IJNS
International Journal of Network Security

ISSN 1816-353X (Print)
ISSN 1816-3548 (Online)

Vol. 23, No. 3 (May 2021)

INTERNATIONAL JOURNAL OF NETWORK SECURITY

PUBLISHING OFFICE

Min-Shiang Hwang
Department of Computer Science & Information Engineering, Asia University, Taichung 41354, Taiwan, R.O.C.
Email: mshwang@asia.edu.tw

International Journal of Network Security is published both in traditional paper form (ISSN 1816-353X) and in Internet (ISSN 1816-3548) at http://ijns.jalaxy.com.tw

PUBLISHER: Candy C. H. Lin

Volume: 23, No: 3 (May 1, 2021)

International Journal of Network Security

Research on Detection and Prevention of Mobile Device Botnet in Cloud Service Systems

Hung-Wei Yang[1], Li-Chin Huang[2], and Min-Shiang Hwang[1,3]
(Corresponding author: Min-Shiang Hwang)

Department of Computer Science & Information Engineering, Asia University[1]
500, Lioufeng Rd., Wufeng, Taichung 41354, Taichung, Taiwan, R.O.C.
Department of Information Management, Executive Yuan, Taipei 10058, Taiwan[2]
Department of Medical Research, China Medical University Hospital, China Medical University, Taiwan[3]
Email: mshwang@asia.edu.tw
(Invited Paper; First Online Apr. 26, 2021)

Abstract

Nowadays, Botnets may be the greatest threat on the Internet, whether distributed denial of service attacks (DDoS) or sending Email spam, or stealing private information. Such attacks caused economic losses of up to several billion dollars every year. As the enhanced computing capacity of the mobile devices (smartphones and tablet PCs) and the user population growth of the mobile network, as well as mobile devices usually stores a lot of personal privacy information and used for online micropayments, attacker focused on the mobile platform. However, there was no media event of any related attacks. With the popularity of the mobile network, Mobile Botnet will be a major threat in the future. Therefore, we intend to use the cloud computing platform to establish a Botnet detection and prevention mechanism to help telecommunications service providers (TSP) detect whether the SMSs of users are sent from malware by using cloud computing platforms warning messages to both TSP and user. The mechanism will also learn from user experience to avoid the system make mistakes in judgment. And further, to establish another mechanism to provide mobile users to detect whether the mobile application contains malicious code to prevent users from Mobile Botnet infections in time.

Keywords: Cloud Computing; DDoS; Mobile Botnet; Smartphone; SMS

1 Introduction

In recent years, due to the development of information and communication technology and the popularization of the Internet, using the Internet to search for information, receive various new knowledge, and communicate with relatives and friends can enjoy the convenience of the Internet [13]. But at the same time, it also attracted the attention of some attackers. These attackers developed a set of malicious software for certain computer operating systems or browser vulnerabilities. When users download and use this malware, it appears to be a harmless tool or game software, but these programs will run silently in the background and be controlled by a remote attacker, and become a zombie computer. A botnet refers to a network composed of these zombie computers controlled by computer hackers [16, 18]. When there are major vulnerabilities in the computer operating system or browser, users usually contact the Internet. A malicious attack that harms computer hardware or personal information leakage. However, compared with being infected by worms and viruses, anti-virus software protection will be provided, while bots are good at hiding themselves and operating behind the scenes. For ordinary users, it is not easy to detect that they have become part of a botnet, which facilitates attacks. This person achieved the purpose of the attack. Currently, the biggest threat facing the Internet may be botnets, whether it is a distributed denial-of-service attack (DDoS) or sending spam (SAPM), personal information theft, phishing, and other attacks [2, 11]. Every year, these cybercrimes cause huge economic losses.

The use of the BotNet virus is similar to a Trojan horse program. Still, the Trojan horse will only attack specific targets and will not attack other computers through the computer host where the Trojan horse is implanted. On the contrary, botnets will not only attack other computers but also have the characteristics of worms. It will slowly crawl into cyberspace. Once it encounters a vulnerable computer host, it will launch an attack on its own. As shown in Figure 1, the botnet routing consists of three entities: Bot Herder, Command and Control Server, and Bot.

Hackers use command and control channels (C&C) to remotely control infected hosts and launch network attacks, including stealing private data, spreading spam,

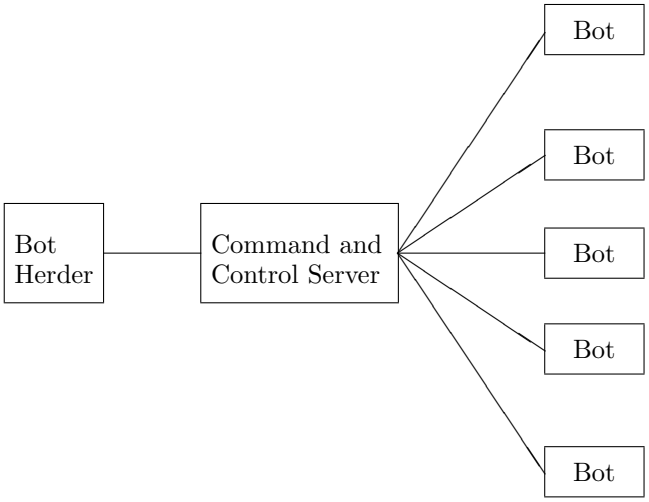


Figure 1: Botnet diagram

and starting blocking services [3]. Botnets have the characteristics of self-replication and active transmission, making it difficult to detect infected hosts. Recently, hackers created a variant of the BotNet virus, and antivirus software is more difficult to detect. When a botnet consisting of thousands, tens of thousands, or even millions of computers is formed, attackers can trick spam businesses. Through the servers they control, they order automated computers to send spam to the mail server. The annual report of New York computer security company MessageLabs pointed out that more than 80% of spam now comes from botnets.

In 2003, Clark of Oregon, USA, used the BotNet botnet to launch a DDoS attack on more than 20,000 computers on the shopping site eBay at the same time. In January 2005, the use of the BotNet botnet paralyzed a 20-year-old in the United States. The medical system of Seattle Northwestern Medical Center caused approximately $1.5 million in damages and endangered patients' lives in the hospital. In 2008, hackers from China launched a series of botnet botnets. DDoS attacked the well-known Bahamut gaming community website and asked the website to cooperate with its propaganda work. Otherwise, it will continue to block its services; the FBI pointed out that there are approximately 1 million in the world. One computer is controlled by a bot and becomes a member of the BotNet botnet. Hackers have unknowingly used your computer to accomplish whatever they want.

In recent years, the purpose of discovering malicious programs on personal computers has gradually changed from destroying computers to seeking commercial benefits, such as stealing credit card accounts and establishing botnets, personal data, address books. Personal information is stored in smart mobile devices. . This type of private information is more diverse than personal computers and can be used to make online micropayments. Under the trend of huge commercial interests, although there have been no reports of related attacks so far, mobile botnets will become the main threat that must be guarded against in the future with the popularization of the mobile Internet. In the next era of the information industry and the Windows operating system for personal computers, there are operating systems for smart mobile devices, such as Android and iOS. Each operating system will produce different information security problems for different management strategies.

In 2014, Lin *et al.* proposed a framework to automatically analyze and classify bot binary files [15]. Their system uses the longest common subsequence between system call traces to classify bot binaries. Their framework can effectively handle obfuscated robot binary files. In 2017, Cheng and Fu proposed a social bot detection scheme based on shared friends (SBDSF) [5]. SBDSF uses the feature of social graphs to detect social bots. The evaluation shows that SBDSF can achieve higher accuracy and precision using neural network classifiers. Rawat *et al.* presented a comprehensive survey of the evolution, functionality, modeling, and development life cycle of P2P botnets in 2018 [23]. They studied various P2P botnet detection methods and discussed key research challenges that are useful for research projects.

This paper is organized as follows. Section 2 introduces the steps used by mobile botnets to lure users into malicious programs. In Sections 3 and 4, we will propose two research issues for mobile botnet detection system and cloud smart mobile botnet App detection service. Finally, a conclusion is conducted in Section 5.

2 The Mobile Bonets

The following describes the steps used by mobile botnets to lure users into malicious programs:

1) Trick to click:
The baits used by previous botnets were popular social events, such as "free live broadcast of World Cup news...", "recommended guest quotes and live images of popular TV shows..." and so on. These newsletters are included—malicious link. The development of this botnet lies not in popular social activities but in the use of upgrade procedures required by our mobile phone software. The content is like "Re-publish the five-star N81 game for free immediately, click the URL to download and install http://nokia.xxx", which attracts you to click the hyperlink.

2) Virus behavior:
After clicking Install, three sub-programs will be installed without your approval, namely Ovi Update, Ovi Store Installer, and Ovi Store. Ovi Update is responsible for sending text messages containing virus links for further dissemination; Ovi Store installation programs can prevent users from uninstalling; Ovi Store is responsible for collecting user mobile phone information. These three subroutines cooperate to

achieve the purpose of stealing user privacy and expanding dissemination.

3) Communication method:
The virus will send about 20 text messages with the following content: "The five-star N81 game is now reissued for free, please click to download and install on the website http://nokia.xxxxxxxx.com/xx.sis". As the virus continues to spread, zombie phones with new locations will continue to join the network. Hackers have improved the learning ability of botnets and made them more concealed. As the learning experience improves, it will bring more harm to mobile phones:

   a. Send SMS automatically: The virus will send about 20 text messages containing links to the virus for further spread.
   b. Background execution: Secretly connect to the Internet, leading to a sharp increase in Internet costs.
   c. Privacy theft: These installed plug-ins will collect the user's mobile phone information and send it back to the attacker.
   d. Uninstallation: To prevent users from manually uninstalling, the virus has a self-protection mechanism and deletes related installation and communication records.
   e. From the start: The Ovi update subroutine is automatically installed and executed when it is started.

However, unlike mobile botnets developed on personal computers in the past, mobile botnets still face some challenges that must be overcome [20, 25]:

1) The power capacity of mobile devices is limited;
2) The cost of surfing on mobile devices is higher;
3) If the hacker's commands and control methods cause abnormal network traffic and cause huge expenditures, users can easily detect;
4) Mobile devices use public IP and frequently switch network connections.

The above four shortcomings make it impossible for robots deployed on personal computers to be directly applied to mobile networks. However, with the lure of huge profits, it seems only a matter of time before mobile botnets invade mobile networks. Therefore, this research topic will do its best to prevent the deployment and expansion of mobile botnets and help users avoid malicious attacks.

This research will be divided into two research topics. The first research topic is to study the parameter relationships and statistical methods of anomaly detection and apply anomaly detection technology to detect and analyze SMS traffic data to identify and summarize the possibility of spam and botnets. In the second research topic, a set of detection and analysis mechanisms are established to provide users with services to determine whether mobile device applications contain malicious code to prevent the infection and spread of botnets in the first place.

3 Topic 1: Research on Mobile Botnet Detection System Combining Statistics and Fuzzy Theory

Cloud computing benefits scientific and engineering applications, such as computing financing, data mining, and many other data-intensive activities, by supporting the paradigm shift from local to network-centric computing and network-centric content. It enables customers with limited computing resources to outsource large-scale computing tasks to the cloud and facilitates access control to big data [1, 17, 24, 28].

Because of the high threat risk of mobile botnets, this research topic of designing an SMS data access and statistical analysis system in a cloud environment to provide telecommunication service providers (TSP) with cloud services for detecting and preventing mobile botnets [19]. The service access process is shown in Figure 2.

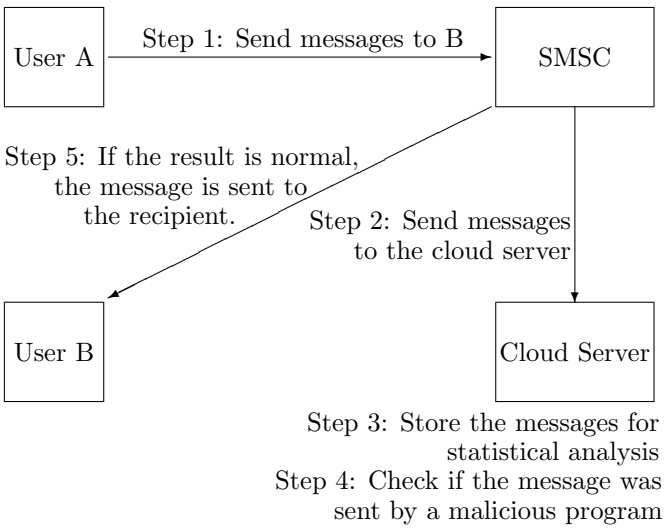


Figure 2: Cloud service access flow chart for mobile device BotNet detection and prevention

As shown in Figure 2, when user A sends a short message to user B, the content of the short message and the phone number of user B are first sent to the TSP operator. The short message service center (SMSC) of TSP receives the short message. And using it, the phone number of user A is transmitted to the cloud service. After the cloud server analyzes the various traffic parameters of user A and user A's various user behavior patterns and user cluster behavior patterns, the results obtained through

fuzzy inference are marked on the SMS data packet and returned. If the security level of the SMS from the cloud service tag is within the acceptable security level of the SMSC, the SMS can be sent to user B.

The research topic focused on studying parameter relationships and statistical methods for anomaly detection [26]. Applied anomaly detection technology to detect and analyze SMS traffic data to identify and summarize the possibility of spam and botnets. First, the combination of statistics and social network analysis techniques is used to build behavioral models of individual mobile phone users and mobile device user clusters. These behavior patterns are used to establish the normally expected behaviors of mobile users and then compare them with current or recently used data to detect changes in abnormal communication behaviors.

To construct the behavioral pattern of individual mobile phone users, it is necessary to study further the mobile devices infected by mobile botnets. The infected mobile devices will have the behavioral characteristics of SMS [26]. The following are the behaviors discovered through preliminary research:

1) Number of recipients per SMS: Bot Master sends SMS with the same text content to each infected mobile device, so the number of SMS recipients is usually no less than one.

