different providers. This opens a new promising paradigm in data security and privacy – multi cloud storage [18]. Since multiple cloud storages from different providers are used, one provider can no longer have full access to the data and that makes the end user the only person who is able to get all the data. From the intruder's perspective, access to the full data is also becoming problematic, with the exclusion of MITM and end-user device attacks, they need to gain access to several cloud providers. But this basic concept is still vulnerable to data guessing and is still able to disclose some part of the original data. Although those issues could be resolved by using data encryption, there arises 2 more issues which need to be resolved: (i) how scheme stores/derives the encryption key; (ii) the rising cost of the overall solution, due to the requirements in applying for multiple subscriptions.

In order to answer those issues and guarantee data privacy in cloud storage, this paper proposes a new design of using a multi-cloud environment and encryption which on one hand will utilise a single cloud to store the actual data and use password protection, but on the other hand involves a multi-cloud paradigm to prevent password brute force attacks. The proposed design provides multiple levels of protection while having low cost of use, because data is actually stored in a single cloud. This work mainly aims at maintaining privacy of individual or small business users, although with some further adaptation it could be proposed as a business solution. That achieved by combing existing well-known solutions (encryption, key-derivation) in specific order boosting security by using multiple clouds to store keyed information. New scheme compared with existing solutions from perspective of complexity crypto operations and from perspective of resistance on existing threads.

The remainder of the paper is formed as follows. Section 2 describes the overview of the related work in the field. Section 3 describes proposed scheme. Section 4 discussed implementation

aspects and performing threat and complex assessments. Section 5 concludes the report and future work.

2. LITERATURE SURVEY

A performed literature survey has shown a lot of research studies in a field of data security and privacy in a cloud throughout the past 2 decades. Although research studies have progressed especially in introducing using multiple clouds to reach the goals of data security and privacy, they avoid scenarios for personal usage purposes ([18], [19], [20], [21], [22], [23]). The survey results shown that only two studies have aims in design to reach individual users and another one has the perspective of being used as a possible solution, but with different intentions.

Studies [24], [25], [26] have performed state-of-art surveys and outlined possible threats, issues and ways of protecting data. Possible threats which were listed in those works are the following: Password cracking or Brute Force; Inconsistent Use of Encryption; Catastrophic Hardware Failure; Malware; DDoS; Man in the middle attack; Data leakage/Side channel attacks; Data Disclosure;

In [25], [26] also highlighted main aspects of data protection in clouds such as: Data Confidentiality; Data Integrity; Data Availability; Non-repudiation of Actions; Fine-Grained Access Control; Secure Data Sharing in Dynamic Group; Leakage-Resistance; Complete Data Deletion; Privacy Protection.

Study [26] declares Data Confidentiality as the ability to prevent the active attack of unauthorized parties on users' data, and ensure that the information received by the data receiver is completely consistent with the information sent by the sender; Data integrity as the reliability of the data, that is, the data cannot be arbitrarily tampered with and replace; Data Availability as Data availability emphasizes that data can be accessed normally at any time and Privacy protection as the ability to guarantee sensitive data protection under curious adversaries and malicious employees of cloud service providers. [25] describes Non-repudiation of Actions as (aspect which) ensures that neither party will be able to deny the occurring transaction.

A discussion about possible solutions (in-terms of fully backwards recoverable data) to these threats in [24] - [26] contains the following: use of strong passwords; hierarchical role-based access control; data encryption; using security level and data classification; tokenisation; Identity-Based Encryption; Attribute-Based Encryption; Homomorphic encryption; Searchable Encryption. Despite this, from a mathematical perspective, strong passwords represent a good enough solution and have the ability to protect data, [24] states that choosing and the proper usage of such kinds of passwords became not just a technical issue, but a behaviour one.

Other solutions mainly propose ways of how to perform Secure Data Sharing and building access systems. And the last is raising questions about the ability to find documents through performing confidential cloud searches.

Study [1] demonstrates one of the possible scenarios of usage – data collection from sensors located on/in factory's equipment. Individual users can experience similar scenarios of use with heavy growth of the Internet of Things. But this proposed solution requires installing an intermediate server which could not being suitable for individual usage.

Study [27] proposes a way of building cloud storage, but it also includes a proposal of using "boot code" technology, which uses a password and file name to generate an encryption key.

However, this method is equally vulnerable as just using passwords, because the filename is not hidden in this case.

Studies [18], [19], [20], [21], [22], [23] use a new paradigm – multiple clouds. Some of the works are just slicing files and then storing chunks in different clouds [20], the rest of the works include encryption of data chunks. Study [19] described an implemented prototype for eGovernment purposes in a conference. As one of suggestion was proposed to look into the field of mix networks and the security modelling used there. But none of the work is responding to the question about how shall the encryption key be located.

Study [28] addresses an issue about availability of such a solution for individual users. The basic idea is pretty similar as outlined in [18] - [23] – original file encrypting, splitting, compressing and storing in multiple clouds, but small pieces of the beginning and the end of files are stored locally at the end user device. This approach allows protecting data in the cloud, because in cases of using streaming symmetric ciphers there would no ability to properly decrypt a file, but this way placed restrictions on data accessibility and in cases of losing users' devices the data will be lost.

Study [29] suggests using a single location for encrypted data and later sharing information of the key by using Shamir's secret sharing and later storing it in different locations. But this work is not aiming at providing an end solution for individual users. Also, study [30] stated that this scheme is partially trusted and has a high computation cost.

Study [31] proposes using biometry, an identity server and user's auxiliary devices in order to protect the encryption used for encrypting backups from end-users' mobile devices. The main aim of the scheme is to provide safe backups of users' IDs and payment information and later the ability to restore them in case the device is lost.

Study [32] discusses issues of providing secure access to the file and meanwhile changing the encryption key. This solution is inventing proxy-servers which are performing re-encryption.

Study [33] provides the opportunity to protect secrets with distribution parts of the secrets in several servers, but the algorithm binds to private and public keys which is not suitable for public clouds where servers are not maintained by the end user.

3. PROPOSED DESIGN

3.1. Design Overview

Data security and privacy, as it was shown in 2. Literature Survey section of this work, are widely discussed. Different studies are proposing different approaches to cover different aspects of data protection. But as it was shown, mainstream studies are concentrated on providing business-oriented solutions. Some of them have potential issues with data availability due to introducing additional intermediate or end-users' servers which could not be as scalable as a cloud provider's infrastructure. Those approaches can also increase networking round-trip times due to ineffective network routing.

This work proposes a fully-multi-cloud solution where the password is left outside of the cloud and must be remembered by user. Meanwhile, protection of end users' privacy is boosted by using a multiple cloud solution. From another perspective, the current proposal only requires a single cloud to store protected data.

In addition, with proper configuration, current design allows protection against a broken encryption algorithm by allowing usage of several independent algorithms in a row.

3.2. Design Requirements and Notation

This section describes a list of features and requirements for the proposed scheme and introduces notations used later in the study. The proposed scheme shall:

- Use single cloud storage to store the main data;
- Use multiple cloud storages to enhance security;
- Provide the ability for the end-user to access data from new devices only by specifying a password and granting access to all clouds;
- Provide the ability to choose different encryption algorithms in order to protect data.

Used notation is described in Table 1.

3.3. Initialisation and Input Data

This section covers the initial and input data required for the scheme. To begin its operation, the scheme requires to be specified:

Table 1. The notation

Symbol	Description
P	Payload to protect
S_j	Cloud Storage
T	Amount of cloud storages
M	Master password
N	Amount of encryption layers
E_i	Symmetric encryption algorithm for protecting payload
R_l	Symmetric encryption algorithm for protecting key gamma sequences[1]
K_i	Session encryption key used in payload encryption
K_M	Encryption key based on master password
G_j	Random transformation gamma sequence[1]
G_{K_i}	Random key-deriving gamma sequence[1]
G_{C_i}	Transformed G_{K_i} with set of G_j
$B_i\ (G, G)$	Gamma-sequence reversible blend function
X_i	Key-extraction function from gamma sequence[1]
C_{P_i}	Encrypted payload
C_{K_i}	Encrypted gamma sequences[1]
A_i	Salt
$h(x, A)$	One-way password transformation function

- P
- M
- List of S_j with size of T
- List of E_i with size of N
- List of R_l with size of $N + T$
- $h(x, A)$

[1]In this study Gamma Sequences is considered as a random sequences of bytes.

- List of B_i (G, G) with size of N
- List of X_i with size of N

Although the list contains more than just payload, password and cloud storages, the rest of the things could be implemented as settings and could be shipped with factory-prepared values. As a result, for the end user it would not be mandatory to specify those values.

For the first launch, the scheme is also required to randomly generate G_j and .G_{K_i}. In other cases, the system needs to extract and decrypt C_{K_i}. This step is described later in Key Information Decryption and Recovering section.

All blend functions must be reversible, i.e. if $X = B(Y, Z)$ then $Y = B\ (X, Z)$ shall be true.

3.4. Protecting Payload

Payload protection basically represents layered encryption of the original payload as shown in Figure 1. The steps are performed as present the equations 1-3 by their numbers.

1. Extracting encryption keys K_i from key-deriving gamma sequences G_{K_i} with extraction function X_i, as in equation 1.

$$K_i = X_i\left(G_{K_i}\right), i \in [1, N] \tag{1}$$

2. Encrypting the payload P several times to form layered encryption with extracted keys K_i with encryption function E_i, as in equations 2 and 3 respectively.

$$C_{P_1} = E_1(P, K_1) \tag{2}$$
$$C_{P_i} = E_i\left(C_{P_{i-1}}, K_i\right), i \in [2, N] \tag{3}$$

3. Store the encrypted payload C_{P_N} to the main cloud storage S_1.

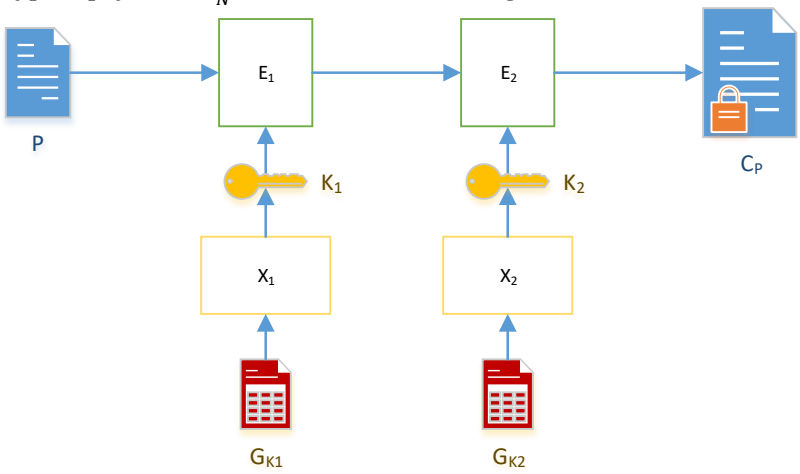


Figure 1. Encryption sequence flow

3.5. Protecting key information

Key information is protected by (i) blending them together and (ii) encrypting it with a master password given by the user. Later this key information could be stored into different cloud storages. The part of the scheme is illustrated in Figure 2.

1. Blend key-deriving gamma sequences G_{K_i} with a transformation gamma sequences G_j with blend function B, as in equations 4 and 5 respectively

$$G_{C_i} = G_{K_i}, \forall\, i \in [1, N] \tag{4}$$
$$G_{C_i} = B\left(G_{C_i}, G_j\right), \forall\, j \in [1, T] \tag{5}$$

2. Encrypt transformed key-deriving G_{C_i} and transformation gamma sequences G_j with the master password M, one-way password transformation function h and encryption function R_l, as in equations 6 and 7 respectively.

$$K_{M_l} = h(M, A_l) \tag{6}$$
$$C_{K_l} = R_l\left(G_l, K_{M_l}\right) \tag{7}$$

where l is either $j \in [1, T]$ or C_i, $i \in [1, N]$; the salt A_l is a randomly generated sequence of bytes, which, in the current scheme, is different for each gamma sequence; K_{M_l} is gamma's sequence's generation's encryption key based on password M and salt A_l.

3. Store encrypted gamma sequences and their salts into different clouds S_i.

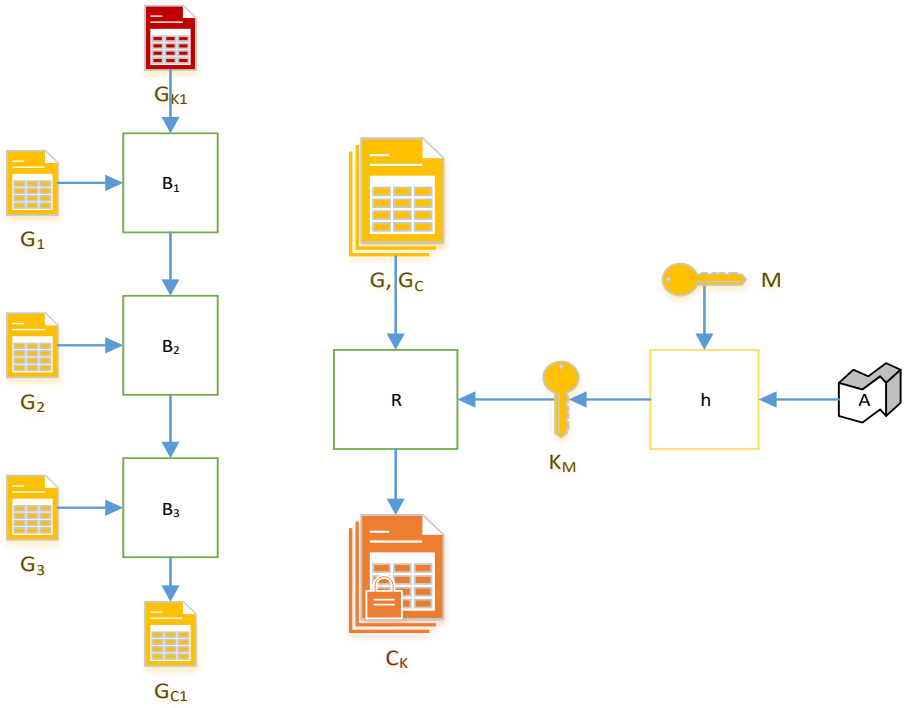


Figure 2. Gamma Sequence transformation and encryption flow

3.6. Key Information Decryption and Recovering

Keys recovering and decryption basically represents backward action presented in equations 4-7. The process of key reconstruction of following steps:

1. Read encrypted C_K and salt A from the cloud storages.

2. Generate an encryption key K_{M_l} from the master password M for each gamma sequence G_l as described in equation 6.

3. Decrypt gamma sequences G_l with given keys K_{M_l} by encryption function R_l as in equation 8.

$$G_l = R_l\left(C_{K_l}, K_{M_l}\right) \tag{8}$$

4. Perform backward blending with the transformed gamma sequences G_{C_i} with a transformation gamma sequences G_j by blend function B to get key-deriving gamma sequences G_{K_i}, as shown in equations 9-10.

$$G_{C_i} = B\left(G_{C_i}, G_j\right), \forall\, j \in [T, 1]\,, i \in [1, N] \tag{9}$$

$$G_{K_i} = G_{C_i} \tag{10}$$

3.7. Payload Decryption

Payload decryption simply consists of (i) loading the encrypted payload from the cloud; (ii) extracting encryption keys by equitation1 and (iii) performing decryption in reverse order for each E_i.

4. DISCUSSION

4.1. Implementation Aspects and Considerations

Some security aspects of the scheme depend heavily on the chosen implementation details and this section intends to highlight possible facts and aspects which need to be considered for the scheme implementation.

The best suited symmetric algorithm mode for this scheme is streaming mode, such as CBC mode at least [34]. The ECB mode is not recommended to use due to its ability to contain vulnerabilities and because it can lead to possible information disclosure [35]. The cryptography strength is dependent on the strength of the strongest used algorithm.

As function h could be considered function PBDKF2 which makes brute forcing the password hard due to its high computational cost. This function is required only in cases of either encryption or decryption keyed information, but does not directly involve generating keys for payload encryption and therefore has little computational cost impact in terms of payload encryption.

Blend function B could just be XOR. Since gamma sequences are random, XOR would be enough to provide a proper XOR cipher scheme.

Extraction function X could be a function which just slices the first bytes from key-deriving gamma sequences, although it could be any other function, including PBDKF2. Since keys are static for all payloads they would be needed to be extracted only once and this shall not significantly increase computational burden.

Also, for the purposes of increasing unpredictability, cipher Message Authentication Code (MAC) could be considered to discard and, instead, implement one's own MAC at the top level before the encryption process starts. This increases the difficulty to guess encryption keys for each layer independently and will require decrypting all layers in order to confirm that all keys and the decrypted payload are valid, meanwhile it still confirms data integrity. As an alternative approach to implementing own MAC, the MAC from the first algorithm could be left in encrypted stream, while the rest of them are discarded.

Meanwhile for gamma sequences, the MAC could be omitted and instead of it, a test file could be used which would be encrypted with multi-layered encryption as a usual file. This makes it impossible to guess a derived key from one stolen set of gamma sequences. The intruder would have to combine all sequences, generate $N + T$ gamma keys, decrypt all sequences, generate N payload keys, decrypt the test payload and check it in order to confirm success.

An additional measure of increasing unpredictability could be considered using randomised start indexes in gamma sequences, which makes it hard to know which section of bytes need to be targeted. Example of such approach shown in

Figure 3.

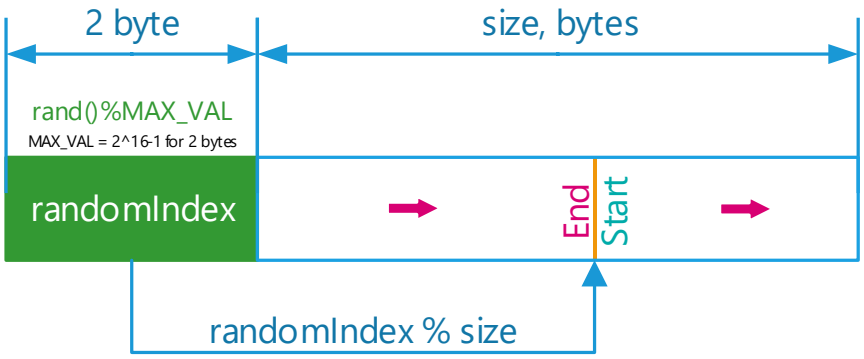


Figure 3. Gamma sequence container with randomised start index

Gamma sequences could be cached in end user devices which reduces risks that this information could be stolen. Additionally, this information on end user devices could be protected by biometric sensors.

The other measure which could be applied to encrypted key-deriving gamma sequences is to avoid/deprioritise using main cloud to store them.

4.2. Threats and Data Protection Aspects Analysis

Current section of present study intends to analyse threats and data protection aspects discussed in this study. Meanwhile it also needs to be considered that providing security at end-user device is out of scope for current research study and this study assumes that end user devices is secure.

4.2.1. Password Brute Force

As was discussed earlier some of the solutions are vulnerable to password brute forcing. It is obvious that full password brute forcing is complex and typically impossible with good length and alphabet, but using password dictionaries makes this task much easier to do. In this case, even if the password was brute forced, the intruder needs to collect all gamma sequences from different cloud storages in order to extract all encryption keys.

Also section 4.1. Implementation aspects and considerations proposes an implementation technique where just owning encrypted keys on the one hand gives no ability to brute force passwords, and, in another significantly slows that process by proposing to use [16] and different salts for each key. Therefore, from this perspective of this threat, the scheme provides Data Confidentiality.

4.2.2. Inconsistent Use of Encryption

The proposed scheme itself could contain such vulnerability, but it was discussed in section 4.1. Implementation aspects and considerations. If proposals would be implemented, those issue would be mitigated.

Furthermore, this scheme allows mixing different algorithms which enables taking the advantages of each algorithm and in case of algorithm cracking still provide security for its content

Therefore, from this perspective of this threat, the scheme provides Data Confidentiality.

4.2.3. Catastrophic Hardware Failure, DDoS

In terms of hardware failure, studies consider that cloud storages typically have strict policies [36], [37], [38] about good hardware redundancy, and most of them store data across several data centres, therefore it very unlikely to that cloud storage will have such enormous failures. Besides, it needs to be acknowledged that individual users typically use a single cloud to store their data and this way does not introduce additional risks, especially if the key cache scheme was implemented on end users' devices providing the ability to recover them -or- end users' can duplicate some keyed information inside cloud storage groups (instead of saving keyed information directly to the cloud – save it into group, like RAID disk groups).

Failure of intermediate communication lines also does not introduce more risks in data loses. Most of them are temporary and won't cause big issues for individual users. Meanwhile, permanent connection loss to keyed information could be mitigated with same ways as cloud storage failure.

The most likely scenario of catastrophic hardware failure would be the end-user's device's unrecoverable inoperability for example device loss. In this case, the user would be able to access their data by accessing their cloud storage accounts and entering their master password. From this perspective, this scheme provides the end-user high Data Availability, even in cases of device loss or an immediate need to access from a new device.

4.2.4. Man in The Middle Attack

Currently all communication between end-user and the cloud storage provider are secure, typically by using the TLS protocol. From this perspective, this study is not considering that a MITM attack could have taken place near the end-user's device.

A MITM attack could be performed inside a cloud provider's data centre, but this does not provide any advantages for the intruder since they need to access keyed information across different cloud providers thereby continuing to provide Data Confidentiality.

4.2.5. Data Leakage/Side Channel Attacks and Data Disclosure

As discussed above, to actually provide an intruder with any advantages of stolen data they need to access several cloud storages. Meanwhile it's very unlikely that several cloud data storage providers will leak information at the same time and it has the same applications on Data Confidentiality as discussed above.

Also, post-incident mitigation measures could be applied – generating new gamma sequences and re-encrypt all content in the background in order to protect it with new keys.

4.2.6. Malware

Malware could be on either the cloud provider's side or end-user device.

Cases of malware on the cloud provider's server are similar to sections 4.2.4. Man in the middle attack and 4.2.5. Data leakage/Side channel attacks and Data Disclosure a scenario that is very unlikely to occur at the same time in different clouds. Cases of malware on the end-user device is out of scope for this study and relates to more specific security tools such as anti viruses.

4.3. Assessing Scheme Complexity

For the purposes of assessing scheme complexity, hereinafter it will be assumed that: (i) proposals from "4.1. Implementation aspects and considerations" are implemented; (ii) single password test iteration is a pair of one key-deriving and decryption operations; (iii) single key test iteration is a decryption operation. In opposition to the current scheme used, simple scheme consists of: (i) key-deriving from the password; (ii) decryption; (iii) MAC/integrity check. Both schemes are equally likely to use same key-derivation and decryption functions and, hereinafter, it is assumed that they are using the same set of functions for each layer independently (if applicable). This section is looking into scheme complexity from two perspectives: (i) password brute forcing; (ii) key brute forcing.

4.3.1. Password Brute Force Complexity

This assessment calculates the amount of password test iterations I which needs to be done to traverse through the entire list of passwords with length equal to $L.L$ is an input value and its calculations are out of scope for this study.

It is obvious that simple schemes take one password test iteration per single password and the overall amount can be found as in equitation 11.

$$I = L \tag{11}$$

The proposed design takes $N+T$ password test iterations to decrypt gamma sequences; N iterations to decrypt payload (if consider B and X functions as a key-deriving function) per single password. The overall amount of password test iterations can be found as in equitation 12. As the proposed design takes $2N+T$ times more iterations. Because of the possibility to choose different algorithms, the actual taken time could be more than that, but, in the worst case, it would not be less than $N+T+1$ times more. This time can be improved by performing Q layered encryption for gamma sequences, where each single layer will have its own salt, presented in equitation 13.

$$I = L(2N+T) \tag{12}$$
$$I = L(N[Q+1]+QT) \tag{13}$$

With an example setup of $T=3$ clouds and $N=2$ layers of encryption, password brute forcing will take 7 times more password test iterations than simple scheme and for $Q=2$ layered encryption of gamma sequences it will be 12 times more iterations.

It is worthwhile to say that this advantage is possible by implementing the following proposals: custom implementation of MAC at top level; removing MAC from gamma sequences. Without them, the complexity could be decreased as specified in equitation 11, but not less than that.

4.3.2. Key Brute-Force

This attack is much harder to be performed, because of the amount of combinations, but due to key information distributed across different cloud services it could be hard for the attacker to get all the necessary information in order to perform a password based attack. In this assessment, the amount of key test operations U would be calculatedwith overall key amount W.

Simple scheme takes one key test operation per each key. This is shown in equitation 14.

$$U = W \tag{14}$$

The proposed design offers N layered encryption and requires to individually pick an encryption key for each encryption level. Also, it need to be considered that it is unlikely that 2 random keys would have identical values (for example, the probability of 2 identical random keys for AES-256 is equal to $P(K_1K_2) = \left(\frac{1}{2^{256}}\right)^2 = \frac{1}{2^{512}}$) and thereby it can be considered that keys are not reused. As a result, the overall amount of iterations would be calculated as k-permutation of n and shown in equitation15.

$$U = A_W^N = P(W,N) = \frac{W!}{(W-N)!} \tag{15}$$

With an example setup of $N=2$ layers of encryption this way will have $W-1$ times more key test operations in order to reach content.

4.4. Known Limitations of the Scheme

The proposed scheme significantly boosts Data Privacy protection inside cloud storages, there exist some limitations of the scheme:

- Requirement of accessing to, at least, 2 cloud storages;

- Scheme itself does not offers resilient against access loss of one cloud, but mitigation strategy was being briefly discussed;
- File search became much harder due to computation cost and network delays;
- Inability to share data with another use.

5. CONCLUSIONS

This work showed that Data Confidentiality in the cloud storages is still experiencing lack of solutions for individual or small business users. To resolve this issue, current study proposed new scheme which is offering a solution for protecting user's data privacy, while not increasing the cost of such protection and utilising advantages of the multiple cloud paradigm. This solution uses combination of well-established security techniques, such as password-based key deriving, symmetric encryption functions, meanwhile involving multiple clouds to place keyed information across them in order to enforce privacy protection and as a result offer multiple levels of Data Privacy protection in the cloud storages. Performed analysis of possible threats and data protection aspects has shown that the scheme offers high levels of protection and mitigates listed threats; analysis of scheme complexity has shown improved complexity against brute force attacks. Also, the study proposes some considerations for possible implementations which can further enhance protection.

Although, this study proposes a ready-to-use scheme and some possible enhancement was discussed earlier in the paper, there still exists the question of implementing a group sharing scheme to allow users to safely share their files.

ACKNOWLEDGEMENTS

I would like to thank a few people who helped me to accomplish this work. First my thanks to my supervisor (Ivan Nechta), who got interested in my idea and work and provided feedback on this paper. Big thanks to my friends from Canada (Jordin McEachern) and UK (Ahmed Elakehal) who reviewed and helped me fix my English misspelling and errors. Also, big thanks to my cat, Cisco. She was sitting around me during my entire first and third days of writing this paper.

And special thanks for the Organizers of ADCO for providing the opportunity to publish this work.

REFERENCES

[1] J. Liu, C. Yuan, Y. Lai and H. Qin, "Protection of Sensitive Data in Industrial Internet Based on Three-Layer Local/Fog/Cloud Storage," Security and Communication Networks, vol. 2020, p. 2017930, 04 2020.

[2] B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," 21 05 2018. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/.

[3] Dropbox, Inc., "2018 Press Release," Q4 2018. [Online]. Available: https://dropbox.gcs-web.com/news-releases/news-release-details/dropbox-announces-fourth-quarter-and-fiscal-2018-results.

[4] Dropbox, Inc., "2019 Press Release," Q4 2019. [Online]. Available: https://dropbox.gcs-web.com/static-files/28e269b2-3442-43bb-b635-53f72e3d26f3.

[5] Dropbox, Inc., "Q4 2020 DBX Investor Presentation," Q4 2020. [Online]. Available: https://dropbox.gcs-web.com/static-files/37497899-90c7-44a1-91da-be8d1c6ed333.

[6] Dropbox, Inc., "Q4 2021 DBX Investor Presentation," Q4 2021. [Online]. Available: https://dropbox.gcs-web.com/static-files/b1e5d02b-6a9f-452d-be29-28a5632a3c2d.

[7] Dropbox, Inc., "Fourth Quarter 2021 Earnings Release," Q4 2021. [Online]. Available: https://dropbox.gcs-web.com/static-files/9ba48281-f330-451c-add8-5e28fded2ef6.
[8] T. Coughlin, "Digital Storage Projections for 2019, Part 3," 27 12 2018. [Online]. Available: https://www.forbes.com/sites/tomcoughlin/2018/12/27/digital-storage-projections-for-2019-part-3/.
[9] Dropbox, "Yesterday's Authentication Bug," 20 06 2011. [Online]. Available: http://web.archive.org/web/20110721173153/https://blog.dropbox.com/?p=821.
[10] BBC, "Dropbox hack 'affected 68 million users'," 31 08 2016. [Online]. Available: https://www.bbc.co.uk/news/technology-37232635.
[11] J. Raphael, "Google Docs Glitch Exposes Private Files," 09 03 2009. [Online]. Available: https://www.pcworld.com/article/527794/google_docs_glitch_exposes_private_files.html.
[12] S. Larson, "Data of almost 200 million voters leaked online by GOP analytics firm," 19 06 2017. [Online]. Available: https://money.cnn.com/2017/06/19/technology/voter-data-leaked-online-gop/index.html.
[13] R. Lakshmanan, "A Google Docs Bug Could Have Allowed Hackers See Your Private Documents," 29 12 2020. [Online]. Available: https://thehackernews.com/2020/12/a-google-docs-bug-could-have-allowed.html.
[14] Guardian News & Media Limited, "NSA Prism program taps in to user data of Apple, Google and others," 2013. [Online]. Available: https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.
[15] BBC, "Edward Snowden: Leaks that exposed US spy programme," 17 01 2014. [Online]. Available: https://www.bbc.co.uk/news/world-us-canada-23123964.
[16] The Internet Society, "PKCS #5: Password-Based Cryptography Specification Version 2.0," September 2000. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc2898.
[17] National Institute of Standards and Technology, "SP 800-57. Recommendation for Key Management: Part 1 – General," National Institute of Standards and Technology, Gaithersburg, MD, USA, 2020.
[18] M. B. Vaidya and S. Nehe, "Data security using data slicing over storage clouds," in 2015 International Conference on Information Processing (ICIP), Pune, 2015.
[19] P. Chiaro, S. Fischer-Hübner, T. Groß, S. Krenn, T. Lorünser and e. al., "Secure and Privacy-Friendly Storage and Data Processing in the Cloud," in Privacy and Identity Management. The Smart Revolution, M. Hansen, E. Kosta, I. Nai-Fovino and S. Fischer-Hübner, Eds., Ispra, Springer, Cham, 2018, pp. 153-169.
[20] R. Pottier and J.-M. Menaud, "Privacy-aware Data Storage in Cloud Computin," in 7th International Conference on Cloud Computing and Services Science (CLOSER 2017), 2017.
[21] P. Xu, X. Liu, Z. Sheng, X. Shan and X. Shan, "SSDS-MC: Slice-based Secure Data Storage in MultiCloud Environment," in 2015 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE), Taipei, 2015.
[22] Dr.K.Subramanian and F. John, "Secure and Reliable Unstructured Data Sharing in Multi-Cloud Storage using the Hybrid Crypto System," IJCSNS International Journal of Computer Science and Network Security, vol. 17, no. 6, pp. 196-206, 06 2017.
[23] K. Subramanian and F. L. John, "Dynamic and secure unstructured data sharing in multi-cloud storage," International Journal of Advanced and Applied Sciences, vol. 5, no. 1, pp. 15-23, 01 2018.
[24] F. Yahya, R. J. Walters and G. B. Wills, "Protecting Data in Personal Cloud Storage with Security Classifications," in 2015 Science and Information Conference (SAI), London, 2015.
[25] S. amamou, Z. trifa and M. khmakhem, "Data protection in cloud computing: A Survey of the State-of-Art," in Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES2019, Budapest, 2019.
[26] P. Yang, N. Xiong and J. Ren, "Data Security and Privacy Protection for Cloud Storage: A Survey," IEEE Access, vol. 8, pp. 131723-131740, 07 2020.
[27] R. Wang, "Research on data security technology based on cloud storage," in 13th Global Congress on Manufacturing and Management, GCMM 2016, Hulunbuir, 2016.
[28] Alqahtani and D. H. Saad, A novel approach to providing secure data storage using multi cloud computing, Bedfordshire: University of Bedfordshire, 2019.
[29] Z. Huang, Q. Li, D. Zheng, K. Chen and X. Li, "YI Cloud : Improving user privacy with secret key," in 2011 IEEE 6th International Symposium on Service Oriented System (SOSE), Irvine, 2011.
[30] N. M. Joseph, E. Daniel and N. A. Vasanthi., "Survey on Privacy-Preserving Methods for Storage in Cloud Computing," in IJCA Amrita International Conference of Women in Computing, Amritanagar, 2013.

[31] O. Mir, R. Mayrhofer, M. Hölzl and T.-B. Nguyen, "Recovery of Encrypted Mobile Device Backups from Partially," in ARES 2018: Proceedings of the 13th International Conference on Availability, Reliability and Security, Hamburg, 2018.
[32] J. Shen, X. Deng and Z. Xu, "Multi-security-level cloud storage system based on improved proxy re-encryption," EURASIP Journal on Wireless Communications and Networking, p. 277, 12 2019.
[33] M. Abdalla, M. Cornejo, A. Nitulescu and D. Pointcheval, "Robust Password-Protected Secret Sharing," in Proceedings of the 21st European Symposium on Research in Computer Security (ESORICS '16), Heraklion, 2016.
[34] M. Vaidehi and B. J. Rabi, "Design and analysis of AES-CBC mode for high security applications," in Second International Conference on Current Trends In Engineering and Technology - ICCTET 2014, Coimbatore, India, 2014.
[35] D. Jayasinghe, R. Ragel, J. A. Ambrose, A. Ignjatovic and S. Parameswaran, "Advanced modes in AES: Are they safe from power analysis based side channel attacks?," in 2014 IEEE 32nd International Conference on Computer Design (ICCD), Seoul, Korea (South), 2014.
[36] Microsoft Corporation, "Data Resiliency in Microsoft 365 - Microsoft Service Assurance | Microsoft Docs," 18 11 2021. [Online]. Available: https://docs.microsoft.com/en-us/compliance/assurance/assurance-data-resiliency-overview.
[37] Google LLC, "Disaster recovery planning guide," 24 08 2018. [Online]. Available: https://cloud.google.com/architecture/dr-scenarios-planning-guide.
[38] Dropbox, "Dropbox Business Security: A Dropbox Whitepaper," 2019. [Online]. Available: https://aem.dropbox.com/cms/content/dam/dropbox/www/en-us/business/solutions/solutions/white_paper/dfb_security_whitepaper.pdf.

AUTHORS

Artem Matveev, M.S. at the Siberian State University of Telecommunication and Information Science in Computer Science; Lead Engineer at the IT-Department and Teacher in Buryat Institute of Info communication.
His area of interest is wide and includes system and networking administration, software development, cyber security and electrical engineering.

FRAUD DETECTION SYSTEM BASED ON ARTIFICIAL IMMUNE SYSTEM

Vitaly Krokhalev

Siberian State University of Telecommunication and Information Science,
Novosibirsk, Novosibirsk Oblast, Russian Federation

ABSTRACT

Nowadays, one of the most important problems for financial companies is fraud related to online transactions. It is becoming increasingly sophisticated and advanced, leading to financial losses on the part of both customers and companies. Based on this, my company was tasked with creating a fraud detection system that is scalable and adaptable to change. This research aims to create a solution that can be used to identify differences in customer behavior patterns and detect fraud. The artificial immune system model proposed in this article, combined with certain informative features, is simple to implement and can describe customer behavior patterns.

KEYWORDS

Fraud Detection, Artificial Immune System, Informative Features, Machine Learning, Information Security.

1. INTRODUCTION

A fraud transaction detection system, a fraud monitoring or antifraud system, is a system designed to evaluate financial transactions on the Internet for suspicion of fraud and offer recommendations for their further processing. Currently, a significant number of monitoring systems are aimed at detecting certain fraud signatures. This approach detects only fraudulent operations described in the signatures, and for a fairly short period of time - attackers adapt, find new vulnerabilities and use new tools for the next attacks.

The most pressing and significant problem in building a fraud detection system is the imbalance of classes. The array of analyzed data is very large, assorted, and imbalanced, so there is a problem of processing a very large volume of data. According to statistics, the number of fraudulent transactions per total number of transactions does not exceed 0.01%. Such a huge imbalance significantly complicates the detection of fraudulent operations. It follows that the problem of detecting such operations (the problem of classification) should be solved in the context of anomaly detection.

In anomaly detection, we assume that there is a "normal" distribution of data points, and anything that sufficiently deviates from this distribution is an anomaly. If we transform the classification problem into an anomaly detection problem, we can consider the majority class as a "normal" distribution of points, and the minority as anomalies.

The anomaly detection problem is a complex problem and belongs to the class of poorly formalized. Various heuristic algorithms, including bioinspired algorithms, are used to solve such

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 83-94, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120707

problems. Heuristic algorithms do not guarantee finding the optimal solution, but they allow one to obtain solutions of acceptable quality quite fast. Bioinspired algorithms are based on the use and modeling of the principles of organization and functioning of various natural systems, such as neural networks of the brain, the immune system, the evolution of living organisms, the genetic laws of heredity and variability, and swarm intelligence.

Artificial Neural Network (ANN) models are some of the most common bioinspired algorithms nowadays. ANNs allow to solve very complex data processing problems, so their application in the context of anomaly detection was considered first. In "Credit Fraud. Dealing with Imbalanced Datasets"[1], the author implements a neural network with three fully connected layers, describes in detail various errors connected with the class imbalance, and shows the necessity of data preprocessing before feeding into a neural network. This means that using the ANN model is not the most preferable solution for our problem.

Alternatively, we can consider a relatively new class of bioinspired algorithms, the Artificial Immune Systems (AIS) class. AIS continue to be actively studied and are increasingly used in various fields, including information security. Some AIS models, such as the negative selection model, show high efficiency in finding anomalies.

One of the main differences between AIS models and ANN models is their learning. ANN learning is implemented by a special algorithm focused on the type of task and the type of a given ANN, by presenting images of different classes followed by adjusting the weights of the links. In AIS learning is implemented by creating recognition elements - detectors, by positive or negative selection of information units of images of only one class "of their own". Thus, the AIS model can work well in conditions of strong class imbalance.

In the general case, the negative selection model is limited by the possibility of binary classification "friend or foe". However, the task of monitoring fraud involves the use of a degree of confidence in the prediction, which characterizes the probability that a given transaction is fraudulent.

While using the AIS model to create a fraud detection system, the following questions arose:

- How to implement a confidence analysis of the prediction made by the AIS;
- How to identify informative features based on customer transaction information.

This article considers the selection and creation of informative features for the AIS model within the frame of solving the task of fraud transactions detection.

2. LITERATURE SURVEY

The literature review showed that a significant number of publications have been devoted to AIS, noting the successful application of AIS in various application areas. Although research is moving forward, detailed information on the application of AIS in fraud detection is currently insufficient, as financial companies do not disclose specific details of the implementation of antifraud systems.

The article "Artificial immune systems: review and current state"[2] reviews the current state of artificial immune systems. Their problems, advantages and disadvantages, current developments in the field of artificial immune systems, and their areas of application are considered. This article also provides a comparative table of the following heuristic bioinspired algorithms:

- Genetic algorithms
- Neural networks
- Artificial immune systems

For each of the algorithms, the corresponding components are listed. The table shows that the main families of algorithms from the class of biological algorithms that are used have a lot in common. The article notes a significant advantage of AIS over genetic algorithms and artificial neural networks is the ability to learn and the availability of memory.

The article "How to choose the antifraud system?"[3] considers the main principles of modern antifraud systems, actual problems of antifraud systems, the efficiency of different methods of fraud detection, as well as the most popular and effective machine learning algorithms. The following problems are cited as the most significant problems:

- The imbalance of legitimate and fraudulent transactions; •The necessity of detecting fraudulent activities in real-time.
- Dynamically changing fraudulent behavior;
- Significant differences in customer behavior patterns.

The article notes the particular effectiveness of machine learning methods compared to signature rule-based approaches, noting the widespread use of machine learning-based antifraud systems in recent years. A comparative table of key machine learning algorithms is provided. Algorithms were compared based on their frequency of use and the following evaluation criteria:

- The algorithm should have high accuracy in detecting fraudulent actions when processing large amounts of data - "accuracy";
- The algorithm must cover the maximum number of possible fraud scenarios - "coverage";
- The algorithm must be the least expensive in terms of both time and money - "cost".

The article "Machine learning against fraud in banking"[4] describes the historical approach and practical experience in detecting fraudulent transactions, describes the basic principles that must be followed when solving the problem of creating a fraud detection model. In addition, the main problems encountered in solving this problem are considered. According to the cybersecurity service practice described in the article, in order to build an effective fraud detection system, it is necessary to create additional features describing customer behavior in addition to existing transaction data. It is noted that a wide range of different aggregations and mathematical functions are usually used when creating features: percentiles, averages and deviations, sliding windows, and many others. Part of the article is devoted to metrics for estimating model efficiency.

The article "Credit card fraud detection using neural network and geolocation"[5] proposes a system based on a neural network that facilitates the detection of fraudulent transactions by analyzing the location of the customer, while taking into account the structure of his expenses. The main value in the article is the classification and description of various fraudulent schemes.

The article "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective"[6] describes current technologies in credit card fraud detection, lists and compares a large number of different machine learning techniques (including most AIS algorithms), and classifies these techniques into two main approaches to fraud detection: abuse detection (using teacher learning) and anomaly detection (using teacherless learning).

3. PROPOSED DESIGN

3.1. Implement a Confidence Analysis

The analysis of the prediction confidence of the AIS is associated with the use of numerical features, which leads to the expediency of using the appropriate detectors - numerical detectors. In this case, the numerical features of the analyzed data are represented by points in the feature space, which can be interpreted as vectors emanating from the origin of the feature space. Numerical detectors are characterized by the coordinates of their centers in feature space and radii. The comparison of the features of the analyzed data with the detectors is carried out on the basis of proximity measures between the corresponding vectors in the feature space.

The information units that the AIS algorithm characterizes as representatives of the "foreign" class must fall into the area of one of the detectors, and the representatives of the "own" class must not fall into any detector. Then the probability that the transaction is fraudulent should be greater than 0.5 if the point is inside the detector. If the point is at the detector boundary, then the probability will be equal to 0.5. We need a function that will match the received number with another number from 0 to 1 (this is what we will interpret as a probability). As such a function, you can use the sigmoid function. When working with detectors, the algorithm will operate on the distances between the test point (the current transaction) and the center of the detectors. Let's define the extreme cases: if the point hits the center of the detector directly, then the result is 0, that is, the sigmoid argument is -∞. When hitting the boundary of the detector, the argument is 0, and when moving away from the detectors, the argument is +∞. Thus, the closer the point is to the center of the detector, the lower the degree of transaction reliability.
The general formula will be as follows:

$$p = sigmoid\left((1-y)*\frac{1}{k}\sum_{i=1}^{k} abs(R_i - r_i)*r_i + y*\frac{-(R-r)}{r}\right)$$

where $y = \{0, 1\}$ is a binary identifier of a point belonging to some detector.

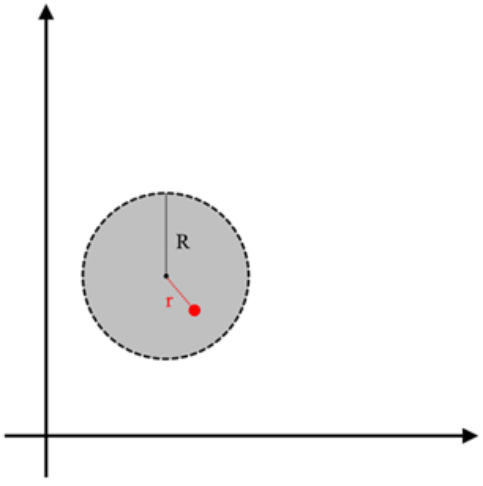


Figure 1. Pont entering the detector area

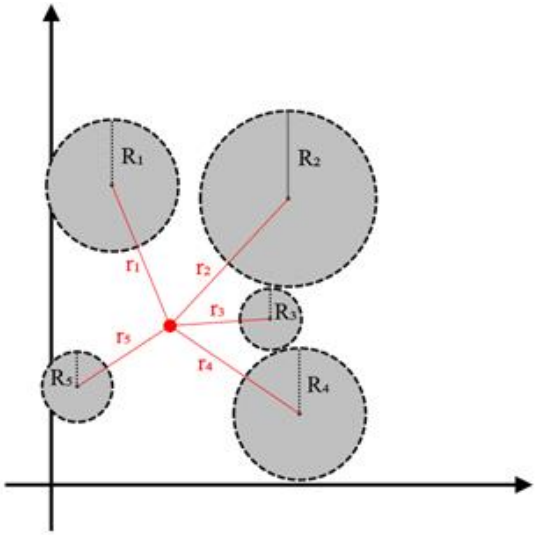


Figure 2. The point is outside the detector areas

A common example of a sigmoid function is the logistic function defined by the formula:

$$S(x) = \frac{1}{1 + e^{-x}}$$

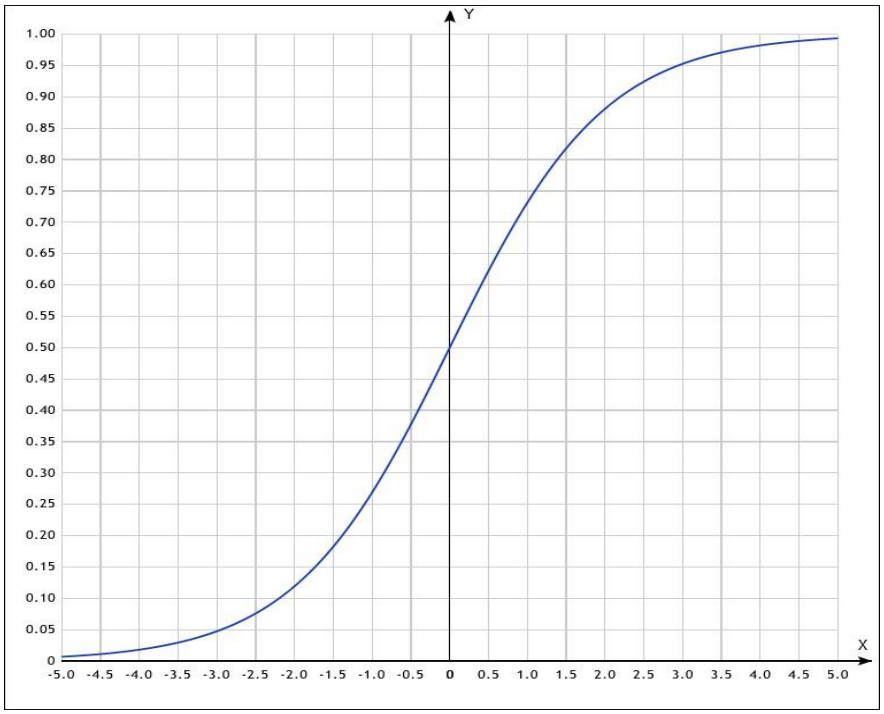


Figure 3. Sigmoid curve of the logistic function

When the behavior of the function changes, the end result will also change: when the point approaches the center of the detector or moves away from the detector boundary, the decrease or increase in the degree of prediction confidence will change more sharply. For the logistic function, you can enter the parameter α, which affects the nature of the function, making it flatter

or sharper. This approach makes it possible to perform a more precise adjustment of the detector parameters, which is optimal for a specific task.

The formula will be as follows:

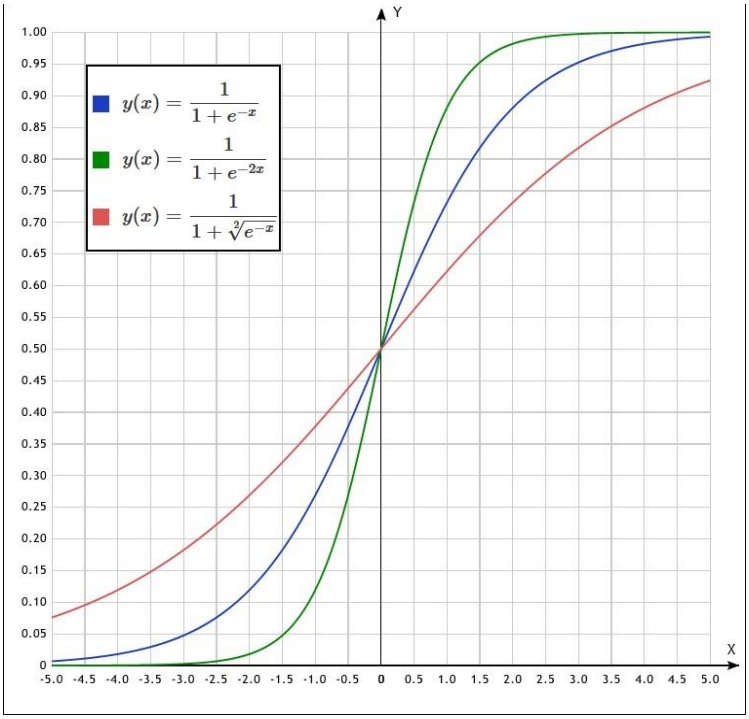


Figure 4. Sigmoid curve of the logistic function with different α

3.2. Creating Informative Features

The application of AIS in the field of fraud detection is described quite often, but the bulk of the information falls on the comparison of AIS with other algorithms, excluding the details of system implementation, in particular, the selection and creation of informative features. We will consider the features for the AIS model in more detail.

Denote the information about the transaction that comes into the system. The data structure representing information about a transaction includes the following fields:

- id (string) - unique identifier of the transaction
- merchant_id (string) - unique merchant identifier
- service (string) - unique service name (mobile bank, PayPal, etc.)
- amount (floating point number) - transaction amount
- currency (string) - currency code
- createdAt (date and time) - date and time of transaction initiation
- customer_id (string) - unique identifier of the client
- status (string) - unique identifier of the current transaction status
- history (history object array) - contains the history of the transaction statuses

History object includes the following fields:

- status (string) - unique identifier of the current transaction status
- date (date and time) - date and time of the status assignment

Of the presented data fields, the following ones are informative features:

- merchant_id
- service
- amount
- currency
- time

Some of the already extracted features are of the string type, and some are of the numeric type. The AIS algorithms can handle both the first type and the second, but for simplicity, it would be reasonable to convert the initial and further obtained features into a numeric format. Such a conversion is justified by the fact that it is not important for us, for example, what specific currency or service was used, what matters is the presence of anomalous behavior: that is, the same currency (or service) is used for this particular transaction as before, or it is the first time this happens.

The practice described in the articles shows that in addition to the existing transaction data, it is necessary to create additional features describing the client's behavior. Many new features can be generated from the existing source features by applying various techniques.

To begin with, we have to condition ourselves on the existence of such a concept as time periods. There will be several of them, they are usually needed for the sliding window. Suppose there will be 4:

t_{p1} = 1 hour, t_{p2} = 24 hours, t_{p3} = 7 days, t_{p4} = 30 days, t_p – unspecified period.

Based on the information from the articles, we propose some new features computed on the basis of the original ones.

Amount

- The average value of *amount* for t_{p1} during t_{p2} (t_{p3}), divided by the total value of *amount* for a year (or other long period);
- The average value of *amount* for t_{p2} during t_{p3} (t_{p4}), divided by the total value of *amount* for a year (or other long period).

Service

Percentage of transactions with this *service* for period t_p.

Amount + Service

- Percentage of *amount*, spent using this *service* for the period t_p;
- Average value of *amount*, spent using this service in period t_p, divided by the annual total value of *amount* (or other period).

Currency

Percentage of transactions using this particular *currency* for the selected period.

Currency + Amount

For each currency used you can add the features from the *Amount* section.

Tnx (transactions)

The average number of transactions in period t_{p1} (t_{p2}) during t_{p2} (t_{p3}). This value can be normalized if the variation of transaction frequency is too different from client to client.

Tnx + Amount

The average number of the amount per transaction during the period t_p, divided by the total amount per year (or other long period).

Time

For time, the following feature is suitable: whether the time of transaction initiation falls into the confidence interval. To do this, you need to do a little preprocessing of the data: calculate the average, deviation, and set the probability for the interval. The peculiarity is that in this case the mean and deviation are calculated in a different way because for the time it is necessary to calculate the periodic average by the following formula:

$$\mu_{vM}(D) = 2\tan^{-1}\left(\frac{\sum_{t_j \epsilon D}\sin(t_j)}{\left(\sqrt{\left(\sum_{t_j \epsilon D}\cos(t_j)\right)^2 + \left(\sum_{t_j \epsilon D}\sin(t_j)\right)^2} + \sum_{t_j \epsilon D}\cos(t_j)\right)}\right)$$

Then the deviation is calculated:

$$\sigma_{vM}(D) = \sqrt{\ln\left(\frac{1}{\left(\frac{1}{N}\sum_{t_j \epsilon D}\sin(t_j)\right)^2 + \left(\frac{1}{N}\sum_{t_j \epsilon D}\cos(t_j)\right)^2}\right)}$$

The calculated values are substituted to calculate the von Mises distribution as follows:

$$x_i^{time} \sim vonmises\left(\mu_{vM}(S_{per}), \frac{1}{\sigma_{vM}(S_{per})}\right)$$

The final formula for calculating the von Mises distribution will be as follows:

$$f(x \mid \mu, k) = \frac{e^{k\cos(x-\mu)}}{2\pi I_0(k)}$$

Then, by setting the probability, we find out whether the time of initiation of the given transaction is within the confidence interval. To perform mathematical operations, the time is converted to floating point numbers.

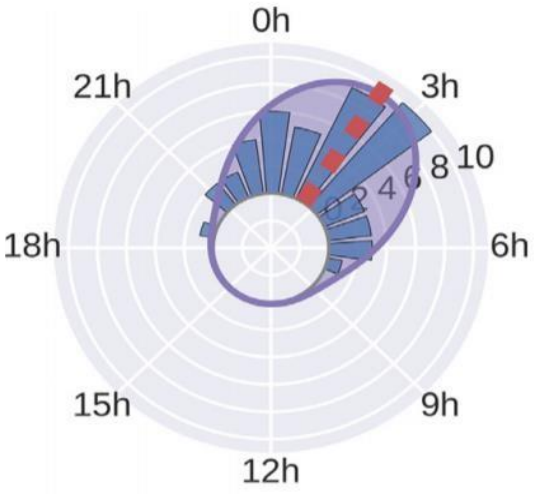


Figure 5. A visual representation of the described case. The red line is the periodic average, the purple outline is the resulting distribution. The columns represent transactions.

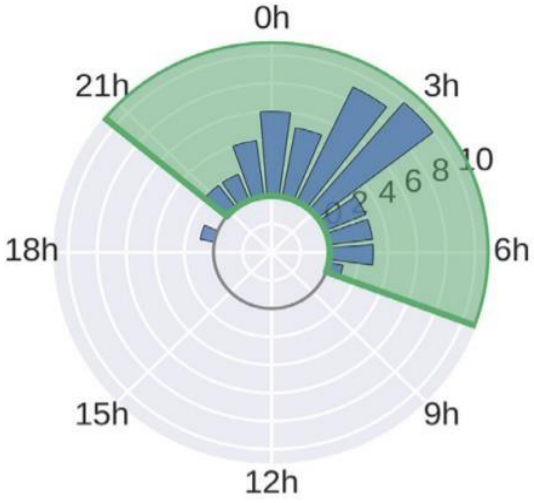


Figure 6. A visual representation of the described case. The transaction columns inside the green sector fall within the confidence interval

4. DISCUSSION

For the above features, it is assumed that the number of transactions in the history of each client is sufficient. However, in cases where the user has a small number of transactions, false triggers of the system become possible. For some features, this problem is solved by rationing, for example, in the case of the amount. However, for some features, there are separate cases, for example, for such a feature as the percentage of transactions with use (which is already a normalized value). If the user's history contains four transactions using one service, and a new transaction will use a different service, in this case, the percentage of use for the new service will be zero, which can affect the model prediction, although obviously there is a short history of

transactions. In this case, an additional summand can be introduced for such values. An example is shown in the following table.

Table 1. Example with an additional summand

a	0,01				
k	0,05				
tnx	1	2	3	4	5
service	s_1	s_1	s_1	s_1	s_2
% s_1	100	100	100	100	80
s_1 count	1	2	3	4	4
% s_2	0	0	0	0	20
s_2 count	0	0	0	0	1
% s_1'	99,9500	99,9750	99,9833	99,9875	79,9900
% s_2'	4,9500	4,9750	4,9833	4,9875	22,9900
softmax s_1	0,7211	0,7211	0,7211	0,7211	0,6388
softmax s_2	0,2789	0,2789	0,2789	0,2789	0,3612

Table 2. Description of the used variables

tnx	Number and quantity of transactions
service	Type of service
s_1 count	Number of transactions with the first service type
s_2 count	Number of transactions with the second service type
% s_1, % s_2	Initial current percentage of transactions with this service
% s_1, % s_2	Recalculated current percentage of operations with this service
softmax	Normalized recalculated current percentage of operations with this service

The formula for recalculation from $s_i\%$to $s_i\%'$:

$$s_i^{\%\prime} = s_i^{\%\prime} - \frac{k * s_{top} - s_{curr} + a}{tnx}$$

Where s_{top} – is the number of operations with the most frequent view, s_{curr} – is the number of operations with the current view, tnx – total number of transactions, k, a – constants less than one. Strictly speaking, after recalculating the percentage, it is no longer percent, since its sum is not equal to 1, so it should be further normalized in some way. In this case, this function $softmax$ was given as an example.

5. FUTURE ENHANCEMENT

In order for the AIS to be more accurate, it is possible to select the optimal parameters for calculating the degree of confidence. For example, by sorting through various options for the sigmoid parameter, you can find the optimal value at which the degree of prediction error will be minimal.

When training AIS, you can change the size of the window that analyzes the data, which affects the distribution and size of the detectors. It is also possible to reduce the degree of prediction error by examining different window sizes. For a more detailed analysis of the transaction, you can enter learning for different time periods, such as a regular day for a certain time, a regular day of the week, a day of the month, and so on. The AIS can be retrained over time so that more recent data is used and outdated data no longer affect the prediction result.

In addition to the listed features, we can add another group of features with *merchant_id* by analogy with the previous features. For example, *Tnx + Amount*, taking into account transactions with a specific merchant.

6. CONCLUSION

Fine-tuning of detectors and the proposed methods of obtaining informative features necessary to build a fraud detection system allow to describe the client's behavior in detail. The use of additional features helps to identify complex or previously unknown fraud patterns using all available parameters, as well as to adapt to changing fraud schemes.

ACKNOWLEDGEMENTS

I express my gratitude to my supervisor (Ivan Nechta), who helped to correct various shortcomings and gave a review on this article. Special gratitude is expressed to the organizers of ADCO for giving the opportunity to publish this work.

REFERENCES

[1] Kaggle, "Credit Fraud. Dealing with Imbalanced Datasets". [Online]. Available: https://www.kaggle.com/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets.

[2] Chernyshev Yu.O., Grigoriev G.V., Ventsov N.N., "Artificial immune systems: review and current state," in *Software products and systems*, vol. 4, pp. 136-142, 2014.

[3] Nikita Andreyanov, "How to choose the antifraud system?" in *Journal IT-Manager*, 2019. [Online]. Available: https://www.it-world.ru/cionews/manage_secure/148780.html

[4] Andrey Pinchuk, "Machine learning against fraud in banking," in *Journal IT-Manager*, 2017. [Online]. Available: https://www.it-world.ru/cionews/manage_secure/118786.html

[5] Aman Gulati, Prakash Dubey, MdFuzailC, Jasmine Norman, Mangayarkarasi R, "Credit card fraud detection using neural network and geolocation," in *IOP Conference Series: Materials Science and Engineering*, pp. 1-6, 2017.

[6] SamanehSorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective," *Cornell University*, 2016. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1611/1611.06439.pdf

AUTHOR

Vitaly Krokhalev M.S. at the Siberian State University of Telecommunication and Information Science in Computer Science; Android developer at Elementpay international company. His area of interest includes software development for Windows, Linux, and Android.

SENTIMENT ANALYSIS OF CYBER SECURITY CONTENT ON TWITTER AND REDDIT

Bipun Thapa

College of Business, Innovation,
Leadership, and Technology, Marymount University, USA

ABSTRACT

Sentiment Analysis provides an opportunity to understand the subject(s), especially in the digital age, due to an abundance of public data and effective algorithms. Cybersecurity is a subject where opinions are plentiful and differing in the public domain. This descriptive research analyzed cybersecurity content on Twitter and Reddit to measure its sentiment, positive or negative, or neutral. The data from Twitter and Reddit was amassed via technology-specific APIs during a selected timeframe to create datasets, which were then analyzed individually for their sentiment by VADER, an NLP (Natural Language Processing) algorithm. A random sample of cybersecurity content (ten tweets and posts) was also classified for sentiments by twenty human annotators to evaluate the performance of VADER. Cybersecurity content on Twitter was at least 48% positive, and Reddit was at least 26.5% positive. The positive or neutral content far outweighed negative sentiments across both platforms. When compared to human classification, which was considered the standard or source of truth, VADER produced 60% accuracy for Twitter and 70% for Reddit in assessing the sentiment; in other words, some agreement between algorithm and human classifiers. Overall, the goal was to explore an uninhibited research topic about cybersecurity sentiment.

KEYWORDS

NTLK, NLP, VADER, Sentiment Analysis, API, Python, Polarity, Evaluation Metrics.

1. INTRODUCTION

Through rapid digitization across the globe that produces a voluminous amount of public data, mostly through social media [1], an area of research that is burgeoning is sentiment analysis or opinion mining [2]. The use of NLP [3] to understand the sentiment of public opinion often presents a prelude to a bigger picture, invaluable and prescriptive information in the digital age. Organizations, political parties, technology, and dependents to the public sentiment or opinion benefit from the foresight [4]. In this context, descriptive research on the perception of cybersecurity in the public domain would provide meaningful feedback to the industry and the entities involved [5]. Presumptively, cybersecurity is often viewed through cynical lenses, and with the frequent unfolding of negative events in the news media [6], it would be insightful to understand if a similar sentiment persists in the social media platforms. Leveraging Twitter data to analyze public sentiment in the modern research literature is common [7]; however, it is partial to another popular platform, Reddit, which has shown to be influential in its rights [8]. The research area of sentiment analysis is relatively young, less than 15 years, albeit experiencing a surging growth due to data availability, with most of the earlier studies focusing on the most optimum algorithms for classification [9]. The objectives of sentiment analysis could be for understanding customer feedback [10], perception of the healthcare system [11], or to improve education quality from an educator's point of view[12], among many other things. Similarly, this

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 95-107, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120708

research is intended to focus on cybersecurity, which through initial analysis is an uninhibited domain in sentiment analysis and hence can potentially provide insights for actors involved. To conduct the research, social media content about cybersecurity, the topic, on Twitter and Reddit will be collected and analyzed through VADER (Valence Aware Dictionary for sEntiment Reasoning), a classification tool [13]. The documents from each source, Twitter and Reddit, will be collected to analyze the sentiment of the content. Concurrently, a small sample size of the content from Twitter and Reddit is human-classified for comparison. Human classification of the content could be utilitarian in understanding the correlation between machine-produced classification and human-produced classification [14], and to understand the efficacy of the algorithm. Sentiment, for this research, is labelled to be either 'positive', 'negative' or 'neutral', for both human and machine classification techniques. In essence, this research primarily aims to explore an inquisitive query, which is to know the opinion of cybersecurity through social media content for a specific time duration. In addition, the collected content will be preprocessed, which is to clean and normalize irregular content to yield better algorithm performance [15], and assessed for the sentiment classification. The classification evaluations metrics [16] will be used to measure the performance of the VADER against human labelling and frequenting entities in the content will be determined.

The three research objectives (RO) are,

RO1: To understand the sentiment of cybersecurity content posted on Twitter and Reddit in a given timeframe.
RO1.1: To understand the sentiment of cybersecurity content posted on Twitter and Reddit in a given time frame without text preprocessing.
RO2: To determine the most mentioned entities on Reddit and Twitter.
RO3: To evaluate VADER against human classification.

The methodology to fulfill the research objectives above are listed below.

2. SIGNIFICANCE, ASSUMPTIONS AND LIMITATIONS

This precedes explanatory research in sentiment analysis of Cybersecurity content on social media while assuming that human classification is accurately reflected for comparison. The limited sample size of human classifiers in the survey could have established biased opinions.

3. LITERATURE REVIEW

Often classifying a sentiment (positive, negative or neutral) of opinions is considered to be a difficult classification problem to solve even with the machine learning integration. The insistence on understanding, and modelling, to understand the sentiment of opinion, however, has not been abated; frequently new researches offering higher performance or unique approaches are presented. Digitization has yielded an enormous amount of data, and an abundance of publicly accessible communication mediums (mainly social media) allows for researchers to observe the pulse of the public sentiment. With this research area being fluid and ever-changing, for literature review, the focus was on peer-reviewed journal articles, conference papers, and API libraries, post-2010 with exception of an article from 2013. It was important to reference, learn and identify gaps from recent work due to the dynamic nature of the field. The literature review observed existing practice and their outcomes, positive and negative. Thereafter, novel methods of refining existing research ideas and methodologies were identified.

Regarding machine learning algorithms, two popular approaches are; supervised and unsupervised learning algorithms [17]. Supervised learning starts with labeled input data, which have defined features or attributes. With the features, algorithms are presented with the relationship of the data, then trained and asked to perform. Unsupervised learning is without the labeled input data; the algorithm will identify inherent structure or relationship that cannot be easily and manually replicated. Naïve Bayes, Decision Trees, and Support Vector Machine are examples of supervised learning which have been used to identify sentiments.

Whilst not a machine learning algorithm in the traditional sense, the Lexicon-Based Approach (LBA), VADER, provides an alternative to understanding sentiments; it has a compilation of previously classified sentiment terms that it will compare with and yield a polarity score to determine its sentiment. LBA is further divided into the Dictionary-Based Approach (DBA) and Corpus-Based Approach (CBA). DBA gathers a smaller set of words without context and classifies their polarity, making DBA somewhat flawed. CBA uses statistical or semantic tools to find context to the words, but it is not as nearly efficient in gathering the data and will take a much longer time to develop a large set [18].

Carlos Costa et al., to address their explanatory research objective of Portuguese parties (RO2: Identify the global sentiment per political party in Twitter communication), used rule-based VADER to classify the sentiment of the tweets. The tweets were first translated into English then assessed via VADER for polarity score, which then was aggregated by the party for comparative score [19].

U. Yaqub et al., conducted subjectivity and polarity analysis, sub-domain of sentiment analysis, on Twitter data of US and UK elections. As their exploration showed, the online sentiment in many ways reflected real public opinion. They postulate that being able to understand the online (Twitter) sentiment can greatly predict the public opinion or decision that is forthcoming regardless of differences in data collection [20]. A larger tweet dataset and inclusive study of all states (US), not just ten populous states, would be more beneficial to determine the outliers.
Another research tries to find a relationship, in the form of correlation, between the sentiment analysis of StockTwits, a microblogging site, and the stock price to accurately predict the direction of the price [21]. The researchers used supervised machine learning algorithms and featurization techniques with a positive correlation and accuracy of more than 61 percent in the five companies examined.

Due to sentiment analysis being a complex classifier problem, it has lagged behind in producing accuracy compared to categorization problems by almost 10 percent [22], and in order to yield higher results, preprocessing or cleaning of input data is important. Haddi et al. recommend data transformation/filtering, classifying, and evaluating. In the first stage, data is isolated from tags and removed stop words. For the second step, using the 4:1 ratio for training: testing with 10 folds cross-validation and lastly evaluating the performance of the model. Some or all combinations of these can greatly reduce the noise, which hinders producing strong sentiment analysis results.

Numerous researches have been conducted to find sentiment analysis of Twitter data using various algorithms and methodologies. Reddit, another popular social media outlet that is abreast with fresh news, is fairly uninhibited by the research community. The sentiment classification has been mostly done through automated computing without the correlation of results with human sentiment analysis to see if they differ significantly. The gap this research can attempt to fulfill, therefore, is an analysis of sentiments from the same data topic (cybersecurity) derived from Twitter and Reddit, which would then be classified by humans, labeling it with a sentiment. The

latter will be inferential in nature, simply validating random classifications from the algorithm to find commonality or discord when compared against human classification

4. METHODS

A descriptive design is adopted for this research, which notes the observation of the phenomenon but will not be able to conclude why it is occurring [23]. In this research, social media contents can be classified but its causation cannot be confirmed by corroborating evidence. The content from social media, Twitter and Reddit, is collected, analyzed, and presented to provide insight into a research area currently uninhibited, potentially providing a precursor to deeper and narrower research in the future.

Figure 1 provides a general overview of the research methodology with data preprocessing. The research dataset is created by filtering Reddit (posts) and Twitter (tweets) for cybersecurity content that falls under specific criteria (Table 3). From those two platforms, two different comma-separated values (CSV) files, twitter.csv, and reddit.csv, are extracted using API libraries praw and tweepy. Those files are preprocessed to be legible for sentiment analysis and fed through the VADER (Valence Aware Dictionary for sEntiment Reasoning) algorithm, which classifies the content of the CSV file appropriately. A small sample of 10 items from each file, which have been analyzed by VADER is collected and presented to the human surveyors for annotation through the form of a survey. The VADER classification and human classification are then compared to find differences, if any.

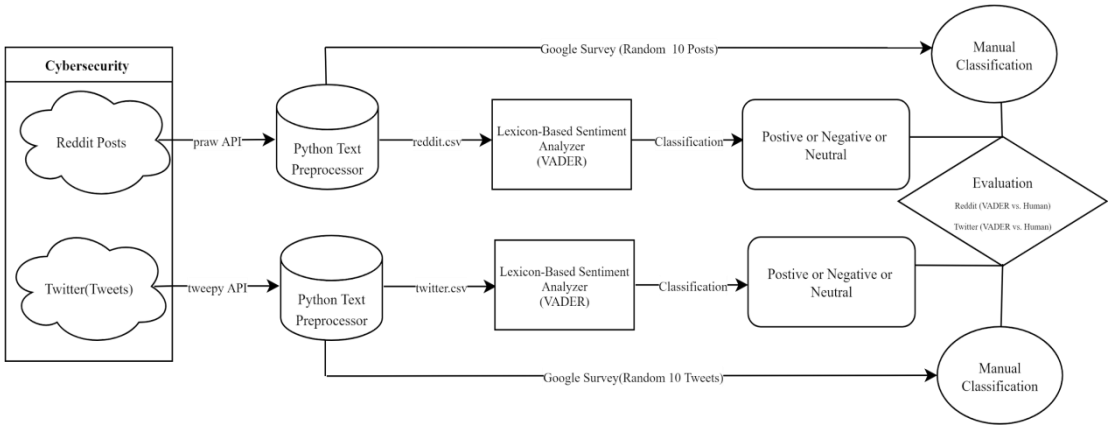


Figure 1. Proposed Method (With Python Text Preprocessing)

In Figure 2, text preprocessing is removed; as demonstrated by the creators of VADER [13], the extra step might not provide much value to the analysis of the sentiment.

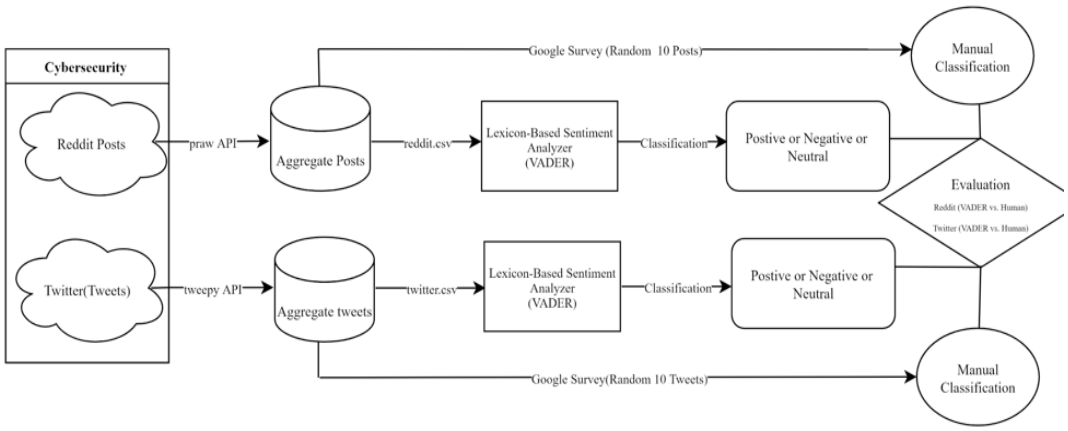


Figure 2. Proposed Method (Without Python Text Preprocessing)

4.1. PRAW (The Python Reddit API Wrapper)

The PRAW is a read-only API wrapper to extract Reddit posts through Python. It requires developer enrollment in Reddit, which provides credentials (Client ID, Client Secret, and User-Agent); this information authenticates the request against the Reddit server to gather posts from subreddits as desired. For this methodology, 'top' and 'hot' are gathered, which identify trending posts on Reddit[24]

4.2. Tweepy (Python Library for Accessing the Twitter API)

Similar to PRAW, Tweepy requires a developer application, when approved, will provide the users with credentials (consumer_key, consumer_secret,access_token, and access_token_secret) to authenticate and collect tweets depending on various parameters[25]. For this methodology, tweets with certain hashtags were collected.

4.3. Text Preprocessing

Text preprocessing is often recommended for data optimization when working with a large number of unstructured entries. The impetus is high on social media data that contains informal language, repetitions, URLs, and abbreviations [26], which creates noise, and as such clouds the 'real' sentiment behind the opinion. Figure 3 illustrates common steps for Text Preprocessing but can be expanded depending on the data type**.**

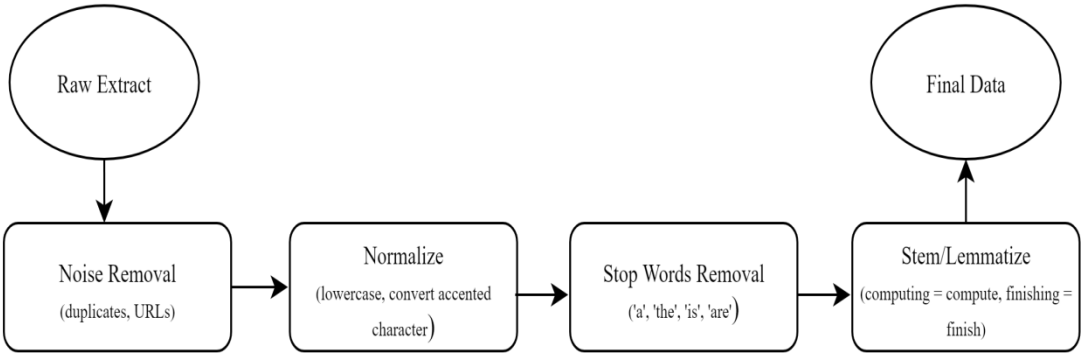


Figure 3. Standard Text Preprocessing for Sentiment Analysis

However, Hutto and Gilbert, the developers of VADER, state that sentiment heuristics play an important role in estimating the writer's mood. The punctuation, capitalization, tri-grams can amplify the mood [13]. Therefore, the dataset (reddit.csv and twitter.csv) for RO1.1 is directly fed into the VADER tool to understand the significance of

4.4. VADER (Valence Aware Dictionary for sEntiment Reasoning)

The foundation of this open-sourced, NTLK algorithm is a dictionary with corresponding sentiment features. The content, words, and phrases included are rated for polarity and intensity based on its built-in comprehension that recognizes more than 7500 features from -1 to +1, which the algorithm assesses for polarity, 'negative', 'positive, and 'neutral' scores, yielding a final 'compound' score. For a sentence, a final classification is based on 'compound' from the words. A positive compound score is 'positive', a negative 'compound' score is 'negative', and '0' is neutral. VADER's results in the past research have had outstanding accuracy at (F =0.96) compared to human classification, which was at (F1 = 0.84) [13], which makes this a popular tool of choice. Table 1 illustrates the VADER classification of individual words with an appropriate sentiment rating.

Table 1. VADER Sentiment Classification (Words)

Word	Positive	Neutral	Negative	Compound	Final
'sentiment'	0	1	0	0	Neutral
'dangerous'	0	0	1.0	-0.47	Negative
'excellent'	1	0	0	0.57	Positive

Table 2 classifies the sentiment of the entire sentence; similarly, a large data set can be processed in bulk to find the aggregate classification.

Table 2. VADER Sentiment Classification (Sentences)

Word	Positive	Neutral	Negative	Compound	Final
'data uses python.'	0	1	0	0	Neutral
'sentiment is interesting'	0.57	0.42	1.0	0.40	Positive
'security is difficult'	0.40	0.16	0.42	-0.02	Negative

4.5. Survey

There were thousands of tweets and posts collected via the framework to create the dataset, and it would not be possible to classify these manually. To find a comparative baseline, using Python's random.sample() function, ten tweets and ten Reddit posts were extracted. Using Google Form, the participants were asked to classify the tweets and posts into three sentiment classifications (positive, neutral, and negative). The survey was close-ended and anonymous, with participants required to classify the content with just one answer to each question. The survey adopted Snowball Sampling Method, where the form was posted on social media, and participants were requested to recruit other participants to increase the sampling size. The participants could originate from any sector, and upon completion of the survey, the human-classified sentiment could be used to be compared against VADER-generated analysis. The responses for each post and tweet would be assessed for polarity, with the majority score providing designation for final classification.

4.6. Data Extraction Criteria

The twitter.csv and reddit.csv datasets were created over a one-week window. Reddit has multiple subreddits (sections) dedicated to cybersecurity content; the three most popular ones were picked. Similarly, for tweets, the same hashtags bearing the name of the subreddits were picked for consistency. For Reddit, only posts that were 'hot' or 'top' were chosen, and on Twitter, tweets with at least one 'like' was chosen. This was to ensure that there was no favorability of the content by another user. Filtering using this method would substantially decrease the observations in the dataset.

Table 3. Dataset Creation Criteria

	Window for Collection	Section from Social Media	Content Filtering
Reddit	10/27/2021 - 03/11/2021	Subreddits(Cybersecurity, computer security, privacy)	only 'top' or 'hot' posts
Twitter	10/27/2021 - 03/11/2021	Hashtags(#cybersecurity, #computersecurity, #privacy)	only tweets with likes >0
Survey	04/11-2021- 11/11/2021	N/A	N/A

5. RESULTS AND DISCUSSIONS

5.1. To Understand the Sentiment of Cybersecurity Content Posted on Twitter and Reddit in A Given Timeframe

The preprocessing of posts and tweets included removing stop words, stemming, normalizing, as shown in Figure 3. It is intended to clean up the noise, which potentially could alter the sentiment of the contents. There were 32481 and 1205 observations made from Twitter and Reddit, respectively within the aforementioned timeframe and criteria. From Table 4 below, the Base Polarity, which is the standard scale, indicates that there were more positive sentiments than negative, both in Twitter and Reddit Posts. The Moderate Polarity expands the classification criteria, where the content has to be over the .25 threshold (negative or positive); this is to identify firmer sentiments. Consequently, this increased the neutral distribution, but still, there were more positive sentiments on both platforms than negative. Lastly, Extreme Polarity measures strong sentiment toward the content where the polarity threshold was .75 (negative or positive). Again, positive sentiments supersede negative sentiments. Twitter was more conducive to positive sentiments than Reddit, albeit the latter has a significantly smaller sample size. Most Reddit posts were neutral, whereas Tweets were either positive or neutral. Cumulatively, the sentiment on cybersecurity content is mostly positive or neutral, while negative sentiments last in every assessment.

Table 4. Polarity Classification for Preprocessed Data

	Obs.	Base Polarity 0=neu,>0=pos,<0=neg	Moderate Polarity >0.25=pos,-.25 <0=neg	Extreme Polarity 0.75=pos,-.75 <0=neg
Twitter	32481	49%= pos 28% = neu 22.5% = neg	42%= pos 41% = neu 16.5% = neg	8%= pos 90% = neu 1.7% = neg
Reddit	1205	26.5%= pos 56% = neu 17% = neg	22.5%= pos 64.5% = neu 13% = neg	1%= pos 98.5% = neu 0.4% = neg

Note: Obs. = Observations, 'pos' = positive, 'neu' = neutral and 'neg' = negative

5.2. To Understand the Sentiment of Cybersecurity Content Posted on Twitter and Reddit in a Given Time Frame without Text Preprocessing

Since the goal was to find the overall sentiment of the observations in percentage, duplicated observations would alter the representation; hence, one exception was made to raw data; only duplicates were removed. The standard preprocessing (Figure 3) was avoided and observations were directly fed in VADER to yield the polarity assessment. The unprocessed Twitter content produced similar sentiment scores compared to processed data; it had a similar narrative and the deviation between the comparable scores was less than 2%, with most of the sentiment being positive or neutral, indicating that preprocessing of Twitter content didn't shift the narrative. The Reddit posts were much more affected by the preprocessing; a significant percentage of neutral content became positive. A minor increase in negative scores also occurred but was not as significant. Overall, unprocessed content was in line with the narrative that most cybersecurity contents on these platforms were mostly positive or neutral, as listed in Table 5.

Table 5. Polarity Classification without Preprocessed Data

	Obs.	Base Polarity 0=neu,>0=pos,<0=neg	Moderate Polarity >0.25=pos,-.25 <0=neg	Extreme Polarity >0.75=pos,-.75 <0=neg
Twitter	32481	48%= pos 29% = neu 22.5% = neg	41.5%= pos 42% = neu 16.5% = neg	8.5%= pos 89.5% = neu 1.6% = neg
Reddit	1205	34%= pos 45% = neu 20% = neg	30%= pos 55% = neu 15% = neg	1.8%= pos 97% = neu 0.74% = neg

Note: Obs. = Observations, 'pos' = positive, 'neu' = neutral and 'neg' = negative

5.3. To Determine the most Mentioned Entities on Reddit and Twitter

As per Merriam-Webster dictionary, an entity is defined as something with independent existence [27]. An entity for this objective could be a company, a specific product or technology, and a uniquely identifiable government. To address this, the two datasets were split into singular words, and the frequency was counted to identify the most discussed entities on Twitter and Reddit. Microsoft, Facebook, and Apple made it to the top ten list as the companies; GitHub, Node.js, and Chrome appeared as technologies, with the United States being the lone government. Figure 4 provides a graphical illustration by ranking.

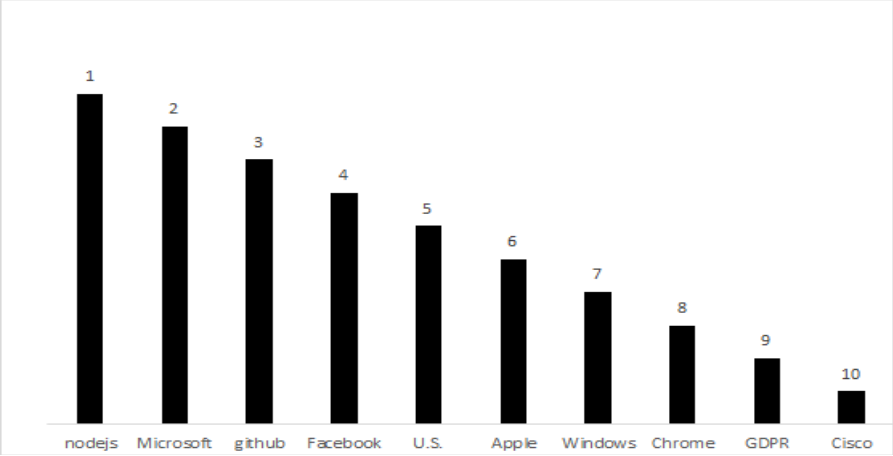


Figure 4. Most Discussed Entities on Twitter

Figure 5 lists the top ten entities from the Reddit cybersecurity posts by ranking. Reddit is most discussed, which could potentially be noise and not relevant to cybersecurity content. The big techs like Facebook, Microsoft, and Google are mentioned, along with the United Kingdom and China. Yubikey stands out as it is not a household name, but this could be due to its popularity during the data collection time.

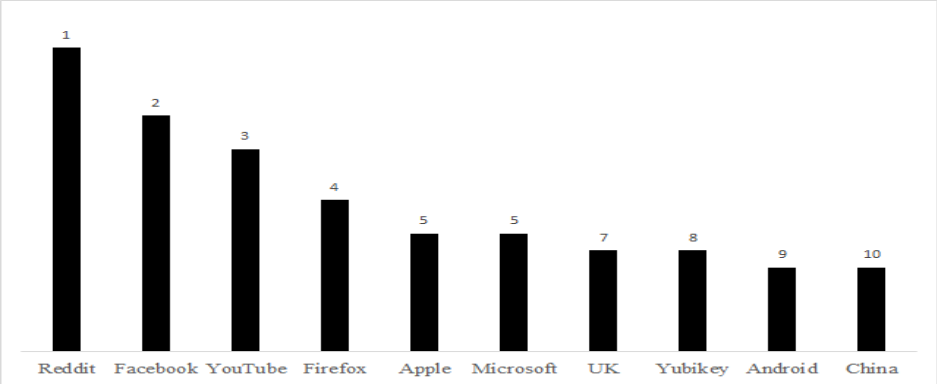


Figure 5. Most Discussed Entities on Reddit

The entities that appeared on Reddit and Twitter were, as expected, a combination of big technology firms, popular products, and governments that are at the forefront of cybersecurity development.

5.4. To Evaluate the VADER Algorithm against Human Classification

The survey conducted requested human participants to classify Twitter and Reddit contents. The majority classification ('neu', 'pos', or 'neg') of a post or a tweet to an appropriate polarity would be the baseline classification for VADER to be compared against. If the VADER assigned the same classification as human participants, the accuracy would be 100%.

Evaluations metrics are important to address the potency of the model[18], in this case, VADER. Four evaluation metrics are primarily used in classification models to get an indicator of how well the model is performing based on how it classifies, which is listed in Table 6.

Table 6. Evaluation Metrics for VADER

	Formula	Explanation
Accuracy	(TP+TN)/(TP+FN+TN+FP)	The ratio of correctly predicted and total observations
Precision	(TP)/TP+FP)	The ratio of correctly predicted positive and total positive observations
Recall	(TP)/(TP+FN)	The ratio of correctly predicted observations and to all observations in the domain
F1-Score	(2xPrecisionxRecall)/(Precision+Recall)	Weighted harmonic mean of precision and recall.

Note: TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative

The evaluation was conducted by comparing the results of VADER with human classified content. The sample size or support was ten tweets, and out of those, nine were labeled positive,

and one was labeled neutral by the human classifiers. The VADER accuracy was 0.60, or it identified 60% of the polarity correctly as indicated in Table 7. The precision, recall, and f-score for negative and neutral were nil because the sample was not present or a correct prediction wasn't made. The model had an 86% detection rate for positive observations.

Table 7. Evaluation Metrics of VADER for Twitter

	precision	recall	f1-score	support
neg	0.0	0.0	0.0	0
neu	0.0	0.0	0.0	1
pos	0.86	0.67	0.75	9
accuracy	-	-	0.60	10
macro avg.	0.29	0.22	0.25	10
weighted avg.	0.77	0.60	0.68	10

Note: 'pos' = positive, 'neu' = neutral, and 'neg' = negative

The human-classified sample or support yielded all positive polarities and VADER was able to identify 70% of the total observations correctly (Table 8). It correctly identified all positive observations, hence the precision score of 1.00 or 100% but could not label negative classification.

Table 8. Evaluation Metrics of VADER for Reddit

	precision	recall	f1-score	support
neg	0.0	0.0	0.0	0
pos	1.00	0.70	0.82	10
accuracy	-	-	0.70	10
macro avg.	-	0.35	0.41	10
weighted avg.	1.0	0.70	0.82	10

Note: 'pos' = positive, 'neu' = neutral, and 'neg' = negative

The performance of VADER was less than satisfactory, albeit due to the small sample size for the comparison hence making it inconclusive to the final designation. However, it was competent in classifying positive observations correctly, with a precision score of 86% for Twitter and 100% for Reddit.

6. CONCLUSIONS

Contrary to the inceptive opinion that presumed cynicism, the cybersecurity content in the social media domain exhibited mostly positive or neutral sentiments. The standard and extreme negative opinions were lower than expected. Popular big techs and competent cybersecurity governments were often discussed on the platform. Due to encouraging NLP advancements, this research framework is modular and can be replicated or altered for other varying content and subject area. The accuracy of the VADER is not convincing partly due to the small sample size, but an opportunity for improvement by increasing the number of participants is possible. Future explanatory research that explains the polarity of content based on variables like length of the content, platform, the time it was posted, etc., could provide further clarity and insight.

REFERENCES

[1] P.-L. Chen, Y.-C. Cheng, and K. Chen, "Analysis of Social Media Data: An Introduction to the Characteristics and Chronological Process," in Big Data in Computational Social Science and Humanities, S.-H. Chen, Ed. Cham: Springer International Publishing, 2018, pp. 297–321. doi: 10.1007/978-3-319-95465-3_16.

[2] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[3] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," J. Am. Med. Inform. Assoc., vol. 18, no. 5, pp. 544–551, Sep. 2011, doi: 10.1136/amiajnl-2011-000464.

[4] Kauffmann, Peral, Gil, Ferrández, Sellers, and Mora, "Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the Main Product Features Using Text Data Mining," Sustainability, vol. 11, no. 15, p. 4235, Aug. 2019, doi: 10.3390/su11154235.

[5] I. Agrafiotis, J. R. C. Nurse, M. Goldsmith, S. Creese, and D. Upton, "A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate," J. Cybersecurity, vol. 4, no. 1, Jan. 2018, doi: 10.1093/cybsec/tyy006.

[6] J. M. Haney and W. G. Lutters, "'It's Scary...It's Confusing...It's Dull': How Cybersecurity Advocates Overcome Negative Perceptions of Security," p. 16, 2018.

[7] V. A. and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," Int. J. Comput. Appl., vol. 139, no. 11, pp. 5–15, Apr. 2016, doi: 10.5120/ijca2016908625.

[8] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer, "Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics," Soc. Media Soc., vol. 7, no. 2, p. 205630512110190, Apr. 2021, doi: 10.1177/20563051211019004.

[9] O. Ahlgren, "Research on Sentiment Analysis: The First Decade," in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, Dec. 2016, pp. 890–899. doi: 10.1109/ICDMW.2016.0131.

[10] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," IEEE Access, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[11] L. Abualigah, H. E. Alfar, M. Shehab, and A. M. A. Hussein, "Sentiment Analysis in Healthcare: A Brief Review," in Recent Advances in NLP: The Case of Arabic Language, vol. 874, M. Abd Elaziz, M. A. A. Al-qaness, A. A. Ewees, and A. Dahou, Eds. Cham: Springer International Publishing, 2020, pp. 129–141. doi: 10.1007/978-3-030-34614-0_7.