2) User SMS sending frequency: Users have a special habit of sending SMS. Some users have a large amount of SMS transmission each month, and some users may not communicate via SMS transmission. Therefore, the amount of change in transmission frequency is also one indicator of the possibility of infection.

3) The ratio of the transmission volume of users who send and receive SMS: SMS is one of the communication channels for people. Because it is a communication method, the transmission volume ratio between sending and receiving SMS should be close to 1:1. If the number of transmissions sent to received is high, it may also be one of the indicators of the possibility of infection.

Therefore, through these characteristics, the six observation parameters that can establish user behavior patterns are summarized as follows:

1) Weekly SMS traffic;

2) Changes in SMS sending traffic every week (consistency);

3) Delay time for sending SMS;

4) Average delay time for sending SMS;

5) Traffic received via SMS every week;

6) Changes in SMS traffic received every week.

The research topic decided to study a mathematical formula based on the autocorrelation function (ACF), which can reflect the degree of correlation between the values of the same sequence at different times. Comparing the random variable of an ordered sequence with itself, this is the definition of the autocorrelation function in statistics. Each sequence without phase difference is similar to itself. That is, in this case, the autocorrelation function has a maximum value. If the various components in the sequence are correlated with each other (no longer random), the value calculated by the correlation value equation is no longer zero, and these components are autocorrelated. Derived from this concept in the time series, the formula for calculating the relevant value of the flow at different specific times is as follows [12, 14]:

$$r = \frac{\sum_{i-1}^{N-1}(x_i - \bar{x_1})(x_{i+1} - \bar{x_2})}{\sqrt{\sum_{i=1}^{N-1}(x_i - \bar{x_1})^2 \sum_{i=1}^{N-1}(x_i - \bar{x_2})^2}}$$

where $x_1, x_2, \cdots, x_N$, are a set of N observations.

$$\bar{x_1} = \sum_{i=1}^{N-1} x_i/(N-1)$$

is the mean of the first $N-1$ observations, and

$$\bar{x_2} = \sum_{i=2}^{N} x_i/(N-1)$$

is the mean of the last $N-1$ observations. In this equation, the interval value of r is between -1.0 and +1.0. The closer the result value is to 0, the greater the variability of the autocorrelation value and the more unstable the fluctuation of the time series. For example, incorporate this concept into the data parameters analyzed in this research, and set the time series unit to month. Suppose the time series of user A is from January to May. In this case, the amount of SMS sent is 4, 6, 6, 8, and 8 packets, and the autocorrelation function calculated by the equation is $r = 0.7$; if user B's The time series is 1 to 5, then the number of SMS sent in May is 4, 4, 100, 4, 4. The autocorrelation function calculated by this formula is $r = 0.3$. This means that user B's monthly SMS volume is more variable than user A, and maybe a victim or attacker infected by a mobile botnet.

Problems faced by the ACF calculation method:

1) Do not consider the impact of the user's emergency: If the user has an emergency (such as the release of an emergency), which causes the user to send a large number of SMS to notify the user's stakeholders at a specific time, this will result in an autocorrelation function. The result is often false touch warnings indicating infection.

2) Failure to define a clear numerical range: Although the variability of the SMS transmitted by the user can be seen from the results of the ACF function, the

statistical results represented by r are actually very abstract, and the data results of r cannot be clearly distinguished. There are two types of "infected" and "uninfected."

Therefore, using abnormal behavior patterns is what criteria should be used to determine abnormal behavior. One way to identify abnormal behavior is to use mathematical models. The problem with the mathematical method used is that "normal" and "abnormal" are usually distinguished by the range of data points [22]. If the deviation of the "normal" mode is within an acceptable range, it will cause the anomaly detector to trigger false alarms when the actual "normal" network activity produces unexpected conditions. Considering the analysis of user behavior that delimits between "normal" and "abnormal," as well as the uncertainty of user traffic behavior between instructions or action sequences, this is an ideal fuzzy system application [4, 21].

The so-called fuzzy concept means that expanding the concept is uncertain, or its expansion is unclear and vague. For example, the concept of "youth" is obvious to us, but its extension, that is, what age young people are, I am afraid it is difficult to express love because there is a difference between "young" and "not young." There is no clear boundary between them [9, 10].

Fuzzy theory helps computers parse vague or unclear vocabulary by simulating the method used by humans through fuzzy sets [14]. Like a traditional set, a fuzzy set has its elements, but you can talk about how far each element belongs to a fuzzy set. It is usually represented by a number between 0 and 1 (from low to high). Fuzzy set theory is proposed by Ruffit Zeder (1965), which is an extension of classical set theory. In classical set theory, the so-called dichotomous condition stipulates that each element can only belong to or not belong to a certain set (hence, fuzzy sets are not sets). It can say that the membership degree of each element to each set can only be 0 or 1. Each fuzzy set has a membership function whose value allows any real number in the closed interval [0,1] (unit interval) to indicate the degree to which the element belongs to the set [8, 27].

The results of structured data from fuzzy inference are used to compare two different time periods [26], namely the experimental time period and the observation time period. The index matrix generated during the experiment is as follows:

$$x_{ij} = f_j(u_i), \text{ for } i = 1, 2, \cdots, k; j = 1, 2, \cdots, 6.$$

i represents user 1 to user k; j represents 6 observation parameters, which can be used to establish the above user behavior pattern. Through the 6 index scoring mechanism f, the index data x is output according to fuzzy reasoning. Therefore, the indicator matrix of the observation period is as follows:

$$y_{ij} = f_j(u_i), \text{ for } i = 1, 2, \cdots, k; j = 1, 2, \cdots, 6.$$

After constructing two periods of index data, calculate the difference between the two matrices, as shown in the following equation:

$$d_{ij} = y_{ij} - x_{ij}, \text{ for } i = 1, 2, \cdots, k; j = 1, 2, \cdots, 6.$$

Then take the average z value, as shown in the following equation:

$$z_i = (d_{i1}, d_{i2}, d_{i3}, d_{i4}, d_{i5}, d_{i6}), \text{ for } z_i \in [0, 1].$$

After fuzzy inference of the z value, the closer the data interval of the z value from 0 to 1 is, the higher the possibility of infection.

By comparing fuzzy inference and autocorrelation function, the six observation parameters that users can establish in two periods are used to build index matrices x and y, analyze the difference between the two matrices d, and take the average value. The z-value is used to indicate the likelihood of a user being infected.

4 Topic 2: Research on Cloud Smart Mobile Botnet App Detection Service

The development of various mobile operating systems, whether it is Android system [2], iOS, or even others, is due to third-party developers' openness. The development of mobile device applications (Apps) enables users to download and use various applications on the market. In addition to bringing entertainment and convenience to users, it also brings hidden dangers. Due to the insufficient review of these applications by various mobile operating system providers, there are some malicious programs in thousands of applications. The purpose of these malicious programs may be: Entraining spam, collecting user privacy data, and destroying mobile phone hardware. Destroying user data and possibly deploying bot programs, these attacks will threaten the safety of users, and according to the research of Cui and other scholars [6], there are not many mobile antivirus software that can detect these malicious programs.

Therefore, we provide a cloud service to analyze and detect the presence of malicious code by placing the application in a sandbox or virtual machine and reminding users to avoid malicious program attacks. The so-called sandbox refers to a security technology that places some programs whose sources are untrustworthy, destructive, or whose intent cannot be determined in an isolated execution environment, observes the program's behavior, and checks whether it is infected with an unknown virus or contains. The code of the malicious code because is tested in a restricted simulation area, so the malicious code will not harm the user's computer.

The sandbox has all the resources of a real computer. The virus can perform all attacks in the sandbox but will closely monitor and record these operations. After mastering all operation methods and steps of malicious programs, security personnel can track these records, extract characteristic values of malicious codes to update

antivirus software, decompose the running process of malicious programs, and analyze malicious programs' running status appropriate countermeasures.

As shown in Figure 3, when a user downloads an application App from the Internet, it avoids accessing malicious applications and causing infection threats. You can use the mobile device BotNet App detection application service to upload the downloaded App to the cloud platform for detection. The cloud uses virtual machines to install and execute applications and monitor the data accessed by the applications and the commands executed. According to Symantec's analysis, it published a "Recent Android Malware Motivation" research report, which lists the 7 most profitable mobile malware:

1) Mobile phone billing (Premium Rate Number Billing) Internet fraud: Make a pay call through a malicious APP.

2) Spyware: Steal data from the user's mobile phone.

3) Search engine poisoning: Tampering with search engine rankings to deceive click behavior.

4) Pay-per-click Internet fraud: Fraudulently obtain click-through rates through mobile phone value-added services.

5) Pay-per-install plan: Malicious APP installs other software to deceive payment.

6) Adware Adware: Insert adware into APP to achieve the purpose of advertising exposure.

7) mTAN stealing: Stealing consumers' online transaction ID verification codes.

Establish 7 kinds of module detection measures for 7 kinds of main malicious software and return the detection and analysis results to users to accurately grasp App access behavior.

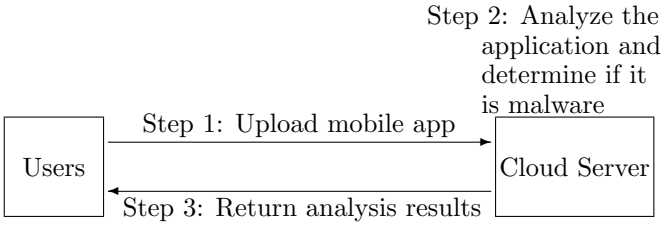


Figure 3: Botnet application detection service architecture diagram

5 Conclusions

It is expected that the research will develop a cloud service suitable for mobile botnet detection mechanisms for telecom service providers. The research also combines users' personal experience to prevent the system from misjudging important user messages and reducing telecom services through cloud platforms. The additional burden of the industry is to improve resource utilization efficiency; besides, a detection mechanism has been developed to allow users to detect whether a mobile application contains malicious code and combine these two mechanisms to defend against mobile botnet attacks.

Based on forward-looking considerations, this research provides a set of security mechanisms to defend against mobile botnets. Through the implementation of this research, users will be able to ensure the security of personal information and mobile business transaction processes; also because of the establishment of a mechanism to prevent the further expansion of mobile botnets, actively prevent the behavior of mobile botnets, and prevent enterprises or individuals from being attacked and causing the economic loss, reduces the protection costs that companies or individuals need to pay, and promotes economic growth.

We summary the main works of these research issues as follows:

1) Establish a cloud platform simulation environment;

2) Establish a mobile botnet detection and statistical analysis mechanism;

3) Analyze the performance and security of the SMS detection and analysis mechanism;

4) Establish a service access mechanism for the cloud platform;

5) Establish a malicious code detection and analysis mechanism;

6) Analyze the accuracy and performance of the application program and analyze the results.

Acknowledgments

The Ministry of Science and Technology partially supported this research, Taiwan (ROC), under contract no.: MOST 108-2410-H-468-023 and MOST 108-2622-8-468-001-TM1.

References

[1] M. H. R. Al-Shaikhly, H. M. El-Bakry, and A. A. Saleh, "Cloud security using Markov chain and genetic algorithm," *International Journal of Electronics and Information Engineering*, vol. 8, no. 2, pp. 96–106, 2018.

[2] Tanweer Alam, "Middleware implementation in MANET of Android devices," *International Journal of Electronics and Information Engineering*, vol. 12, no. 2, pp. 66-75, 2020.

[3] S. Balram and M. Wilscy, "User traffic profile for traffic reduction and effective bot C&C detection", *International Journal of Network Security*, vol. 16, no. 1, pp. 46-52, 2014.

[4] P. Barthakur, M. Dahal, M. K. Ghose, "Adoption of a fuzzy based classification model for P2P botnet detection", *International Journal of Network Security*, vol. 17, no. 5, pp. 522-534, 2015.

[5] B. L. Cheng, J. M. Fu, "Social bots detection on mobile social networks", *International Journal of Network Security*, vol. 19, no. 1, pp. 163-166, 2017.

[6] X. Cui, B. Fang, L. Yin, X. Liu, T. Zang, "Andbot: Towards advanced mobile botnets," in *Proceedings of the 4th USENIX Conference on Large-scale Exploits and Emergent Threats*, 2011.

[7] A. Dewanje and K. A. Kumar, "A new malware detection model using emerging machine learning algorithms," *International Journal of Electronics and Information Engineering*, vol. 13, no. 1, pp. 24–32, 2021.

[8] R H. Dong, B. B. Ren, Q. Y. Zhang, and H. Yuan, "A lightweight user authentication scheme based on fuzzy extraction technology for wireless sensor networks," *International Journal of Network Security*, vol. 23, no. 1, pp. 157-171, 2021.

[9] T. H. Feng, N. Y. Shih, M. S. Hwang, "Safety relay selection algorithms based on fuzzy relationship for wireless sensor networks", *The Journal of Supercomputing*, vol. 75, pp. 4601–4616, 2019.

[10] T. H. Feng, N. Y. Shih, and M. S. Hwang, "A safety review on fuzzy-based relay selection in wireless sensor networks," *International Journal of Network Security*, vol. 17, no. 6, pp. 712-721, 2015.

[11] Z. Guo, "Cryptanalysis of a certificateless conditional privacy-preserving authentication scheme for wireless body area networks," *International Journal of Electronics and Information Engineering*, vol. 11, no. 1, pp. 1-8, 2019.

[12] J. Y. Kim, H. K. Choi, "Spam traffic characterization," in *The 23rd International Technical Conference on Circuits/Systems, Computers and Communications*, pp. 961–964, 2008.

[13] M. S. Hwang and I. C. Lin, *Introduction to Information and Network Security (6ed, in Chinese)*, Taiwan: Mc Graw Hill, 2017.

[14] M. J. H. Lim, M. Negnevitsky, and J. Hartnett, "A fuzzy approach for detecting anomalous behaviour in email traffic," in *Proceedings of the 4th Australian Digital Forensics Conference*, pp. 36–49, 2006.

[15] Y. D. Lin, Y. T. Chiang, Y. S. Wu, and Y. C. Lai, "Automatic analysis and classification of obfuscated bot binaries", *International Journal of Network Security*, vol. 16, no. 6, pp. 477-486, 2014.

[16] C. Y. Liu, C. H. Peng, and I. C. Lin, "A survey of botnet architecture and batnet detection techniques", *International Journal of Network Security*, vol. 16, no. 2, pp. 81-89, 2014.

[17] L. Liu, Z. Cao, C. Mao, "A note on one outsourcing scheme for big data access control in cloud," *International Journal of Electronics and Information Engineering*, vol. 9, no. 1, pp. 29–35, 2018.

[18] M. Mahmoud, M. Nir, and A. Matrawy, "A survey on botnet architectures, detection and defences", *International Journal of Network Security*, vol. 17, no. 3, pp. 272-289, 2015.

[19] M. S. Malhi, U. Iqbal, M. M. Nabi, and M. A. I. Malhi, "E-learning based on cloud computing for educational institution: Security issues and solutions," *International Journal of Electronics and Information Engineering*, vol. 12, no. 4, pp. 162-169, 2020.

[20] C. Mulliner, J. P. Seifert, "Rise of the iBots: Owning a telco network," in *Proceedings of the 5th IEEE International Conference on Malicious and Unwanted Software*, pp. 71-80, 2010.

[21] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd edt., Addison Wesley, 2005.

[22] M. Negnevitsky, M. J. H. Lim, J. Hartnett, L. Reznik, "SMS communications analysis: How to use computational intelligence methods and tools?" in *Proceedings of the 2005 IEEE International Conference Computational Intelligence for Homeland Security and Personal Safety (CIHSPS'05)*, pp. 16–23, 2005.

[23] R. S. Rawat, E. S. Pilli, and R. C. Joshi, "Survey of peer-to-peer botnets and detection frameworks", *International Journal of Network Security*, vol. 20, no. 3, pp. 547-557, 2018.

[24] S. Rezaei, M. A. Doostari, M. Bayat, "A lightweight and efficient data sharing scheme for cloud computing," *International Journal of Electronics and Information Engineering*, vol. 9, no. 2, pp. 115–131, 2018.

[25] P. Traynor, M. Lin, M. Ongtang, V. Rao, T. Jaeger, T. La Porta, and P. McDaniel, "On cellular botnets: Measuring the impact of malicious devices on a cellular network core," in *ACM Conference on Computer and Communications Security (CCS'09)*, pp. 223–234, 2009.

[26] I. Vural, H. Venter, "Mobile botnet detection using network forensics," in *Future Internet Symposium*, Lecture Notes in Computer Science, vol. 6369, pp. 57–67, 2010.

[27] Y. Yan, B. Wang, L. Zhang, and X. Gao, "Information aggregation method of intuitionistic fuzzy set pair analysis in multi-attribute privacy risk decision-making," *International Journal of Network Security*, vol. 23, no. 1, pp. 22-32, 2021.

[28] C. Yang, Q. Chen, Y. Liu, "Fine-grained outsourced data deletion scheme in cloud computing," *International Journal of Electronics and Information Engineering*, vol. 11, no. 2, pp. 81–98, 2019.

Biography

Hung-Wei Yang received B.S. in Industry Engineer From Da-Yeh University, Taiwan in 2001; M.S. in

Information Management, Chao Yang University, Taiwan in 2009; Doctoral Program of Information Engineering, Asia University, Taiwan from 2016 till now. From 2012 to 2014, he was the manager in International Business Machine. From 2014 to 2015, he was the manager in Cisco Systems, Inc. Taiwan branch. From 2016 to 2019 he is the sales director of China branch in Syntron Technology Co. Ltd. Taipei Taiwan .From 2020 he is channel director in M-Power Co. Ltd., Taipei Taiwan.

Li-Chin Huang received the B.S. in computer science from Providence University, Taiwan, in 1993 and M.S. in information management from Chaoyang University of Technology (CYUT), Taichung, Taiwan, in 2001; and the Ph.D. degree in computer and information science from National Chung Hsing University (NCHU), Taiwan in 2001. Her current research interests include information security, cryptography, medical image, data hiding, network, security, big data, and mobile communications.

Min-Shiang Hwang received M.S. in industrial engineering from National Tsing Hua University, Taiwan in 1988, and a Ph.D. degree in computer and information science from National Chiao Tung University, Taiwan, in 1995. He was a distinguished professor and Chairman of the Department of Management Information Systems, NCHU, during 2003-2011. He obtained 1997, 1998, 1999, 2000, and 2001 Excellent Research Awards from the National Science Council (Taiwan). Dr. Hwang was a dean of the College of Computer Science, Asia University (AU), Taichung, Taiwan. He is currently a chair professor with the Department of Computer Science and Information Engineering, AU. His current research interests include information security, electronic commerce, database, data security, cryptography, image compression, and mobile computing. Dr. Hwang has published over 300+ articles on the above research fields in international journals.

NFC-Defender: SVM-based Attack Detection on NFC-enabled Mobile Device

Zhiqiang Wang[1,2,3], Wentao Wang[3], Jianyi Zhang[1], and Tao Yang[4]
(Corresponding author: Zhiqiang Wang, Jianyi Zhang)
Beijing Electronic Science and Technology Institute, Beijing, China[1]
State Information Center, Beijing 100045, China, Beijing, China[2]
Guangdong Hangyu Satellite Technology Co., Ltd, Guangdong, China[3]
Cyber Security Center, the Third Research Institute of the Ministry of Public Security, Shanghai 200031, China[4]
Email: wangzq@mail.besti.edu.cn
(Received Feb. 11, 2019; Revised and Accepted Feb. 5, 2020; First Online Feb. 16, 2021)

Abstract

NFC-enabled mobile devices are suffering from increasing threats since various applications widely used them. Since NFC and Webview based attack vectors were introduced on the Android platform, it's imperative to detect the attack vectors. Conventional approaches usually maintain a blacklist or rule-based to detect attacks. And recent years have witnessed many studies on attack detection using machine learning. However, effective frameworks for NFC-enabled devices are not proposed. This paper proposes a novel NFC-based attack model and presents an SVM-based method to detect attacks. Then we design a framework called "NFC-Defender" for detecting attacks on NFC-enabled devices. Finally, we evaluate the performance of our work. Experimental results show that NFC-Defender can realize the high accuracy (94%) and lower the security risk significantly.

Keywords: Attack Detection; Near Field Communication (NFC); Support Vector Machine

1 Introduction

NFC (Near Field Communication) is a promising communication protocol based on RFID (Radio Frequency Identification) and widely used in many fields. NFC is a wireless communication protocol with a short working distance (usually 10cm) [**?**, 6]. The detailed description is contained in ISO 18092.

Nowadays, NFC technology is almost exclusively used for mobile payments. And in the near future, a vast majority of cell phones will be NFC-enabled [15, 17]. The facts also draw more and more attention from attackers. A series of studies demonstrate that NFC suffered from eavesdropping, data corruption, Man-in-the-Middle attack and so on [**?**, **?**, 4, 16].

Besides, the software components of NFC also suffer from a variety of security vulnerabilities,such as Man-in-the-middle attack, spoofing of tag content, NFC-based worm and Denial-of-Service. Thus, it's imperative to develop techniques and tools to enhance the security and privacy of NFC-enabled devices. Asaduzzaman *et al.* proposed a security-aware architecture, which provides better security by detecting certificate modification, message modification, *etc.* [1]. Asaduzzaman *et al.* proposed a security protocol for protecting near field communication devices from networking attacks [2]. Zhuang *et al.* analyzed and optimized an NFC Security Authentication Algorithm [19]. Though many works have been studied and a number of approaches were proposed. An effective attack detection framework for NFC-enabled devices still remains to be developed. In summary, our main contributions are listed below.

- We propose an NFC-based attack model.
- We present an SVM-based model to detect NFC-based attacks and design a framework called "NFC-Defender" on Android platform.
- We evaluate the performance of our work and conclude results.

2 Background

In this section, we will introduce the NFC Data Exchange Format, Support Vector Machine and Naive Bayes technique.

2.1 Android Network Security

In the past decade, security issues on the Android system have been widly studied [3, 8, 10, 18].

Flaws of closed-source driver components makes wireless communication on Android platform vulnerable [9].

Thereby, the network stacks, *e.g.*, Bluetooth, NFC, cellular network and WLAN, in the Android system is becoming an important sources of security issues [14]. Currently, extensive vulnerabilities have been found on these network stacks, but to the best of our knowledge, the defence technology of NDEF protocol is only little studied [13].

2.2 NFC Data Exchange Format

NDEF (NFC Data Exchange Format) is a binary message format for data exchange between NFC-enabled devices and NFC tags. The payload is composed of a type, a length and an optional identifier. As is shown in Figure 1, an NDEF message is a sequence of NDEF records with a begin marker in the first and an end marker in the last record.

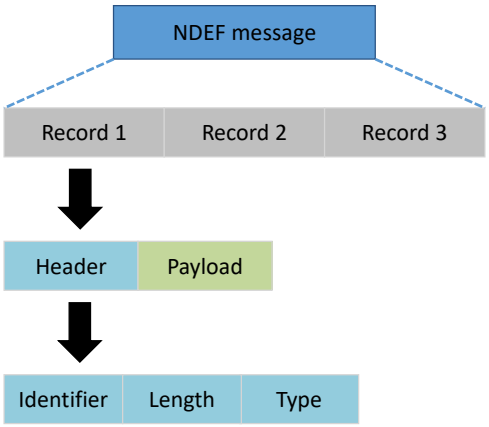


Figure 1: NFC data exchange format message

2.3 Support Vector Machine

Support vector machine (SVM) is developed from classical learning approaches for linearly separable data. Since Corinna Cortes and Vapnik proposed it in 1995 [5], SVM is applied in many fields. SVM doesn't require many samples and performs well on nonlinearly separable data. Besides, SVM can deal with data with high dimension. In a degree, machine learning is an approximation to the real world. Thus, generalization error bounds were proposed to evaluate the error, which is defined in the following inequality [7].

$$R(w) \leq Remp(w) + \phi(\frac{n}{h})$$

where $R(w)$ is the real risk, $Remp(w)$ is the empirical risk, $\phi(\frac{n}{h})$ is the confidence risk and $Remp(w) + \phi(\frac{n}{h})$ is the structural risk.

The objective of SVM is to minimize the structural risk. In a given high-dimensional data, SVM will find an optimal hyperplane based on the structural risk minimization principle (SRM). An optimal hyperplane is defined as follows.

$$\boldsymbol{w}_0^T \boldsymbol{x} + b_0 = 0$$

where $\boldsymbol{x}$ is input feature vector, $\boldsymbol{w}_0^T$ is weight vector and b_0 is bias.

Hence, we have the discriminant function $g(\boldsymbol{x}) = \boldsymbol{w}_0^T \boldsymbol{x} + b_0$. The hyperplane not only accurately distinguish data, but also maximize the distance between the hyperplane and its nearest point. The former ensures the minimum empirical risk, and the latter ensures the minimum confidence risk. The distance between the hyperplane and its nearest point is called interval, which is calculated by the following formula.

$$\rho = \frac{2}{||\boldsymbol{w}_0||}$$

Then, we formally define the constrained optimal problem: Given sample set $\{(x_i, d_i)\}_{i=1}^N$, find a weight vector $\boldsymbol{w}$ and bias b that satisfy following condition:

$$\min \Phi(w) = \frac{1}{2}\boldsymbol{w}^T \boldsymbol{w}$$

s.t.

$$d_i(\boldsymbol{w}^T \boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \cdots, N$$

According to the Lagrange multiplier method, the function to solve the largest interval is equivalent as follows.

$$\max_{\alpha}[\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)]$$

s.t.

$$\sum_{i=1}^{m} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \cdots, m$$

where α_i is the corresponding Lagrange multiplier of $(\boldsymbol{x}_i, y_i)$, $\mathcal{K}(\cdot, \cdot)$ is the kernel function and constant C is slack variable.

With kernel function, SVM maps nonlinear data to the high dimensional space and makes it linearly separable in high dimensional space.

2.4 Naive Bayes

Naive Bayes is a simple technique for constructing classifiers with the independence assumptions between predictors. The Naive Bayes models assign class labels to problem instances, which will be represented as vectors of feature values. Abstractly, Naive Bayes is a conditional probability model and can be described as the following expression.

$$p(C_k | x_1, x_2, \cdots, x_n)$$

where C_k is class variable and x_i is the i-st feature of target. And we call $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ Feature Vector.

According to Bayes' theorem, the expression above can be decomposed as:

$$p(C_k | \boldsymbol{x}) = \frac{p(C_k) p(\boldsymbol{x} | C_k)}{p(\boldsymbol{x})}$$

Thus, using Bayesian probability terminology, the formula can be written as:

$$posterior = \frac{prior \times likelihood}{evidence}$$

Obviously, the denominator does not depend on the class variable C and the value of $\boldsymbol{x}$ is given. Thus, the denominator can be considered as a constant and the numerator is equivalent to the joint probability model: $p(C, \boldsymbol{x})$. According to the chain rule and the independence assumptions, the expression can be rewritten as follows.

$$\begin{aligned} p(C_k, \boldsymbol{x}) &\propto p(C, \boldsymbol{x}) \\ &\propto p(C) \prod_{i=1}^{n} p(x_i|C) \\ &= \frac{1}{Z} p(C) \prod_{i=1}^{n} p(x_i|C) \end{aligned}$$

where the evidence $Z = p(\boldsymbol{x}) = \sum_k p(C_k)p(\boldsymbol{x}|C_k)$ is a scaling factor which only depends on $\boldsymbol{x}$. In other words, Z is a constant if the feature vector is known.