[12] N. Altrabsheh, M. M. Gaber, and M. Cocea, "SA-E: Sentiment Analysis for Education," p. 10.

[13] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," p. 10, 2014.

[14] S. Russo et al., "The value of human data annotation for machine learning based anomaly detection in environmental systems," Water Res., vol. 206, p. 117695, Nov. 2021, doi: 10.1016/j.watres.2021.117695.

[15] F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," Knowl. Inf. Syst., vol. 33, no. 1, pp. 1–33, Oct. 2012, doi: 10.1007/s10115-011-0463-8.

[16] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," Int. J. Data

Min. Knowl. Manag. Process, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.
[17] A. Mittal and S. Patidar, "Sentiment Analysis on Twitter Data: A Survey," in Proceedings of the 2019 7th International Conference on Computer and Communications Management, Bangkok Thailand, Jul. 2019, pp. 91–95. doi: 10.1145/3348445.3348466.
[18] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," J. Travel Res., vol. 58, no. 2, pp. 175–191, Feb. 2019, doi: 10.1177/0047287517747753.
[19] C. Costa, M. Aparicio, and J. Aparicio, "Sentiment Analysis of Portuguese Political Parties Communication," in The 39th ACM International Conference on Design of Communication, Virtual Event USA, Oct. 2021, pp. 63–69. doi: 10.1145/3472714.3473624.
[20] U. Yaqub, N. Sharma, R. Pabreja, S. A. Chun, V. Atluri, and J. Vaidya, "Location-based Sentiment Analyses and Visualization of Twitter Election Data," Digit. Gov. Res. Pract., vol. 1, no. 2, pp. 1–19, Apr. 2020, doi: 10.1145/3339909.
[21] R. Gupta and M. Chen, "Sentiment Analysis for Stock Price Prediction," in 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Shenzhen, Guangdong, China, Aug. 2020, pp. 213–218. doi: 10.1109/MIPR49039.2020.00051.
[22] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis," Procedia Comput. Sci., vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.
[23] M. Shuttleworth, "Descriptive Research Design - Observing a Phenomenon." https://explorable.com/descriptive-research-design (accessed Oct. 28, 2021).
[24] "Quick Start — PRAW 7.4.0 documentation." https://praw.readthedocs.io/en/stable/getting_started/quick_start.html (accessed Nov. 01, 2021).
[25] "Getting started — tweepy 4.2.0 documentation." https://docs.tweepy.org/en/stable/getting_started.html#models (accessed Nov. 01, 2021).
[26] D. Ramachandran and R. Parvathi, "Analysis of Twitter Specific Preprocessing Technique for Tweets," Procedia Comput. Sci., vol. 165, pp. 245–251, 2019, doi: 10.1016/j.procs.2020.01.083.
[27] "Entity | Definition of Entity by Merriam-Webster." https://www.merriam-webster.com/dictionary/entity (accessed Nov. 09, 2021).

AUTHOR

Bipun Thapa is a doctoral candidate in Cybersecurity at Marymount University. His research interests are in the areas of application of Artificial Intelligence in the Information Technology space (threat classification, sentiment detection, proactive security).

Referring Expressions with Rational Speech Act Framework: A Probabilistic Approach

Hieu Le[1], Taufiq Daryanto[2], Fabian Zhafransyah[2], Derry Wijaya[1], Elizabeth Coppock[1], Sang Chin[1]

[1] Boston University , Boston, MA, USA

[2]Institut Teknologi Bandung, Bandung, Indonesia

Abstract

This paper focuses on a referring expression generation (REG) task in which the aim is to pick out an object in a complex visual scene. One common theoretical approach to this problem is to model the task as a two-agent cooperative scheme in which a 'speaker' agent would generate the expression that best describes a targeted area and a 'listener' agent would identify the target. Several recent REG systems have used deep learning approaches to represent the speaker/listener agents. The Rational Speech Act framework (RSA), a Bayesian approach to pragmatics that can predict human linguistic behavior quite accurately, has been shown to generate high quality and explainable expressions on toy datasets involving simple visual scenes. Its application to large scale problems, however, remains largely unexplored. This paper applies a combination of the probabilistic RSA framework and deep learning approaches to larger datasets involving complex visual scenes in a multi-step process with the aim of generating better-explained expressions. We carry out experiments on the RefCOCO and RefCOCO+ datasets and compare our approach with other end-to-end deep learning approaches as well as a variation of RSA to highlight our key contribution. Experimental results show that while achieving lower accuracy than SOTA deep learning methods, our approach outperforms similar RSA approach in human comprehension and has an advantage over end-to-end deep learning under limited data scenario. Lastly, we provide a detailed analysis on the expression generation process with concrete examples, thus providing a systematic view on error types and deficiencies in the generation process and identifying possible areas for future improvements.

1. Introduction

Presented with a scene involving two dogs, where one has a frisbee in its mouth, native speakers of English will effortlessly characterize the lucky dog as *the dog with the frisbee*. Computers are not so good at this yet. The task in question is called referring expression generation (REG).

A common approach to REG is modeling the problem as a two-agent system in which a speaker agent would generate an expression given some input and a listener agent would then evaluate the expression. This modeling method is widely applied, for example in [1].

In the last few years, many attempts at REG have applied deep learning to both the speaker and listener agents, utilizing the advantage of big datasets and massive computation power. As with many other NLP tasks, deep learning has been shown to achieve state-of-the-art results in REG. For example, [2] applied supervised learning and computer vision techniques to referring expression. Nevertheless, explainability remains a problem as it is difficult to fully understand how a deep learning model can generate some texts

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 109-120, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120709

given an image and a target.

On the other hand, recent developments in computational pragmatics have yielded probabilistic models that follow simple conversational rules with great explanatory power. One important example is the Rational Speech Act framework (RSA) by [3], where probabilistic speakers and listeners recursively reason about each other's mental states to communicate—speakers reason about probability distribution over utterances given a referent object, while listeners reason about probability distribution over objects in the scene given an utterance. While [3] and [4, i.a.] have shown that RSA can generate sentences that are pragmatically appropriate, the datasets are small, with simple examples that are carefully crafted with perfect information. [5] extend RSA to real world examples of reference games by using simple shallow models as building blocks to build the speaker and listener agents. [5]'s approach is intractable though, as the speaker model has to consider all possible utterances.

[6] resolves this issue by using a character level LSTM to predict one character at a time, thus reducing the search space. At each step, instead of generating one utterance, the speaker model generates one character. This method greatly reduces the search space and make the neural RSA system more efficient with harder examples. However, their method is applied to the task of generating a referring expression for an image given several other images instead of a referring expression for an object in a scene. In addition, by performing RSA on a character level [6] partially compromises the explainability of RSA as it is harder to reasoning why at each step, the model would prefer one character over another on describing the target.

To extend on the work of [5] and [3] we want to explore a different approach from [6] that would not compromise on the explainability of RSA. In this paper, we introduce a novel way to apply the RSA framework to real world images and a large scale dataset. Specifically, our contributions are as follows:

1. We tackle the intractability problem that [5] faced, we use Graph R-CNN [7] and Detectron2 [8] to extract textual information about objects and their properties i.e., types and attributes, and relations to other objects. This step vastly reduces the search space when generating utterances.
2. We use the world view generated from the previous step to constrain the utterance space; we also sequentially update the utterance prior and the prior over objects in the scene as each descriptor in the utterance is produced. To our knowledge, this is the first attempt to use iterative update of the both the utterance and object probability distributions.
3. We evaluate our framework on refCOCO and refCOCO+ [9], and evaluate generated expressions in terms of accuracy (with human evaluation)—whether the expressions are distinctive—and automatic metrics.
4. We provide a detailed analysis of the result, specifically on the types of error based on the human evaluation. This deviates from standard evaluation process where they key metric is the comprehension accuracy (i.e is the expression distinctively describe the target) and provides a new angle in analysing expression quality.

2. Background

2.1. RSA

RSA, first introduced by [3] encapsulates the idea that pragmatic reasoning is essentially Bayesian. In the reference game scenario studied by [3], the domain consists of a set of objects with various qualities that are fully available to two players. The *speaker* will

describe one targeted object unknown to the *listener* by creating a referring expression and the *listener* needs to reason about which object the expression is referring to. As laid out by [10], RSA is a simple Bayesian inference model with three components: *literal listener*, *pragmatic speaker* and *pragmatic listener*. For a given object o and utterance u:

$$\textit{literal listener } P_{L_0}(o|u) \propto [\![u]\!](o) \cdot P(o) \tag{1}$$

$$\textit{pragmatic speaker } P_{S_1}(u|o) \propto \alpha U(u, o) \tag{2}$$

$$\textit{pragmatic listener } P_{L_1}(o|u) \propto P_{S_1}(u|o) \cdot P(o) \tag{3}$$

where $[\![u]\!]$ is the literal meaning of u, either true (1) or false (0). The *literal listener* thus interprets an utterance at face value, modulo the prior probability of referring to that object $P(o)$, which we take to correspond to the object's salience. The *pragmatic speaker* decides which utterance to make by using the utility function $U(u, o)$, which is a combination of literal listener score and a cost function and the α term denotes the rationality scale of the speaker. Lastly, the *pragmatic listener* infers the targeted object by estimating the likelihood that the *speaker* would use the given utterance to describe it. [3] showed that RSA can accurately model human listener behavior for one-word utterances in controlled contexts with few objects and few relevant properties. Since then, a wealth of evidence has accumulated in support of the framework; see [10] for some examples. Still, most RSA models use a very constrained utterance space, each utterance being a single lexical item. [4] explore RSA models with two-word utterances where each utterance is associated with its own (continuous) semantics. But it remains a major open question how to scale up RSA models for large-scale natural language processing tasks.

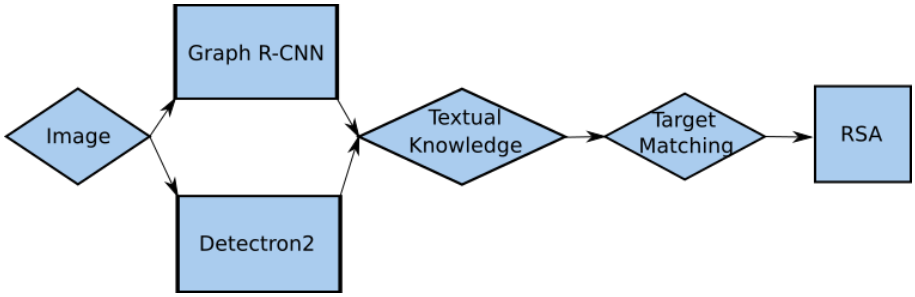


Figure 1: The workflow of Iterative RSA where the input image is passed through information extraction algorithm (Graph R-CNN/Detectron) and pre-processed before given to Iterative RSA.

2.2. Detectron2 and Graph R-CNN

The RSA framework requires prior knowledge about the images and targets in order to generate expressions. Most approaches that use RSA and the speaker/listener model acquire this knowledge through a deep learning model that learns an embedding of the image and the target object, represented as a bounded box or bounded area inscribed on the image then use these embeddings to generate expressions. Instead of using embeddings, we decided to take a different route by generating the symbolic knowledge in the form of scene graph obtained from the image using Detectron2 and Graph R-CNN, which contains objects, properties, and relations, all in a lingual format, which is the ideal input for an RSA model.

Detectron2 is the state-of-the-art object detection model developed by [8] that utilizes multiple deep learning architecture such as Faster-RCNN [11] and Mask-RCNN [12] and is applicable to multiple object detection tasks. Graph R-CNN [7] is a scene graph generation model capable of detecting objects in images as well as relations between them using a graph convolutional neural network inspired by Faster-RCNN with a relation proposal network (RPN). RPN and Graph R-CNN is among the state-of-the-art architecture in objects' relation detection and scene graph generation.

3. Method

As discussed in [10] and [3], RSA requires a specification of the utterance space and background knowledge about the state of the 'world' under consideration. Thus, we view the problem of generating referring expressions as a two-step process where, given an image and a targeted region, we:

(1) Acquire textual classifications (e.g. *car*) of the objects inside the image and the relations between objects in the image;
(2) Generate a referring expression from the knowledge acquired from step (1).

In step (1), most previous work falls into two categories. [3] and [4] assume the information about objects and their properties are known to the agent generating the expression. On the other hand, [5] and [6] use deep learning to obtain embeddings of the image and the targeted region. [13] combine the embedding extraction step with the referring expression in one single model.

In step (1), we neither assume the availability of descriptive knowledge of the images like [3] nor do we use an image and region embedding like [5]. Instead, we generate both the utterance space and the literal semantics of the input image by applying Graph R-CNN to obtain objects' relations and Detectron2 to obtain objects' properties. This idea is motivated by the intractable problem that [5] face when considering a vast number of utterances at every step. By extracting the symbolic textual information from images, we vastly reduce the number of utterances per step since the number of objects, their relations, and properties are limited in each image. Specifically, Detectron2 outputs objects and the probability that some property is applicable to those objects. For example, a given object categorized as an elephant might have a high probability of having the property *big* and a lower probability of having the property *pink*. Graph R-CNN outputs pairs of objects and probabilities of how true some predefined relation is to some pair of objects.

One challenge in merging computer vision systems with datasets like RefCOCO is matching the target referent in the dataset to the right visually detected object (assuming it is found). RefCOCO provides a bounding box around the target referent, and Detectron2 and Graph R-CNN may or may not identify an object with the same position and dimensions. One simple approach is to use the most overlapped detected object with the target box as the subject for the generation algorithm. However, there is no guarantee that the most overlapped detected object is the target. We overcome this problem by combining feature extraction with target feature extraction from Detectron2. We first let Detectron2 identify all the objects it can in the image (call this the *context*). We then instruct Detectron2 to consider the target box an object and classify it. If there is an object in the context that overlaps at least 80% with the target box *and* is assigned the same class, then we leave the context as is; otherwise we add the target box to the context.

To enrich object relations beyond binary relations in Graph R-CNN, we also implemented a simple algorithm to generate ordinal relations. We do so by sorting detected objects of the same category (e.g all dogs in an image) by the x-axis and assign predefined ordinal

	box_alias	image_id	ann_id	ref_id	salience	x1	y1	w	h	TYPE_yolk	...	ATTR_thin	ATTR_decorative
0	train-1	348639	174253	20030	0.249653	107.052635	124.379616	378.298889	181.191986	0.000624	...	0.002494	0.002494
1	building-1	348639	174253	20030	0.167918	0.351816	0.757685	214.786270	214.648575	0.000624	...	0.002494	0.002495
2	people-1	348639	174253	20030	0.015229	573.829712	207.348175	47.538757	87.952179	0.000624	...	0.002492	0.002492
3	light-1	348639	174253	20030	0.000809	476.127625	239.816071	17.229279	12.895142	0.000624	...	0.002492	0.002493
4	tracks-1	348639	174253	20030	0.090863	41.790394	321.809692	260.171417	95.888214	0.000624	...	0.002495	0.002494
5	platform-1	348639	174253	20030	0.050403	551.599915	257.860382	88.156311	156.979919	0.000624	...	0.002494	0.002494
6	windows-1	348639	174253	20030	0.014572	471.407471	181.153656	92.367615	43.315598	0.000624	...	0.002494	0.002494
7	tracks-2	348639	174253	20030	0.089103	314.089294	289.973694	206.937500	118.219421	0.000624	...	0.002495	0.002494
8	building-2	348639	174253	20030	0.282810	257.849670	0.218195	379.708557	204.494568	0.000624	...	0.002495	0.002496

Figure 2: An example of the textual knowledge acquired from Detectron2: each row corresponds to all information about a suggested bounding box, which contains box name, dimension, location, and the likelihoods of types and attributes.

relations such as *left*, *right*, or *second from left*.

The product of these image analysis methods are used in the literal semantics, which are categorical, although they are based on the gradient output of Detectron2 and Graph R-CNN, which assigns objects to properties and relations with varying degrees of certainty. Since Detectron2 and Graph-RCNN output likelihood values for attributes and types for each object as shown in Figure 2, the last step in the textual extraction process is using a cutoff threshold to decide what level of likelihood make one attribute belongs to a particular object. If the threshold is too low, then objects would contain many irrelevant attributes; if the threshold is too high, there may not be enough attributes to uniquely describe some objects. Currently, we use a hard-coded value that is slightly higher than the minimum value where most of the irrelevant attributes and types are, as examined by hand.

Thus, in the spirit of [14], we assume a threshold θ to decide whether a given type or attribute holds of a given object. Let F be a function that assigns: to each attribute and type, a function from D to [0,1]; and to each relation, a function from $D \times D$ to [0,1], where D is the set of objects in the image. F represents the output of the Detectron2 and Graph R-CNN. For each type, attribute, and relation symbol u, $\theta(u)$ is a threshold between 0 and 1 serving as the cutoff for the truthful application of the type, attribute, or relation to the object(s). Then $[\![u]\!](o) = 1$ iff $F(u)(o) \geq \theta(u)$, etc. Ultimately we plan to learn these thresholds from referring expression training datasets such as RefCOCO. Currently, they are fixed by hand: one uniform threshold for types/attributes and relations, respectively. Using categorical semantics rather than the gradient semantics that would be obtained directly from the Detectron2 avoids the well-known problems of modification in fuzzy semantics, a proper solution to which would require conditional probabilities that are unknown [15].

Our key contribution with respect to step (2) is at the speaker level. We introduce *iterative RSA*, described in the Algorithm 1 below. Iterative RSA takes as input the domain of all objects D, a prior $P(d)$ over all objects $d \in D$, the referent object o and list of possible 'utterances' U. Although an utterance may consist of multiple words, each 'utterance' here is a single predicate (e.g. *dog*, *second from left*, *wearing black polo*). We will use the word 'descriptor' instead of 'utterance' in this setting, because the strings in question may be combined into a single output that the speaker pronounces once (a single utterance, in the proper sense of the word). Again, we take the prior over objects to be proportional to salience (which we define as object size). Our RSA speaker will iteratively generate one descriptor at a time and update the listener's prior over objects at every step until

either (i) the entropy of the probability distribution over objects reaches some desirable threshold K, signifying that the listener has enough information to differentiate o among objects in D, or (ii) the maximum utterance length T has been reached.

input : o, D, U, P_D^0
initialization: $E = []$;
while $t < T$ *&* $Entropy(P_D^{t-1}) < K$ **do**
  u = sample(Speaker $P_{S_1}(u|o, P_D^{t-1}, U_E)$);
  P_D^t = Literal listener $P_{L_0}(o|u, P_D^{t-1})$;
  add u to E;
end
output: E

Algorithm 1: Iterative RSA

In standard RSA, the utility function $U(u, o)$ is defined as $U = \log(P_{L_0}(o|u)) + \text{cost}(u)$ [10]. We define ours as:

$$U_E = \log(P_{L_0}(o|u) + P_{ngram}(u|E)) + \text{cost}(u) \tag{4}$$

where P_{ngram} is the probability of u following the previous n words in E. Specifically, we use a 3-gram LSTM model (n=3). Figure 1 outlines our overall workflow.

4. Experiment and Result

The framework is implemented in Python and will be made publicly available. In the implementation of Algorithm 1, we set $T = 4$. This value for maximum utterances per expressions come from the average length of the expressions from our target dataset, both RefCOCO and RefCOCO+ have average length less than 4 utterances per expression. We evaluate our framework on the test set of RefCOCO and RefCOCO+ datasets released by [9]. For these two datasets, each data point consists of one image, one bounding box for a referent (the *target box*) and some referring expressions for the referent. We used pre-trained weights from the COCO dataset for Graph R-CNN and Detectron2. Additionally, we experiment separately with finetuning Detectron on RefCOCO referring expressions. Finally, we test the framework with RefCOCO *Google* split test set and RefCOCO+ *UNC* split test set.

We evaluate the generated expressions on the test dataset with both automatic overlap-based metrics (BLEU, ROUGE and METEOR) and accuracy (human evaluation) (Table 2). Specifically, we run human evaluation through crowdsourcing site Prolific on the following scheme: our IterativeRSA, RecurrentRSA [6] and SLR [16] trained on 0.1%, 1% and 10% of the training sets of RefCOCO and RefCOCO+. For each scheme, we collected survey results for 1000 randomly selected instance from the RefCOCO test dataset from 20 participants and 3000 instances from RefCOCO+ test dataset from 60 participants. Each image is preprocessed by adding 6 bounding boxes on some objects in the image, one of which is the true target. The boxes are chosen from 5 random objects detected by Detectron2 an the true target object. Each participant is asked to find the matching object given expression for 50 images through multiple choice questions. In addition, we also manually insert 5 extra instances where the answer is fairly obvious and use those instances as a sanity check. Data from participants who failed more than half of the sanity checks (i.e 3/5) was not included in the analysis. Since our referring expressions are generated based on extracted textual information about individual objects and not the raw image as a whole, there are cases where Detectron2 does not recognize the object in the

target box or the suggested bounding box from Detectron2 is different in size compared to the target box. In such cases, our algorithm ended up generating an expression for a different observable object than the targeted one. To understand the different types of errors our model makes, we also included additional options in cases where the testers cannot identify a box that matched the expression. Specifically, we added three categories of error when no (unique) matching object is identified:

1. nothing in the picture matches the description
2. several things match this description equally well
3. the thing that matches the description best is not highlighted

Despite the simplicity of our proposed method, it achieves comparable performance in terms of METEOR score to the Speaker-Listener-Reinforcer(SLR) [16]. More importantly, our method outperforms SLR in human comprehension under low training data scheme and RecurrentRSA with both RefCOCO and RefCOCO+.

	True	False	Under-informative	no-match	not-highlighted	adjusted-accuracy
IterativeRSA	27.25	13.03	11.59	44.49	3.64	52.54
Iterative RSA + f-Det2	26.52	15.1	13.79	38.51	6.07	47.86
SLR-10% [16]	26.95	15.71	15.98	36.51	4.85	45.96
SLR-1% [16]	14.3	16.24	11.96	51.13	6.37	33.65
SLR-0.1% [16]	6.1	18.14	10.48	59.73	5.55	17.57

Table 1: Human evaluation response on RefCOCO+ images across IterativeRSA and SLR trained under restricted data.

Beside raw accuracy, we also report the accuracy rate using the formula $adjusted - accuracy = True/(True + False + Underinformative)$ where $Underinformative$ counts instances where the expressions correctly refer to the referent objects but are not distinctive enough. Our human evaluation accuracy is slightly less than that of MMI [9] and while our METEOR score is higher. However, our performance measures fall short when compared to the state-of-the-art extensively trained end-to-end deep neural network model by SLR [17]. This is to be expected as our method was not trained and does not require training on the specific task of referring expression generation or comprehension. Further performance analysis will be given in the next sections.

5. Comparison with Recurrent RSA and SLR trained with limited data

As discussed above, to see the advantages and drawbacks of Iterative RSA, we run human evaluation on generated expressions from RefCOCO and RefCOCO+ datasets and compare Iterative RSA with RecurrentRSA-another RSA approach as well as SLR. From Table 3, Iterative RSA outperforms RecurrentRSA with 28% compared to 26.9%. On the other hand, to make a fair comparison with a deep learning end-to-end approach like SLR, we decided to train SLR with limited training data as Iterative RSA does not require

	bleu	rouge	meteor
MMI [9]	0.37	0.333	0.136
SLR [17]	0.38	0.386	0.16
rerank [13]	0.366	0.354	0.15
Iterative RSA	0.18	0.125	0.11

Table 2: NLP metric comparisons to some previous approaches on RefCOCO+ dataset.

	RefCOCO	RefCOCO+
Iterative RSA	28.05	**27.25**
Iterative RSA + f-Det2	41.3	26.52
Recurrent RSA [6]	26.9	-
SLR-10% [16]	**66.2**	26.95
SLR-1% [16]	49.85	14.3
SLR-0.1% [16]	38.5	6.1

Table 3: Raw accuracy of referring expression comprehension evaluated human evaluation on Iterative RSA, Iterative RSA with finetuned Detectron2 (f-Det2), Recurrent RSA and SLR trained with limited data of 0.1%, 1% and 10% of RefCOCO and refCOCO+ training set.

any direct training process. From Table 3, the Iterative RSA (no training) outperforms all SLR models trained with 0.1%, 1% and 10% training data for refCOCO+ dataset and outperform SLR model trained with highly limited training data (0.1%) on RefCOCO. Furthermore, when examining the SLR-generated expressions, we observed that for the model trained and tested on RefCOCO dataset, a lot of the expressions contains positional property of objects such as *left, right*, which makes identifying the target easier when the expression is low quality and incomplete (as a result of training on limited data). Thus, we can see that SLR performs better on RefCOCO than RefCOCO+. On the other hand, IterativeRSA performs more consistently, especially when used without any training or observation of the data. Finetuning the Detectron2 model for object detection with RefCOCO expressions improve the performance on the corresponding dataset, however, using the same model on the RefCOCO+ dataset does not show any significant change in accuracy. Figure 3 is an example of referring expression generated with RSA compared to SLR

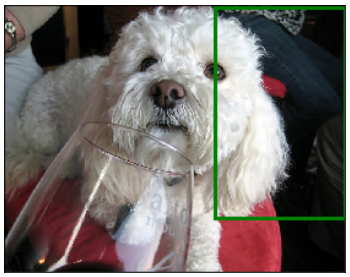


Figure 3: IterativeRSA: *jeans*, SLR-1%: *man in black*, SLR-10%: *woman in red*

trained with limited data. For the RSA expression, it clearly shows that the model explains Gricean maxim of quantity by generating the shortest possible word to describe the target which are the jeans, whereas SLR shows the overfitting behavior when generating unrelated expression to the target.

6. Analysis of the human evaluation

As mentioned above, in our study, aside from letting users choose one of the objects surrounded by bounding boxes given the generated expression, we also give additional options to handle the case where survey participants cannot find a sensible object to match the description. Overall, we observe that incorrect responses can be divided into the following categories: *under-informative expression*, *not highlighted*, *no match* and *false*. These categories of error help in identifying the sources of deficiency in our approach. If the expression is under-informative, there are two possibilities. The first is that the textual data extraction step (i.e., Detectron2) was able to identify multiple objects of the same

type, but the algorithm is unable to differentiate between the target and the rest of the objects. In this case the problem is on the linguistic side of our model. Another possibility is that not all objects of the relevant type were detected, which is the deficiency of our visual system (Detectron2).

Another type of visual system deficiency happens when the described object is not the highlighted one or if there is no match. In these cases, the visual system (Detectron2) mis-classified the object in the bounding box. As shown in Table 3, about 48% of the recorded instances belong to these two categories.

6.1. Under-informative expressions

One type of error is when the generated expression is under-informative. This occurs when the expression correctly indicated the type of the target object but failed to differentiate between the target and other objects of the same type in the picture. For example, in Figure 4, the algorithm was able to correctly identify the type of object in the bounding box but the modifier (*cooking*) failed to differentiate the target from the other instance of that type.

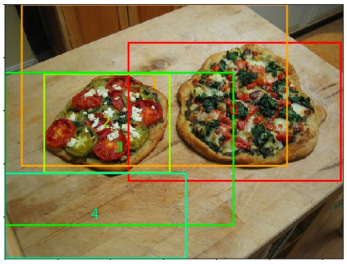


Figure 4: Generated Expression: *cooking pizza*, Gold Label: *pizza on left*, Target box: *2*. i.e., the light green box surrounding the smaller pizza.

6.2. Object not highlighted

Another type of errors revealed through human evaluation is when the matching object is not highlighted as the target. This type of deficiency is due to the textual extraction

Figure 5: Example of *not highlighted* response Generated Expression: *laying down man*, Gold Label: *guy bottom left*, Target box: *2* i.e., the light green box at the bottom left of the image

component (Detectron2) not observing all objects of the same type. In Figure 5, Detectron2 can only observe four instances of the category *man*, which are all highlighted in this image with box $1, 2, 3, 5$. When comparing the available attributes for these *man*s, target *man* in box 2 (i.e., the light green box at the bottom left of the image) is assigned a distinctive attribute that others do not have: *laying down* (although he is sitting, not

laying down). The use of this modifier increases the salience of the target relative to the other individuals that are detected. It is quite possible that participants assumed *laying down man* refers to the only person at the bottom center of the image who is actually laying down. However, that individual is not detected by Detectron2 and thus there is no highlighted box.

6.3. High quality expression

When the participants correctly identify the target object by choosing the right bounding box, we observe that the textual extraction step provides sufficient information for the algorithm to work correctly. Figure 6 is an example where we observe that the system works well when the extracted textual information is accurate and sufficient. Specifically, Detectron2 found all the objects of the type *train* in box 4 and 5. Furthermore, the *train* objects have fairly sensible attributes, including *the left* and *the right*.

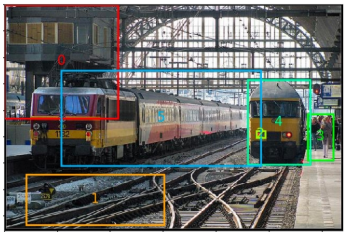


Figure 6: Generated Expression: *the right train*, Gold Label: *right train*, Target box: *4*.

7. Discussion

The Iterative RSA introduced in this paper is able to generate multiple-modifier descriptions, which goes far beyond the vanilla RSA speaker described by [10] and [3], and our RSA speaker has even gone past the two-word stage of [4]. While the result is not at the level of the state-of-the-art end-to-end model, Iterative RSA outperforms Recurrent RSA and SLR trained under limited data. We can clearly explain how our model comes up with the referring expressions it generates. The explainability of our model is a contrast feature when compare with RecurrentRSA. While RecurrentRSA also applies the RSA model to generate expressions, its expression generation by recursively generate characters makes it hard to explain why at each step, why one character is a feasible choice that helps identify a target object. Furthermore, to our knowledge, we are the first attempt to apply pure probabilistic RSA model without any neural network components in the expression generation step of the referring expression generation from image task. From the analysis of the human evaluation and concrete examples, it is clear that the performance of Iterative RSA is tightly coupled with the performance of the textual extraction model, particularly Detectron2. When Detectron2 detects enough information, including the objects in a given image as well as their probable attributes, we observe that our proposed Iterative RSA can create high quality expressions with distinctive modifiers.

Another key strength and also a weakness of our proposed iterative RSA is the size of the vocabulary of descriptors. Currently, this vocabulary is limited to the attributes and types vocabulary that Detectron2 possesses. While this vastly reduces the search space of all possible descriptors, it also limits the possible descriptors that RSA can choose from, given a target. The textual extraction step (Detectron2 in this case) can be analogized to the act of "observing" and the Iterative RSA algorithm to "reasoning". One cannot reason about objects or aspects of objects that are not observed.

On the other hand, in terms of efficiency, our proposed method is fast because Iterative RSA does not require training data and can be applied directly on the fly with any given

textual extraction system. In addition, our application of Detectron2 and Graph-RCNN also does not require training as it utilizes pre-trained weights. Experiments with fine-tuning Detectron2 with RefCOCO data does show better accuracy on the test set of RefCOCO dataset but does not show any major improvement when tested on RefCOCO+ as shown in Table 1. Thus, the base Iterative RSA is more generalized and consistent across different datasets.

Minimal reliance on training data has other advantages: That property makes our approach a promising one for low-resource languages where labeled data for training, especially for vision-language tasks such as referring expression generation/comprehension, are virtually non-existent [18] for languages other than English.

8. Conclusion

In this paper, we have explored the possibility of decomposing referring expression generation into a two-component process of symbolic knowledge acquisition and expression generation, adapting the RSA framework to real world scenes where textual information is not available. We also introduce two promising innovations that help to address the intractability problem of applying RSA to real world scenes in previous work, which includes (1) constraining the utterance space using the output of object recognition and scene graph generation systems, and (2) proposing a simple yet intuitive and explainable model for referring expression generation called iterative RSA, which incrementally outputs referring expression one predicate at a time. Lastly, our method allows for easy analysis and understanding of each individual expression, and provides clear explanations as to why the system generates the expressions it does.

9. References

[1] W. Monroe and C. Potts, "Learning in the Rational Speech Acts model," *CoRR*, vol. abs/1510.06807, 2015.

[2] S. Kazemzadeh, V. Ordonez, M. Matten, and T. Berg, "ReferItGame: Referring to objects in photographs of natural scenes," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Doha, Qatar), pp. 787–798, ACL, Oct. 2014.

[3] M. C. Frank and N. D. Goodman, "Predicting pragmatic reasoning in language games," *Science*, vol. 336, no. 6084, p. 998, 2012.

[4] J. Degen, R. X. D. Hawkins, C. Graf, E. Kreiss, and N. D. Goodman, "When redundancy is rational: A Bayesian approach to 'overinformative' referring expressions," *CoRR*, vol. abs/1903.08237, 2019.

[5] J. Andreas and D. Klein, "Reasoning about pragmatics with neural listeners and speakers," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (Austin, Texas), pp. 1173–1182, ACL, Nov. 2016.

[6] R. Cohn-Gordon, N. D. Goodman, and C. Potts, "Pragmatically informative image captioning with character-level reference," *CoRR*, vol. abs/1804.05417, 2018.

[7] J. Yang, J. Lu, S. Lee, D. Batra, and D. Parikh, "Graph R-CNN for scene graph generation," *CoRR*, vol. abs/1808.00191, 2018.

[8] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2." `https://github.com/facebookresearch/detectron2`, 2019.

[9] L. Yu, P. Poirson, S. Yang, A. C. Berg, and T. L. Berg, "Modeling context in referring

expressions," *CoRR*, vol. abs/1608.00272, 2016.

[10] G. Scontras, M. H. Tessler, and M. Franke, "Probabilistic language understanding: An introduction to the rational speech act framework," 2018.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[12] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow." https://github.com/matterport/Mask_RCNN, 2017.

[13] R. Luo and G. Shakhnarovich, "Comprehension-guided referring expressions," *CoRR*, vol. abs/1701.03439, 2017.

[14] D. Lassiter and N. Goodman, "Adjectival vagueness in a Bayesian model of interpretation," *Synthese*, vol. 10, pp. 3801–3836, 2017.

[15] D. Edgington, "The philosophical problem of vagueness," *Legal Theory*, vol. 7, pp. 371–378, 2001.

[16] L. Yu, H. Tan, M. Bansal, and T. L. Berg, "A joint speaker-listener-reinforcer model for referring expressions," 2017.

[17] J. Kim, H. Ko, and J. Wu, "CoNAN: A complementary neighboring-based attention network for referring expression generation," in *Proceedings of the 28th International Conference on Computational Linguistics*, (Barcelona, Spain (Online)), pp. 1952–1962, International Committee on Computational Linguistics, Dec. 2020.

[18] P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury, "The state and fate of linguistic diversity and inclusion in the NLP world," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Online), pp. 6282–6293, Association for Computational Linguistics, July 2020.

INSTANTANEOUS FREQUENCY AND AOA ESTIMATION OF MULTICOMPONENT SIGNALS BASED ON BORN-JORDAN DISTRIBUTION

Gan Quan, Tang Jie, Song Huan Huan, Wen Hong

Institute of Aeronautics and Astronautics,
University of Electronic Science and Technology of China, Chengdu, China

ABSTRACT

Based on the spatial sample of the multi-component signals by the array antenna, using Born-Jordan distribution which is a kind of typical Cohen type time-frequency distribution, the detection and estimation approach of the spatial multi-component signals' angle-of-arrival and instantaneous frequency is proposed by comparison of the instantaneous frequency of the signals at the same time from the point of digital signal processing. The simulation results show that the proposed method not only has high noise performance, but also enhances the real-time performance of the spatial signal detection.

KEYWORDS

Angle-of-arrival Estimation, Time-frequency Distribution, BJD, Instantaneous Frequency.

1. INTRODUCTION

Based on the array signal processing, the time-frequency analysis has been introduced to estimate the angle of arrival and instantaneous frequency parameters of the spatial non-stationary signals in the recent years. It has become one of the hotspots of the signal processing research [1-4]. S. Ouelha et al [5,6] proposed a method to estimate the angle of arrival (AOA) of the non-stationary signals using quadratic spatial time-frequency distribution in the array signal processing. The paper [7] constructed the array signal model in the time-frequency domain through the time-frequency distribution with the symmetrical array, and estimated the AOA and instantaneous frequency parameters of the spatial signal by the eigenspace decomposition of the data correlation matrix of the adjacent time-frequency points. Wang Shu et al [8, 9] proposed a method to construct correlation matrix by combining the spatial sampling data with the temporal sampling data. This kind of research combines the concept of the spatial spectrum estimation with the method of the time-frequency analysis, and effectively solves the problem of the AOA and instantaneous frequency estimation of the spatial non-stationary signals. However, because the algorithm is based on the analysis of the signal spectrum, the eigenvalue decomposition of matrix and the search of one-dimensional or two-dimensional spectrum peak are needed, which leads to large amount of calculation and low real-time performance in the parameter estimation. For the multi-component signals, the estimation method based on the Wigner Ville distribution used in the reference [5] has some other problems such as the cross term interference etc.

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 121-128, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120710

2. SPATIAL SAMPLING MODEL OF THE MULTIPLE SIGNALS

It is assumed that there is a symmetrical uniform linear array with elements $2L+1$, and the spacing between elements is d. We set the reference element as the center element and number the elements from left to right as $-L,-(L-1),\cdots,L$. Consider M signal sources are incident to the linear array at the same time from the far field, as shown in Figure 1. Then the output of the Nth element at the t_0 time can be expressed as:

$$x_n(t_0)=\sum_{i=1}^{M} s_i(t_0+n\tau_i)+v_n(t_0) \qquad n=-L,-(L-1),\cdots,L\ ;\ i=1,2,\cdots M \quad (1)$$

$$\tau_i=\frac{d}{c}\sin(\theta_i) \quad (2)$$

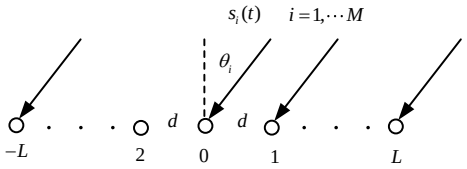


Figure 1. symmetrical linear array antenna

Where, $\{s_i(t)\}_{i=1}^{M}$ are the outputs of M unrelated signal sources in the reference element, and $\{\theta_i\}_{i=1}^{M}$ are the AOAs of the incident signals. τ_i is the delay of signal $s_i(t)$ between the adjacent elements, $\{v_n(t)\}_{n=-L}^{L}$ are the independent white Gaussian noise outputs of the $2L+1$ elements, which are independent of the input signals.

Let the output $\{x_n(t_0)\}_{n=-L}^{L}$ of $2L+1$ elements be the signal sequence obtained by spatial sampling of the linear array at time t_0. It can be seen from formula (1) that the spatial sampling signal $\{x_n(t_0)\}_{n=-L}^{L}$ at time t_0 can be regarded as the sum of M $2L+1$ point sequences and the noise signal, which are sampled by M signal sources $s_i(t)$ with $1/\tau_i$ as the sampling frequency in the short time window of $t\in[t_0-L\tau_i,t_0+L\tau_i]$. The sampling frequency of signal $s_i(t)$ is

$$f_s^{(i)}(\theta)=1/\tau_i=\frac{c}{d\sin(\theta_i)} \quad (3)$$

In order to eliminate the direction ambiguity, the array spacing should meet the half wavelength condition of $d\le\lambda_{\min}/2$, so the array output sequence naturally meets the sampling condition of sampling frequency $f_s^{(i)}\ge 2f_{\max}$. At the same time, it can be seen that the sampling frequency $f_s^{(i)}(\theta)$ of the signal $s_i(t)$ is directly related to the incident angle θ_i of the signal, which provides a basis for us to estimate the arrival angle θ_i of the signal source $s_i(t)$ using the method of time-frequency analysis.

3. INSTANTANEOUS FREQUENCY EXTRACTION OF THE SPATIAL SIGNAL

The instantaneous frequency (IF) is an important parameter to characterize the non-stationary signal, and the time-frequency distribution of the signal is an important means to obtain the

instantaneous frequency. For multi-component signals, the cross terms of the time-frequency distribution must be suppressed in the process of the instantaneous frequency acquisition. Born Jordan distribution [10] (BJD), as a typical Cohen class time-frequency distribution, has strong cross term suppression ability and has been widely studied and applied [11]. BJD is defined as:

$$\mathrm{BJD}_s(t,f)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}s(u+\frac{\tau}{2})s^*(u-\frac{\tau}{2})\psi(t-u,\tau)e^{-j2\pi f\tau}dud\tau \tag{4}$$

Where $\psi(t,\tau)$ is the kernel function of BJD

$$\psi(t,\tau)=\begin{cases}1/|\tau| & |\tau|\geq 2|t|\\ 0 & |\tau|<2|t|\end{cases} \tag{5}$$

Sampling the signal $s(t)$ composed of M components and taking the time window with length of $2L+1$, the frequency distribution $\mathrm{C}(t_0,k)$ of the signal $s(t)$ at time t_0 can be obtained by using BJD:

$$\begin{aligned}\mathrm{C}(t_0,k)&=\mathrm{BJD}(t_0,k)\\&=2\sum_{m=-L}^{L}\sum_{n=-(L-|m|)}^{(L-|m|)}s(t_0+nT_s+mT_s)s^*(t_0+nT_s-mT_s)\psi_0(-n,2m)e^{-4\pi km/N}\end{aligned} \tag{6}$$

Where, T_s is the sampling period, N is the number of sampling points of the digital frequency in the frequency domain, and $\psi_0(n,m)$ is the discrete kernel function of BJD,

$$\psi_0(n,m)=\begin{cases}1/|m| & |m|\geq 2|n|\\ 0 & |m|<2|n|\end{cases} \tag{7}$$

By using the multi-peak detection method [10] for the frequency distribution $\mathrm{C}(t_0,k)$ of signal $s(t)$ at time t_0, the instantaneous frequency $k_1,k_2\cdots,k_M$ of the M signal components at time t_0 can be obtained.

According to the above analysis, the Nth element output of the linear array at time t_0 : $x_n(t_0)=s(t_0+nT_s)$, it is substituted into equation (6) to obtain:

$$\mathrm{C}(t_0,k)=2\sum_{m=-L}^{L}\sum_{n=-(L-|m|)}^{L-|m|}x_{n+m}(t_0)x_{n-m}{}^*(t_0)\psi_0(-n,2m)e^{-4\pi km/N} \tag{8}$$

In the formula (8), if set

$$S(m)=2\sum_{n=-(L-|m|)}^{L-|m|}x_{n+m}(t_0)x_{n-m}{}^*(t_0)\psi_0(-n,2m)$$

we can get:

$$\mathrm{C}(t_0,k/2)=\sum_{m=-L}^{L}S(m)e^{-2\pi km/N} \tag{9}$$

The formula (8) and (9) show that $C(t_0, k/2)$ can be obtained by the FFT transformation of N points by the vector $\{S(m)\}_{m=-L}^{L}$, and $C(t_0, k)$ is the frequency distribution curve of the multi-component signal $s(t)$ at time t_0. Using the multi peak estimation algorithm [12] for $C(t_0, k/2)$, we can get the twice digital instantaneous frequency k_i of the M component signals at time t_0, $i = 1, \cdots, M$. Since the sampling signal of $2L+1$ point is expanded to N point for FFT transformation, the corresponding unit quantity of k_i in frequency domain is $f_s^{(i)}(\theta_i)/N$, and the instantaneous frequency corresponding to k_i shall be $k_i f_s^{(i)}(\theta_i)/N$. The instantaneous frequency extraction process based on BJD can be composed of the steps shown in the Figure 2.

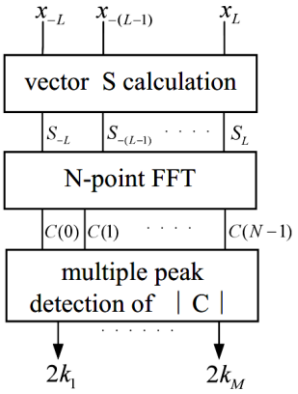


Figure 2. IF extraction based on BJD

The digital instantaneous frequencies $\{k_i\}_{i=1}^{M}$ of the M signal components obtained by the spatial sampling contain the arrival angle information of each component signal. So we can calculate and estimate the arrival angle θ_i of each signal source by comparing the instantaneous frequency at the same time.

4. AOA DETECTION BASED ON THE INSTANTANEOUS FREQUENCY COMPARISON

In order to obtain the arrival angle θ_i of each spatial signal source, the actual instantaneous frequency $f_i(t)$ of each component signal $s_i(t)$ must be obtained. We can analyze the output signal of any array element in time domain. Set the output signal of the reference array element as $s(t)$. Using the signal processing process shown in Figure 3, the instantaneous frequency $f_i(cT_s)$ of each component signal at time cT_s can be obtained through the short-time signal sampled in the time domain.