3 NFC-based Attack Model

Since Mulliner introduced NFC-based attack, research on exploiting NFC-enabled devices becomes popular.

The vulnerabilities which are involved in the attack model are design defect on Android implement of NFC and CVE-2013-4710[1]. The design defect on Android implement of NFC is that Android opens URL on NFC tag without any security check. Thus, attackers can induce users to tap NFC tag, and Android will automatically open the browser and visit the URL even it refers to a malicious page. CVE-2013-4710 is that Android (3.0-4.1.x) does not properly implement the WebView class, which allows remote attackers to execute arbitrary methods of Java objects or cause a denial of service (reboot) via a crafted web page. Figure 2 demonstrates a demo malicious web page based on CVE-2013-4710 which creates a reverse shell towards 210.77.8.1:33301 and upload the image at "/sdcard/DCIM/child.jpg". Please note that CVE-2013-4710 is just an example for better readability and any remote code execution vulnerability can be exploited in proposed model (such as CVE-2016-6754, CVE-2014-7224). Figure 3 shows an exploited device and three remote control windows of attackers. We simulated a user visiting our malicious webpage and the remote control established. Then we had the full command of users' device, and a photo was uploaded in the case. Figure 4 shows the content of the malicious NFC tag. The attack model can be described as follows.

Step 1. Set up a web server and deploy malicious web pages.

Step 2. Generate malicious URL content and write it into NFC tag.

Step 3. Induce target victims to touch malicious NFC tag and trigger WebView vulnerabilities.

Step 4. Control and command.

4 NFC-Defender

4.1 Overall Architecture

As is shown in Figure 5, we design a model to detect NFC-based attacks. Firstly, the reading module reads the NFC tag content. Then, the parsing module loads the content and using the corresponding detection module to evaluate the content. If it's secure, the execution module will execute the content at last. In the detection module, a known content database is analyzed by the machine learning module. Concrete algorithm in this module can be variable. In this work, we choose SVM as a powerful machine learning algorithm to classify normal content and malicious content. Then the feature of the known database is learned by a series of steps, including feature extraction, feature selection, model training and classification. The detailed process is described in Section 4.2.

4.2 Feature Engineering

Many malicious URLs [11, 12] disguise as normal URL by sub domains. For example, the URL in Figure 6 uses a sub-domain of "majdifamily2.ir" and disguise a legal PayPal URL to mislead users. Thus, text features of URL are supposed to be an important index to evaluate whether it's a malicious URL.

Besides, some studies used "bag-of-word" as a part of features. However, the position of token/keywords also plays a vital role in malicious content detection. As a result, we proposed two new features of URL. One is the occurrence of the famous brand name, and the other is the occurrence of high-frequency keywords. The brand name is gained from the domains of the Alexa Top 500 website. We count the top 23 high-frequency keywords in malicious samples and choose the occurrence of such keywords in target URLs as a feature. Another series of features is the popularity of target URLs, including PageRank, Alexa Rank and number of in-site links. These data can be accessed at Google and Alexa.

Hence, we mainly use two types of features in NFC tag content, including text feature and popularity. The detailed features are listed in Table 1. And Figure 7 demonstrates an example case of feature extraction.

5 Evaluation

We choose linear function as kernel function and "2" as constant C in our work. Besides, the implement of our

[1] http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2013-4710

```
<script>
function execute (cmdArgs)
{
        for (var obj in window) {
                if ("getClass" in window[obj]) {
                        return window[obj] .getClass() . forName ("java. lang . Runtime")
                                . getMethod ("getRuntime" ,null) . invoke (null ,null) . exec
(cmdArgs) ;
                }
        }
}
        execute(["/system/bin/sh","-c","nc 210.77.8.144 33301/system/bin/shInc 210.77.8.144 33302"]) ;
        execute(["/system/bin/sh","-c","nc 210.77.8.144 33303 </sdcard/DCIM/child.jpg"]) ;
        alert("An android shell has been returned") ;
</ script>
```

Figure 2: A malicious webpage based on CVE-2013-4710

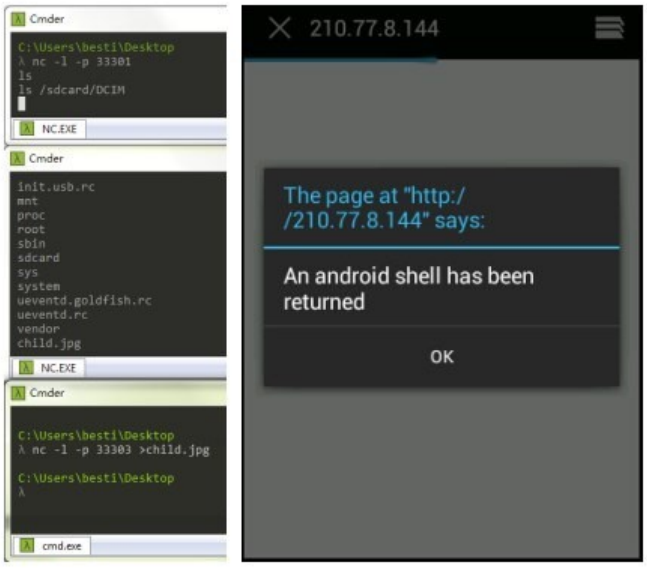


Figure 3: An exploited device

Table 1: Features

Number	Feature
1	length of hostname
2	length of URL
3	number of delimiters in hostname
4	number of delimiters in path
5	number of digits in hostname
6	number of digits in path
7	number of upper cases in URL
8	average length of tokens in hostname
9	average length of tokens in path
10	length of the longest token in hostname
11	length of the longest token in path
12	occurrence of brand name
13	occurrence of high-frequency keywords
14	PageRank
15	Alexa Rank
16	number of in-site links

NFC Tag Content

D1 01 14 55 01 6E 66 63 2E 63 6F 6D 40 6D 61 6C 69 75 72 6C 2E 63 6F 6D

Explanation

D1: MB, ME, SR, NFC Forum well-known type

01: Type Length

14: Payload Length

55: Type-U, URI

01: URI Identifier, http://www.

6E 66 63 2E 63 6F 6D 40 6D 61 6C 69 75 72 6C 2E 63 6F 6D
Value: -nfc.com@maliurl.com

Figure 4: Malicious NFC Tag Content

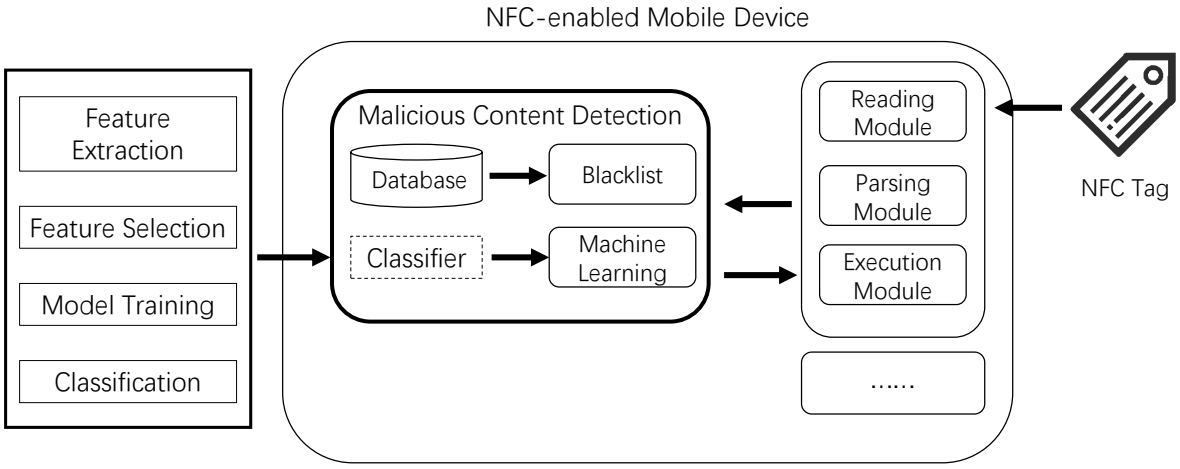


Figure 5: Overview of NFC-defender

http://paypal.com.update.your.accont.paypa
l1.majdifamily2.ir/PayPal/Login.php?login

Figure 6: Example malicious URL

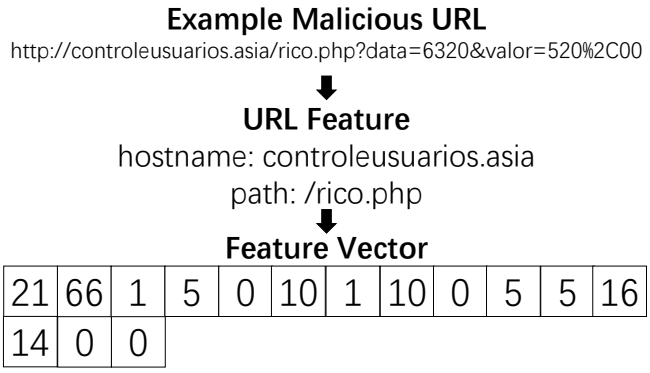


Figure 7: Example case of feature extraction

model is based on the famous open-source library Libsvm[2]. In this work, normal samples are accessed at Open Directory Project DMOZ[3] and malicious samples are accessed at Phishtank[4] (see Table 2).

Table 2: Dataset

Category	Sample Size
Normal Sample	4000
Malicious Sample	3630

Besides, we also implement the Naive Bayes and SVM-RBF model to compare the performance of our model. The results are validated by the ten folds cross-validation method. Specifically, we divide the dataset into ten subsets. Then, we use 9 of these sub-sets as the training set and 1 of these sub-sets as the testing set in turn. Finally, we will test the samples ten times and average the rates 10x as an estimation of the target algorithm. Figure 8 shows the results of the different model, including Naive Bayes, SVM-RBF and Linear SVM. Results demonstrate that our proposed method performs best among the above models(94%).

6 Conclusions

In this paper, we propose a novel NFC-based attack vector. The experiment shows the practicability of our model. Then we present an SVM-based model to detect NFC-based attacks. Finally, we design a series of experiments to evaluate our method with other baseline techniques. Results show that our approach outperforms the baseline techniques.

[2]https://www.csie.ntu.edu.tw/ cjlin/libsvm/
[3]http://www.dmoztools.net/
[4]https://www.phishtank.com/

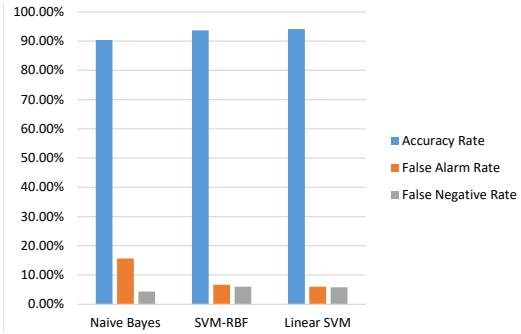


Figure 8: The results of our model and other models

In the future, we will further study the lightweight detection method to overcome the weakness of mobile computing.

Acknowledgments

This research was financially supported by the National Key Research and Development Plan (2018YFB1004101), Key Lab of Information Network Security, Ministry of Public Security (C19614), Special fund on education and teaching reform of Besti (jy201805), the Fundamental Research Funds for the Central Universities(328201910), China Postdoctoral Science Foundation(2019M650606), 2019 Beijing Common Construction Project-Teaching Reform and Innovation Project for Universities in Beijing, key laboratory of network assessment technology of Institute of Information Engineering, Chinese Academy of Sciences. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] A. Asaduzzaman, S. Mazumder, S. Salinas, and M. F. Mridha, "A security-aware near field communication architecture," in *International Conference on Networking, Systems and Security (NSysS'17)*, pp. 33–38, Jan. 2017.

[2] A. Asaduzzaman, S. Mazumder, and S. Salinas, "A promising security protocol for protecting near field communication devices from networking attacks," *International Journal of Security and Networks*, vol. 13, no. 2, pp. 98–107, 2018.

[3] P. Bhat and K. Dutta, "A survey on various threats and current state of security in android platform," *ACM Computing Surveys (CSUR'19)*, vol. 52, no. 1, pp. 21, 2019.

[4] C. H. Chen, I. C. Lin, and C. C. Yang, "NFC attacks analysis and survey," in *The Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pp. 458–462, July 2014.

[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[6] V. Coskun, B. Ozdenizci, and K. Ok, "A survey on near field communication (NFC) technology," *Wireless Personal Communications*, vol. 71, no. 3, pp. 2259–2294, 2013.

[7] S. S. Haykin, *Neural Networks and Learning Machines*, vol. 3, 2009. (`http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf`)

[8] S. Karthick and S. Binu, "Android security issues and solutions," in *International Conference on Innovative Mechanisms for Industry Applications (ICIMIA'17)*, pp. 686–689, 2017.

[9] H. Meng, V. L. L. Thing, Y. Cheng, Z. Dai, and L. Zhang, "A survey of Android exploits in the wild," *Computers & Security*, vol. 76, pp. 71–91, 2018.

[10] A. Misra and A. Dubey, *Android Security: Attacks and Defenses*, 2016. ISBN 13: 978-1439896464.

[11] E. U. Opara and O. A. Soluade, "Straddling the next cyber frontier: The empirical analysis on network security, exploits, and vulnerabilities," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 10–18, 2015.

[12] A. A. Orunsolu and A. S. Sodiya, "An anti-phishing kit scheme for secure web transactions," in *The 3rd International Conference on Information Systems Security and Privacy*, pp. 15–24, 2017.

[13] M. Roland, J. Langer, and J. Scharinger, "Security vulnerabilities of the NDEF signature record type," in *The Third International Workshop on Near Field Communication*, pp. 65–70, 2011.

[14] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, fog *et al.*: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, vol. 78, pp. 680–698, 2018.

[15] O. Sajid and M. Haddara, "NFC mobile payments: Are we ready for them?," in *SAI Computing Conference (SAI'16)*, pp. 960–967, July 2016.

[16] S. M. Shariati, A. Abouzarjomehri, and M. H. Ahmadzadegan, "Investigating NFC technology from the perspective of security, analysis of attacks and existing risk," in *The 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI'15)*, pp. 1083–1087, Nov. 2015.

[17] P. Stirparo, "A fuzzing framework for the security evaluation of ndef message format," in *The Fifth International Conference on Computational Intelligence, Communication Systems and Networks*, pp. 165–170, June 2013.

[18] Y. Zhauniarovich, *Learn Android Security Stack*, 2018. ISBN: 978-1-4842-1681-1.

[19] Z. Zhuang, J. Zhang, and W. Geng, "Analysis and optimization to an NFC security authentication algorithm based on hash functions," in *International Conference on Wireless Communication and Sensor Network*, pp. 240–245, Dec. 2014.

Biography

Zhiqiang Wang He is a lecturer of Beijing Electronic Science and Technology Institute and a post-doctoral of State Information Centre on China. His research interests include cryptography, vulnerability discovery and privacy preserving.

Wentao Wang He is an engineer of Guangdong Hangyu Satellite Technology Co,. Ltd.

Jianyi Zhang He is a lecturer of Beijing Electronic Science and Technology Institute. His research interests include the internet security, data security, and privacy.

Tao Yang He is a researcher of Cyber Security Center, the Third Research Institute of the Ministry of Public Security.His research interests include network security and system security.

An Efficient Biometric Authenticated Protocol for Arbitrary-domain-server with Blockchain Technology

Hongfeng Zhu and Zexi Li

(Corresponding author: Hongfeng Zhu)

Software College, Shenyang Normal University
No.253, HuangHe Bei Street, HuangGu District, Shenyang, P.C 110034 - China
Email: zhuhongfeng1978@163.com; 895197771@qq.com

(Received Aug. 7, 2019; Revised and Accepted May 23, 2020; First Online Apr. 24, 2021)

Abstract

In the biometric cross-domain authentication based on blockchain technology, to improve the scheme's security and efficiency, this paper proposes an efficient user login on an arbitrary-domain-server scheme based on a Blockchain node. Firstly, the two users login to the client and use the biometric fuzzy authentication technology to register personal biometric information. Then the server sends the information to the non-tamper blockchain node. When the user performs the key verification phase, they can log in on any domain server under the blockchain node, saving time from accessing the original server and improving efficiency. Compared with the relevant literature in recent years, the scheme proposed by us has higher efficiency and functionality. It can prevent replay attacks with a timestamp to achieve the purpose of verification of reception. Finally, we verify the security and efficiency of the proposed scheme.

Keywords: Authentication; Biometric; Blockchain; Cross Domain; Fuzzy Extraction Technology

1 Introduction

Block chain technology was first described in Bitcoin: A peer-to-peer electronic cash system [13], which was published by Bencong in mid-2008. It did not attract enough attention in the early period when Bitcoin was proposed, but with the stable operation and development of Bitcoin network for many years, Bitcoin [8] has become popular all over the world. Its prominent advantage lies in the decentralized design of distributed database. Its prominent advantage lies in the decentralized design of distributed databases, through the use of timestamps, Merkle tree structure, asymmetric key encryption algorithm, consensus algorithm and incentive mechanism, to achieve decentralized credit transactions. It also provides a new computing paradigm to solve the problems of poor reliability and inefficiency of the centralized model. Block chains are characterized by decentralization, transparency and credibility, tamper-proof, forgery-proof and high reliability. So its unique technical characteristics can effectively solve the key management, trust, security and privacy problems in identity authentication and management, and it's also provide credible, transparent, distributed storage support for identity authentication and management. Therefore, it is safer for us to keep the biometric key in an immutable block chain. In the traditional ID-based digital signature scheme [10, 14], if a particular user wants to obtain the private key of its formation, he needs to "prove" that the identity is his own to the authority and submit relevant certificates to prove his identity, which is also a huge time and space overhead.

Compared with other methods, using biometrics as identity can save a lot of space and time. Extracting key from biometrics [5, 18] is the most direct and secure method. Signers do not need to prove their identity to authoritative institutions. Moreover, because of the uniqueness of individual biometrics, the key extracted from biometrics is also unique. Biological characteristics are remarkable characteristics that distinguish one from others, such as face shape, fingerprint, palm shape, voice, iris, infrared heat and other congenital characteristics are physiological characteristics. Biological characteristics also have the advantages of safety, confidentiality, convenience, not easy to forget, good anti-counterfeiting performance, not easy to create or be stolen, and portable and available at any time and anywhere.

There are many studies on biometric identity authentication, mainly focusing on two directions [11]:

1) Based on the traditional biometric identification technology: Users submit their own biometric data, and compare them with the data stored in biometric template, the matching degree of the two determines whether they are legitimate users and whether they

carry out the authentication process.

2) Key generation technology based on biometrics: This model based on keystroke dynamics, extracts binary string from user keystroke mode, and then combines it with user password to form a stronger password; Hao and Chan design a key generation system.

Biological characteristics [12, 20] under ideal conditions should have the following properties:

1) Universality: Ideally, all people have this biological characteristic;

2) Uniqueness: In all populations, each person's biological characteristics are different;

3) Stability: Ideally, the biological characteristics are immutable, or there is a smaller one within the class Change;

4) Collectibility: Under certain conditions, the biological characteristics can be accurately obtained.

In view of the complexity of cross-domain authentication process in traditional authentication system, traditional biometric authentication must be carried out under the same domain server, which leads to complex process and slow authentication process. So we can store people's biometric information on the blockchain nodes [2, 3, 22] where the data is not easily tampered. Literature [7] proposed to apply blockchain technology to identity authentication based on biological characteristics, and store fingerprints to form an identity authentication system based on blockchain. However, this scheme directly stores unencrypted fingerprints on the blockchain, which poses security and privacy threats. On the basis of the original, this paper proposes a new efficient user login mode based on block chain in arbitrary domain server mode by using the theory of fuzzy extraction and block chain technology. When the user executes the key authentication level, at the same time, they can log on to any domain server under the block chain node, which saves the time of accessing the original server and improves the efficiency. The rest of the paper is organized as follows: Some preliminaries are given in Section 2. Next, an efficient user login on arbitrary-domain-server scheme based on blockchain is described in Section 3. Then, the security analysis and efficiency analysis are given in Section 4 and Section 5. This paper is finally concluded in Section 6.

2 Preliminaries

2.1 Fuzzy Extraction Technology

The Fuzzy extraction technology proposed by Dodis *et al.* [5]. Has two concepts: Fuzzy Extractor and Secure Sketch.

In order to simplify the process of fuzzy extraction technology, the following is a formal description of fuzzy extraction technology:

1) Random character generation algorithm $\mathrm{Gen}(\omega) \rightarrow (R, P)$: Input the user's biometric information ω and output a string R corresponding to the user's biometric as the user's random key and a public information value P through the random character generation algorithm.

2) Random character recovery algorithm $\mathrm{Rep}(\omega', P) \rightarrow R$: Input the user's biological characteristics ω', and the corresponding user's public information P. If the error of the biological characteristics satisfying the two inputs is within the given allowable range, that is, $\mathrm{dis}(\omega', \omega) \leq \varepsilon$, output the string R corresponding to the user's biometrics [18].

2.2 Blockchain Technology

Block chain is not a single technological innovation, but a distributed book-keeping technology realized by the deep integration of P2P [6] network technology, asymmetric encryption technology [9], consensus mechanism, and chain script. Blockchain technology [1, 4, 19] mainly contains three concepts: Transaction, block and chain. Block chains store data through data blocks and chain structures. Each data block consists of two parts, block head and block body. Each part has a unique hash value corresponding to the block address. The current block is connected to the previous block by storing the hash value of the previous block, thus forming the chain structure. Information such as the hash value, timestamp, and Merkle root value of the previous block chain is encapsulated in the block header. Each transaction is digitally signed by the transaction party. This ensures that the data is not falsified and cannot be tampered with, and that each completed transaction is permanently recorded in the block. The blockchain transaction is a data structure that describes payment details such as the receiving parties of bitcoin and the amount of the transaction. Transactions are the equivalent of certificates that prove the ownership of bitcoin, which is the blockchain records.

The account addresses of both receiving parties on a transaction are generated by cryptography algorithms: A pair of public and private key pairs (pk, sk) are obtained by calling ECDSA's key generation algorithm. The public key pk is calculated by hashing twice according to the formula to obtain the bitcoin account address X. The owner of the private key sk can use the bitcoin in X. $RIPEMD160()$ and $SHA256()$ both the hash Xfunction. Formula is as follows: $X = RIPEMD160(SHA256(pk))$.

3 An Efficient User Login on Arbitrary-domain-server Scheme

The concrete notations used hereafter are shown in Table 1.

Table 1: Description of the scheme symbol

Symbols	Description
$Client_A, Client_B$	The client of the $User_a$ and $User_b$ domain
ID_A, ID_B	The ID of User_a and User_b
ω_a, ω'_a	The biometric characteristics of the User_a registry/authentication
ω_b, ω'_b	The biometric characteristics of the User_b registry/authentication
$Gen()$	The random key generating function of fuzzy extraction technology
$Rep()$	The recovery algorithm of the fuzzy extraction technology
$H()$	The hash function
R, R'	The biometric random keys of the user registry/authentication
P	The biometric public information of the user registry/authentication
T_1, T_2, T_3, T_4	A timestamp
SK	A session key
PK_{sc}/PR_{sc}	The public/the private key of the server S_c
$\longrightarrow$	A secure channel

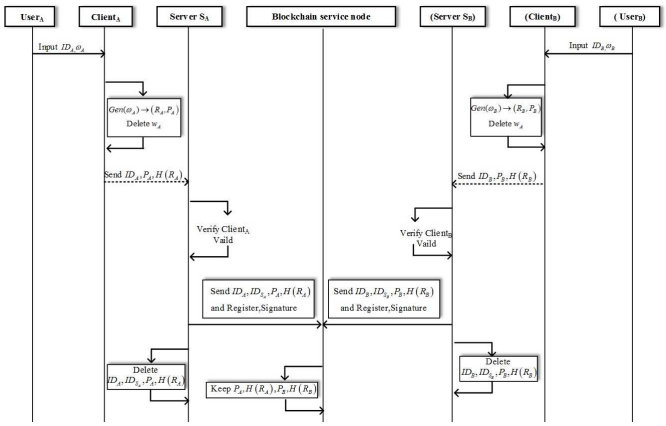


Figure 1: A premium user registration phase

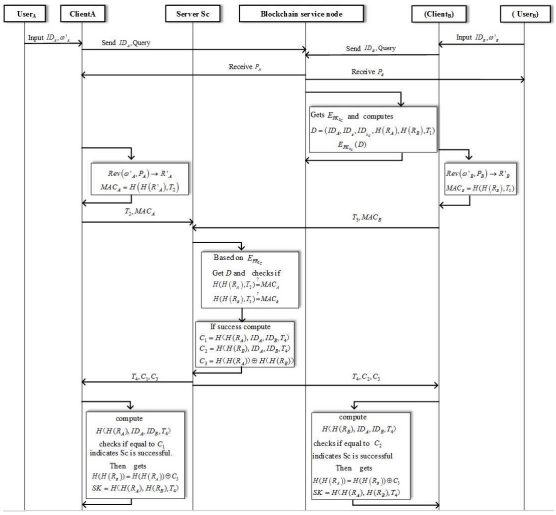


Figure 2: A premium user authentication phase

3.1 User Registration Phase

Figure 1 illustrates the user registration phase.

Step 1. The $User_A$ inputs his identity ID_A on the local client $Client_A$ and collects the biological feature ω_A through the biometric collector. $User_B$ is same as $User_A$, it's also inputs his identity ID_B on the local client $Client_A$ and collects the biological feature ω_B through the biometric collector.

Step 2. The $Client_A$ uses the random key generating function $Gen()$ of fuzzy extraction technology to process the input biological template ω_A, generates the random key R_A and public information P_A, and then deletes the biometric template ω_A. Then the client submits $ID_A, H(R_A), P_A$ to the local authentication server S_A via a secure channel. S_A also verifies the value from $Client_A$ is correct.

Step 3. The $Client_B$ uses the random key generating function $Gen()$ of fuzzy extraction technology to process the input biological template ω_A, generates the random key R_B and public information P_B, and then deletes the biometric template ω_B. Then the client submits $ID_B, H(R_B), P_B$ to the local authentication server S_B via a secure channel. S_B also verifies the value from $Client_B$ is correct.

Step 4. Server S_A sends ID_A, ID_{S_A}, $H(R_A)$, P_A, $Register, Signature$ to the blockchain service node, and deletes ID_A, ID_{S_A}, $H(R_A)$, P_A.

Step 5. Server S_B sends ID_B, ID_{S_B}, $H(R_B)$, $P_B, Register$, $Signature$ to the blockchain service node, and deletes $ID_B, ID_{S_B}, H(R_B), P_B$. The blockchain service node keeps P_A, $H(R_A)$, P_B, $H(R_B)$.

3.2 User Authentication Phase

Figure 2 illustrates the user authentication phase.

Step 1. The $User_A$ inputs his identity ID_A on the local client $Client_A$ and collects the biological feature ω'_A through the biometric collector. $User_B$ is same as $User_A$, it's also inputs his identity ID_B on the local client $Client_B$ and collects the biological feature ω'_B through the biometric collector.

Step 2. The $Client_A$ inquires the blockchain data to the service node, gets the public information P_A. The $Client_B$ inquires the blockchain data to the service node, gets the public information P_B at the same time. Then the blockchain service node gets the public key PK_{s_c} of server S_C, and calculated $D = (ID_A, ID_B, ID_{s_c}, H(R_B), T_1)$ and gets $E_{PK_{s_c}}(D)$ with the public key.