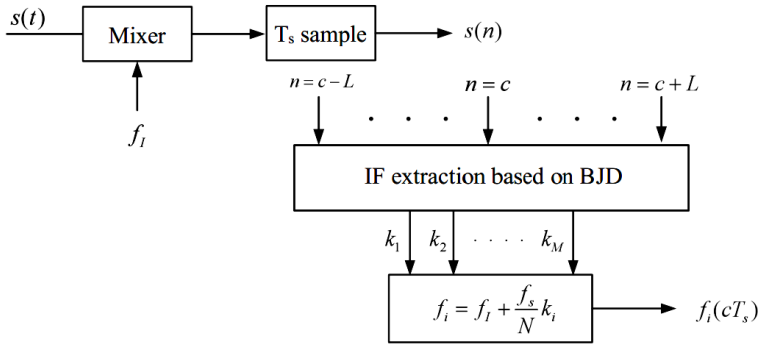


Figure 3. IF detection of $s(t)$

In Figure 3, T_s is the sampling period of the signal after mixing, and f_s is the sampling frequency. The number of the sampling signals selected is $2L+1$, which is equal to the number of the array elements. When the sampling time of the array signal is $t_0 = cT_s$, the digital instantaneous frequency k_i of each component signal at t_0 time can be obtained by using the short-time signal sampled in spatial domain. For the same signal source $s_i(t)$ at the same time cT_s , the instantaneous frequency obtained by spatial sampling signal and temporal sampling signal should be equal. Therefore, the following formula is established:

$$k_i \frac{f_s^{(i)}(\theta_i)}{N} = f_i(cT_s) \tag{10}$$

Therefore, from the formula (3), the arrival angle θ_i of the signal can be calculated according to $f_i(t_0)$,

$$\theta_i = \arcsin\left[(\frac{k_i c}{Nd}) / f_i(cT_S) \right] \tag{11}$$

The experimental results show that the error $\Delta\theta_i$ of the DOA estimation is mainly affected by the length of the array elements and the quantization error Δf_s of the digital frequency in the frequency domain. In addition, the incident angle θ_i of the signal also affects $\Delta\theta_i$ to some extent. For the spatial signal of the incident angle $0^\circ < \theta_i \le 90^\circ$, a large number of sampling points N in the frequency domain is used in the time-frequency transformation process of the instantaneous frequency extraction, and the number of linear array elements is increased appropriately, which can achieve satisfactory results.

5. EXPERIMENT AND ANALYSIS

Exp.1 It is assumed that there is a LFM signal s_1 and a single frequency signal s_2 whose incident angle changes at a constant speed with time in the airspace. They are incident at angles θ_1 and θ_2 on a symmetrical uniform linear array with $L+1$ elements. Where $s_1(t) = \exp(j2\pi(f_1 t + 0.5at^2))$, $s_2(t) = \exp(j2\pi f_0 t)$, $f_1 = 520\text{MHz}$, $a = 100\text{MHz/s}$, $f_0 = 650\text{MHz}$, $\theta_1 = 30^\circ$, $\theta_2(t) = \theta_0 - bt$, $\theta_0 = 40^\circ$,

$b = 5^{\circ}/\mathrm{s}$, the element interval $d = 0.1\mathrm{m}$, Intermediate frequency $f_I = 500\mathrm{MHz}$, sampling frequency $f_s = 600\mathrm{MHz}$, Sampling points in the frequency domain $N = 10000$. In the noise environment with SNR of 0dB, we use the proposed method to take different linear array element length L, and carry out the detection simulation experiment on the arrival angle and instantaneous frequency of s_1 and s_2. The results are shown in Figure 4 and Figure 5.

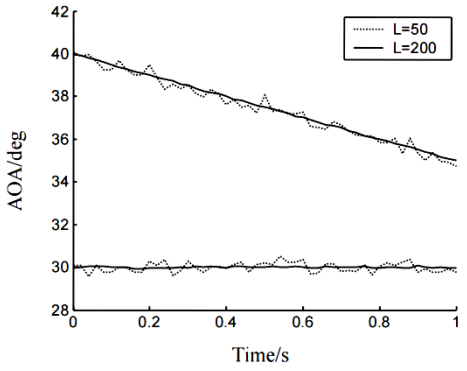


Figure 4. AOA estimation of S_1 and S_2

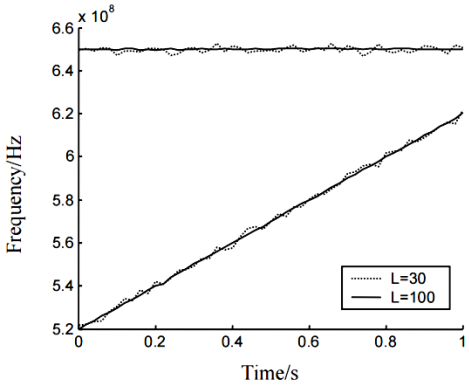


Figure 5. IF estimation of S_1 and S_2

It can be seen from Figure 4 and Figure 5 that in the noise environment with low SNR , the detection and comparison of instantaneous frequency of the spatial sampling signal and the time-domain sampling signal can analyze and judge the change of the instantaneous frequency and the arrival angle of each signal component in the spatial multi-source signal with the time. The estimation accuracy of the instantaneous frequency and the angle of arrival can be improved by increasing the array element number of the linear array, because with the increase of the sampling signal window length, the time-frequency aggregation performance of BJD distribution is also enhanced.

Exp.2 In order to study the influence of noise on the detection results, it is assumed that there are single frequency signal s_1 and LFM pulse signal s_2 in the airspace. s_1 and s_2 are simultaneously incident on the symmetrical uniform linear array with $L+1$ elements at angles $\theta_1 = 30^{\circ}$ and $\theta_2 = 40^{\circ}$ respectively. We have simulated the instantaneous frequency and angle of arrival of signal s_1 and s_2 at time $t = 0$. Where $s_1(t) = \exp(j2\pi f_1 t)$, $s_2(t) = \exp(j2\pi(f_2 t + 0.5bt^2))$, $f_1 = 600\mathrm{MHz}$, $f_2 = 550\mathrm{MHz}$, $b = 100\mathrm{MHz}/\mu\mathrm{s}$, $d = 0.1\mathrm{m}$, $f_I = 500\mathrm{MHz}$, $f_s = 600\mathrm{MHz}$, $N = 10000$. Take the number of array elements as $L = 50$ and $L = 100$, and carry out 100-200 Monte Carlo experiments in each noise environment. Figure 6 and Figure 7 show the curve of RMS error of the arrival angle and the instantaneous frequency estimation varying with SNR.

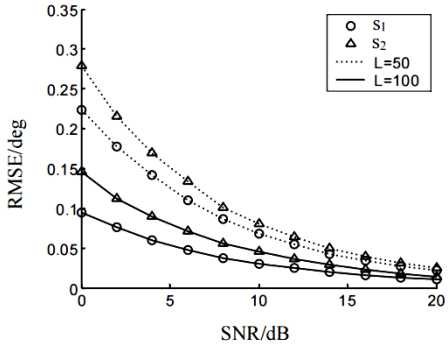


Figure 6. RMSE of AOA estimation

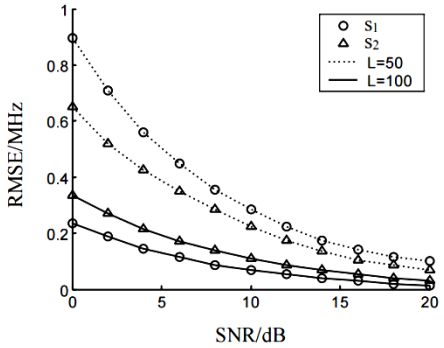


Figure 7. RMSE of IF estimation

It can be seen that the RMSE of the estimation method for the angle of arrival and the instantaneous frequency of the signal basically shows an exponential decay trend with the increase of the SNR. The increase of the number of array elements L can significantly reduce the error of the angle and frequency estimation when the SNR is low. It can also be seen from the figure that the estimation errors of the single frequency signal s_1 and the LFM signal s_2 do not differ greatly in the same case. This shows that the estimation method of the angle of arrival and the instantaneous frequency based on the BJD is not obviously affected by the time-frequency characteristics of the signal.

6. CONCLUSION

Based on the spatial sampling model of the linear array antenna for the incident signal, this paper proposes a method to detect the arrival angle and the instantaneous frequency parameters of the spatial multi-source signal by the digital signal process using the Born Jordan distribution which has good time-frequency aggregation characteristics and cross term suppression ability. The method realizes the estimation of the angle of arrival by means of the instantaneous frequency comparison, and reduces the computation of time-frequency spatial spectrum estimation. This method maintains the noise suppression ability of the second-order time-frequency distribution, and is not affected by the time-frequency characteristics of the incident signal. The simulation results show that the method has not only a good real-time performance, but also maintains a strong noise suppression ability.

REFERENCES

[1] Khan N.A., Ali, S., (2021)"Multi-component instantaneous frequency estimation in mono-sensor and multi-sensor recordings with application to source localization", Multidimensional Systems and Signal Processing, Vol. 32, No. 3, pp. 959-973.

[2] Akram, J., Khan, N.A., Ali, S., Akram, A., (2020)" Multi-component instantaneous frequency estimation using signal decomposition and time-frequency filtering", Signal Image Video Process., Vol. 14, No. 8, pp. 1663-1670.

[3] Abdoush, Yazan; Garcia-Molina, Jose A.; Corazza, Giovanni E. (2019)"Adaptive instantaneous frequency estimation based on time-frequency distributions with derivative approximation", SIGNAL PROCESSING, Vol. 160, pp. 99-105.

[4] B. Boashash and S. Ouelha, (2017) "An improved design of high-resolution quadratic time–frequency distributions for the analysis of nonstationary multicomponent signals using directional compact kernels", IEEE Trans. Signal Process., Vol. 65, No. 10, pp. 2701-2713.

[5] S. Ouelha, A. Aïssa-El-Bey and B. Boashash, (2017)"Improving DOA estimation algorithms using high-resolution quadratic time-frequency distributions", IEEE Trans. Signal Process., Vol. 65, No. 19, pp. 5179-5190.

[6] A. Belouchrani , M. G. Amin. (1999)"Time-frequency Music". IEEE Signal Processing Lett., Vol. 6, pp:109-110.
[7] A. Belouchrani, M. G. Amin, N. Thirion-Moreau and Y. D. Zhang, (2013)"Source separation and localization using time-frequency distributions: An overview", IEEE Signal Process. Mag., Vol. 30, No. 6, pp. 97-107.
[8] WANG S, ZHOU X. (1999) "Extending MUSIC algorithm by using virtual array technique". IEEE Southeast CON Proceedings. pp. 82-85.
[9] WANG S,ZHOU X. (1999) "Extending ESPRIT algorithm by using virtual array and Moore-Penrose general inverse techniques ". IEEE Southeast CON Proceedings. pp. 315-318.
[10] Cohen L. (1995)"Time-Frequency Analysis", Englewood Cliffs, NJ: Prentice-Hall.
[11] Ivanovic, V.N.; Dakovic, M.; Stankovic, L. (2013)"Performance of quadratic time-frequency distributions as instantaneous frequency estimators". IEEE Trans. Signal Processing, Vol.51, No. 1, pp. 77 – 89.
[12] Stankovic, L.J.; Djurovic, I. et al. (2003)"Instantaneous frequency estimation by using Wigner distribution and Viterbi algorithm". Proceedings 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol.6, pp.121-124.

AUTHORS

The author has left the postdoctoral workstation of University of Electronic Science and Technology of China in signal processing since 2010. He is now an associate professor at the school of Aeronautics and Astronautics of UESTC. His main research fields are signal detection, array signal processing, radar signal processing, etc.

IMPROVING THE DIGITAL SECURITY OF SMART ENERGY SYSTEMS WITH SMART CONTRACTS

Pekka Koskela, Jarno Salonen and Juha Pärssinen

VTT Technical Research Centre of Finland,
P.O. Box 1000, FI-02044 VTT, Finland

ABSTRACT

Smart grids are evolving towards intelligent electricity grid where the operation of systems is distributed and automatised. Technical solutions to achieve these future needs are proposed using blockchain with smart contracts in many studies, where smart contracts enhance automation. Fundamentally smart contracts will increase security because of their distributed nature and since it inherits the security of blockchain. However, smart contracts are software components, which have special features like the unstoppable nature of applications and may use special languages like Solidity. Our aim in this paper is to get a holistic review in the smart contract life cycle, what potential new vulnerabilities and threats will they introduce and how can they be prevented, and what smart contract specific issues programmers should focus on. We also propose a future direction to achieve more secure smart contracts in smart energy systems.

KEYWORDS

Blockchain, Smart contracts, Smart energy systems.

1. INTRODUCTION

As the energy industry develops towards smart and micro grids, where a large amount of energy producers and consumers interact with each other using heterogeneous devices. These new activities require new and economically viable solutions for trustworthy communication between the devices and trading of energy and storage capacity. A solution of the trading has been proposed to utilize blockchains and smart contracts [1],[2],[3],[4].

Smart contracts can help solve multidimensional problems related to achieving and securing the integrity and reliability of distributed, complex energy events and information exchange, as well as systems optimization. Blockchain-based smart contracts help eliminate the need for third parties to build mutual trust between different parties and enabling automation facilitating the deployment and commercialization of decentralized energy transactions and exchanges, both in terms of energy flows and financial transactions [5],[6].

Recently there have been many broad reviews concerning the use of blockchain in future smart grids analysing the applicability of blockchain technology [7] and identifying possible applications of blockchain in smart grid [8]. There have also been numerous studies covering the analysis of the security threats and vulnerabilities of smart contracts [9],[10],[11], some of which have been concerned with machine learning detection [12]. The vulnerabilities are strongly

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 129-138, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120711

dependent on the construction and implementation of specific systems and thus depends on the case. For instance the public and private blockchains differ from each other in terms of permission to participate, access to the network and consensus methods. In the public blockchain everybody is permitted while in the private blockhain permission is granted only to specific user groups. Public blockchains such as Ethereum use proof of work (PoW) that is a probabilistic consensus algorithm, in the Ethereum virtual machine environment while private blockchains such as Hyper Ledger Fabric use Raft, which is a deterministic consensus algorithm, in the docker container environment. Many existing papers study Ethereum-based approaches, whereas in this paper we try to form a more holistic overview based on the smart contract life cycle. Our focus is more on what secure improvements and new threats the smart contracts will bring and estimate the risk level of threats and how to prevent them.

This article is organized as follows. In section 2 we describe smart contracts, their features and benefits. Section 3 provides an overview of past studies promoting smart contracts in smart energy systems. Section 4 presents what new threats are posed by smart contracts during their life cycle, including their potential damage and risk and generic means to prevent the aforementioned threats. In section 5 we discuss the future directions on how to achieve more secure smart contracts and finally, we conclude the article in section 6.

2. SMART CONTRACT

A smart contract is a program component attached to a blockchain. As such, the program component is not smart and may not even be a contract, but only code that performs some pre-programmed tasks. The smart contract related software code is connected to the blockchain, which guarantees the integrity, i.e., that the program code has not been altered after the connection. This provides code that the various parties can rely on when the code is executed. The basic features of the smart contract are the following:

- It is stored in a blockchain
- It can be checked by everyone and the operating logic and results are visible to everyone
- It can no longer be changed after it is in the blockchain
- Its operation “cannot” be blocked and the results “cannot” be modified
- It performs the functions assigned to it, programmed automatically and always in the same way, i.e. certain input values always produce the same output result.

In the cybersecurity environment, the smart contracts are connected to the blockchain and the blockchain is connected to cloud services as depicted in the figure below (fig1).

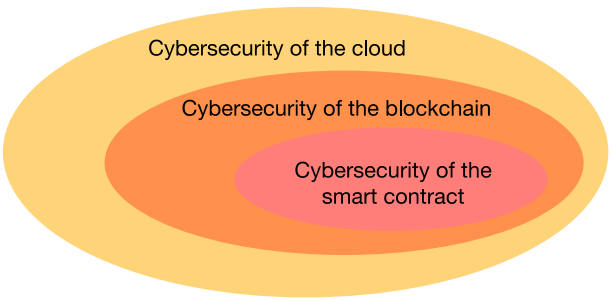


Figure 1. The cybersecurity environment of smart contract.

Because the smart contract is connected to a blockchain, its security is substantially affected by the security of the blockchain and how the blockchain implementation has been done. The following questions are important when determining the security and risk level of the smart contract and the blockchain ”platform”: What are the transaction consensus mechanisms? How decentralized is the blockchain, in other words, what is the number of “block approvers”? Is it a private or public blockchain? One of the basic principles of the blockchain and also the smart contract is that they operate in a decentralized manner, thus avoiding a single point vulnerability, which would exist in a centralized system. Therefore, an essential part of blockchains and thus also smart contracts for information security is how effectively the decentralization has been implemented. The benefits of using a smart contract linked to a blockchain include the following: Increased security based on decentralization and the inherited security of the blockchain, ability of building mutual trust between unknown parties, automation of contract-based functions like sales, outlets, approval chains, material, product and quality monitoring, and eliminating the need to use trusted third parties, see fig 2.

The security of blockchain is based on cryptography where every new block contains a cryptographic hash code generated from the block itself and the previous block. Thus if some malicious third party wanted to create a fake block, they would need to change both the hash code of the previous and the current block, which in practice means that all hash codes of the blockchain would need to be changed. This is difficult to do at a limited time before a new block is accepted and when the blockchain is long enough as it will be in a normal case.

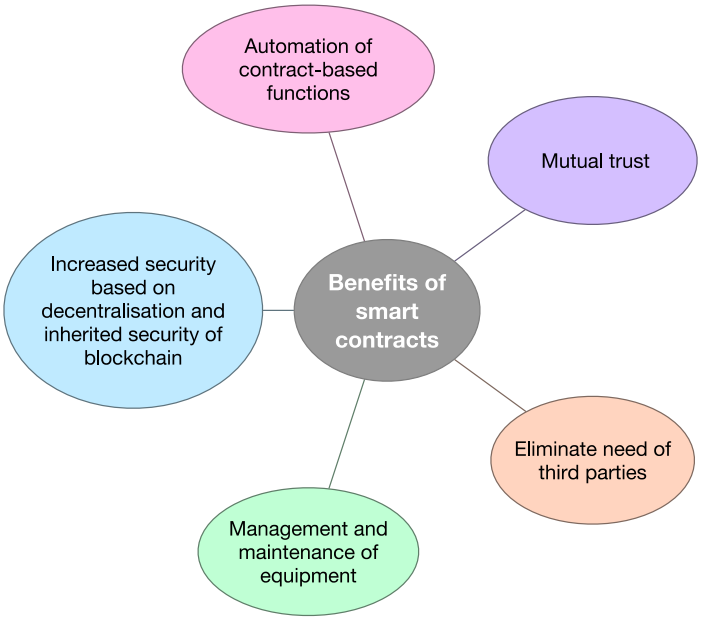


Figure 2. The benefits of smart contracts.

Mutual trust between participants is commonly based on the knowledge that information hasn’t been modified after it has been saved in the blockchain (integrity) and secure consensus mechanisms guaranteeing that the saved information is mutually accepted and any cheat attempt is detected and excluded.

Basically the smart contract is small program, like any other conventional program and can be programmed to consist of any functions like sales, web shops, approval chains, material, product and quality monitoring, which automatise the process. Because the smart contract is connected to

a blockchain, it cannot be modified after implementation and it will always return exactly the same output to a certain input. There will be no additional need of monitoring the results or acceptance by a third party.

We can use the blockchain to form a digital identity for persons, devices and software. When all actors of a system have a digital identity it will be easier to manage resources and devices as well as make maintenance operations in the security point of view. With smart contracts maintenance activities like software updates and resource provisioning can be done automatically.

3. USE OF SMART CONTRACTS

The exploitation of smart contracts and smart grid technologies in the energy sector has been extensively studied and several test environments have been made in connection with the subject [13],[14].

In the energy networks of the future, smart grids and micro-networks, one challenge is to create an economically viable and efficient trading place between the parties connected to the grid. Challenges include among others:

- How to build trust between different actors?
- How are privacy and security issues handled?
- How to make the system work as automatically as possible?
- How to make the system work as efficiently and economically as possible?
- How to optimize energy production, using and storing?

As a solution to the challenges, an online store based on blockchains and smart contracts has been proposed, where the sale and purchase of energy can be done automatically between the parties connected to the grid [1],[2],[3],[4]. We could also create a similar system which includes only the electricity producers and consumers and where sales are automated through smart contracts without a separate sales agency.

In the energy networks of the future, one challenge will be how to charge electric cars economically. A solution to the problem has been proposed to use smart contracts in order to reduce power variation and charging costs [15],[16]. In addition to selling and buying energy, the marketplace could sell storage capacity such as electric car batteries and household energy storage capacity [4],[17].

Similarly, automated mechanisms based on smart contracts can be created for the management and maintenance of equipment, which alert on equipment failures and allocate the cost of maintenance measures according to a service contract taking into account, for example, warranty periods, the cause of the fault and the time taken for maintenance [17].

Solutions based on smart contract security can increase the security and resilience and prevent cyberattacks on distributed network devices, network peripherals, and related infrastructure [6].

When secure digital identity and remote attestation support is formed based on blockchain technology [18] many operations and functions of a system can be automatised and it also allows the development of new services. These services can among others be warranty follow-up, covering following up the operation time, operation conditions and interruptions of a specific device and providing automatised maintenance services and compensations according to the smart contract. Furthermore when devices can be identified we can include automatic provisioning to the system, software updates, calibration, certificate verification and data source

identification. Based on blockchain technology the online store can be established for sale and purchase of energy, energy storage and devices, that uses smart contracts for automatised functionalities. Fig 3 presents the different possible functions which may be automatised with smart contracts.

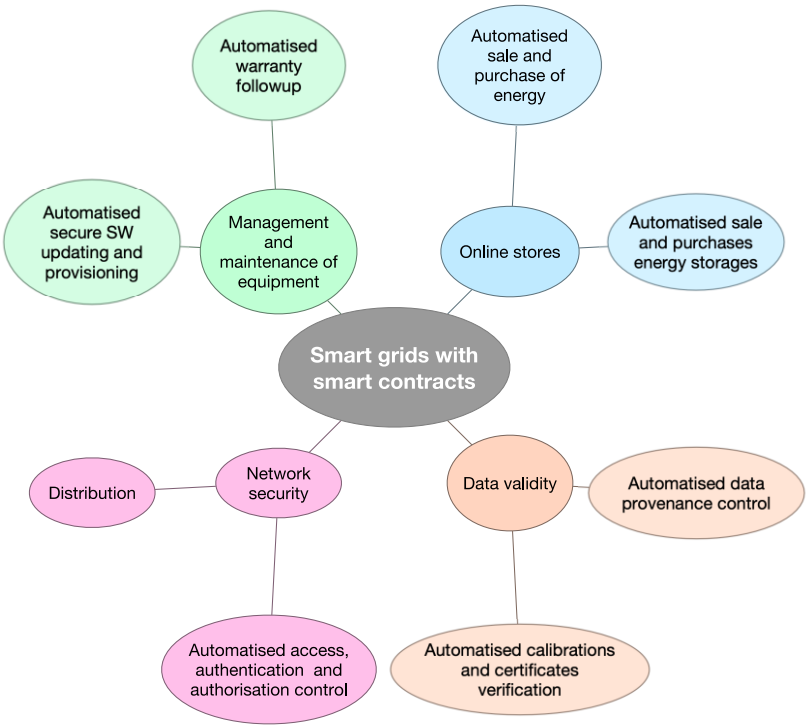


Figure 3. The exploitation of smart contract in the smart energy sector.

4. THREATS TO SMART CONTRACTS

Smart contracts are connected to blockchains that act as a cloud service on virtual machines. As a result, smart contracts face the same threats as cloud services and blockchains. In addition, the smart contracts themselves introduce their own threats to the system. This study does not address the threats to cloud services and blockchains, but only what additional threats are posed by smart contracts.

In fig 4 we present the life cycle of smart contracts which consist of creation, deployment, execution and termination. During the creation phase participants negotiate and make mutual agreement on the content of the smart contract. When the content is clear the smart contract can be designed, programmed, tested and verified. Because the programming work may be demanding due to strange programming languages and syntax, accurate testing is critical in terms of security and operation of smart contract. After verification the contract will attached and stored in the blockchain. During the execution of a smart contract there will be injected input values to the contract, which produce output values. If the contract is unbroken the same input values must provide always same output values. Finally there is a need to design and implement how to terminate the smart contract in a controlled manner, because the code will never disappear in a blockchain.

Figure 4. The life cycle of a smart contract, the creation-deployment-execution-termination.

During the life cycle of a smart contract, the creation-deployment-execution- termination process is subject to various threats. The realization of the smart contract threats is primarily influenced by the chosen blockchain platform such as the decision between private or public blockchain and its implementation, i.e., how well the code is tested. In terms of the extent of the damage, it largely affects what the smart contract does, e.g., how many users does the contract have, and how long the attack proceeds before the system can be repaired. Preventing smart contract related threats may include smart contract, blockchain, and data integrity. We will discuss more about the accuracy of different threats, their damages and risk and how to prevent them in the next chapters.

4.1. Creation

Threat: In the creation phase a programming error in a smart contract can be either intentional, accidental or due to a security bug in the programming environment [19],[20].

Damage and risk: The error can cause the smart contract to malfunction, enable a security breach, or cause operational disturbance of the blockchain. The error risk increases, when the source code is not available or it lacks a well-known language like Solidity which have familiar-like syntax, but in the semantics level there can be significant differences. Obviously the risk level of programming errors will depend on the competences of programmers and testing processes. In general we assumed the risk of errors to be between low and middle, where low risk represents using well known languages and middle risk not so well known ones. Some examples of attacks have been discussed in [21].

Prevention actions: Smart contracts should be carefully tested before commissioning and the first implementations should be done with a smaller test group. The idea is to use techniques and tips which are available [7],[8] and use programming languages/methods which support good human readable source code when they are available [9],[12]. The correctness of the contract can be tested with many tools [11],[22],[23] and methods [26]. Furthermore there are many tools for analysing the vulnerability of smart contracts [16],[22].

4.2. Deployment

Threats: During deployment the smart contract cannot be approved, which is due to attack, e.g., distributed denial-of-service (DDoS), against the blockchain. As a result deployment of the smart contract is not possible. Another possible threat is as a result of the attack against the blockchain, when the attacker modifies an existing smart contract or installs his own smart contract. As a result of the attack, a false or erroneous smart contract will be used. An example of these attacks is decentralised autonomous organisation (DAO)-attack, where an erroneous smart contract is used to steal crypto currency ETH [22]. A possible method to change the smart contract in the blockchain that is based on work of proof or same kind of consensus is, if attackers, i.e. poll, have enough computing power to solve the nonce fairly quickly, create two or more consecutive blocks and then propose and accept their own malicious smart contract into the blockchain.

Damage and risk: The first attack will prevent the start of activities related to the smart contract, like trading whereas in the second attack the smart contract doesn't work properly and will do incorrect things. The risk of attacks like DDoS depends among others on how distributed the implementation or how protected the network is. In general the public blockchain can be assumed as more distributed and therefore more protected than the private one. However, private blockchains have a better control over access and network operations and in that sense the private chain can be considered to be safer against DDoS attacks than the public one.

Generally we assume the risk of DAO attack to be similar to the risk of a programming error and it will depend on a case basis on what kind of a blockchain platform is used and what are the specific vulnerabilities of that platform. However, we generally assume private blockchains to be safer against DAO kind of attacks, because the operations in blockchain can be more restricted and authenticated, which is not the case in public blockchains. This is because public blockchains are managed by the community whereas private blockchains management is limited and more controlled by a pre-selected group and therefore the recovery from an attack will be faster and easier when compared to the public chain.

Because private blockchains can be more controlled and better monitored we assume that it may react to both the DDoS and DAO type of attacks faster than public blockchains, which means that the potential damage will be smaller.

Prevention actions: One should use sufficiently distributed and fast blockchain platforms to prevent DDoS type attacks . Private chains can use fast consensus algorithms and mechanisms and are therefore often faster than public ones. In order to prevent DAO type attacks, monitoring the operation and integrity of smart contracts should be implemented and the blockchain itself should include control programs [23]. Private blockchains have better access and more strict operational control, which is missing from public blockchains.

4.3. Execution

Threats: During the execution, slowing down the operation of a blockchain by an attack may result in a slowdown or total blocking of smart contracts. Another attack affecting the result of the smart contract may be initiated either by changing the initial input or output values which results to an incorrect result.

Damage: The first attack will prevent or slowdown activities related to the smart contract for instance in trading. The actual damages will depend on the tasks of the contract and how efficient the attack is. We assume the risk of this attack being low, but the risk will increase by poor design and/or implementation of the blockchain. In case of an attack manipulating either the input or output values of the smart contract, results into the smart contract not working properly and therefore producing wrong results. In this case the smart contract will be missing sufficient integrity of the input or output values. In general we argue that the case where integrity needs to be improved afterwards, the operation will be much easier in the private than the public chain, because the public blockchain forms a rather loose community which does not have any easy means to make any required updates at once.

Prevention actions: The efficiency of the DDoS attack will mainly depend on the distribution level of blockchain implementation and in general the public blockchain can be assumed more distributed and therefore more protected than the private chain. However, the efficiency of DDoS attacks will affect also the privacy of the operational pace of blockchain. In this case the private blockchain consensus algorithms and therefore the operations can be made faster than in the public chain. We argue that private blockchains which often are faster, have better control with

regards to the access and operation control than public chains. Therefore it can be assumed that private blockchains will be safer against DDoS attacks than public blockchains.

To prevent the integrity and data modification attacks, there is a need to monitor the contract activity, to have input and output data encryption and integrity control programs. Flow control and interaction of other smart contracts can be monitored among others using graph-based analysis and path-searching [24]. We argue that since private blockchains may form a more controlled environment, they can be considered to be safer and the responses to attacks are faster than in public chains and therefore the potential damage might have smaller influence.

4.4. Termination

Threat: Due to the attack against the blockchain or the programming error, the smart contract cannot be terminated at the desired time.

Damage: Influencing the expiry date of an intellectual property contract results into the contract activities being terminated too early or too late. We assume that in general there is a low risk of the termination threat, which depends on the termination process and its implementation.

Prevention actions: Influencing the expiry date of an intellectual property contract is planned in such a way that the termination process of the contract takes into account any possible attacks. We argue that in general detecting the attack and fixing any issues is faster in private chains than public ones, because the private chains have a more controlled environment.

5. FUTURE DIRECTIONS TO MORE SECURE SMART CONTRACTS

Based on previous chapter most of the attack sources are not platform issues but result from smart contract programming errors. One possible means to prevent any errors is to develop smart contracts using simple programming languages. which are often considered non-touring complete [26]. Another means to prevent errors is using integrated development environments, where the operations that are not vital or even harmful in the security point of view -like recursion - are removed or prevented. In case of the smart grid domain, we should conduct an extensive study on what kind platform(s) [26] and language(s) [26], [27] will be used and what tasks the smart contracts are intended to do. If smart contracts need operations like interconnection with other smart contracts or the use of random number generators, time stamps, which must be update identically to all distributed apps, we must study how this can be done in a secure way so that we can keep the apps consistent during an attack. Similarly when smart contract is connected with AI deep learning, we need to take care that all smart contracts have the same data input all the time to maintain the consistency of the contracts.

6. CONCLUSION

As energy networks will become increasingly complex towards smart and micro grids, new solutions will be needed for their management in an efficient, economical and secure manner. One proposed solution has been to utilize blockchains and smart contracts. The exploitation of both technologies has been extensively studied in the energy sector and several test environments have been made in connection with this subject. The possibilities of blockchains and smart contracts are wide in the energy networks of the future, and several solution options have been presented for their utilization. Since the security depends on the implementation, we might obtain a more accurate picture of the security of the different solutions by making small scale test platforms of the most interesting use cases both in the laboratory and the real environment. In

parallel to this testing work we should also study what tasks smart contracts need to do and develop programming tools, languages and environments, which enable secure programming methods, that minimise the possibility of harmful operations. When taking into account the critical nature of energy services, where good cybersecurity is an uttermost important property, our recommendation is to start blockchain implementations with with private blockchains, because they are more manageable and secured environment than public chains.

REFERENCES

[1] C. Zhang, J. Wu, C. Long, and M. Cheng, "Review of existing peer-to-peer energy trading projects," *Energy Procedia*, vol. 105, pp. 2563–2568, 2017.

[2] Z. Guan, X. Lu, W. Yang, L. Wu, N. Wang, and Z. Zhang, "Achieving efficient and privacy-preserving energy trading based on blockchain and abe in smart grid," *Journal of Parallel and Distributed Computing*, vol. 147, pp. 34–45, 2021.

[3] I. Perekalskiy, S. Kokin, and D. Kupcov, "Setup of a local p2p electric energy market based on a smart contract blockchain technology," *in 2020 21st International Scientific Conference on Electric Power Engineering (EPE)*. IEEE, 2020, pp. 1–4.

[4] S. Kushch and F. P. Castrillo, "A review of the applications of the blockchain technology in smart devices and dis-tributed renewable energy grids," ADCAIJ: *Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 6, no. 3, p. 75, 2017.

[5] M. Andoni, V. Robu, D. Flynn, S. Abram, D. Geach, D. Jenkins, P. McCallum, and A. Peacock, "Blockchain technology in the energy sector: A systematic review of challenges and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 143–174, 2019.

[6] M. Mylrea and S. N. G. Gourisetti, "Blockchain for smart grid resilience: Exchanging distributed energy at speed, scale and security," in *2017 Resilience Week (RWS). IEEE*, 2017, pp. 18–23.

[7] C. Yapa, C. de Alwis, M. Liyanage, and J. Ekanayake, "Survey on blockchain for future smart grids: Technical aspects, applications, integration challenges and future research," *Energy Reports*, vol. 7, pp. 6530–6564, 2021.

[8] H. Arezoo, M. H. Seyed, S.-k. Miadreza, and A. Hasan, "Blockchain technology in the future smart grids: A comprehensive review and frameworks[j]," *International Journal of Electrical Power and Energy Systems*, vol. 129, 2021.

[9] N. Ashizawa, N. Yanai, J. P. Cruz, and S. Okamura, "Eth2vec: Learning contract-wide code representations for vulnerability detection on Ethereum smart contracts," in *Proceedings of the 3rd ACM International Symposium on Blockchain and Secure Critical Infrastructure*, 2021, pp. 47–59.

[10] O. Lutz, H. Chen, H. Fereidooni, C. Sendner, A. Dmitrienko, A. R. Sadeghi, and F. Koushanfar, "Escort: Ethereum smart contracts vulnerability detection using deep neural network and transfer learning," *arXiv preprint arXiv:2103.12607*, 2021.

[11] C. Liu, X. Zhang, K. K. Chai, J. Loo, and Y. Chen, "A survey on blockchain enabled smart grids: Advances, applications and challenges," *IET Smart Cities*, vol. 3, no. 2, pp. 56–78, 2021.

[12] F. Mi, Z. Wang, C. Zhao, J. Guo, F. Ahmed, and L. Khan, "Vscl: Automating vulnerability detection in smart contracts with deep learning," in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2021, pp. 1–9.

[13] M. B. Mollah, J. Zhao, D. Niyato, K.-Y. Lam, X. Zhang, A. M. Ghias, L. H. Koh, and L. Yang, "Blockchain for future smart grid: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 18–43, 2020.

[14] A. Bhardwaj, S. Shah, A. Shankar, M. Alazab, M. Kumar, and T. Gadekallu, "Penetration testing framework for smart contract blockchain," *Peer-to-Peer Networking and Applications*, vol. 14, p. 2635–2650, Sep. 2021.

[15] A. Bahga and V. K. Madisetti, "Blockchain platform for industrial internet of things," *Journal of Software Engineering and Applications*, vol. 9, no. 10, pp. 533–546, 2016.

[16] C. Liu, K. K. Chai, X. Zhang, E. T. Lau, and Y. Chen, "Adaptive blockchain-based electric vehicle participation scheme in smart grid platform," *IEEE Access*, vol. 6, pp. 25 657–25 665, 2018.

[17] T. Alladi, V. Chamola, J. J. Rodrigues, and S. A. Kozlov, "Blockchain insmart grids: A review on different use cases," *Sensors*, vol. 19, no. 22, p.4862, 2019.

[18] U. Javaid, M. N. Aman, and B. Sikdar, "Defining trust in iot environments via distributed remote attestation using blockchain," in *Proceedings of the Twenty-First International Symposium on*

Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing, 2020, pp. 321–326.
[19] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 254–269.
[20] "list of known bugs," accessed: 2021-9-14. [Online]. Available: https://docs.soliditylang.org/en/v0.8.7/bugs.html.
[21] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on Ethereum smart contracts (sok)," in *Principles of Security and Trust*. POST 2017. Lecture Notes in Computer Science, M. Maffei and M. Ryan, Eds., vol. 10204. Springer, Berlin, Heidelberg, 2017, pp. 164–186.
[22] A. L´opez Vivar, A. T. Castedo, A. L. Sandoval Orozco, and L. J. Garc´ıa Villalba, "An analysis of smart contracts security threats alongside existing solutions," *Entropy*, vol. 22, no. 2, p. 203, 2020.
[23] Y. Huang, Y. Bian, R. Li, J. L. Zhao, and P. Shi, "Smart contract security: A software lifecycle perspective," *IEEE Access*, vol. 7, pp. 150 184–150 202, 2019.
[24] Z. Zheng, S. Xie, H.-N. Dai, W. Chen, X. Chen, J. Weng, and M. Imran, "An overview on smart contracts: Challenges, advances and platforms," *Future Generation Computer Systems*, vol. 105, pp. 475–491, 2020.
[25] M. Jansen, F. Hdhili, R. Gouiaa, and Z. Qasem, "Do smart contract languages need to be turing complete?" in *International Congress on Blockchain and Applications. Springer*, 2019, pp. 19–26.
[26] ´A. J. Varela-Vaca and A. M. R. Quintero, "Smart contract languages: A multivocal mapping study," *ACM Computing Surveys (CSUR)*, vol. 54, no. 1, pp. 1–38, 2021.
[27] D. Harz and W. Knottenbelt, "Towards safer smart contracts: A survey of languages and verification methods (2018)," 2019.

AUTHORS

Pekka Koskela received the D.Sc. degree in 2018. Over 20 years he has been worked in several research projects both researcher and project leader. Currently he is studying among others the exploitation of quantum, homomorphic and digital ledger technologies.

Jarno Salonen is working as a Senior Scientist in the applied cybersecurity team at VTT He has a professional background of over 20 years in making the digital world a better place for ordinary users especially in the areas of cybersecurity, privacy, resilience and development of electronic services.

Juha Pärssinen obtained his M.Sc in 1998 and Lic.Sc. in 2003 from Helsinki University of Technology, Department of Computer Science and Engineering. He is working as a Senior Scientist in the applied cybersecurity team at VTT. During the years he has participated in multiple European joint research projects and ETSI standardization efforts. His research interest includes system architecture and workflow analysis, industrial control systems security and digital forensics.

SECURITY CONCERNS FOR BLOCKCHAIN BASED SHARING OF MOBILE STUDENT CREDENTIAL

Timothy Arndt

Department of Information Systems,
Cleveland State University, Cleveland, OH, USA

ABSTRACT

Blockchain has recently taken off as a disruptive technology, from its initial use in cryptocurrencies to wider applications in areas such as property registration and insurance due to its characteristic as a distributed ledger which can remove the need for a trusted third party to facilitate transaction. This spread of the technology to new application areas has been driven by the development of smart contracts – blockchain-based protocols which can automatically enforce a contract by executing code based on the logic expressed in the contract. One exciting area for blockchain is higher education. Students in higher education are ever more mobile, and in an ever more agile world, the friction and delays caused by multiple levels of administration in higher education can cause many anxieties and hardships for students as well as potential employers who need to examine and evaluate student credentials. Distance learning as a primary platform for higher education promises to open up higher education to a wider range of learners than ever before. Blockchain-based storage of academic credentials is being widely studied due to the advantages it can bring. As with any network-based system, blockchain comes with a number of security and privacy concerns. Blockchain needs to meet several security-related requirements in order to be widely accepted: decentralization; confidentiality; integrity; transparency; and immutability. Researchers have been busy devising schemes to ensure that such requirements can be met in blockchain-based systems. Several types of blockchain-specific attacks have been identified: 51% attacks; malicious contracts; spam attacks; mining pools; targeted DDoS attacks; and others. Real-world attacks on blockchain-based systems have been seen on cryptocurrency sites. In this paper, we will at look at the specific privacy and security concerns for blockchain-based systems used for academic credentials as well as suggested solutions. We also examine the issues for academic credentials which are stored "off-chain" in such systems (as is often the case).

KEYWORDS

Blockchain, Mobile Education, Higher Education, Privacy, Security.

1. INTRODUCTION

The blockchain idea was introduced in a paper published by Satoshi Nakamoto [1] and deployed in the bitcoin cryptocurrency the following year. Blockchain is an open, distributed ledger that can efficiently record transactions between two parties in a verifiable and immutable (permanent) fashion without the need for a trusted third party (disintermediation).

Bitcoin employs a peer-to-peer architecture and relies on proof-of-work, a piece of data which is time consuming and computationally complex to produce, but which is easy for others to verify (via "miners" who are rewarded with bitcoin for this computational work) and which satisfies

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 139-145, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120712

certain requirements as a consensus mechanism. Consensus mechanisms allow for the correctness or "truth" of a transaction to be confirmed (depending on a set of rules) when multiple distributed actors perform transactions, and some of those actors may be untrustworthy. Subsequent developments have allowed blockchain to be programmed (via smart contracts) to trigger transactions automatically [2]. Transactions can cause code implementing rules which are part of the blockchain to be run, through these blockchain can lead to what have been called Distributed Autonomous Organisations (DAOs).

As a foundational technology, blockchain has been used or proposed in many application areas besides cryptocurrencies [3] including the banking sector [4], land registration (especially in developing countries) [5], the insurance sector [6], and electronic voting [7]. Alternative consensus mechanisms (such as proof of stake and mechanisms based on Byzantine Fault Tolerance) have been developed as well as alternative architectures (e.g. client-server).

In this paper, we will at look at the specific privacy and security concerns for blockchain-based systems used for academic credentials as well as suggested solutions. For further study on blockchain in higher education, [8], [9] (related works section of this work), and [10] will be helpful. We also examine the issues for academic credentials which are stored "off-chain" in such systems (as is often the case).

2. BLOCKCHAIN IN HIGHER EDUCATION

In this section we will take a brief look at a few representative projects in the application of blockchain technology in higher education.

A number of researchers have explored the use of blockchain to store university grades, i.e. university transcripts. A group at the University of Glasgow has developed a functional prototype for storage of student grades at the institution [11]. The platform chosen was Ethereum, hence it was built on a public blockchain. Based upon an exploratory, qualitative evaluation the authors found several tensions between the concepts of a university as an organization and of distributed autonomous organizations (DAOs) in Ethereum. Another project with a prototype implementation is [12], where a private BigChainDB blockchain is used for storage of student transcripts (not grades within a course as in the previously described research, though). Initial results were reported to be promising. Mahamatov et al.[13] describe a prototype implementation of a university transcript system using an Ethereum private blockchain and ERC-20 tokens (a standard for tokens, which are needed to carry out smart contracts, on Ethereum) on that blockchain. Students are able to read their grades, while professors and administrative personnel can record grades.

EduCTX [14] is an ambitious project for the development of a higher education credit platform based on the European Credit Transfer and Accumulation System (ECTS), a framework which has been approved by the EU. The decentralized higher education credit and grading system can offer a globally unified viewpoint for students, higher education institutions, and other potential stakeholders such as prospective employers. A prototype implementation has been built on the ARK blockchain platform [15]. ARK is a public blockchain, but the authors transformed it into a private (permissioned) one by taking advantage of the flexible nature of ARK to change the parameters of the DPoS (delegate proof of stake) consensus algorithm used. Logic to ensure the validity of transactions on the blockchain has been defined. ECTX tokens represent credits that students gain for completing courses (analogous to the way that ERC-20 tokens are used in [13]). The authors will use the prototype system firstly at their home institution, the University of Maribor, and then at a select set of institutions of higher education. They anticipate that this or a

similar system could potentially evolve into a unified, simplified, globally ubiquitous higher education credit and grading system.

A more theoretical investigation is given by [16] in which the architecture for the Disciplina platform for student records is described and an analysis of the main issues arising from storing student records in a blockchain is given. Their platform incorporates both private blockchains (maintained by individual institutions of higher learning) and public blockchains, managed by "Witnesses" who witness the fact that a private block was produced by a valid institution. A good, theoretical discussion of the problems of privacy, provability, and data disclosure in this context is given.

Besides traditional transcripts, blockchain is also being used or proposed for various alternative types of educational credentials. Blockchain enables permanent authentication and storage for a myriad of alternative credentials made up of diverse microcredentials, nanodegrees, MOOCs, and certificates/badges from various types of training programmes. These credentials can then be directly controlled and managed by users [17].

Among the most well-known of higher education blockchain projects was that at MIT's Media Lab which created blockcerts, a mobile app for educational credentialing built on Bitcoin [18]. At the Open University, researchers have developed the OpenLearn system built on the Ethereum public blockchain which awards OpenLearn badges for completing sections of a course and passing assessments [19]. The creators of that system have also developed a blockchain project to create a permanent distributed record of intellectual effort and associated reputational reward that instantiates and democratises educational reputation beyond the academic community. Blockchain for Education [20] is another prototype system supporting the storage, retrieval and verification of certificates via blockchain technology. Certificates are an important means of proving lifelong learning achievement in today's environment, however they are susceptible to forgery. Blockchain helps to solve this problem. The prototype system uses Ethereum and its smart contracts to manage identities of registered certificate authorities and the hashes of certificates which are stored in a separate, centralized document management system, while the profile information of certificate authorities is stored using the Interplanetary File System (IPFS) distributed file system. Storing data off the blockchain allows it to be be deleted as is required, for example, by the European General Data Protection Regulation (GDPR) for personal information.

Other uses of blockchain in higher education have also been contemplated, including motivation, assessment, advising, etc. [21]. A prototype system for a blockchain based learning analytics platform built on Ethereum has been proposed as well [22].

3. SECURITY AND PRIVACY ISSUES IN BLOCKCHAIN

A number of different types of security problems have been noted in blockchain. In this section, we survey a few of these and how they relate to blockchain in education.