Step 3. The $Client_A$ uses the recovery algorithm $Rep(\omega'_A, P_A)$ of the fuzzy extraction technology to recover the random key R'_A. Then $Client_A$ computes $MAC_A = H(H(R'_A), T_2)$. Finally, $Client_A$ sends T_2, MAC_A to the server S_C.

Step 4. The $Client_B$ uses the recovery algorithm $Rep(\omega'_B, P_B)$ of the fuzzy extraction technology to recover the random key R'_B. Then $Client_B$ computes $MAC_B = H(H(R'_B), T_3)$. Finally, $Client_B$ sends T_3, MAC_B to the server S_C.

Step 5. Server S_C based $E_{PR_{S_C}}$, gets D, and exams if $H(H(R_A), T_2) \stackrel{?}{=} MAC_A$ and $H(H(R_B), T_3) \stackrel{?}{=} MAC_B$. If equal, Server S_C would calculates $C_1 = H(H(R_A), ID_A, ID_B, T_4)$, $C_2 = H(H(R_B), ID_A, ID_B, T_4)$, and $C_3 = H(H(R_A)) \oplus H(H(R_B))$. Then sends T_4, C_1, C_3 to $Client_A$, sends T_4, C_2, C_3 to $Client_B$.

Step 6. When receiving T_4, C_1, C_3, $Client_A$ calculates the $H(H(R_A), ID_A, ID_B, T_4)$, checks if $H(H(R_A), ID_A, ID_B, T_4)$ equal to C_1, so indicates S_C is authentication successful. Then $Client_A$ gets $H(H(R_B)) = H(H(R_A)) \oplus C_3$, and computes session key $SK = H(H(R_A), H(R_B), T_4)$.

Step 7. When receiving T_4, C_2, C_3, $Client_B$ calculates the $H(H(R_B), ID_A, ID_B, T_4)$, checks if $H(H(R_A), ID_A, ID_B, T_4)$ equal to C_2, so indicates S_C is authentication successful. Then $Client_B$ gets $H(H(R_A)) = H(H(R_B)) \oplus C_3$, and computes session key $SK = H(H(R_A), H(R_B), T_4)$.

4 Security Analysis

4.1 Security Model

Participants in the protocol include server S_a and server S_b, block chain service nodes, and server S_c with registration information on block chain nodes. During key authentication, we generate session key SK on secure channel. These participants involve many instances. We express the example i of participant U as π^i_U, and use the random prediction model to query the ability of opponent A. Finally, the model is built. The detailed steps are as follows.

4.2 Security Proof

This paper proves that the proposed protocol satisfies the security definition in the stochastic prediction model, and proves the security based on DDH assumptions.

Theorem 1. *We will express $Adv_\tau^{ind-cca}$ as the advantage of cracking public key in time t_1, and Adv_G^{ddh} as the advantage of solving DDH classical cryptography problem in time t_1. Then we can get the formula of breaking AKE protocol by adversary A as follows*

$$\begin{aligned} Adv^{ake}(t', q_0, q_1, q_2, q_3) \le & \frac{q_1^2+q_2^2+q_3^2}{2^{l-1}} + \\ & 2 \cdot \frac{q_1+q_2}{N} + 4 \cdot Adv_\Gamma^{ind-cca}(t_1, q_0, q_1, q_2, q_3) \\ & + 2 \cdot Adv_G^{ddh}(t_2, q_0, q_1, q_2, q_3). \end{aligned} \tag{1}$$

Among them,$t' \le t_1 + (q_1 + q_2) \cdot \tau_1 + 4 \cdot \tau_2$; q_0 represents the number of queries of Send, q_1 represents the number of interactions between S_a and server S_c, q_2 represents the number of interactions between S_a and server S_c, q_2 represents the number of interactions between block chain node A and server S_c, the key is a random number of characters with length of N in the dictionary set, L is a security parameter, τ_1 is the number of computational biometric operations, and τ_2 is the number of exponential operations.

Proof.

GameG0: Because this protocol corresponds to a real and effective attack, we get it by definition

$$Adv^{aie}(A) = |2\Pr[E_0] - 1| \tag{2}$$

GameG1: Because this round considers the use of password guessing attack by rival A, because $ID_A, ID_{S_A}, H(R_A), P_A$ and $ID_B, ID_{S_B}, H(R_B), P_B$ use bio-fuzzy authentication algorithm to encrypt, so they can resist attacks. This means that security possibilities are provided by $ID_A, ID_{S_A}, H(R_A), P_A$, $ID_B, ID_{S_B}, H(R_B), P_B$, and in Send questioning, because S_a and server S_b are secure. Therefore, we have come to the conclusion that

$$|\Pr[E_0] - \Pr[E_1]| \le \frac{q_1+q_2}{N} \tag{3}$$

GameG2: This round simulates the previous round. We use a table list τ to store the number of interactions between S_a and S_c, and between S_a and S_c servers. By calculating the number of interactions between servers, we can log on to any server. Therefore, we get the formula:

$$|\Pr[E_1] - \Pr[E_2]| \le \frac{q_1^2+q_2^2}{2^l} + Adv_\Gamma^{ind-cca}(\mathrm{A}_1) \tag{4}$$

Game G3: In this round, we simulate what may have happened before and express it in Listing H . Because hashing operations involve server S_a and server S_b, **Game**

Table 2: *Send*, *Excute*, *Reveal* and *Test* simulation of the query

Symbols	Definition
$Execute(\pi_U^i, \pi_S^i)$	This query returns the messages that are exchanged during the honest execution of the protocol π.
$Send(\pi_U^i/\pi_S^i, m)$	This query simulates the computation of opponents A, A sending information M to π_U^i, π_U^i that strongly control all communications in the protocol. The calculation results are returned to the protocol. A Starts Sending Questions $(\pi_U^i, "start")$ to π_U^i
$Reveal(\pi_U^i/\pi_S^i)$	This query simulates known key attacks. That is to say, sacrificing one authentication key will not expose other authentication keys.
$Corrupt(U)$	This query simulates the characteristics of full forwarding confidentiality. That is to say, sacrificing the long-term key will not damage the previous session key. Opponent A sends this query to participant U and returns the long-term secret key of U.
$Biometricreveal(U, B_U)$	This query returns the biometric password of the client U.
$RevealHash(x)$	If x is not hashed, select a random number as the hash value of x and store$(x, Hash(x))$, if a hash query has been made on x, the hash table is looked at to find the corresponding hash value returned
$Test(\pi_U^i/\pi_S^i)$	This query measures the advantage of the adversary. The adversary can send only one query of this form to a fresh instance π_U^i/π_S^i . The instance π_U^i/π_S^i flips a coin b and returns the session key if b=1, otherwise, returns a random value drawn from the space of the session key. It is important to note here that the test session cannot be an already queried session $Reveal(\pi_U^i/\pi_S^i)$.

G_2 and Game G_3 do not conflict with the H lists in Game G_3. The rival A has made q_3 queries between S_a and server S_b. Therefore, we can get the result

$$|\Pr[E_0] - \Pr[E_1]| \leq \frac{q_3^2}{2^l} \tag{5}$$

Game G4: In this round, we simulate the prediction model in the previous round. We add $MAC_A = H(H(R'_A), T_2)$, $MAC_B = H(H(R'_B), T_3)$ to $Test(U^i)$ lists, and then calculate these query operations accordingly. We call (S_A, S_B, S_C) as DDH classical problem type. We will simulate a possible session with a simulator S, and imitate (S_A, S_B, S_C) accurately to illustrate the classicality and self-reducibility of Diffie-Hellman problem. Firstly, S establishes a long secret key for S_A, S_B and then selects a random number $i \in [1, q_{se}]$. When a Send query is generated, S calculates $(a_0, H(R'_A))$, $(b_0, H(R'_B))$, $(z_0, H(R'_A), H(R'_B))$. In this way, a_0, b_0 are random numbers in the database. When a Test query is made, S responds to a pre-computed numerical value Z_0, at which time S creates all possible secret keys, so it can respond to Z_0. In addition, in Game G_3, random variables replace free variables in turn Game G_4, so that Game G_3 and Game G_4 are equal, so we get

$$Pr[E_3] = pr[E_4] \tag{6}$$

Game G5: Suppose A_{ddh} tries to solve the DDH problem in the range of G, and A_{ake} tries to destroy the security of the session key. First, A_{ddh} flips a fair coin instance and throws a coin C. It returns the real session key to A_{ake}. If $C = 1$, it returns the session key. Otherwise, it returns the random value that can be extracted from the session key space. A_{ake} Put out his guess and see if $c = c'$. Calculate by detecting the values of $C_1 = H(H(R_A), ID_A, ID_B, T_4)$, $C_2 = H(H(R_B), ID_A, ID_B, T_4)$, $C_3 = H(H(R_A)) \oplus H(H(R_B))$, if they are equal, A_{ddh} ask *Send*,*Excute*, *Reveal* and *Test*, and return to (S_A, S_B, S_C). If A_{ake} can output C, then A_{ddh} will output 0. If (S_A, S_B, S_C) is a real Diffie-Hellman problem, A_{ddh} and A_{ake} will be calculated $Pr[A_{ddh} output1] = Pr[E_5]$ in Game G5, so

$$|Pr[E_4] - Pr[E_5]| \leq Adv_G^{ddh}(A_{ddh}) \tag{7}$$

Since Z_0 is free and independent, no information about C is public, and we can get $Adv^{die}(t', q_0, q_1, q_2, q_3) \leq \frac{q_1^2+q_2^2+q_3^2}{2^{i-1}} + 2 \cdot \frac{q_1+q_2}{N} + 4 \cdot Adv_\Gamma^{jnd-cca}(t_1, q_0, q_1, q_2, q_3) + 2 \cdot Adv_G^{din}(t_2, q_0, q_1, q_2, q_3)$. □

So Theorem 1 is provable.

4.3 Further Security Discussion

1) Resist password guessing attack. An opponent can try to guess the password PW of a legitimate user using the transmitted message. Password guessing attacks can only crack a function with a low-entropy variable, so if we insert at least one large random variable that can resist such attacks. In our protocol, the opponent can only launch an online password guessing attack, because there is no message sent including password as input value. Even if the

opponent obtains confidential information $MAC_A = H(H(R'_A), T_2)$, $MAC_B = H(H(R'_B), T_3)$, without the help of the server, even if he intercepts it T_2, T_3, but he does not know the value $(R'_A)(R'_B)$, so he cannot carry out the first verification. Similarly, the enemy can't get the public information value P through the block chain node, so it can't authenticate the next step. On the other hand, on the online password guessing attack, because the maximum number of invalid attempts to guess the password is only a few times, and then the account registered by the server will be locked.

2) Withstand man-in-the-middle attacks. In our protocol, servers, clients, and users are not allowed to be man-in-the-middle by the enemy because we consider two-way authentication between them. And the user's biometric information and values are stored using asymmetric encryption. Because we encrypt the biological information, the random server S_C's PK_{S_C} is stored on the block chain node, so even if the attacker can intercept the information, he can not tamper with the information.

3) Perfect forward secrecy. The perfect forward secrecy means if the adversary cannot compute the established session key by compromised secret key of any server. The proposed scheme achieves perfect forward secrecy. Because our secret key contains a bio-fuzzy algorithm Rep, Gen to store the biometric values, as well as the session secret key $SK = H(H(R_A), H(R_B), T_4)$. So our protocol has forward security.

4) Withstand replay attacks. Replay attack is an attacker sending a packet that has been received by the destination host for the purpose of spoofing the system, mainly used in the authentication process. The verification messages include the temporary time stamp T_1, T_2, T_3, T_4, the adversary can't impersonate. So our proposed scheme resists the replay attacks.

5) Support privacy-protection. When the user registers, the user transfers the information $H(R_B)H(R_A)$ to the block chain node. Other clients do not need to read the information when accessing, guarantees the privacy of the information. In the user login stage, because of the user's encrypted information value, the server's public key information is transmitted to the block chain node, which ensures that the data can not be changed, and records on the account book of the block chain, which ensures the transparency of the information.

6) Key freshness property. Note that in our scheme, each established session key $SK = H(H(R_A), H(R_B), T_4)$, and C_1, C_2, C_3. The unique key construction for each session shows that proposed scheme supports the key freshness property.

7) Support mutual authentication. In our protocol, $Client_A$ first send $MAC_A = H(H(R'_A), T_2)$ and $Client_B$ respectively send $MAC_B = H(H(R'_B), T_3)$ to S_C. To detect the equivalence, and S_C then check MAC_B, MAC_A and send $T_4, C_1, C_3, T_4, C_2, C_3$ to the client for secondary authentication, so our protocol supports multiple authentication.

8) Resist dictionary attack.

   Case 1: The enemy wants to launch $C_1 = H(H(H(R_A), ID_A, ID_B, T_4))$ $C_2 = H(H(R_B), ID_A, ID_B, T_4)$ $C_3 = H(H(R_A)) \oplus H(H(R_B))$ in an offline dictionary attack, but the equations contain more than two unknown parameters, so the offline dictionary attack is not feasible.

   Case 2: If an adversary wants to perform $MAC_A = H(H(R'_A), T_2)$, in online dictionary attack on the server and user authentication, and if the number of authentication failures exceeds a predetermined threshold, the server or user will know whose password is targeted and can take appropriate action. Therefore, both offline dictionary attacks and untestable (measurable) online dictionary attacks are not feasible.

9) Resist user impersonation attack. In such an attack, an attacker would impersonate a user to negotiate with the server, this attack is not feasible because the attacker cannot receive the information value P_A, P_B returned by the blockchain service node, so the fuzzy algorithm $Rep(\omega'_A, P_A)$, $Rep(\omega'_B, P_B)$ cannot be used to restore R'_A R'_B.