3.1. 51% Attacks

Blockchain relies on a distributed consensus mechanism such as proof of work (PoW) in order to establish trust. Unfortunately, the consensus mechanism is itself vulnerable to a 51% attack. This occurs when a single malicious miner or mining pools controls more than 50% of the total hashing power of the entire blockchain [23]. The malicious miner who controls this much hashing power has basically unlimited power to manipulate and modify the blockchain

information including reversing transactions, changing the ordering of transactions and blocking transactions from being verified. A mining pool for the Bitcoin blockchain at one time reached 42% of hashing power before miners dropped out of the pool in order to stop a 51% attack from becoming possible [24].

51% attacks require a large investment in computing power and coordination in order to be carried out. While the effort might be worthwhile for the monetary gain associated with blockchains used for cryptocurrencies, for large public blockchains the effort will probably not be worthwhile for the advantages to be gained by manipulating educational credentials, unless some group wants to corrupt all of the credentials for an institution (for example), not just those of a single student. It is not to be discounted, though, that such corruption could occur as collateral damage from a 51% attack aimed at some other, more remunerative target.

3.2. Malicious Contracts

Ethereum is an open-source, public, blockchain-based distributed computing platform and operating system featuring smart contract (scripting) functionality [25]. Users of the Ethereum platform are 'pseudo-anonymous' and a single user can have multiple accounts under multiple cryptographic identities. It has been shown [26] how 'criminal smart contracts' can facilitate leakage of confidential information, theft of cryptographic keys, and various real-world crimes.Atzei et. al, [27], have presented a taxonomy of vulnerabilities of smart contracts in Ethereum including such vulnerabilities as stack overflow, exception disorder and unpredictable state. Various countermeasures to these malicious contracts have been proposed, including malicious contract detection using supervised learning [25].

For higher education blockchain, likely the most serious threat from this type of attack is the leaking of private credentials. However, it is also possible that a malicious contract exploiting the vulnerabilities identified above might falsely update a student's credentials, corrupting them.

3.3. Sybil Attacks

In a Sybil attack, a single adversary controls many nodes in a system by creating numerous fake identities in order to gain a disproportionately large influence [28]. A defense against Sybil attacks can rely on validated identities issued by a trusted authority. The requirements for agents to present a trusted identity conflicts with the need for permissionless open membership, though. An example of a blockchain development to defeat Sybil attacks is TrustChain [29] which includes a novel Sybil-resistant algorithm NetFlow to determine the trustworthiness of agents in an online community.

3.4. Eclipse attacks

An eclipse attack will allow the attacker to monopolize all of the victim's connections – both incoming and outgoing [30]. This will isolate the victim from other users on the network. The attacker is able to filter the victim's view of the blockchain or to cause the victim to waste computing power on obsolete views of the blockchain. The attacker is also able to use the victim's computing power to conduct its own malicious acts. In the context of education blockchain, a student might prevent a potential employer from having a complete view of all of the student's credentials.

4. CONCLUSIONS

Blockchain has been increasingly accepted in higher education as an alternative for the storage of academic records. A systematic literature review [10] shows that interest is continuing to grow, with the peak of interest occurring in the last complete year surveyed - 2020. The main advantage of blockchain in this context is that the records are not under the control of several institutions or of a centralized third party.

The current approach has each academic institution maintaining its own records, and students having to collect records from each institution they attended in order to present them to prospective employers. This system is both slow and costly (for students – institutions may make money by charging for the release of academic records). Blockchain solves the problem of speed (in addition to taking the records out of the hands of the institutions, as noted above). Some cost is associated with blockchain, however, much of that cost can be borne by prospective employers who carry out transactions on the blockchain in order to have access to academic records.

An alternative approach could be to have a trusted third party hold all of the academic records, but then we still have the problem of trust. Not all countries have an institution worthy of trust, and even in those that do, not all students may trust them (no universally trusted institution). Blockchain does not require a trusted third party, so it solves this problem.

As was seen in the descriptions of the types of attacks in the previous section, many of the attack strategies require obtaining large amounts of computing power in order to be effective. Such an attack may be considered worthwhile in the cryptocurrency application of blockchain, but in the education sector, no such financial incentive is apparent for attacks on a single student's credentials. Possibly the blackmailing of a large institution poses more of a threat, though given the higher payoffs available to criminals in other blockchain areas, that is still rather farfetched. The threats associated with malicious contracts seem to be more of concern in educational blockchain, since they may corrupt educational credentials.

Given the high cost of storing data on the blockchain, it is likely that educational transcripts themselves will be stored off the blockchain, with only the credentials needed to access and verify them being stored on the blockchain. One popular mode of storing data off the blockchain is to use the InterPlanetary File System (IPFS). IPFS is a peer-to-peer distributed file system, so its architecture and management style is a good match for blockchain-based architectures. For examples of the combination of blockchain and IPFS, see [31], [32] for the use in storage of electronic medical records, and [33] for the use in the automotive insurance sector.

In any case, it will be necessary for users of blockchain-based educational credential systems to keep up to date with the latest security issues in the blockchain area in order to guard against attacks, since those attacks, and the defences against them, are continuously evolving.

REFERENCES

[1] Nakamoto, S. (2008) *Bitcoin: A Peer-to-Peer Electronic Cash System*, [online] https://bitcoin.org/bitcoin.pdf.

[2] Wang, S., Ouyang, L., Yuan, Y., Ni, X., Han, X. & Wang, F. (2019) "Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 49, No. 11, pp. 2266-2277.

[3] Iansiti, M. & Lakhani, K. R. (2017) "The Truth About Blockchain", *Harvard Business Review*, Vol. 95, No. 1, pp. 118-127.

[4] Peters, G. W. & Panayi, E. (2016) "Understanding Modern Banking Ledgers Through Blockchain Technologies: Future of Transaction Processing and Smart Contracts on the Internet of Money," in *Banking Beyond Banks and Money*, Springer, Cham, pp. 239-278.

[5] Underwood, S. (2016) "Blockchain Beyond Bitcoin," *Communications of the ACM*, Vol. 59, No. 11, pp. 15-17.

[6] Lamberti, F., Gatteschi, V., Demartini, C., Pranteda, C. & Santamaria, V. (2017) "Blockchain or Not Blockchain, That is the Question of the Insurance and Other Sectors," *IT Professional*.

[7] Ayed, A. B., (2017) "A Conceptual Secure Blockchain-Based Electronic Voting System," *International Journal of Network Security & Its Applications*, Vol. 93.

[8] Yumna, H. et al. (2019) "Use of Blockchain in Education: A Systematic Literature Review," in *Intelligent Information and Database Systems. ACIIDS 2019, Lecture Notes in Computer Science,* vol. 11432, N.Nguyen, F. Gaol, T.P. Hong, B. Trawiński Eds, Springer, Cham, pp. 191-202.

[9] Yokubov, B. (2018) "Blockchain Based Storage of Students Career," Master Degree Thesis, Politecnico di Torino, [online] https://webthesis.biblio.polito.it/9471/.

[10] Raimundo, R. & Rosário, A. (2021) "Blockchain System in the Higher Education", *European Journal of Investigation in Health, Psychology and Education*, Vol. 11, No. 1, pp. 276-293.

[11] Rooksby, J. & Dimitrov, K. (2019) "Trustless Education? A Blockchain System for University Grades", *Ubiquity: The Journal of Pervasive Media*, Vol. 6, No. 1, pp. 83-88.

[12] Arndt, T. (2018) "Empowering University Students with Blockchain-Based Transcripts," *Proceedings of CELDA 2018,* Budapest, Hungary, October 21-23.

[13] Mahamatov, N., Kuvnakov A. & Yokubov, B. (2020) "Application of Blockchain Technology in Higher Education," *2020 International Conference on Information Science and Communications Technologies (ICISCT),* pp. 1-6

[14] Turkanović, M. et al. (2019) "EduCTX: A Blockchain-Based Higher Education Credit Platform," *IEEE Access,* Vol. 6, pp. 5112-5127.

[15] Košič, K., Černec, R., Barnsley, A. & Thoorens, F., (2018) Building an open-source blockchain ecosystem with ARK, *OTS 2018 Sodobne informacijske tehnologije in storitve*, p. 45.

[16] Kuvshinov, K., Nikiforov, I., Mostovoy, J., Mukhutdinov, D., Andreev, K. & Podtelkin, V. (2018) Disciplina: Blockchain for education. *Yellow Paper. URL: https://disciplina. io/yellowpaper. pdf*.

[17] Selvaratnam, R.M. & Sankey, M. (2021) "An Integrative Literature Review of the Implementation of Micro-credentials in Higher Education: Implications for Practice in Australasia", *Journal of Teaching and Learning for Graduate Employability*, Vol. 12, No. 1, pp. 1-17.

[18] Blockcerts (2019) *The Open Standard for Blockchain Credentials,* [online] https://www.blockcerts.org

[19] Sharples, M. & Domingue, J. (2016) "The Blockchain and Kudos: A Distributed System for Educational Record, Reputation and Reward. In *European conference on technology enhanced learning*. pp. 490-496, Springer, Cham.

[20] Gräther, W., Kolvenbach, S., Ruland, R., Schütte, J., Torres, C. & Wendland, F. (2018) "Blockchain for Education: Lifelong Learning Passport", In *Proceedings of 1st ERCIM Blockchain Workshop 2018*. European Society for Socially Embedded Technologies (EUSSET).

[21] Chen, G., Xu, B., Lu, M. & Chen, N.S. (2018) "Exploring Blockchain Technology and its Potential Applications for Education", *Smart Learning Environments*, Vol. 5, No. 1, pp. 1-10.

[22] Ocheja, P., Flanagan, B. & Ogata, H. (2018) "Connecting Decentralized Learning Records: A Blockchain Based Learning Analytics Platform", In *Proceedings of the 8th international conference on learning analytics and knowledge*, pp. 265-269.

[23] Shanaev, S., Shuraeva, A., Vasenin, M. and Kuznetsov, M. (2019) "Cryptocurrency Value and 51% Attacks: Evidence from Event Studies", *The Journal of Alternative Investments*, Vol. *22*, No. 3, pp.65-77.

[24] Hajdarbegovic, N. (2014) "Bitcoin Miners Ditch Ghash. io Pool Over Fears of 51% Attack", [online] http://www.coindesk.com/bitcoin-miners-ditch-ghash-io-pool-51-attack/.

[25] Kumar N., Singh A., Handa A., & Shukla S.K. (2020) "Detecting Malicious Accounts on the Ethereum Blockchain with Supervised Learning", In: Dolev S., Kolesnikov V., Lodha S., Weiss G. (eds) *Cyber Security Cryptography and Machine Learning*. CSCML 2020. Lecture Notes in Computer Science, vol 12161.

[26] Juels, A., Kosba, A. & Shi, E. (2016), "The Ring of Gyges: Investigating the Future of Criminal Smart Contracts" In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 283-295.

[27] Atzei, N., Bartoletti, M. & Cimoli, T. (2017), "A Survey of Attacks On Ethereum Smart Contracts (sok)", In *International conference on principles of security and trust*, Springer, Berlin, Heidelberg, pp. 164-186.
[28] Douceur, J.R. (2002) "The Sybil Attack", In *International workshop on peer-to-peer systems* (pp. 251-260). Springer, Berlin, Heidelberg.
[29] Otte, P., de Vos, M. & Pouwelse, J. (2020) "TrustChain: A Sybil-Resistant Scalable Blockchain", *Future Generation Computer Systems*, Vol. 107, pp. 770-780.
[30] Heilman, E., Kendler, A., Zohar, A. & Goldberg, S. (2015) "Eclipse Attacks on Bitcoin's Peer-to-Peer Network", In *24th USENIX Security Symposium (USENIX Security 15)*, pp. 129-144.
[31] Zheng, Q., Li, Y., Chen, P. & Dong, X. (2018) "An Innovative IPFS-Based Storage Model for Blockchain", In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 704-708.
[32] Sun, J., Yao, X., Wang, S. & Wu, Y. (2020) "Blockchain-Based Secure Storage and Access Scheme for Electronic Medical Records in IPFS", *IEEE Access*, Vol. *8*, pp. 59389-59401.
[33] Nizamuddin, N. & Abugabah, A. (2021) "Blockchain for Automotive: An Insight Towards the IPFS Blockchain-Based Auto Insurance Sector", *International Journal of Electrical & Computer Engineering* pp. 2088-8708, Vol. 11, No. 3.

AUTHOR

Timothy Arndt is Chair and Professor of the Department of Information Systems, Cleveland State University. He received a Ph.D. in Computer Science from the University of Pittsburgh. He is editor of several journals and has been involved in the organization of several international conferences. His research interests include database systems, software engineering, e-learning, and blockchain.

IMPLEMENTING BLOCKCHAIN TECHNOLOGY IN SUPPLY CHAIN MANAGEMENT

Atul Anand[1], A Seetharaman[2] and K Maddulety[3]

[1]Researcher, SP Jain School of Global Management- Mumbai
[2]Dean Research, SP Jain School of Global Management- Singapore
[3]Deputy Director, SP Jain School of Global Management- Mumbai

ABSTRACT

This paper is aimed at studying the factors influencing the implementation of blockchain in supply chain management to solve the current issues faced in the supply chain ecosystem. Supply chains are part and parcel of every business and have multiple inefficiencies in the system. Some of these inefficiencies can be managed by usage of blockchain Platform .Technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation are critically evaluated for adoption of blockchain in supply chain. A pilot study is conducted in form survey for analysis of these factors. Hypotheses are derived for these factors for quantitative research. Subsequently these hypotheses are examined with the help of ADANCO2.3 for structural equation modelling. As an outcome, it is evident that Innovation and Extrinsic factors are significantly impacting the adoption of blockchain in supply chain management.

KEYWORDS

Blockchain technology, Supply chain management, Technology, smart contract, Intracompany synergies, Intercompany collaboration; Extrinsic Factors, Innovation.

1. INTRODUCTION

In today's world, Supply Chain management has grown into an intricate network of suppliers and partners. It faces a myriad of challenges such as fraudulent transactions, non-traceability of genuine products, counterfeit products, unethical practices of using child labour and many others, resulting in a lack of transparency and trust issues amongst stakeholders. In this paper, blockchain technology is analysed to manage the current issues of Supply chain management. This paper is concluded by laying out research framework for studying the adoption of blockchain technology in Supply chain environment. Based on the literature survey, research problem and question are derived. Based on the research question, the research objective is framed to provide the direction for the studies.

1.1. Research Problem and Question

Below research question is framed based on the research problem at hand:

a) What variables influence the usage of blockchain technology in supply chain management?

The logical reduction of the research question is the research objective, which is to examine this field in more depth.

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 147-160, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120713

1.2. Research Objectives

Research objective is logically deduced from the research question to study the aspects of blockchain adoption in supply chain.

a) To examine the reasons which affect blockchain technology adoption in supply chain ecosystem

Due to time constraints, factors impacting the adoption of blockchain technology in supply chain management are studied in pilot studies. For Pilot studies, assessment of the outcomes are carried out in a time-bound way following the required research practices involving multiple steps from problem identification to the providing recommendation based on data analysis.

2. LITERATURE REVIEW

As part of literature review last 5 years (2018-2022) research papers, journals, industry reports are studied with regards to adoption of blockchain in supply chain management to arrive at the variables of conceptual framework.

2.1. Blockchain Technology

In 2008, Santoshi Nakamoto (pseudonym) wrote a white paper describing Bitcoin and blockchain technology. Blockchain technology can be applied in multiple industries such as supply chain, manufacturing, and finance (for example, Bitcoin is a use case of blockchain technology in finance). Blockchain is a distributed database system that is secured cryptographically and uses a consensus mechanism to store transactions in blocks. Each block in the chain is attached to an earlier block via the hash function, resulting in blockchain's key features of transparency, traceability, immutability, timestamping, and decentralisation [1][2]. The traceability and transparency features of blockchain can contribute to addressing many of the issues of the traditional supply chain [3].

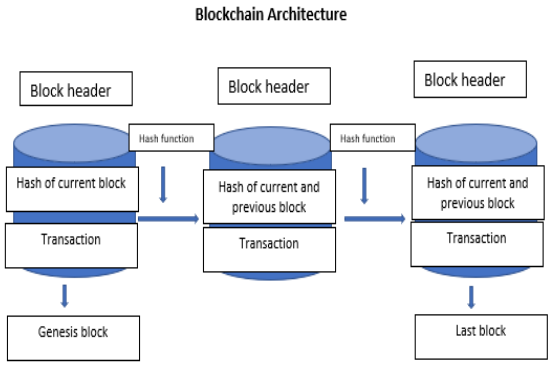


Figure 1. Framework of blockchain architecture (Giri & Manohar, 2021, p. 3)

2.2. Role of blockchain technology in the supply chain

Traditional SCM starts from the raw material supplier to the manufacturer, distributors, wholesalers, retailers and, finally, to the end user, in a linear manner. In contrast, the blockchain-enabled supply chain employs three additional entities which are not part of the traditional supply chain. These entities include registrars, certifiers, and standards organisations, each of which

ensures the blockchain-based platform is building the elements of trust and transparency through the use of smart contracts [4] . In the blockchain-based supply chain model, change of ownership can be executed through a smart contract without any manual intervention. Blockchain-based records can be updated by certifiers and registrars once the change of ownership is completed. In this manner, all the records can be tracked from the time of origination to the end of the delivery chain without the chance of anything getting tampered [5][4].

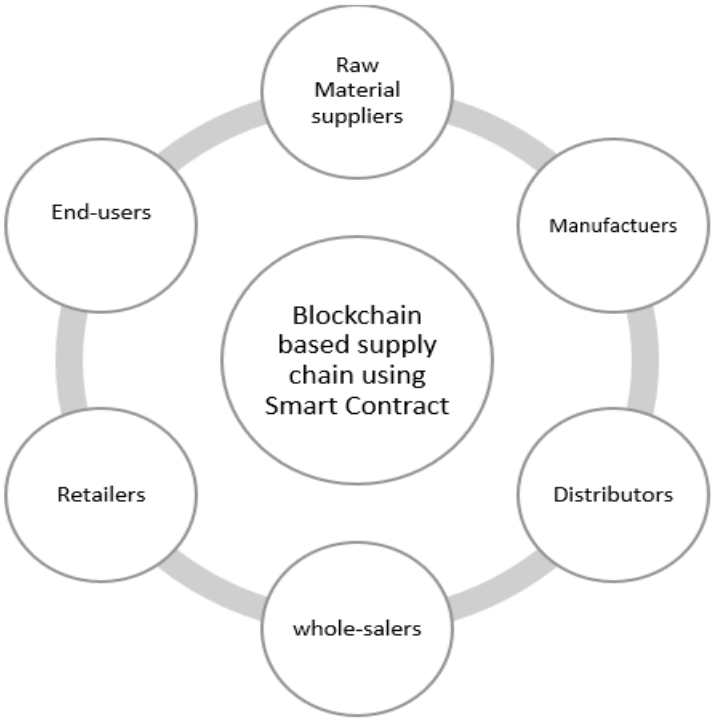


Figure 2. Supply chain transformation through blockchain technology

2.3. Blockchain technology adoption in supply chain

This section is organised into five themes: technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation, which are used for investigating the barriers and challenges to the adoption of blockchain technology [6] [7] [8] [4].

2.3.1. Technology

Technology is constantly evolving in this era and, in decades past, enterprise resource planning, SCM and customer relationship management (CRM) packages were in great demand to address supply chain business problems. Radio-frequency identification (RFID) is considered a cutting-edge technology to address some of the tracking and tracing issues of the supply chain [4]. Technology constructs are focused mainly on security and privacy, technology scalability, technology resilience, design architecture and technology maturity as primary areas in studying the applicability of blockchain in SCM. Information generation, handling and usage need to be secured and protected [4, 7, 9-11]. Security and privacy of data offer a competitive advantage to businesses in the supply chain [12]. Blockchain has features of encryption and hashing mechanisms to protect the data from tampering [2].

2.3.2. Intracompany Synergies

Intracompany synergy elements are internal to the organisation. An organisation's culture, leadership, knowledge and capability, cost of ownership and enterprise strategies are considered internal to the organisation. Top management support is considered as a positive influencing factor demonstrating the plan of small and medium businesses to implement blockchain [10] whereas a lack of support and commitment from management is one of the barriers to the adoption of this technology[4]. Organisations planning to implement blockchain technology need to invest in building required knowledge and capabilities in this area. Being as blockchain is a niche technology, organisations are apprehensive about its overall return on investment. Whilst organisations continue to use legacy technology due to the lack of tools and their hesitation to convert to newer systems, this, consequently, acts as barrier for blockchain technology adoption.

2.3.3. Intercompany Collaboration

Intercompany collaboration refers to coordinating with multiple supply chain partners outside the organisation. Under intercompany collaboration, the elemental attributes of sustainability, traceability, transparency, collaboration, and interoperability are important. The lack of customer awareness and the challenge of incorporating sustainable practices of blockchain technology between different supply chain partners create barriers to the adoption of the technology. Information security and data sharing policies of different organisations are not coherent, adding to further complexity for blockchain adoption [4]. To establish a common culture of transparency and trust amongst different supply chain partners that are geographically dispersed across the world takes much time and effort. It is one of the key challenges to be addressed. The absence of common technical standards amongst intercompany collaboration of stakeholders results in a lack of collaboration and coordination amongst industry players[13] .Collaboration is one of the six themes presented as well [14] and it is a critical success factor for implementation of blockchain in production and operation management. This can be addressed through the roles of registrars and certifiers in the blockchain. The goal of blockchain application is to attain leaner processes resulting in reduced paperwork, improved information sharing and automated processes overall between various stakeholders in different supply chain businesses, such as maritime industries [8].

2.3.4. Extrinsic Factors

The category of extrinsic factors is comprised of extrinsic stakeholders, governments, industries, and institutions that have an impact on blockchain implementation in the supply chain [4]. Governmental policies, decision rights, extrinsic stakeholder involvement, governance of traceability efforts, social challenges, customer influence, market demand, the environment, supply chain practices, global standards, legislations, and regulations are the subconstructs of extrinsic factors[5-11,13-16]. Blockchain adoption is impacted due to non-availability of uniform international standards and a lack of clarity amongst different nation states on the policies and standards of blockchain. Governments across the world are divided in their intention to allow blockchain technology in multiple sectors. This needs to be addressed soon to reach some consensus about this technology.

2.3.5. Innovation

Blockchain technology is currently undergoing tremendous innovation. Innovation in blockchain solutions, along with smart contracts, internet of things (IoT) adoption, big data implementation, artificial intelligence (AI) and machine learning can result in tremendous potential for digital

disruption. In blockchain-enabled supply chain practices, a smart contract ensures the automatic change of ownership of a product once the product moves across various supply chain actors such as manufacturer to distributor and helps in easier tracking and trust-building amongst stakeholders [3] [4].

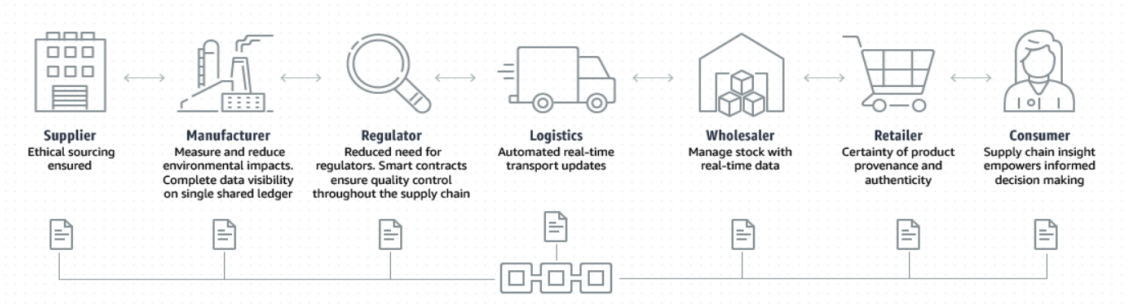


Figure 3. Blockchain for Supply Chain (AWS Amazon,2022)

The development of blockchain comprised of smart contracts and IoT technologies is the real innovation needed for Supply chain management.

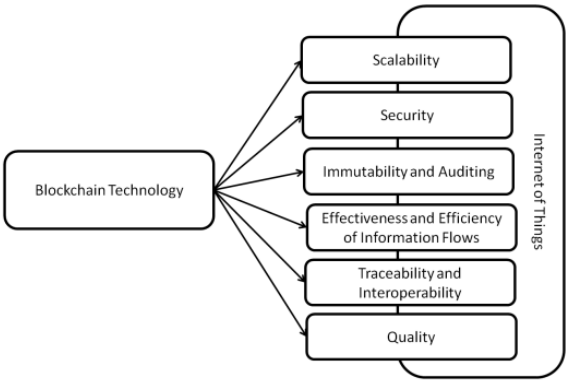


Figure 4. Blockchain and IoT (Rejeb et al., 2019, p. 8)

Characteristics of blockchain technology along with IoT enables scalability, security, immutability, and auditability of the system. Effectiveness and efficiency of information systems, traceability, interoperability, and overall quality of information is maintained by the application of blockchain technology. Algorithms such as proof-of-stake and proof-of-work add to the data security in the blockchain[19] .Organisations in the supply chain are experimenting with various blockchain design choices, such as public and private blockchains, permissioned and permissionless blockchain technologies, distributed ledger technology (DLT), and other blockchain-based platforms and solutions, and are selecting the design choices based on their requirements to ensure acceptance of blockchain along the supply chain [2,6,10,19-21].

3. RESEARCH FRAMEWORK

It is evident that acceptance and applicability of blockchain in the supply chain sector are currently in a nascent phase and need to be researched further to address supply chain issues. Logistics and supply chain businesses are facing multiple challenges in adoption and implementation of blockchain technology. Below is the proposed framework for blockchain adoption and implementation in Supply chain management

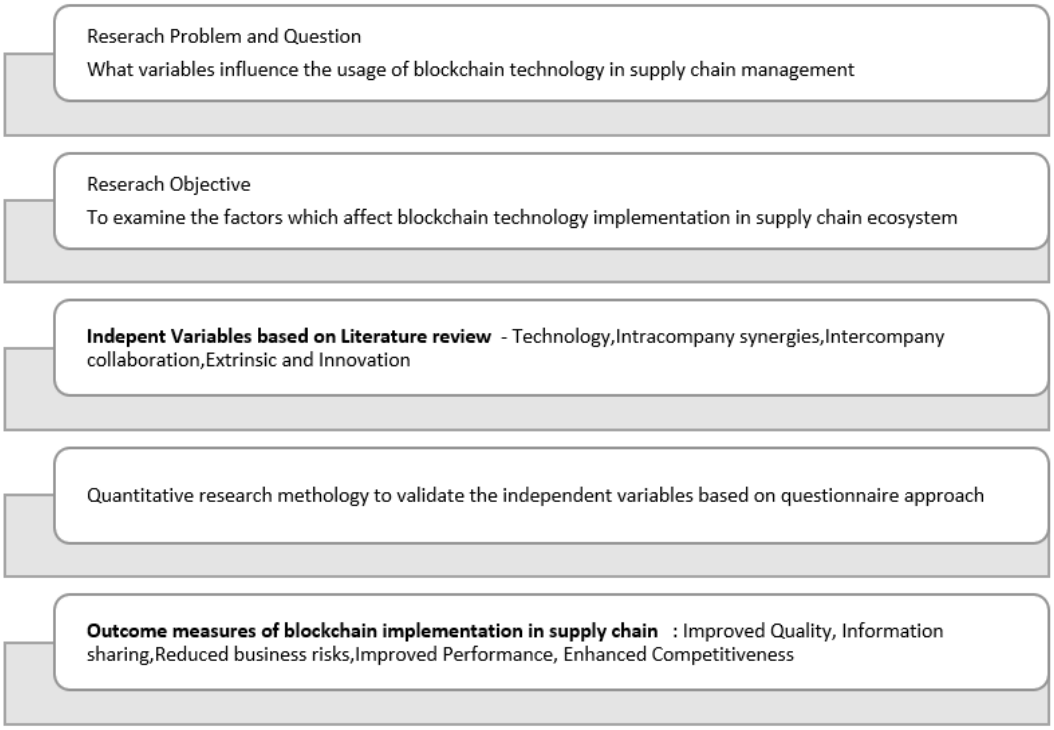


Figure 5. Research Framework for Blockchain Implementation in Supply chain management

Based on the five independent variables of research framework below hypotheses have been formulated for quantitative research

H1: Technology is significantly impacting the adoption of blockchain in supply chain management

H2: Intracompany synergies is significantly impacting the adoption of blockchain in supply chain management

H3: Intercompany Collaboration is significantly impacting the adoption of blockchain in supply chain management

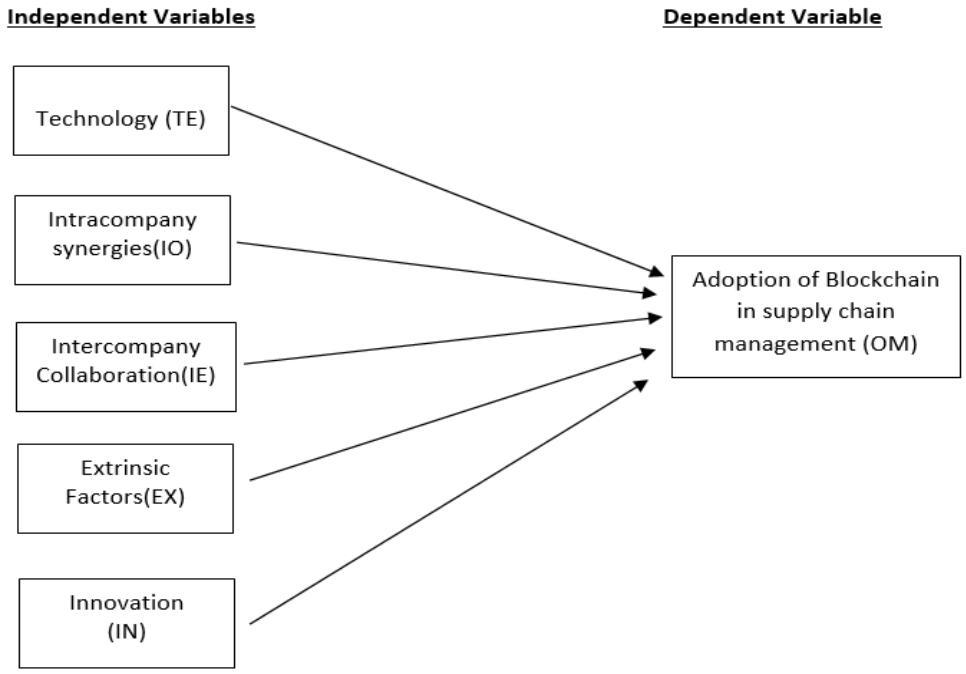


Figure 6. Conceptual model for blockchain adoption in supply chain management

H4: Extrinsic Factors is significantly impacting the adoption of blockchain in supply chain management

H5: Innovation is significantly impacting the adoption of blockchain in supply chain management

The factors influencing adoption of blockchain in the supply chain are categorised under the constructs of technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation. The supply chain and operations can be transformed by usage of blockchain technology [1] because it enhances the safety and security of a product. The outcome measure for blockchain technology results in improvement of overall quality of the supply chain ecosystem by using smart contracts [22]. Blockchain system usage ensures better information sharing by bringing intracompany synergies [23]. Blockchain technology also results in improving interorganisational information sharing amongst supply chain stakeholders [24]. A blockchain-based system helps in reducing business risk and illegal counterfeiting through a risk management structure between supply chain partners [25]. The adoption of blockchain technology in the supply chain ecosystem improves the overall performance of the SCM, owing to intermediaries being removed, and results in addressing sustainable SCM practices. Innovation in the application of blockchain technology for SCM enhances competitiveness by improving operational performance. Innovations such as application of blockchain with IoT results in enhancing the integrity of the supply chain businesses [18].

The challenges in adoption of blockchain in SCM are attributed to organisations lacking interoperability and technological standards for its implementation [5]. For widespread adoption of blockchain in the supply chain, there is heavy dependence on intercompany collaboration amongst various stakeholders across the ecosystem, which is a challenge faced by supply chain

partners. Scalability of blockchain technology to meet the growing demands is a barrier to be addressed in the adoption of blockchain technology [20]. Issues of higher costs, technological complexity, and the lack of skills and capabilities are inhibitors for the adoption of blockchain technology [10] .Leadership support and organisational culture as part of intracompany synergies are required and act an enabler for the adoption of blockchain in SCM.

4. RESEARCH METHODOLOGY AND APPROACH

There are mainly 3 types of research methods: Mono, mixed and multi. For this studies, mono method was used due to time limitation. Mono method uses a single method for data collection and analysis. For example: either Quantitative method or Qualitative method for data collection and analysis. For current studies, mono method is used in form of quantitative method. Online survey questionnaire is used due to limitation of overall time.

Deductive approach is used as part of the current study to arrive at the hypothesis based on literature review. These hypotheses are tested through data collection and quantitative analysis to validate them

5. SOURCES OF DATA

While carrying out the Pilot study, multiple sources of data were considered. The Literature review was based on secondary data, and it was followed by using primary data for preparing the questionnaire.

5.1. Secondary Data

As part of literature survey, secondary research was carried out with the help of various databases search such as google scholar, ProQuest, EBSCO and other online databases, journals, and articles. Secondary data has helped in ensuring the limitations and scope for future research is taken by for further studies by researcher. After thorough analysis of these articles, gap variables were identified from scope of future research. There were more than 185 gap variables discovered as part of secondary research. The frequency distribution for these gap variables were ensured, so that these gap variables can be bucketed into 5 major variables - Technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation - for taking it ahead as independent variables.

5.2. Primary Data

Based on conceptual framework and hypotheses derived from secondary data, questionnaire is used as an instrument for primary data collection. In this pilot studies – google form survey was used for reaching out to the respondent. Initially questionnaire in form of word document was created to get the endorsement from ethical committee. After the required ethical standards are met, google form was created for capturing the response from the users and assurance was provided to the respondents to the safety and security of their data.

6. DATA ANALYSIS

The google form questionnaire was rolled out for 4 weeks' time to get maximum responses as part of this assignment. There were 85 responses received as part of this pilot exercise. The data received was cleaned up and coded so that it can be analysed with the help of ADANCO2.3 for structural equation modelling. The hypothesis was formulated and analysed as part of this

exercise. The results of the Adanco2.3 was analysed, and the output is received in form of different tables. Various tables of construct reliability, convergent validity, loading-based validity, discriminant validity, inter-construct correlations, indicator multicollinearity, direct effects, indirect effects, total effects, t-values for various constructs and so on were analysed and compared to the threshold values generated. A figure is generated having relationship with dependent variables – Outcomes and measure with satisfactory R^2 values for the same.

Construct reliability is meant for measuring consistency of any construct within research. Construct reliability can be measured through values of ρA , ρc and Cronbach's values. As part of analysis, ADANCO 2.3 provide the values for the model for these parameters. Values of Cronbach alpha (α) is fine for Intercompany collaboration, Extrinsic factors, and Innovation. For Technology and Intracompany synergies, value is relatively on lower side

Table 1. Construct Reliability

Construct	**Dijkstra-Henseler's rho (ρ_A)**	**Joreskog's rho (ρ_C)**	**Cronbach's alpha(α)**
Technology (TE)	.5283	.7237	.4554
Intracompany Synergies (IO)	.6249	.7782	.5847
Intercompany Collaboration (IE)	.8162	.8691	.8117
Extrinsic factors(EX)	.7739	.8009	.6983
Innovation (IV)	.8344	.8800	.8275

TE: Technology; IO :Intracompany synergies; IE: Intercompany collaboration; EX: Extrinsic factors; IV: Innovation

Convergent validity is the extent to which a measure correlates positively with alternative measures of the same construct. Convergent validity refers to the degree to which two theoretically related construct measurements are really related [45]. The average variance generated from the model is used to assess its convergent validity (AVE). Average variance extracted is used to compare the degree of variance explained by an unobserved construct to the variance attributed to random measurement error (AVE). A construct with an AVE value larger than 0.5 explains a considerable portion of the variance in the model. For all the constructs AVE value is more than .5 except technology and extrinsic factors.

Table 2. Convergent validity

Construct	Average Variance extracted(AVE)
Technology (TE)	.4781
Intracompany Synergies (IO)	.5435
Intercompany Collaboration (IE)	.5710
Extrinsic factors (EX)	.4701
Innovation (IV)	.5968

The discriminant validity of the construct is confirmed when it has the largest absolute value in each column and row which is at the major diagonal. Since the diagonal values of Average variance extracted are bigger than the non-diagonal squared correlation values of their respective

rows and columns, the model has discriminant validity. Each Construct is distinct and significantly different from each other

Table 3. Discriminant validity : Fornell-larcker Criterion

Construct	TE	IO	IE	EX	IV
Technology (TE)	**.4781**	.1657	.4482	.3466	.3169
Intracompany Synergies (IO)		**.5435**	.2400	.2838	.3003
Intercompany Collaboration (IE)			**.5710**	.3795	.4421
Extrinsic factors (EX)				**.4701**	.2463
Innovation (IV)					**.5968**

In the structural model, R^2 provides the explanatory power of the model. For dependent variable – Blockchain adoption in supply chain (OM) has R(Square) value of .646 which is good value for Pilot study as there are limited number of respondents.

Table 4. Structural Model

R-Squared		
Construct	Coefficient of determination(R^2)	Adjusted(R^2)
Blockchain Adoption (OM)	.6461	.6174

OM: Blockchain Adoption

7. DATA INTERPRETATION

Based on Direct effect inference table analysis (ADANCO 2.3):

- Innovation (IV) is significantly impacting the blockchain adoption (OM) in supply chain management since P value (2 sided) < .01 and t-value (3.374) > 2.59
- Extrinsic factor (EX) is significantly impacting the blockchain adoption (OM) in supply chain management since P value (2 sided) <.01 and t- value (2.9)>2.59 meaning that external (EX) factors such as government policies, customer, global markets have role to play for the adoption of blockchain in supply chain management
- Intercompany collaboration (IE) is not significantly impacting the blockchain adoption (OM) in supply chain management since P value (2 sided) >.01 and t-value (1.5) < 2.59
- Intracompany synergies (IO) is not significantly impacting the blockchain adoption (OM) in supply chain management since P value (2 sided) >.01 and t-value (1.1) < 2.59
- Technology (TE) is not significantly impacting the blockchain adoption (OM) in supply chain management since P value (2 sided) >.01 and t- value (.3562) <2.59

Direct Effects Inference

Effect	Original coefficient	Standard bootstrap results					Percentile bootstrap quantiles			
		Mean value	Standard error	t-value	p-value (2-sided)	p-value (1-sided)	0.5%	2.5%	97.5%	99.5%
IV -> OM	0.3115	0.2960	0.0923	3.3740	0.0007	0.0004	0.0435	0.1107	0.4750	0.5412
EX -> OM	0.2601	0.2803	0.0887	2.9331	0.0034	0.0017	0.0113	0.0806	0.4351	0.4787
IE -> OM	0.2576	0.2751	0.1700	1.5150	0.1298	0.0649	-0.1047	-0.0244	0.6131	0.7025
IO -> OM	0.1074	0.0950	0.0975	1.1016	0.2707	0.1353	-0.1341	-0.0830	0.2958	0.3608
TE -> OM	0.0458	0.0357	0.1285	0.3562	0.7217	0.3608	-0.2847	-0.2105	0.2866	0.3644

Figure 7. Based on ADANCO2.3 analysis - Direct Effect Inference

8. CONTRIBUTION TO PRACTICE

The conceptual framework for adoption of blockchain in SCM include the five constructs of technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation. The outcome measures for these constructs include improved quality, information sharing, reduced business risk, improved performance, and enhanced competitiveness. These enable organisations and stakeholders to solve the existing problems of the supply chain, such as lack of trust, counterfeit products, and malpractices in the supply chain business, resulting in the establishment of sustainability practices in the value chain and supply chain 2.0.

9. CONCLUSION AND RECOMMENDATIONS

In this paper, we propose a conceptual framework for the adoption of blockchain in supply chain management based on the constructs of technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation. Based on these constructs five hypotheses are formulated with respect to the significance of them in adoption of blockchain in supply chain management. A pilot study was conducted through online distribution of questionnaire to 345 individuals for duration for one month. 85 responses were received. These were analysed using ADANCO2.3 (PLS-SEM) tool after data cleaning. Based on the response analysis, it is evident that Blockchain is still in the early phase of adoption in supply chain management. Innovation is significantly impacting adoption of blockchain in supply chain ecosystem. Smart contract, IOT, AI, Cloud Computing are some of the recent innovations which has the potential to impact supply chain. Extrinsic factors are also significantly impacting the adoption of blockchain. Some of these external factors are government policies, overall global markets, governance framework and customer demands. This conceptual framework based on the blockchain system can result in establishing SCM 2.0 practices. Since blockchain technology adoption is still in a nascent stage of application in the supply chain ecosystem, it has a vast future potential for researchers and industry practitioners.

10. LIMITATIONS AND SCOPE FOR FUTURE RESEARCH

Most of the respondents of this survey were from Asia pacific region, hence the result cannot be generalized across geographics .As part of the scope of future research, these constructs of the conceptual framework - technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation must be empirically evaluated for specific geography or country to give accurate results. In this research there were no qualitative techniques such as focused group discussions with industry experts were used. Interviews and focused group discussed should be explored as a scope of future research along with this conceptual framework. Due to limited timelines – no moderation or mediation variables were considered during the analysis. The effect of theories such as TAM, UTAUT, TOE on the conceptual framework variables - technology, intracompany synergies, intercompany collaboration, extrinsic factors, and innovation – should be evaluated as future scope of research.

REFERENCES

[1] Cole, R., Stevenson, M., & Aitken, J. (2019). Blockchain technology: Implications for operations and supply chain management. *Supply Chain Management: An International Journal, 24*(4), 469–483. https://doi.org/10.1108/SCM-09-2018-0309

[2] Giri, G., & Manohar, H. L. (2021). Factors influencing the acceptance of private and public blockchain-based collaboration among supply chain practitioners: A parallel mediation model. *Supply Chain Management: An International Journal, ahead-of-print*(ahead-of-print). https://doi.org/10.1108/SCM-02-2021-0057

[3] Moosavi, J., Naeni, L. M., Fathollahi-Fard, A. M., & Fiore, U. (2021). Blockchain in supply chain management: A review, bibliometric, and network analysis. *Environmental Science and Pollution Research.* https://doi.org/10.1007/s11356-021-13094-3

[4] Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019b). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research, 57*(7), 2117–2135. https://doi.org/10.1080/00207543.2018.1533261

[5] Akhtar, P., Azima, N., Ghafar, A., & Din, S. U. (2021). Barricades in the Adoption of Block-Chain Technology in Supply Chain Management: Challenges and Benefits. *Transnational Marketing Journal, 9*(1). https://doi.org/10.33182/tmj.v9i1.1021

[6] Irannezhad, E. (2020). Is blockchain a solution for logistics and freight transportation problems? *Transportation Research Procedia, 48*, 290–306. https://doi.org/10.1016/j.trpro.2020.08.023

[7] Orji, I. J., Kusi-Sarpong, S., Huang, S., & Vazquez-Brust, D. (2020). Evaluating the factors that influence blockchain adoption in the freight logistics industry. *Transportation Research Part E: Logistics and Transportation Review, 141*, 102025. https://doi.org/10.1016/j.tre.2020.102025

[8] Pu, S., & Lam, J. S. L. (2021). Blockchain adoptions in the maritime industry: A conceptual framework. *Maritime Policy & Management, 48*(6), 777–794. https://doi.org/10.1080/03088839.2020.1825855

[9] Karuppiah, K., Sankaranarayanan, B., & Ali, S. M. (2021). A decision-aid model for evaluating challenges to blockchain adoption in supply chains. *International Journal of Logistics Research and Applications*, 1–22. https://doi.org/10.1080/13675567.2021.1947999

[10] Kumar Bhardwaj, A., Garg, A., & Gajpal, Y. (2021). Determinants of Blockchain Technology Adoption in Supply Chains by Small and Medium Enterprises (SMEs) in India. *Mathematical Problems in Engineering*, 2021, 1–14. https://doi.org/10.1155/2021/5537395

[11] Winkelhaus, S., & Grosse, E. H. (2020). Logistics 4.0: A systematic review towards a new logistics system. *International Journal of Production Research, 58*(1), 18–43. https://doi.org/10.1080/00207543.2019.1612964

[12] Chaudhuri, A., Bhatia, M. S., Kayikci, Y., Fernandes, K. J., & Fosso-Wamba, S. (2021). Improving social sustainability and reducing supply chain risks through blockchain implementation: role of outcome and behavioural mechanisms. *Annals of Operations Research*, 1-33. https://doi.org/10.1007/s10479-021-04307-6

[13] Schmahl, A., Mohottala, S., Burchardi, K., Egloff, C., Govers, J., Chan, T., & Giakoumelos, M. (n.d.). *Resolving the Blockchain Paradox in Transportation and Logistics*. 18

[14] Hastig, G. M., & Sodhi, M. S. (2020). Blockchain for Supply Chain Traceability: Business Requirements and Critical Success Factors. *Production and Operations Management, 29*(4), 935–954. https://doi.org/10.1111/poms.13147

[15] Ghode, D., Yadav, V., Jain, R., & Soni, G. (2020). Adoption of blockchain in supply chain: An analysis of influencing factors. *Journal of Enterprise Information Management, 33*(3), 437–456. https://doi.org/10.1108/JEIM-07-2019-0186

[16] Park, K. O. (2020). A Study on Sustainable Usage Intention of Blockchain in the Big Data Era: Logistics and Supply Chain Management Companies. *Sustainability, 12*(24), 10670. https://doi.org/10.3390/su122410670

[17] AWS, Amazon. 'Blockchain for Supply Chain: Track and Trace'. Blockchain for Supply Chain: Track and Trace, n.d. https://aws.amazon.com/blockchain/blockchain-for-supply-chain-track-and-trace/.

[18] Rejeb, A., Keogh, J. G., & Treiblmaier, H. (2019). Leveraging the internet of things and blockchain technology in supply chain management. *Future Internet, 11*(7), 161. https://doi.org/10.3390/fi11070161

[19] Tan, W. K. A., & Sundarakani, B. (2020). Assessing Blockchain Technology application for freight booking business: A case study from Technology Acceptance Model perspective. *Journal of Global Operations and Strategic Sourcing, 14*(1), 202–223. https://doi.org/10.1108/JGOSS-04-2020-0018

[20] Panos, A., Kapnissis, G., & Leligou, H. C. (2020). The Blockchain and DLTs in the Maritime Industry: Potential and Barriers. *European Journal of Electrical Engineering and Computer Science, 4*(5). https://doi.org/10.24018/ejece.2020.4.5.243

[21] Toyoda, K., Mathiopoulos, P. T., Sasase, I., & Ohtsuki, T. (2017). A Novel Blockchain-Based Product Ownership Management System (POMS) for Anti-Counterfeits in the Post Supply Chain. *IEEE Access, 5*, 17465–17477. https://doi.org/10.1109/ACCESS.2017.2720760

[22] Chen, S., Shi, R., Ren, Z., Yan, J., Shi, Y., & Zhang, J. (2017). A Blockchain-Based Supply Chain Quality Management Framework. *2017 IEEE 14th International Conference on E-Business Engineering (ICEBE)*, 172–176. https://doi.org/10.1109/ICEBE.2017.34

[23] Azogu, I., Norta, A., Papper, I., Longo, J., & Draheim, D. (2019). A Framework for the Adoption of Blockchain Technology in Healthcare Information Management Systems: A Case Study of Nigeria. *In Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance* (pp. 310–316). Melbourne VIC Australia: ACM, 2019. https://doi.org/10.1145/3326365.3326405.

[24] Guggenberger, T., Schweizer, A., & Urbach, N. (2020). Improving interorganizational information sharing for vendor managed inventory: Toward a decentralized information hub using blockchain technology. *IEEE Transactions on Engineering Management, 67*(4), 1074–1085.

[25] Wang, Keyao, Xiuxia Yan, and Kaiying Fu. 'Research on Risk Management of Agricultural Products Supply Chain Based on Blockchain Technology'. *Open Journal of Business and Management* 08, no. 06 (2020): 2493–2503. https://doi.org/10.4236/ojbm.2020.86155.

[26] Wong, L. W., Tan, G. W. H., Lee, V. H., Ooi, K. B., & Sohal, A. (2020). Unearthing the determinants of Blockchain adoption in supply chain management. International Journal of Production Research, 58(7), 2100-2123

[27] Perboli, G., Musso, S., & Rosano, M. (2018). Blockchain in logistics and supply chain: A lean approach for designing real-world use cases. Ieee Access, 6, 62018-62028.

[28] Tijan, E., Aksentijević, S., Ivanić, K., & Jardas, M. (2019). Blockchain technology implementation in logistics. Sustainability, 11(4), 1185.

[29] Jain, G., Singh, H., Chaturvedi, K. R., & Rakesh, S. (2020). Blockchain in logistics industry: in fizz customer trust or not. Journal of Enterprise Information Management.

[30] Dobrovnik, M., Herold, D. M., Fürst, E., & Kummer, S. (2018). Blockchain for and in Logistics: What to Adopt and Where to Start. Logistics, 2(3), 18.

[31] Sivula, A., Shamsuzzoha, A., & Helo, P. (2018, January). Blockchain in logistics: mapping the opportunities in con-struction industry. In International Conference on Industrial Engineering and Operations Management.

[32] Pervez, H., & Haq, I. U. (2019, March). Blockchain and IoT based disruption in logistics. In 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE) (pp. 276-281). IEEE.

[33] Liao, D. Y., & Wang, X. (2018, December). Applications of blockchain technology to logistics management in integrated casinos and entertainment. In Informatics (Vol. 5, No. 4, p. 44). Multidisciplinary Digital Publishing Institute.

[34] Upadhyay, A., Ayodele, J. O., Kumar, A., & Garza-Reyes, J. A. (2020). A review of challenges and opportunities of blockchain adoption for operational excellence in the UK automotive industry. Journal of Global Operations and Strategic Sourcing.

[35] Hirata, E., Lambrou, M., & Watanabe, D. (2020). Blockchain technology in supply chain management: insights from machine learning algorithms. Maritime Business Review.

[36] Montecchi, M., Plangger, K., & Etter, M. (2019). It's real, trust me! Establishing supply chain provenance using blockchain. Business Horizons, 62(3), 283-293.

[37] Wong, L. W., Leong, L. Y., Hew, J. J., Tan, G. W. H., & Ooi, K. B. (2020). Time to seize the digital evolution: Adoption of blockchain in operations and supply chain management among Malaysian SMEs. International Journal of Information Management, 52, 101997

[38] Feng, H., Wang, X., Duan, Y., Zhang, J., & Zhang, X. (2020). Applying blockchain technology to improve agri-food traceability: A review of development methods, benefits and challenges. Journal of Cleaner Production, 260, 121031.

[39] Dubey, R., Gunasekaran, A., Bryde, D. J., Dwivedi, Y. K., & Papadopoulos, T. (2020). Blockchain technology for enhancing swift-trust, collaboration and resilience within a humanitarian supply chain setting. International Journal of Production Research, 58(11), 3381-3398
[40] Bodkhe, U., Tanwar, S., Parekh, K., Khanpara, P., Tyagi, S., Kumar, N., & Alazab, M. (2020). Blockchain for industry 4.0: A comprehensive review. IEEE Access, 8, 79764-79800
[41] Kouhizadeh, M., Saberi, S., & Sarkis, J. (2021). Blockchain technology and the sustainable supply chain: Theoretically exploring adoption barriers. International Journal of Production Economics, 231, 107831.
[42] Koh, L., Dolgui, A., & Sarkis, J. (2020). Blockchain in transport and logistics–paradigms and transitions. International Journal of Production Research, 58(7), 2054-2062.
[43] Vadgama, N., & Tasca, P. (2021). An Analysis of blockchain adoption in supply chains between 2010 and 2020. Frontiers in Blockchain, 4, 8.
[44] Balci, G., & Surucu-Balci, E. (2021). Blockchain adoption in the maritime supply chain: Examining barriers and salient stakeholders in containerized international trade. Transportation Research Part E: Logistics and Transportation Review, 156, 102539.
[45] Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the academy of marketing science, 43(1), 115-135.

AUTHORS

Atul Anand is a research scholar from SP Jain School of global management -Mumbai, having an industry experience of more than 16 years in various sectors. He is an MBA (Global) from SP Jain school as well. He has more than 100 professional certifications and 50+ industry awards. He is a reviewer of the book – "Introduction to blockchain technology" published by Van Haren and author of the book – "Experiencing Life As A YOGI"

Dr Seetha is Dean of Research in SP Jain School of Global Management, Singapore, having experience of 30+ years in research in multi-disciplinary areas and has produced more than 300+ papers for publications.

Dr K Maddulety is Deputy Dean in SP Jain School of Global Management, Mumbai, having experience of 25+ years in industry and research and has produced more than 90+ papers for publications.

AN INTELLIGENT SOCIAL-BASED ASSISTANT APPLICATION FOR STUDY TIME MANAGEMENT USING ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING

Haoyu Li[1], Ryan Yan[2] and Ang Li[3]

[1]Whittier Christian High School, 501 N Beach Blvd, La Habra, CA 90631
[2]Cal Poly Pomona, 3801 West Temple Avenue, Pomona, California 91768
[3]California State University, Long Beach,
1250 Bellflower Blvd, Long Beach, CA 90840

ABSTRACT

The question that we aim to solve is "How will elderly people be able to increase their productivity and remember their tasks?" There are many ways to go about answering this question, but we have devised a simple solution to this question that can directly and quickly have a positive impact on these individuals. Our method to solving this is creating a to-do list in Flutter, which will allow elderly people to have easy access to a list of tasks that they have to complete [5].

Some predicted results of this to-do list is that it can raise productivity for its users. Our to-do list features a ChatBot, which talks to the user through a message-like system in order to prompt user input for specific details such as the time, date, and main description of the task [6]. Then, the ChatBot will take in all this information to produce a clean and concise final task description that takes keywords from the user-inputted description. This provides users of our to-do list with an alternative method of adding tasks, which may be greatly appreciated by those who are less able-bodied or struggle to type. By offering the elderly a way of adding tasks that can take less typing, these individuals may rely on this to-do list as a great convenience to their lives.

KEYWORDS

Flutter, To-Do List, Tasks, Productivity.

1. INTRODUCTION

Our topic is creating a to-do list with Flutter, and this to-do list can create tasks to remind people of what to do each day. This to-do list can be helpful to anyone of all ages, as long as they have a smartphone or a similar portable electronic device that is compatible with Flutter. However, this to-do list is mainly directed towards those who are not as able-bodied, such as senior citizens. Benefits of the Flutter to-do list include providing people with an intuitive and easy-to-use application for storing tasks to be completed later. Some to-do lists that currently exist may be harder to use and might make people frustrated when trying to create and handle their to-do list. Therefore, making a to-do list that is more user-oriented will allow people to live their day-to-day

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 161-171, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120714

lives more conveniently. There is little to no negative consequence for having this to-do list, besides taking extra memory on the device.

Our topic is significant because it allows elderly people to live more productive lives and brings them a convenient way to remember their tasks [7]. Some elderly people may have poor memory, and a to-do list would provide them with a method of remembering what they need to accomplish. A setback that elderly people may come across is that they may have less capable bodies, or they may not be great at typing. This to-do list seeks to improve the lives of these individuals by using keywords in order to retrieve the main ideas of these tasks and delivering them to the users.

There are some existing to-do lists on mobile device application stores that have aimed to achieve the same general goal as we have, which is to improve productivity among our application's users. These to-do lists, for the most part, have the same general format. This format includes a checklist with functioning check boxes, which include the list of tasks that still need completing [8]. The users will press a button that prompts them to type in text that describes what task they will need to complete in the future, and they will check the boxes next to the corresponding task once they have completed it to cross the tasks out.

Although these existing to-do lists may not necessarily have glaringly large issues in the eyes of the general public, those who are less-able bodied may struggle to use these applications. This is because the vast majority of these to-do lists involve having to type out the entire task in order to save it onto the to-do list. However, for those who have poor typing skills or are otherwise unable to easily type, they will likely have a difficult time using these to-do lists. This concern mainly applies with the elderly, whose joints may not work as well as before due to arthritis or other medical conditions. Therefore, an alternative way to add tasks that requires less typing or physical ability is a necessity for some of these individuals. Another issue that lies with some to-do lists is the inability to save tasks after closing the application. The tasks that are typed in are only saved for that specific instance of the application. Most people want their tasks to be saved for the next time they reopen the application, considering that the tasks they typed in were tasks that still needed completing. Saving tasks from the previous session every time the application is opened is a crucial part of a functioning to-do list.

Our to-do list admittedly shares many similarities with other to-do lists that currently exist in mobile application stores. However, our to-do list differs due to having a login and register system. Our to-do list uses Firebase, which is a database that has the capabilities of saving usernames and passwords [9]. The user can use our application by first creating an account, then logging in with this account. Most existing to-do lists do not have a system like this, and the tasks are only saved to the corresponding device. However, with the login and register system of our file, the tasks that are saved to a specific account can be transferred over to another device, as long as the user is using the same account to log in with. This capability provides our to-do list with more flexibility than most others.

Another feature of our to-do list that most existing to-do lists do not have is a ChatBot [10]. This ChatBot is in a second tab of our application, which prompts the user in a message-like system for input on their task. First, it will ask for the user-inputted description of the task. Then, the ChatBot will ask for the date and the time of when the task needs to be completed. Finally, the ChatBot will take keywords from the user-inputted task description and combine it with the date and time, then ask for confirmation from the user if they want to add ChatBot's task to their to-do list. After the user agrees and responds with "yes", the task is successfully added. This provides a more intuitive way for users who may not be as adept in using technology, such as elderly individuals.

We proved our results about the effectiveness of the to-do list by performing two types of experiments. One of these experiments involved testing out the IBM Watson Assistant and recording how accurate it was when extracting the main keywords from a user-inputted description and including it into the final description. In order to have the measurement of accuracy tested more fairly, we had 10 participants download the application and experiment with using the ChatBot at least 20 times each, and we urged them to use as much variety in the test descriptions as possible. Each person would record how many times they had tested this feature in total and how many times the IBM Watson Assistant would properly extract the necessary keywords out of these test descriptions.

The second experiment involved testing their overall productivity boost from using the app. Each of the aforementioned participants would give themselves a rating from 1 to 10 that measured how productive they felt during that week. Then, they would use the to-do application for all of their needs during the next week and record their rating from 1 to 10 of how productive they felt after one week of using the application. Furthermore, each participant would state whether they used the ChatBot feature at all in their day-to-day lives and whether they preferred using the ChatBot to add tasks over the traditional method of typing the task description on their own. This experiment would attempt to measure how effective this application was at improving productivity in its users.

The rest of the paper is organized in sections from 2 to 5. Section 2 dives into the details of what challenges we faced during the process of creating our to-do list application. Section 3 covers a general overview of the solution we proposed to boost the productivity of individuals, as well as each detail and step that was taken to reach the final product. Section 4 describes the experiments performed to prove the effectiveness of the to-do list application. Section 5 goes over a brief summary of three related works and how they compare to this paper. Lastly, Section 6 provides concluding remarks to summarize our paper and consider what could be done in the future regarding this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Lack of knowledge in Flutter

Our first challenge was a lack of knowledge in Flutter. It was the first real experience that we had with using Flutter and Dart, so being able to create a full-fledged application with not only basic to-do list compatibility but also adding a login and register system was a massive accomplishment [11]. At first, we struggled greatly when trying to wrap our head around the Flutter framework, and writing code seemed to involve a lot of unnecessary indenting and formatting here and there that made it difficult to understand what our Dart code was trying to accomplish at its core. However, as we persevered and studied Dart code closer, we were able to figure out the purpose of each line of code and were able to write our own code that could successfully perform to-do list functions and even include our own features to the application. we will be able to use the knowledge and experience from overcoming this challenge in future projects.

2.2. Figuring out how to approach the topic

Our second challenge was figuring out how we could approach the topic of to-do lists in a unique way. When we looked at other papers to figure out how we should write our own, all we saw

were good ideas that were all being taken. Other papers focused on creating a group to-do list that could be accessed and modified by multiple people or comparing the efficiency of electronic to-do lists and paper to-do lists. In order to come up with our own emphasis on to-do lists, we had to dig deeper and think about what we personally valued the most in a to-do list. The answer we came up with was that we wanted a to-do list that could be as convenient and helpful as possible for the user. We are very pleased that this was the concept we chose to focus on, as it was something that we felt passionate about.

2.3. Finding ways to send a signal from a watch to the server

Our third challenge was finding a way to integrate Python code into our Flutter project [12]. Our goal with this to-do list application was to add a ChatBot as an additional option for adding tasks. This ChatBot involved the IBM Watson Assistant, which would help in extracting the necessary keywords from user-inputted descriptions. However, the only way we knew to use this ChatBot was through Python. Therefore, we needed a method to fuse these two different programming languages into one coherent and functional application. After searching online for a while, we discovered that this was possible through integrating the Python code. We progressed steadily with our project, using Flutter primarily as our front-end code and using a few small Python files to assist with back-end code. We would not have learned about such a useful Flutter skill if we had not thought to include this feature into our to-do list application.

3. SOLUTION

The whole system of our program works by running a Flutter application [13]. Most of the application's code is made up of Dart files, which control the formatting and functionalities of the application. The application uses multiple files that control both the application's front-end and back-end. Firebase is used as the database to store the users' usernames and passwords as well as the tasks in their to-do lists. With Firebase, a login system is used to store a user's to-do list and tasks. Furthermore, the Flutter application has Python code integrated into it, allowing even more functionality. This Python integration was specifically added to include the feature of a ChatBot. With the help of the IBM Watson Assistant, the ChatBot gives the user a second way of adding a task to the to-do list with a series of questions that require short answers. First, the user taps on the ChatBot tab to enter a tab that has a user interface that is similar to a messaging app. Then, after a second of staying on the ChatBot tab, the ChatBot will send a message to the user, prompting the user to send a description of the task. The ChatBot will prompt the user for a date once the user sends the description as a message, then the ChatBot will prompt the user for a time once the user sends the date as a message. If the user had added any information for the next prompt in a previous prompt, such as including the time along with the date when prompted for just the date, the ChatBot will take this into account by taking in both the date and time and skipping the prompt asking for the time. Finally, the ChatBot will ask for confirmation from the user. If the user accepts, the task is added to the to-do list successfully.

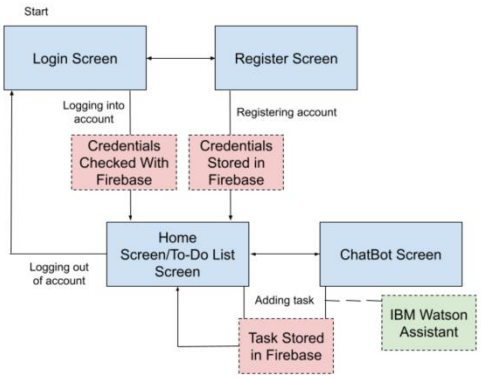


Figure 1. Overview of the system

The database was a significant part of the application, as it was what stored both the login credentials and the tasks in the to-do list that corresponded to each account. We decided to use Firebase as the database due to its compatibility with Flutter and its cost (free to use below a certain threshold of data stored). The database was implemented by first importing the Firebase libraries into the relevant Dart files. In the main Dart file, the application would first initialize Firebase before running. Firebase would handle the login, register, and logout processes for the application as well as adding tasks to the to-do list that corresponded with the user's ID. The logging in, registering, and logging out was handled in a Dart file that created an instance of the Firebase authentication service and called functions from the instance of this service to perform these actions. In the case of logging in and registering, the email and password entered into the text field were used as parameters for the corresponding functions. Registering saves the email and password combination to Firebase.

```dart
void main() async {
  WidgetsFlutterBinding.ensureInitialized();
  await Firebase.initializeApp();

  runApp(MyApp());

}
```

```dart
import 'package:firebase_auth/firebase_auth.dart';

class AuthService {
  final FirebaseAuth _firebaseAuth = FirebaseAuth.instance;
  Myuser? _userFromFirebase(User? user) {
    if (user == null) {
      return null;
    }
    return Myuser(user.uid, user.email!);
  }

  Stream<Myuser?> get user {
    return _firebaseAuth.authStateChanges().map(_userFromFirebase);
  }

  Future<Myuser?> signinWithEmailandPassword(
      String email, String password) async {
    print(email);
    final credential = await _firebaseAuth.signInWithEmailAndPassword(
        email: email, password: password);

    return _userFromFirebase(credential.user);
  }

  Future<Myuser?> createWithEmailandPassword(
      String email, String password) async {
    print(email + password);
    final credential = await _firebaseAuth.createUserWithEmailAndPassword(
        email: email, password: password);

    return _userFromFirebase(credential.user);
  }

  Future<void> signOut() async {
    return await _firebaseAuth.signOut();
  }
```

Figure 2. Screenshot of code 1

Another major component of the application is the ChatBot. Unlike the rest of the code in the application, ChatBot uses Python as its programming language. In order to integrate the Python code, we imported an API called Flask in one of the Python files. In this Python file, we made

two functions, one of them for creating a ChatBot session and the other one for sending messages in the session. For each of these functions, we used "@app.route", which allowed this code to be included in the functionality of the application. The ChatBot worked by importing the IBM Watson Assistant in a Python file and setting up the service for this assistant through an API key. The ChatBot would start a session when the user tapped on the ChatBot tab to enter the ChatBot screen, and the ChatBot would send messages by asking about the task description, date, and time. After going through the task adding process with the user and having the user confirm the task description, the ChatBot would send a final message as a visual notice to the user that the task was successfully added. As each message sent by the ChatBot is checked for whether it includes the string "I would schedule", it would then notice that the final message has been sent and call a function from Firebase that adds a new task with the full task description that is shown in the second half of the ChatBot's final message.

```
from flask import Flask
import ibm_chatbot as bot
import json

app = Flask(__name__)

# Create session.
@app.route("/create_session/<user_id>")
def create_session(user_id):
    response = {}
    response ['session_id'] = bot.create_session()
    bot_init_message = bot.get_response(message='',session_id=response ['session_id'],user_id=user_id)
    response.update(bot_init_message)
    return response



@app.route("/send_message/<message>/<session_id>/<user_id>")
def send_message(message,user_id,session_id):
    response = bot.get_response(message=message,session_id=session_id,user_id=user_id)
    return json.dumps(response)

app.run(host='0.0.0.0')
```

```
print(response)
if response['output']['generic']:
    for generic in response['output']['generic']:
        text = generic['text']
        res['response'] += text + '\n\n'
        if 'I would schedule' in text:
            start_index =  len('I would schedule ')
            firebase.set_context(text [start_index:], user_id)
            res['end_session'] = True
return res
```

Figure 3. Screenshot of code 2

There are four main screens of the application: the login screen, the register screen, the home screen, and the ChatBot screen. The home screen includes the to-do list with tasks. The screens were implemented by creating Dart files for each screen, in which the Dart files had code that both formatted the layout of these screens and implemented functional buttons that either brought the user to another screen or performed some sort of action. When the user opens the application for the first time, they will start at the login screen. From here, the user could enter the credentials of an existing account and tap the Login button to enter the home screen. The user could also tap the Register button to switch to the register screen, create the new account by filling in the credentials, and tap Read Content and then Back to Login to enter the home screen. These screens will also check for any errors in the user input. For example, if the two passwords are not matching in the register screen, the Dart file that handles the register screen will check for this and return the message "password not equal" when it determines that the two password text fields

hold different String values. After logging in or registering, the user can swap between the home screen with the to-do list and the ChatBot screen through a bottom navigation bar in a Dart file for routing that has separate dart files imported for both the home screen and the ChatBot screen. The check boxes in the home screen function by creating a strikethrough effect on the text, filling the check box in blue with a white check mark, and having a boolean variable switch from false to true when an unchecked checkbox is tapped. When a checked checkbox is tapped again, the text and checkbox revert, and the boolean variable becomes false again.

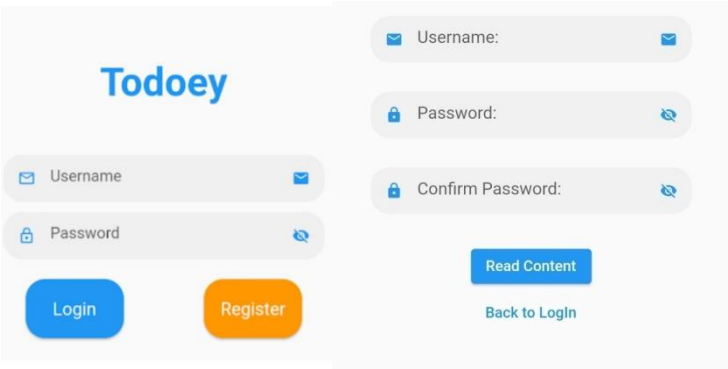


Figure 4. Screenshot of login page

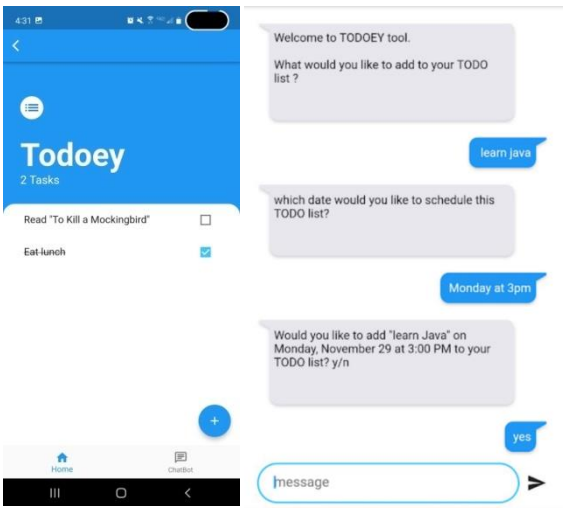


Figure 5. Screenshot of task page

4. EXPERIMENT

4.1. Experiment 1

The problem we aim to solve is creating a to-do list for users who may be slow or unable to type well, such as those who may be elderly and have issues with their joints. The solution we have devised to tackle this problem is creating a ChatBot that can potentially reduce the amount of typing and break the task description creation process down into simple steps. The experiment design involves 10 participants testing the accuracy of the IBM Watson Assistant in the ChatBot regarding the extraction of important keywords from the user-inputted descriptions at least 20 times each, which gives a decent sample size to account for variability. For each time that the IBM Watson Assistant failed to successfully extract the keywords, either through having an

awkward-sounding description or losing crucial information, the participant must provide what exactly they inputted and what exactly the ChatBot outputted.

Participant #	Number of times tested	Number of successful tests	Percentage of successful tests
1	20	20	100%
2	20	20	100%
3	22	21	95.45%
4	20	20	100%
5	21	21	100%
6	20	18	90%
7	25	24	96%
8	21	21	100%
9	20	19	95%
10	20	20	100%

Table 1. Result of Experiment 1

According to the results, the ChatBot was fairly accurate for the majority of the user inputs. However, there are still cases where some of the keywords are cut out of the final description that make the task description sound awkward. An example is when one of the participants entered the following task description: “do math”. Instead of the ChatBot including the full user-inputted description into the final task description, the ChatBot simply decided that the word “math” would be included. The participant who tested this description labeled this as an “ awkward description ”. Since the description still gets its point across without losing any crucial information, this seems to just be a minor issue. All minor issues were also labeled as “ inaccurate ” for the sake of this experiment. Throughout all the tests, there were no major issues involving the loss of any crucial information in any of the ChatBot results. Therefore, the ChatBot seems to be a very accurate and reliable alternative to adding tasks.

4.2. Experiment 2

The problem that our solution aims to solve is that people in their everyday lives, especially the elderly, may forget their tasks and become less productive than they could potentially be as a result. The solution we have designed, a to-do list application that can be accessed on smart devices, would be able to achieve this by giving people a quick and convenient way to remind themselves of their tasks. This second experiment is a survey that the participants will take after using the to-do list application for all of their needs for 7 days. Since the experiment includes 10 participants, there is enough to account for any variability. The participants are also asked at the end of the survey whether they regularly opted to use the ChatBot feature when adding their tasks.

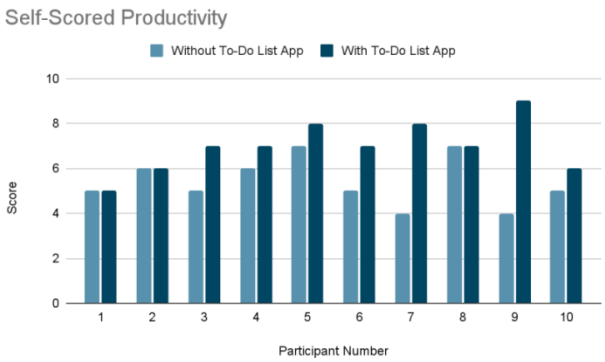


Figure 6. Self-Scored Productivity

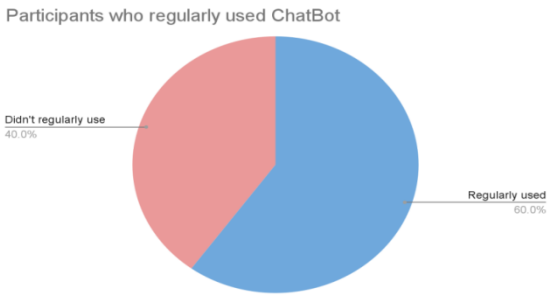


Figure 7. Participants who regularly used ChatBot

The experiment's main measurements were a self-given score of how productive they felt the week before using the to-do list application and the week after using the to-do list application. From the results that were collected from this experiment, it appears that all participants felt that they were either more productive or at the same level of productivity after a week of using this to-do list. Most of the improvements in productivity seemed to be relatively small, but there were two participants who felt that they had their productivity significantly boosted in the week after, with a significant productivity boost being considered as an improvement of 4 or greater from the previous week. Therefore, this to-do list application has been shown to boost productivity among its users. In the feedback, six out of the ten participants used the ChatBot feature regularly in their to-do list application use, indicating that the feature was a significant part of their experience with using the application.

For the first experiment, the results proved that the ChatBot is capable of being a convenient alternative to the traditional method of typing in full task descriptions. Due to being accurate at extracting keywords with the majority of user-inputted descriptions tested, the ChatBot is a fairly reliable tool when setting reminders for tasks. In the second experiment, the results showed that slightly over half of all our to-do list application users will regularly use the ChatBot application. A significant portion of the users, therefore, will treat this feature as a necessary part of the experience when using this application. Furthermore, the second experiment also proves that regular use of the to-do list application can provide, on average, a small boost in productivity to its users. None of the participants that tested the application experienced a reduction in productivity. Based on the experiment results, the to-do list application appears to increase the productivity of its users and provide the ChatBot as an alternative method to add tasks that the majority of its users will take advantage of.

5. RELATED WORK

The related work involved a study that compared the productivity of a smartphone to-do list application and paper-based lists within an intensive care unit. The study found that although the to-do list application received positive feedback from the medical staff, the application could not be proven to be more effective at increasing productivity than paper to-do lists [1]. This related work had its main emphasis on experimenting with the effectiveness of a smartphone application . However, our work focused more on the development of the application itself. Our strengths were detailing our application creation process, while the strengths of the related work were studying how the workflow within an intensive care unit was affected through the use of a to-do list application.

The main topic of the related work is using the Analytic Hierarchy Process, or AHP for short, to prioritize use cases in mobile application development [2]. AHP is a method that utilizes the main problem that is meant to be solved, every possible solution to the aforementioned problem, and the criteria that will be used to measure how preferable each solution is. Through these three factors, AHP can find the best solution and analyze complex decisions through math and psychology. [3]. The related work appears to cover how AHP is used with mobile applications in general, with to-do list applications as a small focus in part of the paper. On the other hand, our work involves exclusively to-do lists and how to make the individual user's experience of a to-do list application as convenient as possible. Overall, the related work covers a broader topic with more information, while our work has a more narrow focus.

In this related work, the to-do list that they have created was meant to appeal to families and those in group projects who were in need of a to-do list that could be edited by multiple people rather than a single person [4]. While our work focused on developing a ChatBot to create tasks for the users, the author of this work had a great emphasis on having a group to-do list as well as a monetary management system built in. This work's strength is improving to-do lists to fit a group setting and allow for collaboration. On the other hand, our work had our greatest strengths in maximizing the convenience of an individual.

6. CONCLUSIONS

The application we propose to help others against forgetfulness and to improve productivity is our to-do list [14]. We believe that having a list of tasks that can be conveniently accessed from a portable electronic device will aid users in their everyday lives. This to-do list application includes the basic features of adding tasks, checking off tasks, and viewing the whole task list [15]. Our application has two ways to add tasks, with the second way being a ChatBot that the user can communicate with through messages to potentially require less typing. Typing less will be incredibly helpful towards the elderly and others who may have poor typing skills. This ChatBot comes with the added benefit of making concise tasks that are easy to understand. Some users may type long, complicated messages to add to their to-do list, then come back and get confused about what they were trying to remind themselves about before. On the other hand, some users may type up task descriptions that are much too simple, and they may not remember important details such as the date and time that a task needed to be completed. The ChatBot aims to solve both of these issues. When the user inputs a description into the ChatBot, the IBM Watson Assistant will extract the most necessary keywords from the description to use later for the final task description. The ChatBot will also explicitly ask the user for a date and a time to ensure that the user will have a clear idea of when a task needs to be completed.

One limitation that our to-do list has is the ability to check off multiple tasks at once. At the moment, our to-do list is capable of only checking off one task at a time, and trying to check off another task will uncheck the previously checked-off task. A second limitation that our to-do list has is the ability to remove tasks. Tasks can only be added, and only one task can be checked off at a time, but there is currently no way in our to-do list to have a task completely removed. With extended use, the to-do list can begin to feel cluttered with many unneeded tasks.

In the future, we plan to solve these limitations by adding a button next to each task that removes the task rather than just checking them off. Furthermore, to solve the issue of only being able to check off one task at a time, we will search for a different method of implementing the checking off of tasks.

REFERENCES

[1] Esposito, M., Rocq, P.-L., Novy, E., Remen, T., Losser, M.-R., & Guerci, P. (2020, January 21). Smartphone to-do list application to improve workflow in an intensive care unit: A superiority quasi-experimental study. International Journal of Medical Informatics. Retrieved February 4, 2022, from https://www.sciencedirect.com/science/article/pii/S1386505619309487

[2] Yildirim, Onur, and Serhat Peker. “Prioritizing Use Cases for Development of Mobile Apps Using AHP: A Case Study in To-Do List Apps.” SpringerLink, 26 July 2019, link.springer.com/chapter/10.1007%2F978-3-030-27192-3_24.

[3] “What Is the Analytic Hierarchy Process (AHP)?” Passage Technology, www.passagetechnology.com/what-is-the-analytic-hierarchy-process.

[4] Ringgau, Diana anak, et al. “Development of to-Do List and Monetary Management System.” International ABEC, abecindonesia.org/iabec/index.php/iabec/article/view/4.

[5] Payne, Rap. "Developing in Flutter." Beginning App Development with Flutter. Apress, Berkeley, CA, 2019. 9-27.

[6] Dahiya, Menal. "A tool of conversation: Chatbot." International Journal of Computer Sciences and Engineering 5.5 (2017): 158-161.

[7] Vergados, Dimitrios, et al. "Intelligent services for assisting independent living of elderly people at home." Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments. 2008.

[8] Gordon, Suzanne, Patrick Mendenhall, and Bonnie Blair O'toole. Beyond the checklist. Cornell University Press, 2012.

[9] Moroney, Laurence. "The firebase realtime database." The Definitive Guide to Firebase. Apress, Berkeley, CA, 2017. 51-71.

[10] Adamopoulou, Eleni, and Lefteris Moussiades. "An overview of chatbot technology." IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, Cham, 2020.

[11] Kuzmin, Nikita, Konstantin Ignatiev, and Denis Grafov. "Experience of developing a mobile application using flutter." Information Science and Applications. Springer, Singapore, 2020. 571-575.

[12] Sanner, Michel F. "Python: a programming language for software integration and development." J Mol Graph Model 17.1 (1999): 57-61.

[13] Joorabchi, Mona Erfani, Ali Mesbah, and Philippe Kruchten. "Real challenges in mobile app development." 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. IEEE, 2013.

[14] Kral, Vojtech Adalbert. "Senescent forgetfulness: benign and malignant." Canadian Medical Association Journal 86.6 (1962): 257.

[15] Benacerraf, Paul. "Tasks, super-tasks, and the modern eleatics." The Journal of Philosophy 59.24 (1962): 765-784.

IMAGE ENCRYPTION ALGORITHM OF CHAOS SYSTEM ADDING COSINE EXCITATION FUNCTION

Zhenzhou GUO[1] and Xintong LI[2]

[1]School of Artificial Intelligence,
Shenyang Aerospace University, Shenyang, China
[2]School of Computer, Shenyang Aerospace University, Shenyang, China

ABSTRACT

In order to increase the chaotic performance of the chaotic system, the chaotic system A-S proposed by Sprott is improved by adding a cosine excitation function to a controller. A series of new chaotic systems are obtained, and the chaotic performance of the improved system is verified. The image is encrypted by the chaotic sequence generated by the improved A chaotic system. In the scrambling part of the image encryption algorithm, the zigzag transformation is improved, and different directions are selected to start the traversal, so that the scrambling process is not easy to be restored. The diffusion part draws on the traditional IDEA algorithm to perform diffusion operations on the image. Finally, the encryption algorithm is analyzed and tested, and the results show that the algorithm has fast encryption and decryption speed, sufficient key space, and can resist statistical analysis attacks well. The algorithm can provide better guarantee for the security of images.

KEYWORDS

Cosine Excitation Function, Three-dimensional Chaotic System, Digital Image Encryption.

1. INTRODUCTION

In today's increasingly prosperous Internet era, to ensure the security of people using the Internet to transmit information, many encryption algorithms have been proposed to ensure the security of the transmission process. However, in the field of image encryption, it has never been a recognized algorithm that can encrypt images according to their own characteristics. The chaotic system is favoured by many scholars because of its good uncertainty, unpredictability and non-repeatability. Among the chaotic systems, low-dimensional chaotic systems like [1], [2], [3], [4] are simple and efficient. But their chaotic range is small, the parameters are few, and they are easier to predict [5], [6], [7], [8]. Therefore, on the basis of the original one-dimensional chaotic system, Zhou et al. proposed to connect two one-dimensional chaotic systems in series to iteratively generate a cascaded chaotic system in [9]. First, the output value of the first chaotic system is used as the input value of the second chaotic system, and then the output value of the second chaotic system is fed back to the first chaotic system as its input value. Compared with ordinary low-dimensional chaotic systems, cascade mapping has more obvious chaotic characteristics, and the chaotic trajectory is more difficult to predict. Taking this as a theoretical basis, WU and HUA et al. proposed some cascaded maps: 2D-HSM, 2D-LSCM, 2D-LASM and 2D-SLMM in [10], [11], [12], [13].

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 173-186, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120715

Even so, low-dimensional maps can no longer meet people's security requirements for chaotic algorithms. On this basis, the high-dimensional chaotic map stands out, it has a larger number of parameters and chaotic range, and also has higher security. To make high-dimensional chaotic maps as efficient as possible while having strong encryption performance, many proposals and improvements for high-dimensional chaotic systems have emerged. [14] improved the traditional Baptista algorithm, combined with the hash algorithm and the cyclic shift function, and proposed a one-time encryption algorithm. [15] proposed a new Logistic dynamic linear and nonlinear hybrid coupled mapping lattice. The image encryption algorithm obtained by random encryption has high sensitivity. A new CGCML custom global coupling map is proposed to encrypt color images in [16]. In addition to improving the chaotic system itself, it is also possible to combine the chaotic system with other technologies, and often get good encryption effects, such as DNA technology, quantum technology and Fourier transform. In 2000, Gehani et al first proposed an algorithm to combine DNA coding with chaotic systems in [17]. The color image is encrypted with DNA technology. Dynamic DNA coding is used to improve the security of the encryption algorithm in [18]. Since three-dimensional and lower-dimensional chaotic systems only have a positive Lyapunov exponent, hyperchaotic systems are proposed to improve the performance of chaotic systems in [19]. Hyperchaotic systems have two or more positive Lyapunov exponents. Therefore, hyperchaotic systems have more complex dynamics and are difficulty to be deciphered. However it has higher time complexity. Therefore, in the field of chaos research, how to design a chaotic system with both high efficiency and security is still a research hotspot.

In this paper, a series of three-dimensional chaotic systems are improved, and a cosine excitation function is added to one of the controllers, and a series of new three-dimensional chaotic systems are obtained. Comparing the Lyapunov exponents before and after improvement, and analyze the dynamic characteristics of one of the chaotic systems. The chaotic sequence is generated by the improved chaotic system, and the encryption key is obtained after passing through. This paper uses an improved zigzag transform combined with line shifting to scramble the plain image. Then use the traditional IDEA algorithm to diffuse the image. Finally, the performance of the proposed encryption algorithm is analyzed, and the algorithm is simulated and tested through different performance indicators.

The section 2 of the full text introduces the improved method of 3D chaotic system. And analyzing its chaotic performance analysis. The section 3 applies the image encryption algorithm designed for 3D chaotic system. The section 4 analyzes the performance of the image encryption algorithm through simulation experiments. The section 5 summarizes the full text.

2. INTRODUCTION TO CHAOS SYSTEM

In [20], Sprott summarized some chaotic systems in 1994. In this paper, the A-S system is improved. The cosine excitation function is introduced and the system control parameters are added, so that the output result of the chaotic system has wide-area convergence. In the new chaotic system, the cosine excitation function plays a role in influencing the transition process, and the characteristics of the original three-dimensional chaotic system still play a major role in the influence of the new chaotic system. Table 1 shows the comparison of the Lyapunov exponents of the better performing systems before and after improvement with the same coefficients. It can be seen from Table 1 that the improved method proposed in this paper has a good enhancement effect on chaotic systems. Therefore, the improved method is feasible.

Table 1. Comparison of chaotic systems

Case	Equation	Improved equation	Lyapunov exponent	New Lyapunov exponent
A	$\begin{cases}\dot{x}=ay\\ \dot{y}=-bx+cyz\\ \dot{z}=d-ey^2\end{cases}$	$\begin{cases}\dot{x}=ay+r\cos(\omega t)\\ \dot{y}=-bx+cyz\\ \dot{z}=d-ey^2\end{cases}$	0.020 0 -0.012	0.044 0 -0.010
B	$\begin{cases}\dot{x}=ayz\\ \dot{y}=bx-cy\\ \dot{z}=d-exy\end{cases}$	$\begin{cases}\dot{x}=ayz+r\cos(\omega t)\\ \dot{y}=bx-cy\\ \dot{z}=d-exy\end{cases}$	0.442 0 -1.019	0.471 0 -0.432
C	$\begin{cases}\dot{x}=ayz\\ \dot{y}=bx-cy\\ \dot{z}=d-ex^2\end{cases}$	$\begin{cases}\dot{x}=ayz+r\cos(\omega t)\\ \dot{y}=bx-cy\\ \dot{z}=d-ex^2\end{cases}$	0.051 0 -0.426	0.240 0 -0.640
D	$\begin{cases}\dot{x}=-ay\\ \dot{y}=bx+cz\\ \dot{z}=dxz+3y^2\end{cases}$	$\begin{cases}\dot{x}=-ay+r\cos(\omega t)\\ \dot{y}=bx+cz\\ \dot{z}=dxz+3y^2\end{cases}$	2.744 0 -8.579	3.422 0 -14.145
E	$\begin{cases}\dot{x}=ayz\\ \dot{y}=bx^2-cy\\ \dot{z}=d-4x\end{cases}$	$\begin{cases}\dot{x}=ayz+r\cos(\omega t)\\ \dot{y}=bx^2-cy\\ \dot{z}=d-4x\end{cases}$	0.076 0 -0.688	0.685 0 -0.455
F	$\begin{cases}\dot{x}=-ay+bz^2\\ \dot{y}=cx+0.5y\\ \dot{z}=dx-ez\end{cases}$	$\begin{cases}\dot{x}=-ay+bz^2+r\cos(\omega t)\\ \dot{y}=cx+0.5y\\ \dot{z}=dx-ez\end{cases}$	0.319 0 -0.828	0.500 0 -0.825
G	$\begin{cases}\dot{x}=0.4x+az\\ \dot{y}=bxz-cy\\ \dot{z}=-dx+ey\end{cases}$	$\begin{cases}\dot{x}=0.4x+az+r\cos(\omega t)\\ \dot{y}=bxz-cy\\ \dot{z}=-dx+ey\end{cases}$	0.202 0 -0.939	0.267 0 -0.896
H	$\begin{cases}\dot{x}=-ay+bz^2\\ \dot{y}=cx+0.5y\\ \dot{z}=dx-ez\end{cases}$	$\begin{cases}\dot{x}=-ay+bz^2+r\cos(\omega t)\\ \dot{y}=cx+0.5y\\ \dot{z}=dx-ez\end{cases}$	0.306 0 -0.764	0.828 0 -0.323
K	$\begin{cases}\dot{x}=axy-bz\\ \dot{y}=cx-dy\\ \dot{z}=ex+0.3z\end{cases}$	$\begin{cases}\dot{x}=axy-bz+r\cos(\omega t)\\ \dot{y}=cx-dy\\ \dot{z}=ex+0.3z\end{cases}$	0.079 0 -0.548	0.119 0 -0.515
L	$\begin{cases}\dot{x}=ay+3.9z\\ \dot{y}=0.9x^2-by\\ \dot{z}=c-dx\end{cases}$	$\begin{cases}\dot{x}=ay+3.9z+r\cos(\omega t)\\ \dot{y}=0.9x^2-by\\ \dot{z}=c-dx\end{cases}$	0.243 0 -4.581	0.597 0 -4.498
N	$\begin{cases}\dot{x}=-2y\\ \dot{y}=ax+bz^2\\ \dot{z}=c+dy-2z\end{cases}$	$\begin{cases}\dot{x}=-2y+r\cos(\omega t)\\ \dot{y}=ax+bz^2\\ \dot{z}=c+dy-2z\end{cases}$	0.164 0 -1.912	0.437 0 -1.895
Q	$\begin{cases}\dot{x}=-az\\ \dot{y}=bx-cy\\ \dot{z}=3.1x+dy^2+0.5z\end{cases}$	$\begin{cases}\dot{x}=-az+r\cos(\omega t)\\ \dot{y}=bx-cy\\ \dot{z}=3.1x+dy^2+0.5z\end{cases}$	0.485 0 -0.928	0.487 0 -0.937
R	$\begin{cases}\dot{x}=0.9-ay\\ \dot{y}=0.4+bz\\ \dot{z}=cxy-dz\end{cases}$	$\begin{cases}\dot{x}=0.9-ay+r\cos(\omega t)\\ \dot{y}=0.4+bz\\ \dot{z}=cxy-dz\end{cases}$	0.311 0 -0.988	0.326 0 0.987

2.1. Trajectory diagram of chaotic system and Lyapunov exponent

In this paper, the improved A system is used to analyze and generate encrypted sequences.

$$\begin{cases} \dot{x} = ay + r\cos\omega t \\ \dot{y} = -bx + cyz \\ \dot{z} = d - ey^2 \end{cases} \tag{1}$$

When the initial value of the system (x, y, z) is (1, 1, 1), the system parameters a=5; b=2; c=1; d=10; e=15; r=-1; ω=1 , the chaotic trajectory of the system is shown in Figure 1.