10) Resist stolen verifier attack. In this scheme, after either party sends the secret key or verifies the secret key, the user's private information will be deleted, and all the information can be recorded in the user's mind and the unchangeable blockchain node, so this scheme can resist the attack of the stolen verifier.

From the Table 3, we can see that the proposed scheme is more secure and has much functionality compared with the recent related scheme.

5 Efficiency Analysis

Table 4 shows performance comparisons between our proposed scheme and relate literatureas

6 Conclusion

In order to improve efficiency on the basis of ensuring security, this paper improves the existing scheme and proposes an efficient user login on arbitrary-domain-server scheme based on Blockchain node. In this scheme, two-party users biometric information is stored in non-tamper

Table 3: Comparison our scheme among and other protocols

	Odelu [15]	**Odelu *et al.* [16]**	**Wang [17]**	**Zhou [21]**	**Proposed**
Provides user anonymity	Yes	Yes	Yes	Yes	Yes
Provides Forward secrecy	Yes	Yes	Yes	Yes	Yes
Provides mutual authentication	Yes	Yes	No	No	Yes
Resists illegal smart card revocation/reissue attack	No	No	No	Yes	Yes
Resists Insider and Stolen verifier attacks	Yes	Yes	Yes	No	Yes
Resists replay attack	Yes	Yes	Yes	Yes	Yes
Resists password guessing attack	Yes	Yes	Yes	Yes	Yes
Resists man-in-the-middle attack	Yes	Yes	No	Yes	Yes
Arbitrary server login	No	No	No	Yes	Yes
User non traceability	No	Yes	Yes	No	Yes
Usage of biometrics	Yes	No	No	Yes	Yes
Resists biometric recognition error	Yes	No	No	No	Yes

Table 4: Efficiency of our proposed scheme

Protocols		**Odelu [15]**	**Odelu *et al.* [16]**	**Wang [17]**	**Zhou [21]**	**Ours**
Computation	User	$4T_h+T_F+T_e$	$6T_h+T_e+T_s$	T_h	$T_h+2T_F+2T_e+2T_s$	$2T_h+2T_F+2T_s$
	SBN	$4T_h+T_F+4T_e$	$6T_h+T_e+T_s$	T_h+2T_s	$T_h+2T_e+2T_s$	$8T_h+T_F+2T_e+3T_x$
	Total	$8T_h+2T_F+5T_e$	$12T_h+2T_e+T_s$	$2T_h+2T_s$	$2T_h+2T_F+4T_e+4T_s$	$10T_h+3T_x+3T_F+2T_e+2T_s$
Communication	Messages	17	16	6	9	18
	rounds	6	3	3	5	5
Design	Concise design	yes	no	yes	no	yes
	Model	Threat model	Threat model	Random Oracle	Random Oracle	Random Oracle

T_h: The time for executing the hash function;
T_e: computational time for encryption/decryption
T_s: computational time for Number of signatures
T_F: computational time for fuzzy extraction
T_x: x denotes the XOR operation;
SBN: Server, Blockchain service node or other Network service node

block chain nodes, which ensuring the authenticity of data. When the users perform the key verification phase, they can log in on any domain server under the blockchain node. Compared with the relevant literatures in recent years, this scheme has not only higher efficiency and functionality, but also can prevent replay attack with time stamp and achieve the purpose of verification of reception.

Acknowledgements

This work was supported by the Liaoning Provincial Natural Science Foundation of China (Grant No. 2019-MS-286), and Basic Scientific Research Project of Liaoning Provincial Department of Education (Grant No. LJC202007).

References

[1] S. D. Angelis, L. Aniello, F. Lombardi, R. Baldoni , "PBFT vs proof-of-authority: Applying the cap theorem to permissioned blockchain," in *Italian Conference on Cybersecurity*, 2017. (https://www.researchgate.net/publication/320619309_PBFT_vs_proof-of-authority_applying_the_CAP_theorem_to_permissioned_blockchain)

[2] L. Axon, *Privacy-Awareness in Blockchain-based PKI*, 2018. (https://ora.ox.ac.uk/objects/uuid:f8377b69-599b-4cae-8df0-f0cded53e63b/datastreams/ATTACHMENT01)

[3] D. S. Baars, *Towards Self-Sovereign Identity using Blockchain Technology*, 2016. (http://purl.utwente.nl/essays/71274)

[4] L. M. Bach, B. Mihaljevic, M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *The 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO'18)*, pp. 1545-1550, 2018.

[5] R. Das, E. Piciucco, E. Maiorana, P. Campisi, "Convolutional neural network for finger-vein-based biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 99, pp. 1-1, 2018.

[6] J. A. D. Donet, C. Pserez-Solra, and J. Herrera-Joancomarts, "The bitcoin P2P network," in *International Conference on Financial Cryptography and Data Security*, pp. 87-102, 2014.

[7] J. S. Hammudoglu, J. Sparreboom, J. I. Rauhamaa, *et al.*, "Portable trust: Biometric-based authentication and blockchain storage for self-sovereign identity systems," *Cryptography and Security*, 2017. (arXiv:1706.03744)

[8] A. Hayes, "Bitcoin price and its marginal cost of production: Support for a fundamental value," *Applied Economics Letters*,vol. 26, no. 7, pp. 1-7, 2018.

[9] B. He, Y. Zhang, S. Qing, "Overview of blockchain technology development (in Chinese)," in *Annual Report on Development of China's Blockchain*, 2019. (https://www.pishu.com.cn/skwx_ps/bookdetail?SiteID=14&ID=11236589)

[10] M. S. Hwang, J. W. Lo, S. C. Lin, "An efficient user identification scheme based on ID-based cryptosystem", *Computer Standards & Interfaces*, vol. 26, no. 6, pp. 565–569, 2004.

[11] C. T. Li, M. S. Hwang, "An online biometrics-based secret sharing scheme for multiparty cryptosystem using smart cards," *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 5, pp. 2181-2188, May 2010.

[12] A. Ludwiczuk, Y. Asakawa, "Fingerprinting of secondary metabolites of liverworts: Chemosystematic approach," *Journal of Aoac International*, vol. 97, no. 5, pp. 1234-1243, 2019.

[13] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2008. (https://bitcoin.org/bitcoin.pdf)

[14] B. Nayak, "A secure ID-based signcryption scheme based on elliptic curve cryptography," *International Journal of Computational Intelligence Studies*, vol. 6, no. 2-3, 2019.

[15] V. Odelu, "IMBUA: Identity management on blockchain for biometrics-based user authentication," *Advances in Intelligent Systems and Computing*, vol. 1010, pp. 1-10, 2020.

[16] V. Odelu, A. Kumar, M. Wazid, and M. Conti, "Provably secure authenticated key agreement scheme for smart grid," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1900-1910, 2018.

[17] W. Wang, N. Hu, "BlockCAM: A blockchain-based cross-domain authentication model," *IEEE Third International Conference on Data Science in Cyberspace*, 2018. DOI: 10.1109/DSC.2018.00143.

[18] Y. Wang, J. Wan, J. Guo, Y. M. Cheung, P. C. Yuen, "Inference-based similarity search in randomized montgomery domains for privacy-preserving biometric identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 7, pp. 1611-1624, 2018.

[19] D. Yaga, P. Mell, N. Roby, K. Scarfone, "Blockchain technology overview," *Technical report, National Institute of Standards and Technology*, 2018. (https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8202.pdf)

[20] L. Zheng, "Isolation, identification and biological characteristics analysis of porcine deltacoronavirus TJ_1, *China Animal Husbandry & Veterinary Medicine*, vol. 45, no. 1, pp. 219-224, 2018.

[21] Z. C. Zhou, L. G. Li, Z. H. Li, "Biometric and password two-factor cross domain authentication scheme based on blockchain technology," *Journal of Computer Applications*, vol. 38, no. 6, pp. 1620-1627, 2018.

[22] G. Zyskind, O. Nathan, "Decentralizing privacy: Using blockchain to protect personal data," in *Proceedings of IEEE Security and Privacy Workshops (SPW'15)*, pp. 180-184, 2015.

Biography

Hongfeng Zhu, obtained his Ph.D. degree in Information Science and Engineering from Northeastern University. Hongfeng Zhu is a full professor of the Kexin software college at Shenyang Normal University. He is also a master's supervisor. He has research interests in wireless networks, social networks, network security and quantum cryptography. Dr. Zhu had published more than 60 international journal and international conference papers on the above research fields.

Zexi Li graduated from Zaozhuang University and is now a master of computer application technology in Shenyang Normal University. His research interests are information security, quantum communication and quantum cryptography, focusing on quantum activities related to human beings. Under the guidance of the teacher, he has published two articles in SCI journals.

Double Circulant Self-Dual Codes From Generalized Cyclotomic Classes of Order Two

Wenpeng Gao[1] and Tongjiang Yan[2]
(Corresponding author: Tongjiang Yan)
College of Sciences, China University of Petroleum[1]
Qingdao, 266555, China
College of Sciences, China University of Petroleum, Shandong Provincial Key Laboratory of Computer Networks[1]
Key Laboratory of Applied Mathematics (Putian University), Fujian Province University[2]
Qingdao, 266555, China; Jinan, 250014, China; Fujian Putian, 351100, China
Email: yantoji@163.com
(Received Oct. 22, 2019; Revised and Accepted Mar. 7, 2020; First Online Apr. 11, 2021)

Abstract

In this article, constructions of some double circulant self-dual codes by generalized cyclotomic classes of order two are presented. This technique is applied to [72, 36, 12] binary highest know self-dual codes over GF(2) and [32, 16, 8] almost high known self-dual codes over GF(4). Based on the properties of generalized cyclotomy, some of these codes can be proved to possess good minimum weights.

Keywords: Double Circulant Code; Generalized Cyclotomic Classes; Self-Dual Code

1 Introduction

Self-dual codes have important applications in transmission [13]. Constructing self-dual codes which have good minimum weight has been an important research problem [1]. And a number of papers devoted to constructing self-dual codes [2, 3]. Most of self-dual codes which are double circulant codes have high minimum weights [14, 22]. The generalized cyclotomic classes have wide applications in constructing sequences [5], cyclic codes [8, 9] and difference sets [11]. There are a number of results about generalized cyclotomic classes of order two [6, 7, 21]. Motivated by the constructions of double circulant codes by cyclotomic classes of order four [20], we give constructions of double circulant codes by generalized cyclotomic classes of order two. All computations have been done by MAGMA V2.20-4 [4] on a 3.40 GHz CPU.

This paper is organized as follows. In Section II, we present definitions and some preliminaries about double circulant self-dual codes and generalized cyclotomic classes. In Section III, three infinite families of double circulant self-dual codes are given. Section IV concludes the paper.

2 Preliminaries

2.1 Self-Dual Codes

A linear code C of length n and dimension k over finite field GF(l) is a k-dimensional subspace of GF(l)^n, where l is a prime power. A generator matrix G of the code C is a $k \times n$ matrix over GF(l). The elements of C are called codewords. The Euclidean inner product is defined by

$$(x, y) = \sum_{i=1}^{n} x_i y_i,$$

where $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$. The Euclidean dual code, denoted by $C^{\perp}$, of linear code C is defined by

$$C^{\perp} = \{x \in \mathrm{GF}(l)^n | (x, c) = 0, \text{for all } c \in C\}.$$

Then we have $dim(C) + dim(C^{\perp}) = n$. C is said to be Euclidean self-orthogonal if $C \subset C^{\perp}$ and Euclidean self-dual if $C = C^{\perp}$. If C is self-dual code, the dimension of C is $n/2$. From now on, what we mean by self-dual is Euclidean self-dual.

The Hamming distance of codewords x and y, denoted by $d(x, y)$, is defined to be the number of places at which x and y differ. The Hamming weight $wt(x)$ of a codeword x is the number of its nonzero coordinates and the minimum Hamming distance $d(C)$ of a code C is equal to the minimum nonzero Hamming weight of all codewords in C. Then for the self-dual codes, we have the following result.

Lemma 1. *[16, 17] Let C be a self-dual code over GF(l) of length n. Then we have*

(i) If $l = 2$, then

$$d(C) \leq \begin{cases} 4\lfloor \frac{n}{24} \rfloor + 4, & \text{if } n \not\equiv 22 \pmod{24}; \\ 4\lfloor \frac{n}{24} \rfloor + 6, & \text{if } n \equiv 22 \pmod{24}. \end{cases}$$

(ii) If $l = 4$, then $d(C) \le 4\left\lfloor \frac{n}{12} \right\rfloor + 4$.

The self-dual code C is called extremal if and only if the above equality holds, and which is called optimal if and only if it can get the highest possible minimum distance. If the self-dual code meets the highest known minimum distance for its parameters, we call it highest known self-dual. All the known optimal and highest known self-dual codes can be found in [12].