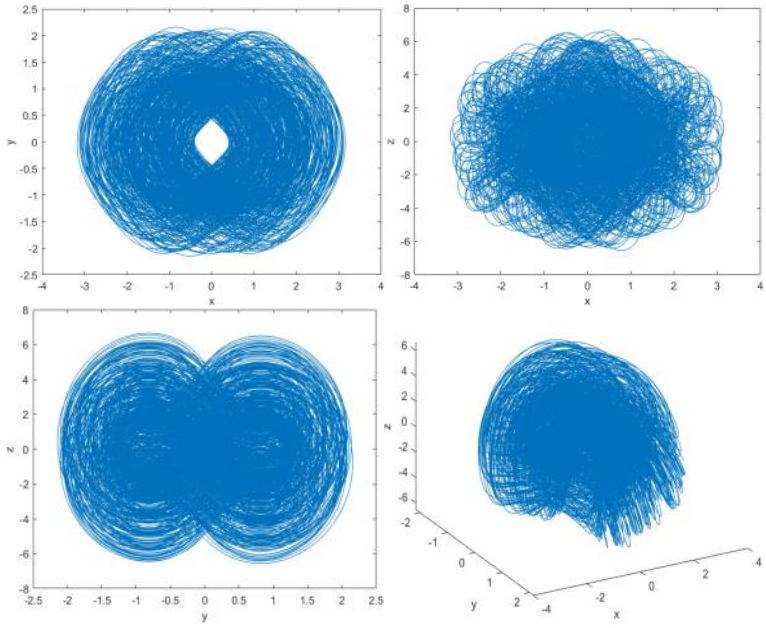


Figure 1. Trajectory diagram of chaotic system

The Lyapunov exponent can be used to analyze the characteristics of the chaotic system. The Lyapunov exponent spectrum of the improved A system is shown in Figure 2. It can be seen from the figure that one of the exponents is always greater than 0. This shows that the chaotic system is chaotic at this time.

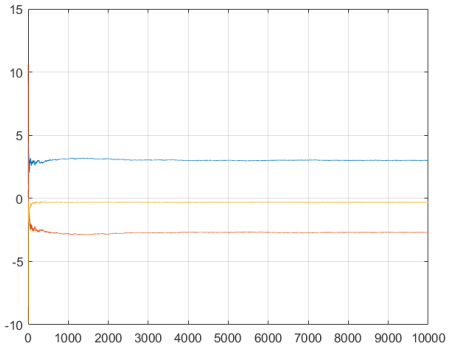


Figure 2. Lyapunov exponent

2.2. Bifurcation diagram

For a chaotic system to exhibit good chaotic dynamics, appropriate system parameters are also required. When the initial values of x, y, and z are (0, 0, 0), draw a bifurcation diagram about parameters a, b, c and d, as shown in Figure 3. It can be seen from the figure that the system has a large parameter range.

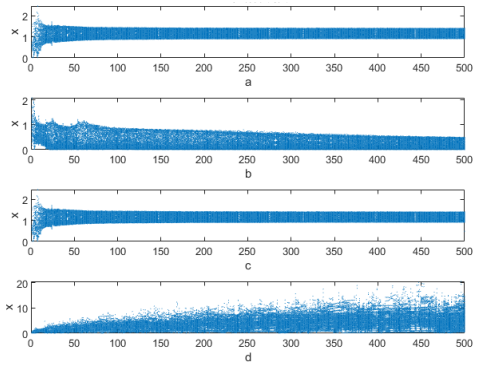


Figure 3. Bifurcation diagram of parameters

2.2. Balance point analysis

Taking system A as an example. Let rcos (ωt)=N and independent of x, y, and z. The system equations are shown below.

$$\begin{cases} \dot{x} = ay + N \\ \dot{y} = -bx + cyz \\ \dot{z} = d - ey^2 \end{cases} \tag{2}$$

Let the right-hand side of the equation be zero:

$$\begin{cases} ay + N = 0 \\ -bx + cyz = 0 \\ d - ey^2 = 0 \end{cases} \tag{3}$$

Solving Eq.3 to get the system equilibrium point (0, -N/a, 0) or (0, $\sqrt{d/e}$, 0), and the Jacobian matrix at the equilibrium point is:

$$J = \begin{bmatrix} 0 & \mathrm{a} & 0 \\ -\mathrm{b} & 0 & \mathrm{c}\sqrt{\mathrm{d}/\mathrm{e}} \\ 0 & -2\sqrt{\mathrm{ed}} & 0 \end{bmatrix} \tag{4}$$

Getting the eigenmatrix of the system:

$$J - \lambda E = \begin{bmatrix} -\lambda & a & 0 \\ -b & -\lambda & c\sqrt{d/e} \\ 0 & -2\sqrt{ed} & -\lambda \end{bmatrix} \tag{5}$$

Let the system characteristic matrix |J-λE|=0. Its expression is shown in Eq.6.

$$-\lambda 3-(2cd+ab)\lambda=0 \tag{6}$$

Solving the characteristic Eq.6 to get three eigenvalues: $\lambda_1=0$, $\lambda_2=\sqrt{2cd+ab}\,i$, $\lambda_3=-\sqrt{2cd+ab}\,i$. All eigenvalues have non-positive real parts. According to the Lyapunov stability method, the system is asymptotically stable at the equilibrium point.

3. IMAGE ENCRYPTION ALGORITHMS

3.1. Generate chaotic sequence

(1) The hash value of the plain image is calculated by the SHA256 algorithm and converted to decimal output. Due to the precision problem of MATLAB, the first 45 bits are selected to obtain three values between 0 and 1, which are used as the initial values x_0, y_0, and z_0 of the chaotic system.

(2) Enter the given system initial values x0, y0, z0, and let the step size l=0.001. Generating three chaotic sequences through the system function ode45, and processing the three chaotic sequences respectively:

$$Xi=\mathrm{mod}(\mathrm{floor}((xi(1001:\mathrm{floor}(M\times N/2)+1000))\times 10\text{^}15),256),\quad i=1,2,3 \tag{7}$$

After processing by Eq.7, we get three integer sequences X1, X2 and X3.

3.2. Image Encryption Algorithm Description

Chaos-based encryption algorithms are generally composed of two ways of permutation and diffusion. In this paper, the image is scrambled by zigzag transformation and line shift, and the whole encryption process is completed through the designed diffusion algorithm.

3.3. Zigzag

The elements in the two-dimensional matrix are traversed from the upper left corner as shown in Figure 4 according to the "zigzag" trajectory, and then the traversed elements are stored in one-dimensional, and then the one-dimensional matrix is converted into a two-dimensional matrix to obtain A transformed new 2D matrix.

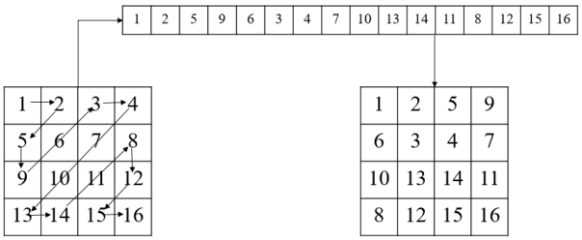


Figure 4. Zigzag transform

However, it can be seen from the above figure that no matter how many times this transformation is repeated, the positions of the first two digits and the last two digits of the matrix have not changed. In the actual encryption, it may leave an opportunity for attackers. Therefore, we

change the starting position of the zigzag, so that the positions of all elements in the matrix change as much as possible.

Figure 5 shows the traversal order of the improved Zigzag. The steps of the scrambling algorithm using Zigzag are as follows:

(1) Read the grayscale image P and the number of iterations n.
(2) From the lower left corner of the two-dimensional image matrix, traverse the entire two-dimensional matrix in the order marked in the figure, and place the traversed elements in a one-dimensional matrix in turn.
(3) Reshape the obtained one-dimensional matrix into a two-dimensional matrix P′, let P=P′.
(4) Repeat steps (2) and (3) until n times are completed.
(5) The image P after the disturbance is output.

In the end, each element in the new matrix P we get is completely different from its position in the original matrix.

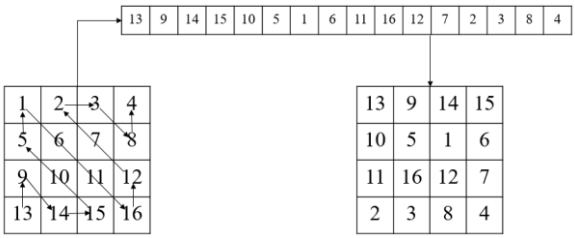


Figure 5. Improved Zigzag transform

We call the matrix to be zigzag transformed as P=a(i, j)M×N, where a(i, j) is the value of the matrix at row i and column j, and M and N are the number of rows and columns of the matrix, respectively number. z is the position of the element in the two-dimensional matrix in the one-dimensional matrix obtained after the transformation.

After comparison, the period of zigzag transformation is much larger than that of traditional Arnold scrambling, and at most four zigzag transformations, its scrambling effect will be better than Arnold scrambling. Meanwhile, the time complexity of the zigzag transformation is $\theta(n^2)$. It can be seen that the improved zigzag transformation algorithm is simple to implement, faster and of better quality.

3.4. row shift

To enhance the scrambling effect of the image, a set of line shifts is added after the zigzag transformation. The first line of the image does not move, and from the second line, the elements of each line are shifted to the right by i-1 bits, where i is the number of lines.

3.5. Diffusion algorithm

The diffusion algorithm proposed in this paper is inspired by the IDEA block encryption algorithm proposed and improved by X.J.Lai and Massey. IDEA algorithm is proposed on the basis of DES algorithm, which is closer to triple DES. This paper simplifies a part of the algorithm, and the execution order of the proposed diffusion encryption algorithm is:

(1) Convert the two-dimensional image matrix P into a one-bit matrix, and divide it into two groups P1 and P2 of equal length.
(2) XOR P1 with the processed chaotic sequence X1.
(3) Add P2 to the processed chaotic sequence X2 and take the modulo.
(4) Add the results of steps (1) and (2) and take the modulo.
(5) XOR the result of step (1) with the processed chaotic sequence X3.
(6) Concatenate the results of steps (4) and (5) to obtain a one-dimensional matrix with a length of M×N, and reshape it into a two-dimensional matrix to obtain the diffused cipher image.

4. PERFORMANCE ANALYSIS

In this section, a series of simulation experiments will be conducted to test the algorithm using several different methods. Three images of Lena, Baboon and Pepper are selected as test images.

4.1. Histogram analysis

Due to the particularity of the image, the pixels of the plain image are unevenly distributed, which makes the image easy to be cracked when subjected to statistical analysis attacks. The histogram can clearly show the distribution of image pixel values. This section analyzes the statistical characteristics of the histograms of plain images and cipher images.
The histograms of Lena's plain and cipher images are shown in Figure 6.

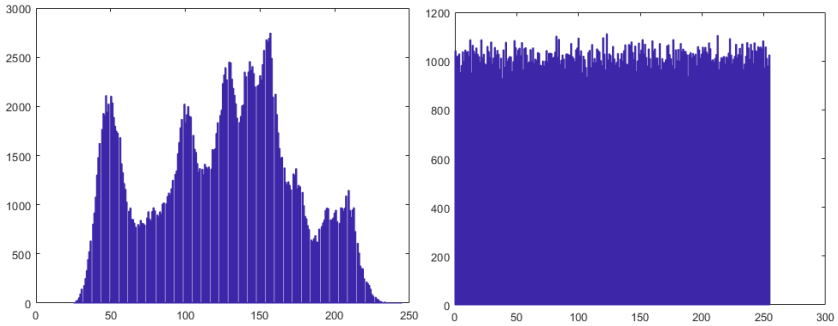


Figure 6. Plain image histogram (left) and cipher image histogram (right) of Lena

4.2. Correlation analysis

Generally speaking, plain images have strong correlations in four directions: horizontal, vertical, positive diagonal, and anti-diagonal, while cipher images have weak correlations in all directions. This paper selects 5000 pixels in Lena image, Baboon image and Pepper image respectively, and calculates the pixel correlation coefficient r in four directions. The closer r is to 1, the higher the correlation between pixels; the closer to 0, the lower the correlation.

The experimental results are shown in Table 2. Figure 11 shows the pixel correlation graphs of Lena plain image and cipher image in horizontal, vertical, positive and anti-diagonal directions.

Table 2. Correlation coefficients

Image		**Horizontal direction**	**vertical direction**	**Diagonal direction**	**Anti-angle direction**
Lena	Plain image	0.9867	0.9734	0.9625	0.9698
	Cipher image	-0.0075	-0.0071	-3.8565×10-4	0.0035

Baboon	Plain image	0.7542	0.8638	0.7050	0.7091
	Cipher image	0.0049	-0.0105	-0.0101	0.0151
Pepper	Plain image	0.9771	0.99781	0.9622	0.9680
	Cipher image	-0.0010	0.0072	-0.0042	0.0164

From the table, we can see that the difference is highly correlated with the pixels of the plain image, and the connection between the adjacent pixels of the cipher image has been reduced to almost non-existent in the encryption process.

4.3. Key space

The key space of the algorithm proposed in this paper is 2^{161}, and the key length of the encryption algorithm with fast encryption speed is at least 2^{128}, hence the key space of the algorithm in this paper can effectively resist the exhaustive attack.

4.4. NPCR and UACI

In image encryption, there are often cases where the difference between two images cannot be observed by the naked eye. The two values NPCR and UACI are generally used to quantify the difference between images. The two images of the same size to be compared are marked as P1 and P2 respectively. The meaning and calculation method of NPCR and UACI are briefly introduced below.

NPCR: Compare whether the values of the pixels at the same position of the two images are the same. The ratio of the number of different pixels to all the pixels is the value of NPCR, and the calculation formula is shown in formula (8) and formula (9).

$$NPCR(P_1, P_2) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|Sign(P_1(i,j) - P_2(i,j))\right| \times 100\% \tag{8}$$

$$Sign(x) = \begin{cases} 1, & x>0 \\ 0, & x=0 \\ -1, & x<0 \end{cases} \tag{9}$$

For two random images, the probability of the same pixel at any position is 1/256, and the probability of difference is 255/256. Therefore, the theoretical expectation of NPCR is about $255/256 \approx 99.6094\%$.

UACI: NPCR measures the number of different values of pixels in two random images, but does not show the degree of difference in the values of pixels in the two images, so UACI is introduced to supplement. It describes the average value of the difference between the values of all pixel points in the same position of the two images to be compared and the ratio of the maximum difference value (255). Its calculation formula is shown in formula (10).

$$UACI(P_1, P_2) = \frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}\frac{\left|P_1(i,j) - P_2(i,j)\right|}{255-0} \times 100\% \tag{10}$$

The expected UACI of two random images is calculated to be $257/768 \approx 33.4635\%$.

Through the values of NPCR and UACI, the degree of difference between the two images can be known. The larger the value, the greater the difference between the two images.

4.5. Key sensitivity analysis

Key sensitivity analysis refers to the difference analysis of the cipher image obtained by encrypting the same plain image with two keys with little difference. Since the chaotic system is very sensitive to the change of the initial value, the purpose of testing the sensitivity of the key is achieved by changing only a small initial value of the chaotic system.

To increase the size of an initial value by 10-15 to generate a new set of chaotic sequences. Encrypt the same plain image through two chaotic sequences respectively. Analyzing the difference between the two cipher images obtained, and calculate the value of NPCR and UACI. The three grayscale images of Lena, Baboon and Pepper are tested respectively. The test results are shown in Table 3.

Table 3. NPCR and UACI for key sensitivity

Image		**Lena**	**Baboon**	**Pepper**	**Expectation**
Proposed	NPCR	99.5899%	99.6094%	99.6014%	99.6094%
	UACI	33.4469%	33.3903%	33.4592%	33.4635%
Ref.[21]	NPCR	99.565%	99.572%	/	99.6094%
	UACI	33.450%	33.448%	/	33.4635%
Ref.[22]	NPCR	99.6002%	99.5903%	99.6112%	99.6094%
	UACI	33.5079%	33.5281%	33.5265%	33.4635%

As can be seen from the table, the key sensitivity of the encryption algorithm is very close to the expectation. The performance on Baboon and Pepper are also better than the algorithms in [21] and [22].

4.6. Plaintext sensitivity analysis

Plaintext sensitivity refers to the difference between the contrasting cipher images when two images with very little difference are encrypted with the same key. If the difference between the two images is large, the plaintext sensitivity of the algorithm is high; otherwise, the plaintext sensitivity is poor.

Therefore, change the value of a random pixel point, and analyze the difference of the cipher image after encryption respectively. Table 4 is the NPCR and UACI values of the images tested. It can be concluded from the table that the algorithm has better plaintext sensitivity.

Table 4. NPCR and UACI for plaintext sensitivity

Image	**Lena**	**Baboon**	**Pepper**	**Expectation**
NPCR	99.5823%	99.6086%	99.5949%	99.6094%
UACI	33.5124%	33.5196%	33.3958%	33.4635%

4.7. Noise attack

Noise is manifested as irrelevant and abrupt pixels in the image, and the noise generated by different methods is also different. Figure 7 shows the noise in the simulated channel, and the decrypted image after adding four kinds of noise to the Lena cipher image. After adding Poisson noise with variance of 0.01, multiplicative noise with variance of 0.04, Gaussian noise with variance of 0.01 and variance of 0.1 salt-and-pepper noise to the cipher image, it is decrypted.

Figure 7. Decrypted image of cipher image after adding noise

It can be seen from the four images that after adding different degrees of noise, there is still a good decryption effect. Even if there are errors in the values of some pixel points, the approximate image can still be seen. It shows that the encryption algorithm has a certain anti-interference ability in the face of noise attack, and can restore the original image better in the face of interference, which provides a certain reliability guarantee for transmission in the channel.

4.8. Shear attack

When transmitting in the channel, not only will it face noise attacks, but sometimes part of the image will be lost. At this time, the encryption algorithm needs to decrypt the part of the cipher image that has been transmitted to obtain a clear plain image as much as possible. Figure 8 shows the cipher images and the decrypted images after 1/8 cutting, 1/4 cutting and 1/2 cutting of the Lena cipher image respectively.

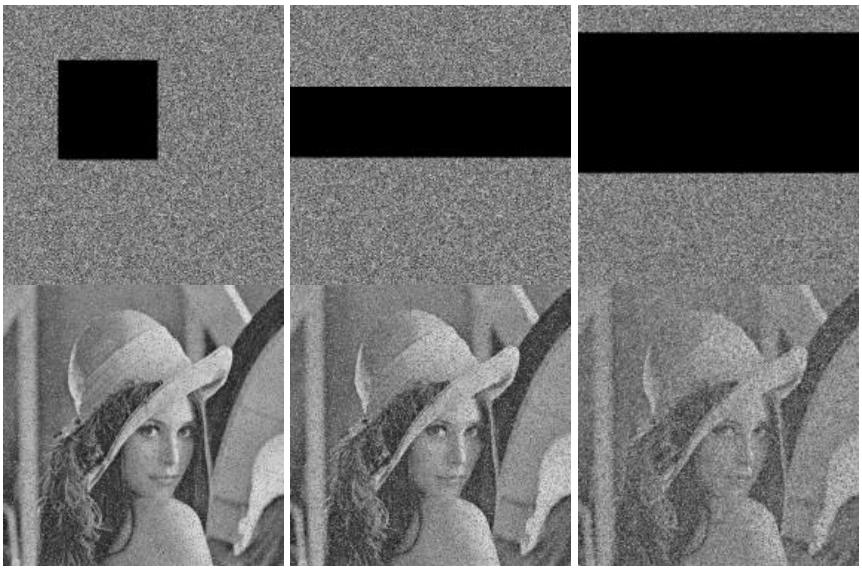


Figure 8. Cut cipher image and its decrypted image

It can be seen from the above figure that as the cut-out part becomes larger and larger, the decrypted image is distorted to a certain extent. Although the decrypted image is not clear enough, the basic outline of the image can still be seen, and sometimes the algorithm proposed in this paper can cope well with the need for real-time transmission. Therefore, the algorithm proposed in this paper can still give a more complete image even if there is data loss.

4.9. Information entropy analysis

The information entropy was drawn and summarized by Shannon in 1948 from the concept of thermal entropy in thermodynamics, which refers to the uncertainty of information and the level of information value. The information entropy of low information degree will be high, and the information entropy of high information degree will be low. The formula of information entropy is shown in Eq.11.

$$H = -\sum_{i=0}^{L} p(i)\log_2 p(i) \tag{11}$$

Among them, L is the number of gray levels of the image (L=256 in this paper). p(i) represents the probability of occurrence of gray value i. In this paper, the theoretical value of information entropy H is 8. Table 4 shows the information entropy of the plain images of Lena, Baboon, and Pepper and their cipher images.

Table 5. Information entropy

Image	**Lena**	**Baboon**	**Pepper**
Plain image	7.4478	7.3579	7.5943
Cipher image	7.9993	7.9994	7.9993
Ref.[21]	7.9993	7.9992	/
Ref.[22]	7.9979	7.9971	7.9974

It can be seen from the table that the information entropy of the image encrypted by the algorithm proposed in this paper is relatively high. The cipher image contains less information, so the algorithm proposed in this paper has better encryption effect.

5. CONCLUSION

In this paper, an improved method for 3D chaotic system by adding cosine excitation function is proposed. The chaotic trajectory diagram, Lyapunov exponent, time series diagram, system bifurcation diagram and equilibrium point analysis of the improved A system all show that the system has obvious dynamic characteristics and good chaos. The proposed scrambling and diffusion algorithm has made some changes on the original basis, which ensures the security and improves the efficiency. The final algorithm test results show that the algorithm can resist some common security attacks, and the value of pixel points is related to more pixels as much as possible, which has better robustness. Therefore, the encryption algorithm proposed in this paper can play a better role in the process of image security communication.

REFERENCES

[1] Zhou Y , Long B , Chen C . A new 1D chaotic sys-tem for image encryption[J]. Signal Processing, 2014, 97(apr.):172-182.

[2] Liu, Wenhao, Sun, et al. A fast image encryption algori-thm based on chaotic map.[J]. Optics & Lasers in Engi-neering, 2016, 84:26-36.
[3] Hui W A , Di X A , Xin C B , et al. Cryptanalysis- and enhancements of image encryption using combinat-ion of the 1D chaotic map - ScienceDirect[J]. Signal P-rocessing, 2018, 144:444-452.
[4] Abd, El-Latif, Ahmed, et al. A novel image encryption scheme based on substitution-permutation network and chaos[J]. Signal Processing: The Official Publication of the European Association for Signal Processing (EURA-SIP), 2016, 128:155-170.
[5] Zhou Y , Bao L , Chen C L P . A new 1D chaotic system for image encryption[J]. Signal Processing, 2014, 97:172–182.
[6] Arroyo D , Rhouma R , Alvarez G , et al. On the security of a new image encryption scheme based on chaotic map lattices[J]. Chaos An Interdisciplinary Jour-nal of Nonlinear Science, 2008, 18(3):033118-113.
[7] Papadopoulos H-E , Wornell G-W . Maximum-likelihoo-d estimation of a class of chaotic signals[J]. IEEE Tran-sactions on Information Theory, 2002, 41(1):312-317.
[8] Wu X , Hu H , Zhang B . Parameter estimation only from the symbolic sequences generated by chaos system[J]. Chaos Solitons & Fractals, 2004, 22(2):359-366.
[9] Zhou Y , Hua Z , Pun C-M , et al. Cascade Chaotic System With Applications[J]. IEEE Transactions on Cyb-ernetics, 2015:2001.
[10] Wu J , Liao X , Bo Y . Image encryption using 2D Hénon-Sine map and DNA approach[J]. Signal Process-ing, 2018, 153:11-23.
[11] Hua Z , Fan J , Xu B , et al. 2D Logistic-Sine-Co-upling Map for Image Encryption[J]. Signal Processing, 2018, 149.
[12] Hua Z , Zhou Y . Image encryption using 2D Logistic-adjusted-Sine map[J]. Information Sciences, 2016, 339.
[13] Hua Z , Zhou Y , Pun C M , et al. 2D Sine Logistic modulation map for image encryption[J]. Information e-nces, 2015, 297:80-94.
[14] Wang X A , Zhu X A , Wu X B , et al. Image encryption algorithm based on multiple mixed hash functions and cyclic shift - ScienceDirect[J]. Optics and Lasers in Engineering, 2018, 107:370-379.
[15] Wang X , Yang J , Guan N . High-sensitivity image encryption algorithm with random cross diffusion based on dynamically random coupled map lattice model[J]. Chaos Solitons & Fractals, 2021, 143(5):110582.
[16] Wang X , Qin X , Liu C . Color image encryption algorithm based on customized globally coupled map lattices[J]. Multimedia tools and applications, 2019.
[17] Gehani A , Labean T , Reif J . DNA-based cryptography[M]. 2000.
[18] Chai X , Fu X , Gan Z , et al. A color image cryptosystem based on dynamic DNA encryption and chaos[J]. Signal Processing, 2019, 155(FEB.):44-62.
[19] Nguyen N T , Bui T Q , Gagnon G , et al. Designing a Pseudo-Random Bit Generator with a Novel 5D-Hyperchaotic System[J]. IEEE Transactions on Industrial Electronics, 2021, PP(99):1-1.
[20] Sprott, J. Some simple chaotic flows[J]. Physical revie-w. E, Statistical physics, plasmas, fluids, and related int-erdisciplinary topics, 1994, 50(2):R647-R650.
[21] Wang X , Zhao H , Feng L , et al. High-sensitivity image encryption algorithm with random diffusion based on dynamic-coupled map lattices[J]. Optics and Lasers in Engineering, 2019, 122(Nov.):225-238.
[22] Wu J , Liao X , Bo Y . Image encryption using 2D Hénon-Sine map and DNA approach[J]. Signal Processing, 2018, 153:11-23.

AUTHORS

Zhenzhou GUO born in 1977, MS, lecturer, his research interest includes Chaos encryption.

Xintong LI, born in 1997, MS candidate, her research interest includes Chaos encryption.

A POSE-BASED IMAGE SEARCHING USING COMPUTER VISION AND POST-ESTIMATE

Hang Wang[1] and Yu Sun[2]

[1]University of California—Berkeley, Berkeley, USA
[2]Cal Poly Pomona, USA

ABSTRACT

As the cost of human forces increases, people in some careers, like the artists, may find some difficulties when models are needed in the processes of making art. Obviously, one alternative solution is to find pictures online, however, when some specific poses are needed, they may also find some difficulties to describe them. This paper develops an application to search for pictures with the target pose described by the user's graphical input. In this project, one hundred images describing distinct actions and activities with various poses and view-angles are collected. Mediapipe is then used to analyze those images in a quantitative way. We also embedded a User Interface that allows the user to imitate the intended pose as well as the viewing angles by simply dragging around body joints of the figure. Different sets of feature points and matching algorithms are also tested to find out the best solution.

KEYWORDS

Computer vision, Pose detection, Image searching.

1. INTRODUCTION

The internet nowadays is ever powerful to connect everyone to any open resources that one may desire. Searching engines, for example, give their clients a way to access those resources by string matching. When we trying to search for an image, those images are categorized in a way so that we can easily find one, for example, by simply typing “a running man” and hit the search button. However, when someone is trying to search for a very detailed pose, the user must know how to describe that pose in text, such as “jumping with the left arm up and right arm down” or “sitting with both hands crossed on the leg”, which is annoying and not necessarily returning a promising result. The users oftentimes need to rephrase their words couple times before they give up or a satisfying result returns.

In this project, we define an innovative way of image-searching by the user's own graphical input. The clients of our application can expect a promising searching experience by simply modifying the provided default body-joint figure to imitate the pose in their mind. Such application is even more powerful when high-degree of accuracy is demanded such as “left arm curved 30 degrees”.

Our image-search engine is accomplished by linking the user input with preprocessed dataset using the matching algorithm.

The rest of this paper is structured as follows: Section 2 will detail the multiple challenges faced in this study and how they were overcome; Section 3 will describe the methodology and solution in greater detail; Section 4 details the experiments that were performed in this study, as well as a

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 187-194, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120716

thorough analysis of the results; Section 5 will list any related works that have also been done regarding volunteering; and lastly, Section 6 will conclude the study and state any future work that may be done.

2. CHALLENGES

2.1. Challenge 1: Aligning User Input for prediction in machine learning does not always translate perfectly. (2D input for user prediction does not translate perfectly to a 3D captured image turned 2D (aka a photo))

A three-dimensional space has much more information than a two-dimensional space. In this project, we are focusing on photographs, which are a 3D image flattened to 2D. The user input is therefore also treated as a 2D object, as controlling a 3D model is not only more complex but prone to error. When matching the user input to a predicted image, the process of estimating for a higher dimension has always been a challenge. Due to the complex nature of imaging, oftentimes user input does not perfectly correlate with an image as we are cutting one dimension down. Photographs, while 2D, are reflections of 3D space. We have to do this very precisely (understanding how a 3D object in different angles is displayed in a 2D image) to have a good matching result. To better match the user's input to the image there are a few options. We can either increase how much the user is going to actually put in, give them the ability to add more detail to their data, or limit their options, handling post-processing for them automatically. In this project, we give the users the ability to adjust the length of selected components to more flexibly demonstrate the depth of a three-dimensional image.

2.2. Challenge 2: Selecting a pattern matching model is a difficult process. (testing the models, comparing accuracy, comparing flexibility, etc.)

Selecting a model is essential for every machine learning project, as selecting a model has a huge impact on the accuracy of the result. Each model has its own strengths and weaknesses. Depending on the problem, a model's accuracy and impact can vary wildly. The process of selecting a suitable model for our computer vision project is specifically tough as there are numerous features that we can collect from an image. Because the data and process of image matching is flexible in implementation, we can end up with a lot of different models that work depending on how things are done. Each choice we make can either be an important feature or just some noise depending on the project itself. The standard way of selecting a suitable model for training is to do comparisons among all selected models. Whichever returns the highest accuracy is the one chosen. Often when a project is much simpler, just picking the most commonly accurate models will suffice as well. In this project, to select a good model to match user input to the actual image in the database, we do the following: for every model we are testing them on a wide array of training data and then splitting our tests to have them undergo cross validation. This will give us the average accuracy and success for each model.

2.3. Challenge 3: Finding out which model parameters are important is a tiresome process

There are numerous models in Machine Learning that are developed well. Among those, most models are flexible in terms of the parameters they are taking in, allowing the developers to tune the models to fit their specific needs. A model's parameters often have a direct correlation to its success rate. Selecting the perfect maximum depth for Random Forest or the optimal kernel size for a SVC model is a very time consuming process. The number of features, defining what a pose is for our project, also needs to be considered as the format of our training data also impacts what

model parameters are important. The process of hyper-parameter tuning is an excellent solution to this problem as we test the models with a wide gamut of parameters so as to see what lends itself best to our problem. Some problem solutions, however, do not require modifying model parameters past the default options. In our project, we use hyper-parameter tuning on the model that initially returns the highest general accuracy without any modification.

3. METHODOLOGY/SOLUTION

3.1. Overview of the solution (whole system)

The two most important things are the preprocessed dataset and the matching algorithm. The preprocessed dataset contains images with different poses as well as skeleton drawings. The matching algorithm is a machine learning algorithm that matches the user's skeleton drawing with the skeleton drawings on the preprocessed images. Once we have these two components, the whole system is constructed in the following way: first, a website with a 2D geometric user interface that shapes like an individual person figure is provided for the user to imitate the pose by simply dragging things around. Once the users finish their drawings, they will click on the "Search" button so that a web API function is called to trigger a HTTP request that is sent to the web server where the pre-trained model is located. The web server then calls this model which is trained on the dataset with all preprocessed images. The model returns the most close image and returns to the web server. The web server then interprets the request and asks to query the database in our database server where we store and provide access to persistent data [our preprocessed images]. The database server receives the query and returns the image that is asked for to the web server. The web server then construct the response, send it back through HTTP response so that the client browser can render the page to show the related image.

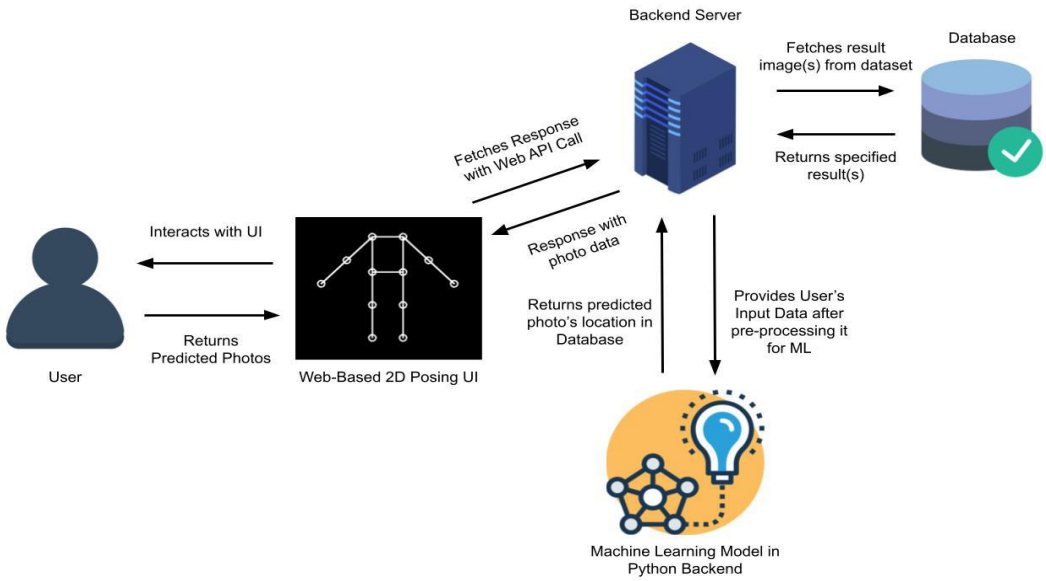


Figure 1. System work flow

3.2. The implementation

When designing our user interface, we created two object classes called Hinge and Joint, representing what they actually mean in the human body. With only one Hinge object originally assigned to the end position of the Joint object instead of both side, we are skipping the problematic case where we have to delete the duplicated Hinge objects at the exactly same location when connecting two Joints together. In our case, when trying to connect two body joints, the Hinge object actually handles the connection through the "add_connection" function. The way of how we achieving this makes much more sense and provides more modularity when we do the experiment. With these two object classes, the full body figure is then created where we assign the default positions of every parts by hard-code, which gives us the following UI:

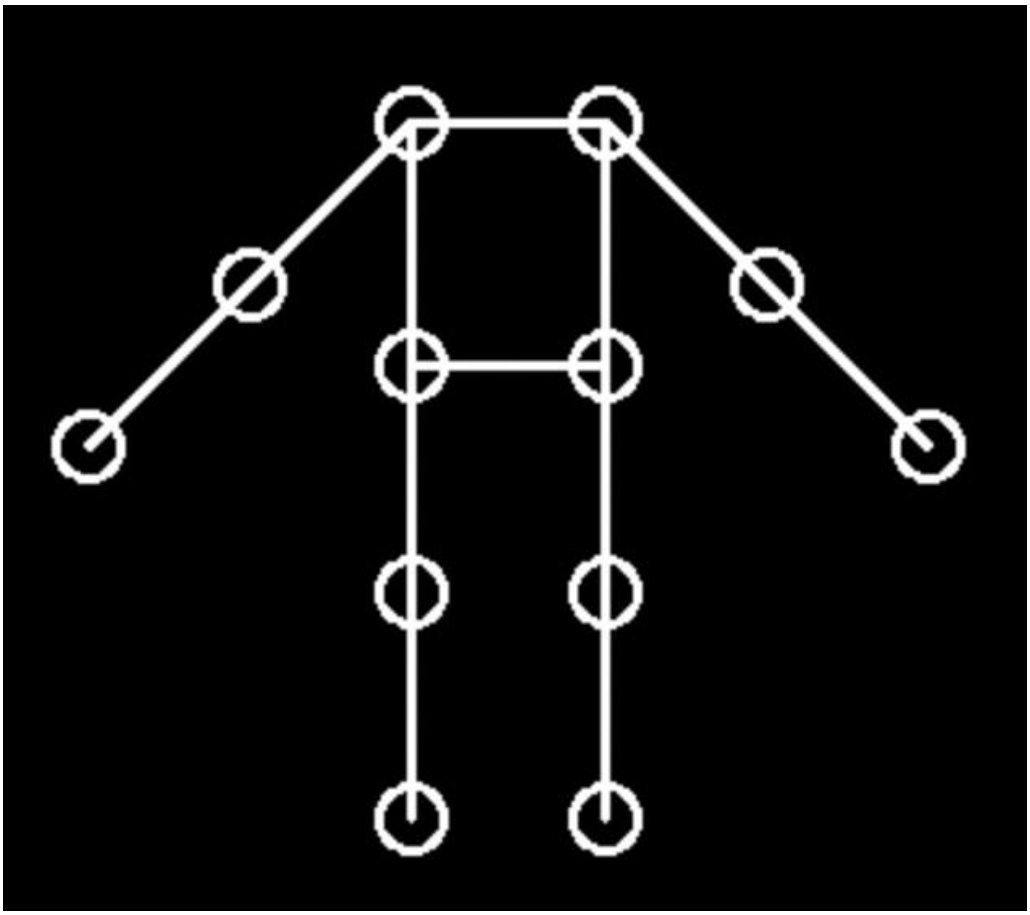


Figure 2. Example user interface

To make our model and dataset work on the website, we use the Python Flask server to organize everything. It takes in the HTTP requests with the drawings, cleans the data, passes it to machine learning, gets responses and then pulls out the correct image from the database.

For the matching algorithms that we are comparing, we used SVM, GaussianNB, Random Forest(maximum depth=2), Random Forest(no maximum depth), and CNNs.

4. EXPERIMENT

4.1. Experiment: Find the best matching algorithm

To find out which machine learning algorithms works the best as our matching algorithm, our group conducted couple experiments with each candidate models testing on different cross validations. The Mediapipe's pose landmark model converts images to skeleton drawings with 32 feature points as shown in the following figure. We would also like to see if our search engine is scalable, so we tested our models on different size of datasets.

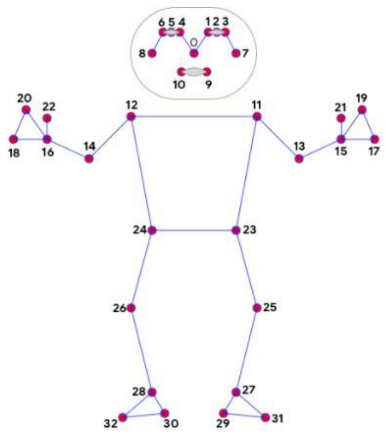


Figure 3. Default pose landmark model

First, we choose 10 images with 5 of them represent running and another 5 of them represent yoga then duplicated the dataset by 4 times in order to do the cross-validation. We used 3-fold cross-validation, SVM model gave the result of being 79.85% accurate, however, GaussianNB model, Random Forest model with maximum depth set to 2, and Random Forest model with unlimited maximum depth all gave 100% accuracy.

Next, we tested those models again on 50 images containing much more distinct activities and poses including sitting, jumping, yoga, running, etc. Again, for cross-validation, we choose to duplicate the dataset by 4 times. We used 3-fold cross-validation, SVM model gave the result of being 72.28% accurate, GaussianNB model and Random Forest model with unlimited maximum depth both gave 100% accuracy, and Random Forest model with maximum depth set to 2 gave us an accuracy of 67.86%.

Finally, with all 100 images, we duplicated the dataset by 8 times so that we can do a 3-fold cross-validation, a 5-fold cross-validation, and a 7-fold cross-validation. For 3-fold cross-validation, SVM model gave the result of being 93.97% accurate, GaussianNB model and Random Forest model with unlimited maximum depth both gave 100% accuracy, and Random Forest model with maximum depth set to 2 gave us an accuracy of 71.66%; For 5-fold cross-validation, SVM model gave the result of being 94.63% accurate, GaussianNB model and Random Forest model with unlimited maximum depth both gave 100% accuracy, and Random Forest model with maximum depth set to 2 gave us an accuracy of 73.45%; For 7-fold cross-validation, SVM model gave the result of being 97.32% accurate, GaussianNB model and Random Forest model with unlimited maximum depth both gave 100% accuracy, and Random Forest model with maximum depth set to 2 gave us an accuracy of 76.34%.

4.2. Analysis

As shown in the experiments, SVM model is the only which returns a good result while not being overfitted. However, we would ask our self a question, does overfitting really matters in our matching problem? The answer is no, since we only care about how accurately a matching model can be to link a user input to one of the images in our database. We finally choose the SVM model as our matching algorithm in our application.

5. RELATED WORK

In the paper "Image Matching Algorithm based on Feature-point and DAISY Descriptor", their team is focusing on an image matching algorithm where they combined SURF algorithm which is based on partial features and DAISY descriptor which is based on the principal direction. Their proposed algorithm, while almost maintaining the same computing speed, improves the SURF algorithm's ability on image rotation. In our research, while trying to increase the confidence level of matching the image where people are standing on a surface with slope or in different rotation to the user inputs, we can also use the idea of how to improve the accuracy of matching images in rotating condition. The main difference between our focus and their group's focus is that we are matching feature points to the images while they are trying to match images to images. One of the strengths of our research is our project's unique approach to matching, centered around matching to a particular image just given some raw data.

In the paper "Classification of yoga pose using machine learning techniques", their team is focusing on using pose detection techniques to identify the posture and thus the accuracy of yoga poses. They used four machine learning algorithms to classify the yoga asana for Sun salutations set of postures as well as a real-time skeleton drawing using pose estimate technique. In our project, instead of real-time skeleton pose drawing with video-capturing, we asked the user to move the body joints we provided as the user interface. Instead of some limited set of yoga poses, we allowed users to search for a large number of poses in their daily routine.

In the paper "A fingerprint recognition algorithm using phase-based image matching for low-quality fingerprints", their team is focusing on finding a fingerprint matching algorithm for low-quality fingerprints. The fact that the fingerprint condition, whether environmental or personal causes, can highly affect the recognition process is the most changeling problem. They suggested a phase-based image matching algorithm where they use the phase components in 2D discrete Fourier transforms of fingerprint images to try to achieve a better performance. In our project, while trying to match the user input to the image, we can also use the same idea to handle the case where some images didn't catch the whole body. The main difference between our focus and their group's focus is that we are matching posing features to the image while they are matching sub-optimal fingerprints to the complete fingerprints. One of the strengths of our research is that we are extracting and modifying the important feature points while matching to the dataset.

6. CONCLUSION AND FUTURE WORK

Our final application is able to return the target image on the right in the dataset given the graphical input on the right as shown in the following example.

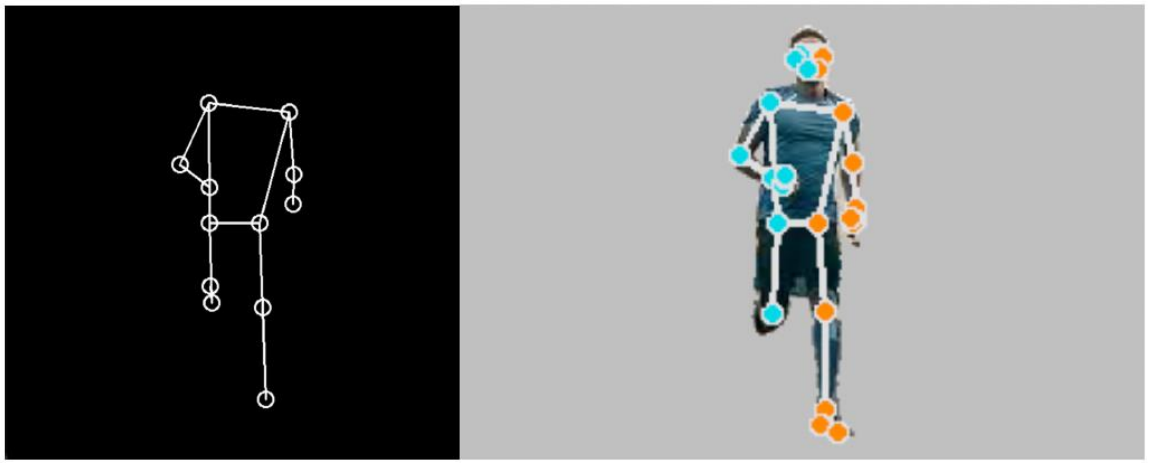


Figure 4. Example result

However, every project has limitations. In our project, some images that represent different actions/poses are probably matched to the same or very similar 2D body-joint data. When the user gives his body-joint inputs in our UI, he might, as a consequence, get many images that he doesn't expect. Another potential issue is that a lot of the practicability depends on a huge variety of images in the dataset which we may lack. A dataset that contains many different poses and actions is highly preferred while also being highly time-consuming. As a small research team, we may not have the ability to collect such a huge valid dataset.

In terms of how the way of our data-processing defines, we may expect that when turning a 3D image to a 2D body-joint data, different images may share the same or very similar data while the images themselves may vary widely. To address this issue in the future, a 3D body-joint data processing is preferred. However, as the accuracy of a 3D body-joint data processing highly depends on the pose-detection model we have, a substitute solution may be to add an extra field in our input data that will let the user select the body's facing position. In terms of how we can approach generating a larger dataset while maintaining its accuracy, having more people to work on it is obviously preferred. As the current size we have, one way of doing this is to set a separate program to automatically collect images from the web.

REFERENCES

[1] Li Li, (2014) Image Matching Algorithm based on Feature-point and DAISY Descriptor, Journal of Multimedia, Volume 9, NO. 6

[2] J. Palanimeera & K. Ponmozhi, (2021) Classification of yoga pose using machine learning techniques, Materials Today: Proceedings, Volume 37, Part 2, Pages 2930-2933

[3] K. Ito, A. Morita, T. Aoki, T. Higuchi, H. Nakajima & K. Kobayashi, (2005) A fingerprint recognition algorithm using phase-based image matching for low-quality fingerprints, IEEE International Conference on Image Processing 2005, pp. II-33

[4] F. Lv and R. Nevatia, (2007) Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching, 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8.

[5] J. Kilner, J. Guillemaut and A. Hilton, (2009) 3D action matching with key-pose detection, 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 1-8.

[6] R. M. Haralick, H. Joo, C. Lee, X. Zhuang, V. G. Vaidya and M. B. Kim, (1989) Pose estimation from corresponding point data, IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 6, pp. 1426-1446.

[7] Y. Yang and D. Ramanan, (2011) Articulated pose estimation with flexible mixtures-of-parts, CVPR 2011, pp. 1385-1392.

[8] V. Belagiannis and A. Zisserman, (2017) Recurrent Human Pose Estimation, 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 468-475.
[9] X. Zhu and D. Ramanan, (2012) Face detection, pose estimation, and landmark localization in the wild, 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2879-2886.
[10] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, (2021) OpenPose: Realtime Multi- Person 2D Pose Estimation Using Part Affinity Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186.
[11] J. Shotton et al., (2013) Efficient Human Pose Estimation from Single Depth Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 12, pp. 2821-2840.
[12] G. Borgefors, (1988) Hierarchical chamfer matching: a parametric edge matching algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 6, pp. 849-865.
[13] S. Gold and A. Rangarajan, (1996) A graduated assignment algorithm for graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 4, pp. 377-388.
[14] Bin Luo and E. R. Hancock, (2001) Structural graph matching using the EM algorithm and singular value decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1120-1136.
[15] L. P. Cordella, P. Foggia, C. Sansone, F. Tortorella and M. Vento, (1998) Graph matching: a fast algorithm and its evaluation, Proceedings. Fourteenth International Conference on Pattern Recognition, pp. 1582-1584 vol.2.
[16] S. Umeyama, (1988) An eigendecomposition approach to weighted graph matching problems, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 5, pp. 695-703.
[17] D.K. Isenor, S.G. Zaky, (1986) Fingerprint identification using graph matching, Pattern Recognition,Volume 19, Issue 2, pp. 113-122.
[18] Gerard Sanromà, René Alquézar, Francesc Serratosa, (2012) A new graph matching method for point-set correspondence using the EM algorithm and Softassign, Computer Vision and Image Understanding, Volume 116, Issue 2, pp. 292-304.

HOW TO ENHANCE THE SHARING OF CYBER INCIDENT INFORMATION VIA FINE-GRAINED ACCESS CONTROL

Jarno Salonen[1], Tatu Niskanen[2] and Pia Raitio[3]

[1]VTT Technical Research Centre of Finland, Tampere, Finland
[2]University of Jyväskylä, Jyväskylä, Finland
[3]Finnish Transport Infrastructure Agency, Helsinki, Finland

ABSTRACT

Industry 4.0 and the ongoing digital transformation along with a large number interconnected machines and devices increase the role of cybersecurity, cyber incident handling and incident response in the factories of the future (FoF). Cyber incident information sharing plays a major role when we need to formulate situational pictures about FoF operations and environment, and respond to cybersecurity threats related to e.g. the implementation of novel technologies. Sharing of incident information has a major drawback since it may reveal too much about the attack target, e.g. in the case of legacy systems and therefore restrictions may apply. We have developed a proof-of-concept service that combines access control and encryption of data at high granularity and a mechanism for requesting access to restricted cyber incident information. The objective was to demonstrate how access to restricted incident data fields could be managed in a fine-grained manner to enhance information sharing.

KEYWORDS

Incident Management, Visualisation, Cybersecurity, Information Sharing.

1. INTRODUCTION

Industry 4.0 promotes the use of emerging technologies such as Internet of Things (IoT), Big Data, artificial intelligence (AI) and cloud computing, supported by cyber-physical systems, as the design to achieve what they call smart factories or factories of the future (FoF). One common characteristic of Industry 4.0 is to decentralise production systems [1] and allow controlling, monitoring, adaptation, and optimisation to be done in real time, based on the large amounts of data available in the factory environment that feeds the use of machine learning (ML) techniques. This so called fourth technological revolution is expected to bring significant gains in productivity, resource savings and lower maintenance costs, as machines will have all the information necessary to operate more efficiently, adaptable and keep up with any fluctuations in demand. Devezas et al. (2017) summarised Industry 4.0 to comprise strong customisation of products under high flexibility mass production that require the introduction of methods for self-organised systems to establish a suitable link between the real and virtual worlds [2].

The rapid human evolution and the search for strategies that facilitate our daily lives, imply a close synergy with technology that must adapt to the user needs. In this context, HMI (Human Machine Interface) interfaces emerge, responsible for human-machine interconnection. HMI, as the name already indicates, consists of software and hardware that allow an operator (human) interact with a controller (machine) [3]. The term "machine" is defined by Merriam-Webster as "*mechanically,*

David C. Wyld et al. (Eds): DMML, SEAS, ADCO, NLPI, SP, BDBS, CMCA, CSITEC - 2022
pp. 195-208, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120717

electrically or electronically operated device for performing a task" [4]. In other words, a machine is a device that has the function of transmitting or modifying energy to perform a certain task that assists the human being. Technically, you could apply the acronym HMI to any kind of screen or visual display that someone uses to interact with a device, but in general, we use it to describe screens used in industrial environments. HMIs display real-time data and allow the user to control the equipment using a graphical user interface. In the industrial environment, the HMI can take many forms. It can be a standalone screen, a panel connected to other equipment, or even a tablet computer. Despite of its looks, the primary purpose of HMI is to allow users to view data about the device operations and control it. Operators can use an HMI, for example, to see which conveyor belts are running or adjust the temperature of an industrial water tank.

In terms similar to human-human interaction, humans as rational beings also need to communicate and interact with technological systems, hence the emergence of artificial intelligence (AI). Artificial intelligence, which we consider today as a domain of knowledge, and define as the type of intelligence similar to that of the human but displayed through mechanisms or software. Thus, the "intelligent agent", which is also often called as autonomous agent, is a system that perceives the entire environment (internal and external factors) via sensors and acts rationally (generally makes decisions to maximise its success) upon that environment with its effectors. Weiss et al. (1999) define intelligent agent being capable of flexible autonomous actions in order to meet its design objectives and where flexible means reactivity, pro-activeness and social ability [5]. The goal of an intelligent system is to perform functions, with the ability to establish logical reasoning (by means of established rules), recognising patterns, learning (acquiring future knowledge future knowledge based on mistakes), and inference (apply reasoning to everyday situations). In the caseof a vehicle, the human being adopts the role of driver and performs the interaction with the entiresystem present in the vehicle. Therefore, this system must be prepared to interact with the driver through the most appropriate means of communication (e.g. haptics, images, and/or sound), and must be able to recognise the interactions used by the driver. The previous can be applied to the factory of the future where the human operator is the driver and the industrial system or multiple (interconnected) systems represent the vehicle(s).

All technological development, for example the fourth technological revolution with its decentralisation, breakthrough technologies such as Industrial IoT and AI, and the convergence of human and machine interaction have also weaknesses. In this case, one major weakness is cybersecurity. The digital transformation of the FoF involves a growing number of interconnecteddevices, which we need to protect from hackers and other malicious third parties. Traditionally theoperational technology (OT) systems of factories were isolated from the (public) internet, but thisseparation is no longer valid. This is due to e.g. robots being maintained remotely while membersof the supply network gain access, not only to the factory information technology (IT), but also toOT systems in order to, e.g. collect orders, acquire production specifications or update their own production information to the system. In addition, customers or service providers may access thesesystems directly in order to obtain production schedules or even optimise factory operations via the use of existing production data.

In order to maintain an adequate level of cybersecurity within the FoF during the aforementioned development, we need to update our cybersecurity strategy to meet the requirements of the FoF. The strategy consists of among others the following items:

- FoF cyber risk and threat management
- Development and implementation of cybersecurity policies for access and trust management within the FoF
- Evaluation of new technologies (e.g. 5G and AI) potential cybersecurity vulnerabilities
- FoF monitoring and incident response (intrusion detection/prevention systems, SIEM,

SOC, etc.)
- Organisation of cybersecurity training and awareness activities
- Planning of decision-aided or autonomous remediation and recovery of assets in case of a cyber incident

This article focuses on the topic of FoF monitoring and incident handling and response by introducing a proof-of-concept for a fine-grained cyber incident information sharing which is also the basis of cyber situational awareness for the FoF. Our research questions are the following:

1. How can we enhance incident information sharing between organisations?
2. How should we modify the IODEF data format to enable encryption and decryption?
3. How to request and enable the sharing of sensitive information in the cyber-incident dashboard (service)?

We developed this proof-of-concept as part of an ongoing Horizon 2020 project. Due to the nature of the PoC and the schedule of the project, the concept does not involve end-user-evaluation or the security and vulnerability testing of the service. This is because the objective was to discover new and innovative possibilities for sharing incident information in a fine-grained manner.

The article is structured as follows. We begin by describing the terminology and existing research behind our concept. Then we describe our methodology related to the IODEF data format update, visualisation aspects and the development of the event generator, which we used to generate random IODEF data content. In chapter 4, we describe the developed incident manager dashboardand some of its functionalities. Finally, we discuss about the findings during our research and futureR&D topics before we conclude the article.

2. THEORETICAL PERSPECTIVE

NIST defines cybersecurity as "*the process of protecting information by preventing, detecting and responding to attacks*" [6]. The prevention, detection and response to cyber-attacks has been studied quite extensively during the two decades. For example, Lee (2015) has studied the prevention of cyber-attacks and their response and listed both technical (firewalls, routers, filtering, etc.) and non-technical (training, information sharing, awareness-raising, etc.) means for different kind of cyber-attacks ranging from Distributed Denial of Service (DDoS) attacks to phishing and data breaches [7]. Indre and Lemnaru (2016) propose a solution against malware and intrusion attacks that is based on intrusion detection and prevention systems [8]. Kholidy has studied autonomous mitigation of cyber risks in cyber-physical systems (CPS) [9] and Zhou et al. have studied a multi-agent-based hierarchical detection and mitigation of cyber-attacks in smart grids [10]. In addition to the scientific contribution, the U.S. National Institute of Standards and Technology (NIST) has introduced its own cybersecurity framework to create a baseline and a toolbox to protect the government, critical infrastructure and individual companies against cyber-attacks, which includes among others learning material, implementation guidance and even models for evaluating the security maturity of the organisation [11].

Two key measures for improving the cybersecurity of the FoF or any other organisation is situational awareness and information sharing. According to Gilson (1995) the term "situational awareness" was first identified by Oswald Boelke during World War I who described it as "*the importance of gaining an awareness of the enemy before the enemy gained a similar awareness, and devised methods for accomplishing this*" [12]. The CNSS Glossary defines situational awareness as "*the perception of an enterprise's security posture and its threat environment*" [13]. Information sharing has been studied by, e.g. Harrison and White (2012), and Steenbruggen and Nijkamp (2012) both highlighting the importance of information sharing and proposing methods

for enabling communities to detect cyber incidents via the use of shared security information [14] and stating that public organisations often have information that is valuable to each other's operations [15]. In general, legislation such as the General Data Protection Regulation in Europe requires cyber incident reporting in the case of a breach of personal data, but it does not cover reporting of other incidents [16]. The same applies to the U.S. and China where the respective privacy acts have been enforced to protect the personal data of citizens, but other incident reporting is often voluntary though strongly recommended. There are some exceptions though. For example, national legislation often forces critical infrastructure providers to report any kind of incidents that influence their operations to the supervising authority.

In cybersecurity, situational awareness and information sharing often conflict with each other. Even though sharing cyber incident information may help others prevent the same kind of cyber-attack, it may reveal other vulnerabilities that may cause new and even larger attacks against the party sharing the incident information. For example, the description of the attack or its countermeasures may reveal that the target was a specific legacy system, which makes it possible for the malicious third party to focus on its known vulnerabilities. This makes the supervising authorities' role as the middle-man quite difficult. After all, they receive the information from all (cyber) incidents, but they necessarily cannot share the information to other relevant parties even within the same sector. The risks of sharing cyber incident information have been studied by Mallinder and Drabwell (2014) [17] and Albakri et al. (2018) [18] while Lawton and Parker (2002) focused on barriers in incident reporting of a healthcare system more than a decade earlier [19]. This information sharing issue has been partially solved by establishing sector-based Information Sharing and Analysis Centers (ISACs). According to the definition by ENISA (2018) ISACs are "*trusted entities to foster information sharing and good practices about physical and cyber threats and mitigation*" [20]. Most countries have their own ISAC networks focused especially in critical infrastructure sectors. In Finland, there are eleven different ISAC groups ranging from food production and distribution to water management and hosted by the National Cyber Security Centre [21].

3. DESIGN AND PLANNING

In this section, we describe the methodology related to the IODEF data format update. Then we will describe the development of the event generator, which we used to generate random IODEF data content so that we can test the incident dashboard with data that resembles actual incident reports and encrypt some of the field contents. Finally we will focus on the main topic, i.e. the incident manager dashboard and its visualisation.

3.1. Applying fine-grained access control to the IODEF data format

In order to demonstrate and visualise fine-grained access control (FGAC) in action, we created an applicable use case. The selected demonstrative use case is the visualisation of fine-grained access control in cybersecurity incident documents, also known as reports, and more specifically, incident documents in the incident object description exchange format (IODEF). IODEF is a data format used to describe cybersecurity incident information for exchange between computer security incident response teams (CSIRTs). IODEF was first defined in RFC5070 [22] and it was later updated to version 2 in RFC7970 [23]. The format is used in multiple software and other real-world applications, such as in the security information and event management system SIEM.

According to the RFC7970, IODEF has the following main information fields that are shown in table 1.

Table 1. IODEF file fields [23]

Field	Multiplicity	Description
IncidentID	One	An incident identification number assigned to this incidentby the CSIRT who creates the IODEF document.
AlternativeID	Zero or one	The incident ID numbers used by other CSIRTs to refer tothe incident described in the document.
RelatedActivity	Zero or one	The ID numbers of the incidents linked to the onedescribed in this document.
DetectTime	Zero or one	Time at which the incident was detected for the first time.
StartTime	Zero or one	Time at which the incident started.
EndTime	Zero or one	Time at which the incident ended.
ReportTime	One	Time at which the incident was reported.
Description	Zero or more	Non-formatted textual description of the event.
Assessment	One or more	A characterisation of the incident impact.
Method	Zero or more	Techniques used by the intruder during the incident.
Contact	One or more	Contact information for the groups involved in theincident.
EventData	Zero or more	Description of the events involving the incident.
History	Zero or more	A log, of the events or the notable actions which tookplace during the incident management.
AdditionalData	Zero or more	Mechanism which extends the data model.

In addition to these main fields, IODEF contains multiple other less used and optional fields. For the scope of this demo, we deal only with the main fields. We can apply fine-grained access control to the IODEF documents when incident information needs to be shared with parties in an organised manner, but the parties have different privileges to the incident information. Fine-grained access control allows the encrypted documents to be uploaded to a central location so the information can be shared as fast as possible. This way the party in charge of distributing the incident information does not need to alter the information in the document for every receiver. Fine-grained access control can also be used to dynamically encrypt individual tags in an IODEF document, based on the privileges of the reader.

To demonstrate the possibilities of fine-grained access control in IODEF documents, a dashboard was created. The dashboard is a demo of a cybersecurity incident management cloud service, where the incident information can be uploaded in the IODEF format. The users of the service can then view these documents and only access the information they are authorised to. The dashboard is developed with Angular 12 typescript framework. It has simple demonstrative login and logout functionalities and allows you to login as four different users in order to demonstrate the fine-grained access control with IODEF documents. You can then view two different documents as each individual user to see the varying access rights in action.

3.2. Event generator and API

For simulation purposes we developed an event generator to allow gathering information about the functionalities and response of the system. By firing stochastic incident events that simulate the incidents happening within a real productive scenario, the event generator works as the base engine of the simulation environment, producing plausible incident data that can visualised across all the dashboards of different actors and locations.

Each of the fired incident events follows the Incident Object Description Exchange Format (IODEF) RFC5070 Standard, which is XML-based data with a human readable structure. Each file contains at least one incident entity, with each entity being comprised by the fields listed in table 1.

The fine-grained access control functionality manages the ability of certain actors to visualise these fields in a different way, and acts in conjunction with the cryptographic keys. The event generator randomly selects a subset of tags within the IODEF document, and encrypts them with a randomly generated key and a certified and secure cryptographic algorithm. The individual teams or users can then request this key, by providing the necessary data such as the unique user identification (UID) and the incident identifier. These encrypted tags contain the FGAC keyword as an XML attribute with the team/user identification numbers, which can access and decrypt them. This way, when the document is shared, the sensitive tags are shared encrypted and, unless using a key, cannot be viewed in plain text.

The event generator system was designed to be object-oriented with a strict segregation of responsibilities between different components. It contains a central component, IODEF generator, responsible for generating the documents and assign them to different products. It also contains an XML processor component, responsible for building valid XML documents from the generated incidents, sharing them with the team. The event generator overall architecture is displayed in figure 1 below.

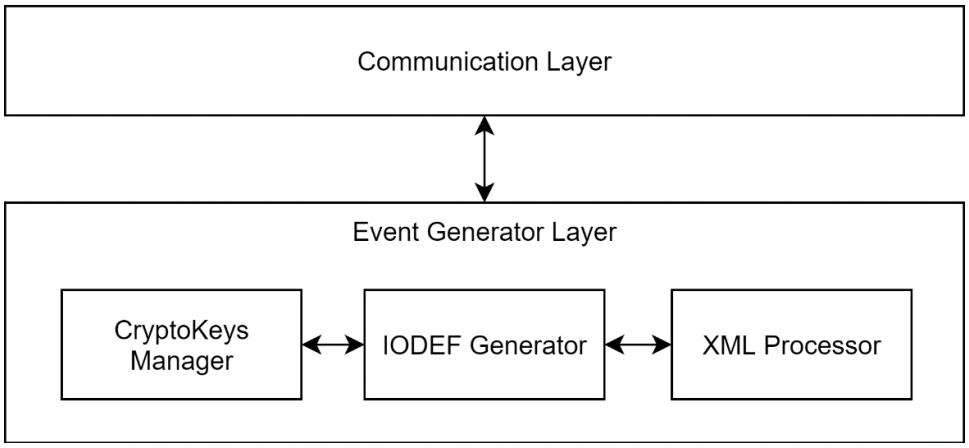


Figure 1. Event generator architecture

Since the dashboards are to be web-based, the event generator has a communications layer using HTTPS protocol, in order to quickly and effectively exchange data with the different dashboards.

The communication methods of the event generator are displayed in table 2 below.

Table 2. Event generator communication methods

Method	**Description**	**Parameters**
/InitSimulation	Initializes/resets the simulation.	None
/GetSimulationData	Returns all the generated IODEF documentsand organizations in both JSON and XML format	None

/GetKey	Returns the key used to decrypt elements thatare encrypted within an FGAC tag	Owner, IncidentID and TagName

The InitSimulation method allows the user to reset the simulation, creating new organisations and IODEF documents. GetSimulationData returns both the list of generated organisations and IODEF, together with the assignments between both entities, in JSON and XML (IODEF standard) format. Lastly, GetKey is used for requesting a cryptographic key that can decrypt a set of sensitive information of an IODEF document. This key is only returned if both the tag and the organisation that is requesting the key match, avoiding giving keys to unwanted entities.

Since this article focuses on the incident dashboard and the visualisation of incident data, we will consider the event generator as a black box with the desired output that the incident manager receives and processes from this point forward and therefore we will not describe the event generator any more.

3.3. Visualisation of incidents in a fine-grained manner

One task of our project, the collaborative monitoring and response task aims to design and implement a collaborative security operations center (SOC) for distributed operations. It collects all relevant information, then orchestrates, analyses it, and finally responds to security incidents in a timely manner. The objective is to enable collaborative incident response on distributed manufacturing environments, shorten the decision making, response and recovery time, and optimise attack and response resources (costs). Information sharing is seen as a key element to support vulnerability, threat and incident management, and thus helping organisations to prepare and prevent cyberattacks. However, it is primarily based on trust, which is also seen as the most significant barrier to organisational information sharing as e.g. Albakri et al. (2018) state in their research [18].

In this work we have focused on boosting the trust in incident information sharing by using fine-grained access control. The aim is to reveal only pre-allowed pieces of incident data to different users or groups. For example, an automotive factory that has been attacked may allow that all incident information may be shared with their named SOC operator who works in close collaboration with them, but other SOC operators do not see the target IP address which was under attack at the factory. The factory may also allow sharing of the incident details to other automotive factories in Europe, which might be targets to a similar attack as it seemed to be targeted to the automotive sector. However the factory allows sharing of only the very basic details to other European factory operators, not revealing, e.g. what was the targeted factory and when did the attack happen. By revealing too much information, it might be possible to count one plus one and figure out what was the targeted organisation or/and system. By sharing too much information about the incident, the information might end up into the hands of a malicious third party who may get insights of how successful their attack was. Thus, using fine-grained access control we try to create a PoC of a system that would boost organisations into sharing incident information in a trustable manner.

In order to visualise the incidents in a fine-grained manner, the IODEF documents are enhanced with fine-grained access control tags. These tags act as attributes to the fields containing the incident information. The users are then granted access to the encrypted information based on these tags. The users can be saved in a database and given two corresponding tag-attributes. One specifies the clearance group the user belongs to and the other is a specific tag belonging to the user. The responsible SOC can then distribute IODEF information based on the clearance groups or by the specific user. The SOC can achieve this by populating the IODEF document with the corresponding

tag-attributes within the information fields. The Angular front-end then compares these tags within the IODEF document to the ones held by the logged in user. When the tags match, the information in the IODEF-document is displayed to the user. Otherwise a placeholder text "You are not authorised for this information" is shown. In the demo, the information is encrypted if the user has not authorisation to view it. In case the user has no authorisation, the system consist of a functionality for requesting permission (namely a decryption key) to the data from the data owner, which can then be used to decrypt the information.

There is already some research conducted on fine-grained access control systems for XML-documents. For example, Luo et al (2004) have presented the QFilter method for fine-grained run-time access control for XML-documents. In the research it is stated that it is crucial to tailor information in XML-documents for various user and application requirements, preserving confidentiality and efficiency at the same time. Thus, it is critical to enforce access control over XML data to ensure that the users only have access to the portion of the data they are allowed to. The research currently presents different access control methods for XML-documents, but only little to no research has been done on visualisation of fine-grained access control. [24]

When choosing how to visualise encrypted data in a fine-grained manner, it is also important to decide on how to convey the data that is not visible to the viewer. In some contexts, the best approach is not to show the data at all. In other contexts, a better approach is to convey that the data exists but hide it from the unauthorised viewer in other ways. Possible approaches to visualising the hidden data include:

- Blurring, overlining and other visual hiding methods
- Placeholder text, such as "You are not authorised to view this information"
- Not showing the hidden information at all

In this case, our chosen approach was a placeholder text. This is because the document does not contain fields whose existence in itself is classified. In some other fine-grained access control situations, the existence of the hidden items may be information that should not be shared with all parties. The existence of the hidden information may be enough for the intruder or other malicious party to get interested and therefore figure out where to intrude.

In order to be able to visualise data in a fine-grained manner with many different users having access to the same document, it is important to structure the document in a way that supports implementing fine-grained access control. This is where structures such as IODEF are helpful. When the structure of the document is predefined, the programming logic behind the fine-grained access control is easier to implement and maintain. If fine-grained access control is to be implemented in a cloud environment, for example for viewing documents, it should be considered to come up with a predefined and commonly agreed structure for the shared documents. Formats such as IODEF can be adapted to other contexts. For example in the automotive industry, predefined formats such as IODEF can be applied to the manufacturing manuals shared in the cloud. The predefined individual tags or fields allow for easy distribution of access control.

4. IMPLEMENTATION OF THE INCIDENT MANAGER

This section describes the implementation of the incident manager dashboard (service) as well as its functionalities. The incident manager dashboard is a service (frontend application) for viewing incident information that is stored in an external (cloud) repository. The service requirements towards the repository in addition to being the source of incident information (IODEF) data are the

implementation of fine-grained access control (FGAC) tags along with the encryption mechanism of any desired individual incident information field. In addition, the repository should offer the possibility to provide the decryption key to the incident manager dashboard based on a request from the manager.

After launching the service from the browser, the user is shown an empty dashboard with instructions to perform user login. After the user has logged into the service, the user clicks the "Fetch Incidents" button to download incident data from the (cloud) repository, which in this case is represented by the event generator. The figure below (Figure 2) displays the incident manager dashboard after the user has logged in and fetched the incident data. As you can see from the figure, the dashboard screen is divided into two sections. The left half of the dashboard shows a list of incidents and when the user clicks on a specific incident, the detailed incident information is displayed on the right hand side. In this case the user can see six incidents. Please also note that the user has full privileges, i.e. he can see all fields in incident number six.

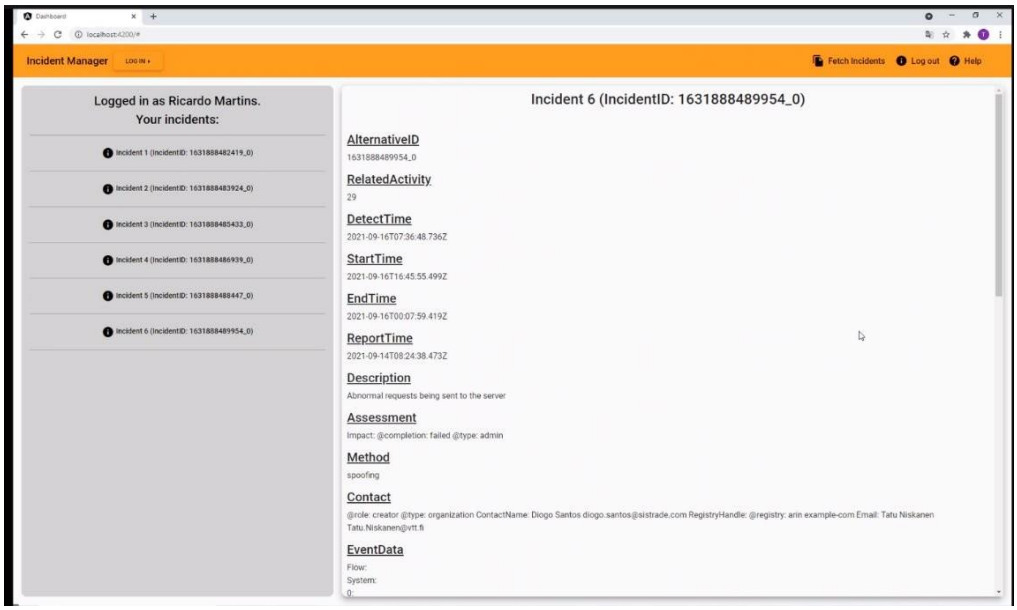


Figure 2. Incident view by user with "all" privileges

The figure (Figure 3) shows the same incident six as before, but with a different user who has a restricted view. Instead of showing the field content, the user sees a placeholder text with "You are not authorized for this information" along with a "Request key" button, which is used for requesting access to that specific field content.

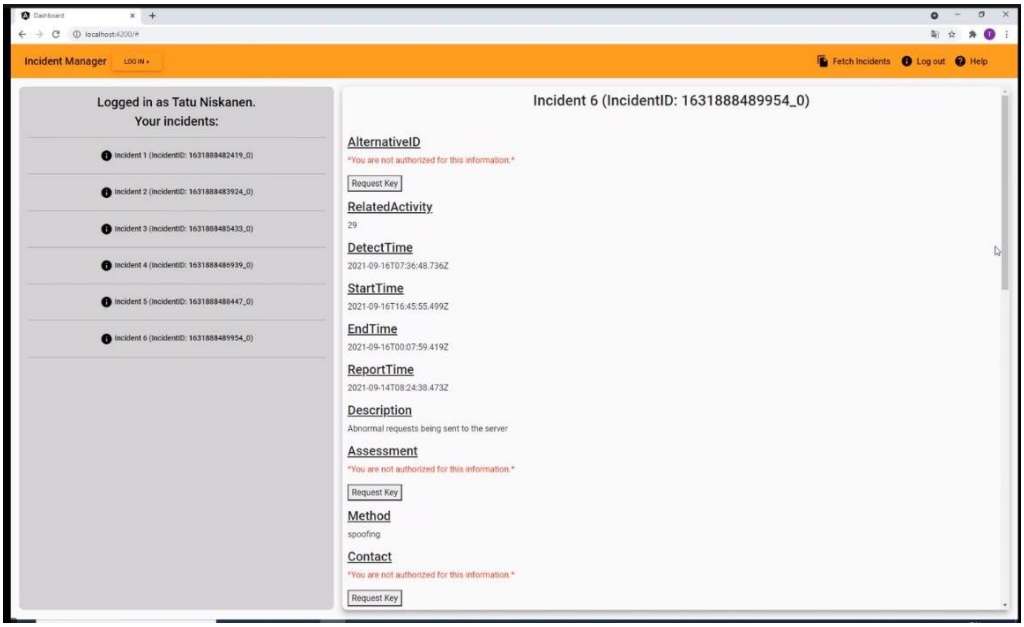


Figure 3. Incident view with restricted access showing both visible and encrypted fields

This functionality highlights the high granularity of the service, i.e. each field has been encrypted with its own key and therefore access is requested for each field separately.

The next figure (Figure 4) shows the incident manager dashboard after the user has clicked the "Request key" button in order to gain access to the encrypted data content.

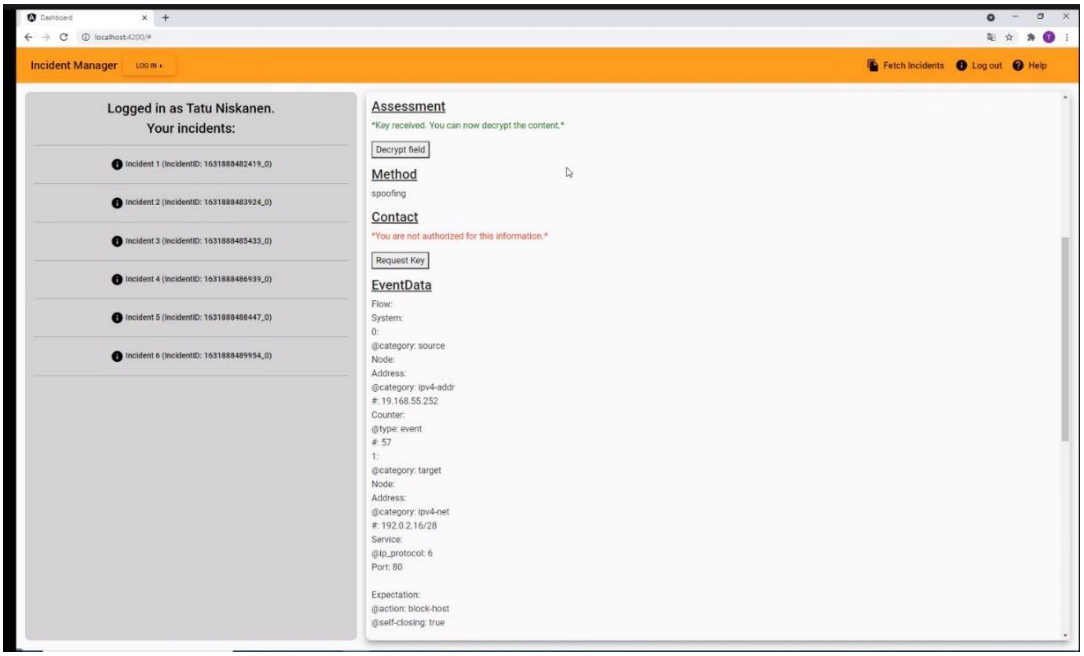


Figure 4. Incident view with decryption key received from the data owner

Clicking the button sends the user and field identifiers to the data owner (which in this case is represented by the event generator) who grants access to the data and therefore sends the decryption

key to the incident manager dashboard. In this case, the user sees a message that the access to the field content has been received and he can then click on the “Decrypt field” button to display the content.

The last figure (Figure 5) displays the incident view with the restricted content now visible to the user. Please note that only the requested field information is shown to the user, i.e. if the user wishes to see the content of other restricted fields, then he should click on the respective “Request key” buttons to request access to them. In other words each incident information field has been encrypted with a unique key and therefore the received decryption key works only for that one specific field. You may also note that the structure of the field content visible in Figure 5 is slightly different from the content in figure 2, i.e. the decrypted data has line spaces between the different words while the user with “all privileges” did not have them (i.e. all field data was in one row). This is mainly due to the technical encryption and decryption process which will be taken care of in the next version of the service.

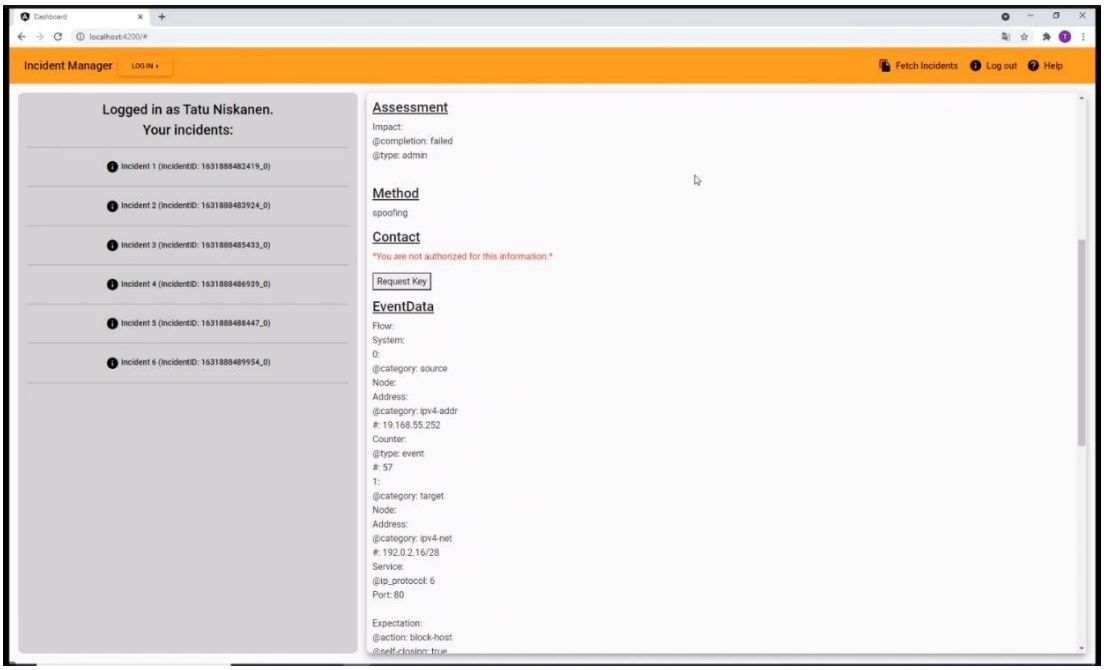


Figure 5. Incident view with decrypted field information visible to the user

5. DISCUSSION AND FUTURE WORK

This section describes the major observations that we discovered during the development of the incident manager dashboard. This list of observations is not an exhaustive one, but rather a review of what kind of functionalities we considered to be useful in future work, i.e. the next version of the service.

The PoC offers only username and password as the only available authentication mechanism, but it would be possible and also highly recommended to use two-factor authentication (2FA) or a similar strong authentication mechanism to identify and authorise the user properly. Since our objective was to demonstrate the incident information sharing functionalities of the dashboard, as well as the visualisation, we decided not to waste time on implementing strong authentication into the service.

The PoC includes only one repository for downloading incident information, but it would be possible to fetch incident information from any available repository included in the list of sources. The "Request key" functionality resulted in automatic granting of access rights from the event generator, i.e. it sends the decryption key to the service immediately. It would be possible to demonstrate an open text for justifying the access request to the incident information owner, but since we had only one repository at use and the user interface of the data owner nor the manual access-granting mechanism was not included in our objectives, they were not implemented. In any case the user interface for the data owner could be implemented either directly to the incident manager dashboard or it could be done e.g. via email. In the latter case the justification message would be sent via email to the owner (using the email listed in the incident contact information) and the message could also include links to approve or reject the request), which might provide added usability to the system.

During the PoC development, we also discussed about the access right (key) validity period. The access right to an individual field could be given for a certain time period, it could be based on the number of times that the user reads the restricted content, or it could also be valid indefinitely. Since this wasn't the objective of this PoC, we decided to grant only one time access rights, i.e. the key was valid only for that specific login session, which was enough for our demonstration. The decision is often dependent on the data owner and therefore we should perhaps offer multiple options for choosing the validity period in the next version of the incident manager.

Like we described in the introduction section, we did not consider an end-user piloting nor evaluation to the PoC. Neither did we consider conducting security testing to the system or cryptographic analysis to the encryption/decryption functionality. These decisions were made intentionally since the objective of the PoC was to test the possibility of sharing incident information in a novel way.

The future research and development activities would consist of a thorough evaluation of the proposed service. It would consist among others an analysis of the service efficiency and applicability, collection and statistical analysis of the views from end-user evaluations, and a comparative analysis covering the advantages and disadvantages of the service compared to other (existing) incident reporting management tools and services. The evaluation would include an implementation of the service to a real incident management reporting use case with actual incident data. In addition we could test the security of the service and the developed encryption mechanism, and try to find a similar tool or service from another sector/topic in order to compare the evaluation results with each other.

Since the project is very close to its end, we do not have plans to develop the incident manager dashboard further at this stage. However in case we find another suitable project, then we might reconsider creating the next iteration of the service. Due to the current trend of cybersecurity information sharing and this topic being in the focus of the current Horizon Europe Cybersecurity topic calls, future development might be highly possible.

6. CONCLUSIONS

This article describes the research and development of a proof-of-concept incident manager dashboard (service) that combines access control and encryption of data at high granularity, and a mechanism for requesting access to restricted cyber incident information. We claim that the PoC will enhance incident information sharing between organisations since it allows the sharing of incident (IODEF) data while applying access control and encryption in a fine-grained manner to individual fields containing sensitive information in the perspective of the information owner. The access rights are implemented into the IODEF in the form of FGAC tags that are defined separately

for each field, enabling the fine-grained access control and encryption functionality. The service also demonstrates a way to request access to one or more restricted (encrypted) IODEF fields within the incident information, while maintaining the full control of data within its owner. The article also discusses about the observations regarding some missing but perhaps useful functions that could be implemented in next iterations of the PoC.

ACKNOWLEDGEMENTS

This article is based on research and development work conducted together with the Portuguese partner Sistrade in the Secure Collaborative Intelligent Industrial Assets (SeCoIIA) project. SeCoIIA aims at securing the digital transition of manufacturing industry towards more connected, collaborative, flexible and automated production techniques. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871967.

REFERENCES

[1] Meissner, H., Ilsen, R., Aurich, J. (2017). Analysis of Control Architectures in the Context ofIndustry 4.0. Procedia CIRP. 62. 165-169. 10.1016/j.procir.2016.06.113.
[2] Devezas, T., Sarygulov, A. (2017). Industry 4.0. Basel: Springer. 10.1007/978-3-319-49604-7
[3] NIST, S. (2015). 800-82 Rev 2. Guide to industrial control systems (ICS) security. 10.6028/NIST.SP.800-82r2.
[4] Definition of Machine. Available from: https://www.merriam-webster.com/dictionary/machine (Accessed 15.2.2022)
[5] Weiss, G. (Ed.). (1999). Multiagent systems: a modern approach to distributed artificial intelligence. MIT press.
[6] Ross, R., Pillitteri, V., Graubart, R., Bodeau, D., & McQuaid, R. (2019). Developing cyber resilient systems: a systems security engineering approach (No. NIST Special Publication (SP) 800-160 Vol. 2 (Draft)). National Institute of Standards and Technology. 10.6028/NIST.SP.800-160v2r1.
[7] Lee, N. (2015). Cyber attacks, prevention, and countermeasures. In Counterterrorism and Cybersecurity (pp. 249-286). Springer, Cham. 10.1007/978-3-319-17244-6.
[8] Indre, I., Lemnaru, C. (2016). "Detection and prevention system against cyber attacks and botnet malware for information systems and Internet of Things," 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), pp. 175-182, 10.1109/ICCP.2016.7737142.
[9] Kholidy, H. A. (2021). Autonomous mitigation of cyber risks in the Cyber–Physical Systems. Future Generation Computer Systems, 115, 171-187. 10.1016/j.future.2020.09.002.
[10] Zhou, T. L., Xiahou, K. S., Zhang, L. L., & Wu, Q. H. (2021). Multi-agent-based hierarchical detection and mitigation of cyber attacks in power systems. International Journal of Electrical Power & Energy Systems, 125, 106516. 10.1016/j.ijepes.2020.106516.
[11] NIST. 2021. Cybersecurity framework. https://www.nist.gov/cyberframework (Accessed:15.2.2022)
[12] Gilson, R. (1995). Situation awareness — special issue preface. Hum. Factors 37 (1), 3–4.
[13] Dukes, C. W. (2015). Committee on national security systems (CNSS) glossary. CNSSI, Fort 1322 Meade, MD, USA, Tech. Rep, 1323, 1324-1325. https://rmf.org/wp-content/uploads/2017/10/CNSSI-4009.pdf (Accessed: 15.2.2022)
[14] Harrison, K., White, G. (2012). Information sharing requirements and framework needed for community cyber incident detection and response. IEEE Conference on Technologies for Homeland Security (HST) (pp. 463-469). IEEE. 10.1109/THS.2012.6459893.
[15] Steenbruggen, J., Nijkamp, P., Smits, J. M., Mohabir, G. (2012). Traffic incident and disaster management in the Netherlands. Challenges and obstacles in information sharing. Netcom. Réseaux, communication et territoires, (26-3/4), 169-200. 10.4000/netcom.975.
[16] European Commission. (2018) Guidelines on Personal data breach notification under Regulation 2016/679 (wp250rev.01). European Commission / Data protection Newsroom. https://ec.europa.eu/newsroom/article29/items/612052 (Accessed 16.2.2022)
[17] Mallinder, J., & Drabwell, P. (2014). Cyber security: A critical examination of information sharing

versus data sensitivity issues for organisations at risk of cyber attack. Journal of business continuity & emergency planning, 7(2), 103-111.
[18] Albakri, A., Boiten, E., De Lemos, R. (2018). Risks of Sharing Cyber Incident Information. In Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES 2018). Association for Computing Machinery, New York, NY, USA, Article 58, 1–10. 10.1145/3230833.3233284
[19] Lawton, R., & Parker, D. (2002). Barriers to incident reporting in a healthcare system. BMJ Quality & Safety, 11(1), 15-18. 10.1136/qhc.11.1.15
[20] ENISA. (2018). Information Sharing and Analysis Centers (ISACs) - Cooperative models. European Union Agency For Network and Information Security. 10.2824/549292
[21] National Cyber Security Centre. (2022). ISAC information sharing groups. Finnish Transport and Communications Agency. https://www.kyberturvallisuuskeskus.fi/en/our-services/situation-awareness-and-network-management/isac-information-sharing-groups (Accessed 16.2.2022)
[22] Danyliw, R., Meijer, J., Demchenko, Y. (2007). RFC5070 - The Incident Object Description Exchange Format. Internet Engineering Task Force (IETF), Network Working Group. December 2007. https://datatracker.ietf.org/doc/html/rfc5070 (Accessed 10.2.2022)
[23] Danyliw, R. (2016). RFC7970 - The Incident Object Description Exchange Format Version 2. Internet Engineering Task Force (IETF). November 2016. https://datatracker.ietf.org/doc/html/rfc7970 (Accessed 10.2.2022)
[24] Luo, B., Lee, D., Lee, W. C., Liu, P. (2004). QFilter: fine-grained run-time XML access control via NFA-based query rewriting. In Proceedings of the thirteenth ACM international conference on Information and knowledge management (pp. 543-552). 10.1145/1031171.1031273.

AUTHORS

Jarno Salonen is working as a Senior Scientist in the applied cybersecurity team at VTT Technical Research Centre of Finland. He has a professional background of over 20 years in making the digital world a better place for ordinary users especially in the areas of cybersecurity, privacy, resilience and the development of electronic services.

Tatu Niskanen is a cybersecurity oriented graduate student from University of Jyväskylä, Finland. He is currently finishing up his master's studies from Hanyang University, South Korea, where he is staying as an exchange student in the school of engineering. He has worked for VTT Technical Research Centre of Finland in multiple research projects during 2021.

Pia Raitio is currently working as a Senior Officer, Cybersecurity Specialist at Fin- nish Transport Infrastructure Agency, where she recently started after a long career as a cybersecurity researcher and project manager at VTT. Her focus is on ensuring the cybersecurity of critical infrastructures and other ICS/OT-systems, covering both cybersecurity governance of the whole infrastructure as well as the very detailed technical aspects - and everything in between.

AUTHOR INDEX