2.2 Generalized Cyclotomy and Generalized Cyclotomic Number

Let $n = pq$, where p and q are distinct odd primes with $\gcd(p-1, q-1) = d$. The Chinese Remainder Theorem [10] guarantees that there exists common primitive roots of both p and q. Let g be a fixed common primitive root of both p and q. Let $e = \mathrm{ord}_n(g)$ denote the multiplicative order of g modulo n. Then

$$\begin{aligned}\mathrm{ord}_n(g) &= \mathrm{lcm}(\mathrm{ord}_p(g), \mathrm{ord}_q(g)) \\ &= \mathrm{lcm}(p-1, q-1) = \frac{(p-1)(q-1)}{d}.\end{aligned}$$

Let x be an integer satisfying

$$x \equiv g \pmod{p}, \quad x \equiv 1 \pmod{q}.$$

Whiteman [18] defined the generalized cyclotomic classes

$$C_i = \{g^s x^i : s = 0, 1, \ldots, e-1\}, i = 0, 1, \ldots, d-1.$$

In this paper, we shall always assume that p, q are distinct primes with $\gcd(p-1, q-1) = 2$.

$$\begin{aligned}P &= \{p, 2p, \ldots, (q-1)p\}, \\ Q &= \{q, 2q, \ldots, (p-1)q\}, \\ R &= \{0\}.\end{aligned}$$

The generalized cyclotomic numbers corresponding the generalized cyclotomic classes of order two are defined by

$$(i, j) = |(C_i + 1) \cap C_j|, \text{for all } i, j = 0, 1.$$

Lemma 2. *[7] $-1 \in C_0$ if $(p-1)(q-1)/4$ is odd; $-1 \in C_1$ if $(p-1)(q-1)/4$ is even.*

Lemma 3. *[7] If $(p-1)(q-1)/4$ is even. Then*

$$\begin{aligned}(0,0) = (1,0) = (1,1) &= \frac{(p-2)(q-2)+1}{4}, \\ (0,1) &= \frac{(p-2)(q-2)-3}{4}.\end{aligned}$$

Lemma 4. *[19] For each $\eta \in P \cup Q$,*

$$|(C_i + \eta) \cap C_j| = \begin{cases} \frac{(p-1)(q-1)}{d^2} & \text{if } j \neq i, \\ \frac{(p-1)(q-1-d)}{d^2} & \text{if } j = i \text{ and } \eta \in P, \\ \frac{(p-1-d)(q-1)}{d^2} & \text{if } j = i \text{ and } \eta \in Q. \end{cases}$$

Lemma 5. *[7] If $(p-1)(q-1)/4$ is even, for each $\eta \in Z_n$,*

$$|(C_0 + \eta) \cap P| = \begin{cases} 0 & \text{if } \eta \in R \cup P, \\ \frac{q-1}{2} & \text{if } \eta \in Q \cup C_0, \\ \frac{q-3}{2} & \text{if } \eta \in C_1. \end{cases}$$

Lemma 6. *[7] If $(p-1)(q-1)/4$ is even, for each $\eta \in Z_n$,*

$$|(C_0 + \eta) \cap Q| = \begin{cases} 0 & \text{if } \eta \in R \cup Q, \\ \frac{p-1}{2} & \text{if } \eta \in P \cup C_0, \\ \frac{p-3}{2} & \text{if } \eta \in C_1. \end{cases}$$

Lemma 7. *[7] If $(p-1)(q-1)/4$ is even, for each $\eta \in Z_n$,*

$$|(C_0 + \eta) \cap R| = \begin{cases} 0 & \text{if } \eta \in R \cup P \cup Q \cup C_0, \\ 1 & \text{if } \eta \in C_1. \end{cases}$$

Lemma 8. *[7] If $(p-1)(q-1)/4$ is even, for each $\eta \in Z_n$,*

$$|(P + \eta) \cap Q| = \begin{cases} 0 & \text{if } \eta \in R \cup P \cup Q, \\ 1 & \text{if } \eta \in Z_n^*. \end{cases}$$

2.3 Double Circulant Codes

Definition 1. *[15] Let $G_n(R)$ and $B_n(\alpha, R)$ be matrices of the forms (I_n, R) and*

$$\begin{pmatrix} & \alpha & 1 & \cdots 1 \\ & -1 & & \\ I_{n+1} & \vdots & & R \\ & -1 & & \end{pmatrix},$$

where $\alpha \in \mathrm{GF}(l)$ and R is an $n \times n$ circulant matrix. The codes with generator matrices $G_n(R)$ and $B_n(\alpha, R)$ are called **pure double circulant codes** *and* **bordered double circulant codes**, *respectively.*

Let m_0, m_1, m_2, m_3 and m_4 be elements of $\mathrm{GF}(l)$. For convenience, we denote $\overrightarrow{m} = (m_0, m_1, m_2, m_3, m_4) \in GF(l)^5$. The circulant matrix $R_n(\overrightarrow{m})$ is the $n \times n$ matrix on $\mathrm{GF}(l)$ with components $r_{i,j}$, $1 \le i, j \le n$,

$$r_{i,j} = \begin{cases} m_0 & \text{if } j = i, \\ m_1 & \text{if } j - i \pmod{n} \in P, \\ m_2 & \text{if } j - i \pmod{n} \in Q, \\ m_3 & \text{if } j - i \pmod{n} \in C_0, \\ m_4 & \text{if } j - i \pmod{n} \in C_1. \end{cases}$$

Define by I_n and J_n the identity and the all-one square $n \times n$ matrices. Then $R_n(1,0,0,0,0) = I_n$ and $R_n(1,1,1,1,1) = J_n$. Denote $P_n = R_n(0,1,0,0,0)$, $Q_n = R_n(0,0,1,0,0)$, $A_1 = R_n(0,0,0,1,0)$ and $A_2 = R_n(0,0,0,0,1)$.

For convenience, let

$$\begin{aligned}&G_n(\overrightarrow{m}) = G_n(m_0 I_n + m_1 P_n + m_2 Q_n + m_3 A_1 + m_4 A_2), \\ &B_n(\alpha, \overrightarrow{m}) \\ &\quad = B_n(\alpha, m_0 I_n + m_1 P_n + m_2 Q_n + m_3 A_1 + m_4 A_2), \\ &R_n(\overrightarrow{m}) R_n(\overrightarrow{m})^{\perp} = a_0 I_n + a_1 P_n + a_2 Q_n + a_3 A_1 + a_4 A_2.\end{aligned}$$

The main theorem of this section is given by

Theorem 1. *Let* $\alpha \in \mathrm{GF}(l)$ and $\overrightarrow{m} \in \mathrm{GF}(l)^5$. Then

1) the code with generator matrix $GP_n(\overrightarrow{m})$ is self-dual over $\mathrm{GF}(l)$ if and only if the following holds:

$$a_0 = -1, a_1 = a_2 = a_3 = a_4 = 0.$$

2) the code with generator matrix $GB_n(\alpha, \overrightarrow{m})$ is self-dual over $\mathrm{GF}(l)$ if and only if the following holds:

$$\begin{aligned}&\alpha + n = -1, a_0 = -2, a_1 = a_2 = a_3 = a_4 = -1,\\ &-\alpha + m_0 + (p-1)m_1 + (q-1)m_2\\ &+ \frac{(p-1)(q-1)}{2}(m_3 + m_4) = 0.\end{aligned}$$

Proof.
The result follows from

$$\begin{aligned}&G_n(\overrightarrow{m})G_n(\overrightarrow{m})^{\perp}\\ =&I_n + a_0 I_n + a_1 P_n + a_2 Q_n + a_3 A_1 + a_4 A_2,\end{aligned}$$

and

$$\begin{aligned}&B_n(\alpha, \overrightarrow{m})B_n(\alpha, \overrightarrow{m})^{\perp}\\ =&I_{n+1} + \begin{pmatrix} \alpha + n & S \cdots S \\ S & \\ \vdots & X \\ S & \end{pmatrix}\end{aligned}$$

where $X = J_n + a_0 I_n + a_1 P_n + a_2 Q_n + a_3 A_1 + a_4 A_2$ and $S = -\alpha + m_0 + (p-1)m_1 + (q-1)m_2 + \frac{(p-1)(q-1)}{2}(m_3 + m_4)$. □

2.4 General Results

Lemma 9. *If* p *is a prime of the form* $4\omega + 1$ *and* q *is a prime of the form* $4\omega' + 3$, *where* ω *is a nonnegative integer. Then*

$$\begin{aligned}A_1 =&A_2^{\perp}, A_2 = A_1^{\perp}, P_n = P_n^{\perp}, Q_n = Q_n^{\perp},\\ P_n A_1 =&A_1 P_n = Q_n + A_1, P_n A_2 = A_2 P_n = Q_n + A_2,\\ Q_n A_1 =&A_1 Q_n = A_2, Q_n A_2 = A_2 Q_n = A_1,\\ P_n Q_n =&Q_n P_n = A_1 + A_2, P_n^2 = P_n, Q_n^2 = Q_n,\\ A_1^2 =&(0,1)A_1 + (0,0)A_2,\\ A_2^2 =&(0,0)A_1 + (0,1)A_2,\\ A_1 A_2 =&A_2 A_1 = Q_n + (0,0)(A_1 + A_2).\end{aligned}$$

Proof. We only prove the last result of this theorem, and the rest can be proved similarly. Let $M = A_1 A_2 = (m_{i,j})$, where $A_1 = (a_{i,j}^{(1)})$ and $A_2 = (a_{i,j}^{(2)})$. Then we can get

$$\begin{aligned}a_{i,j}^{(1)} = 1 \text{ iff } j - i \pmod n \in C_0,\\ a_{i,j}^{(2)} = 1 \text{ iff } j - i \pmod n \in C_1.\end{aligned}$$

It means that $a_{j,i}^{(1)} = 1$ iff $i - j \pmod n \in C_0$. And by Lemma 2, that means $j - i \pmod n \in C_1$. Thus $a_{j,i}^{(1)} = a_{i,j}^{(2)}$. Then $A_2 = A_1^{\perp}$, and $A_1 = A_2^{\perp}$.

According to Lemma 2, we can get $M = A_1 A_2 = A_1 A_1^{\perp}$. Thus

$$\begin{aligned}m_{i,j} &= |(C_0 + i - j) \cap C_0|\\ &= |-(C_0 + i - j) \cap (-1)C_0|\\ &= |(C_1 - i + j) \cap C_1|\\ &= |C_1 \cap (C_1 + i - j)|.\end{aligned}$$

Let $M' = A_2 A_1 = A_2 A_2^{\perp}$. Thus

$$\begin{aligned}m'_{i,j} &= |(C_1 + i - j) \cap C_1|\\ &= |C_1 \cap (C_1 + i - j)|\\ &= m_{i,j}.\end{aligned}$$

Therefore $A_1 A_2 = A_2 A_1$.

Since a product of circulant matrices is a circulant matrix, we can get the matrix M if we get the values of $m(1, j)$, for $j = 1, 2, \ldots, n$. Consider the element

$$m_{1,j} = |(C_0 + 1 - j) \cap C_0|.$$

1) If $1 - j \equiv 0 \pmod n$,

$$\begin{aligned}m_{1,j} = |C_0 \cap C_0| &= \frac{(p-1)(q-1)}{2}\\ &= 4\omega(2\omega' + 1) \equiv 0 \pmod 2.\end{aligned}$$

2) If $1 - j \pmod n \in P$, by Lemma 4,

$$\begin{aligned}m_{1,j} &= \frac{(p-1)(q-1-d)}{d^2}\\ &= \frac{4\omega 4\omega'}{4} = 4\omega\omega' \equiv 0 \pmod 2.\end{aligned}$$

3) If $1 - j \pmod n \in Q$, by Lemma 4,

$$\begin{aligned}m_{1,j} &= \frac{(p-1-d)(q-1)}{d^2} = \frac{(4\omega - 2)(4\omega' + 2)}{4}\\ &= (2\omega - 1)(2\omega' + 1) \equiv 1 \pmod 2.\end{aligned}$$

4) If $1 - j \pmod n \in C_0$, $(1-j)^{-1} \pmod n \in C_0$,

$$m_{1,j} = \left|C_0^{(n)} + 1 \cap C_0\right| = (0,0).$$

5) If $1 - j \pmod n \in C_1$, $(1-j)^{-1} \pmod n \in C_1$,

$$m_{1,j} = \left|C_1^{(n)} + 1 \cap C_1\right| = (1,1).$$

By Lemma 3, we can get that

$$A_1 A_2 = A_2 A_1 = Q_n + (0,0)(A_1 + A_2).$$

□

Lemma 10. *If p is a prime of the form $4\omega+3$ and q is a prime of the form $4\omega^{'}+1$. Then*

$$\begin{aligned}
A_1 =&A_2^{\perp}, A_2 = A_1^{\perp}, P_n = P_n^{\perp}, Q_n = Q_n^{\perp},\\
P_nA_1 =&A_1P_n = A_2, \quad P_nA_2 = A_2P_n = A_1,\\
Q_nA_1 =&A_1Q_n = P_n + A_1, \quad Q_nA_2 = A_2Q_n = P_n + A_2,\\
P_nQ_n =&Q_nP_n = A_1 + A_2, P_n^2 = P_n, Q_n^2 = Q_n,\\
A_1^2 =&(0,1)A_1 + (0,0)A_2,\\
A_2^2 =&(0,0)A_1 + (0,1)A_2,\\
A_1A_2 =&A_2A_1 = P_n + (0,0)(A_1 + A_2).
\end{aligned}$$

Proof.
The proof is similar of Lemma 9. □

Lemma 11. *If $p = 4\omega+1$ and $q =4\omega^{'}+3$. Then*

$$\begin{aligned}
&a_0 = m_0^2,\\
&a_1 = m_1^2 + 2m_0m_1,\\
&a_2 = m_2^2 + m_3^2 + m_4^2 + 2m_0m_2 + 2m_1(m_3 + m_4),\\
&a_3 = a_4\\
&\quad = (m_0 + m_1 + m_2)(m_3 + m_4) + 2m_1m_2\\
&\qquad + (0,0)(m_3^2 + m_4^2) + ((0,0) + (0,1))m_3m_4.
\end{aligned}$$

If $p = 4\omega+3$ and $q =4\omega^{'}+1$. Then

$$\begin{aligned}
&a_0 = m_0^2,\\
&a_1 = m_1^2 + m_3^2 + m_4^2 + 2m_0m_1 + 2m_2(m_3 + m_4),\\
&a_2 = m_2^2 + 2m_0m_2,\\
&a_3 = a_4\\
&\quad = (m_0 + m_1 + m_2)(m_3 + m_4) + 2m_1m_2\\
&\qquad + ((0,0) + (0,1))m_3m_4 + (0,0)(m_3^2 + m_4^2).
\end{aligned}$$

Proof.
The proof is obtained from Lemmas 9 and 10 directly. □

3 Double Circulant Self-Dual Codes Over Fields

In this section, we construct two families of self-dual codes. As a preparation, we have the following two lemmas.

Lemma 12. *Let $p = 4\omega+3$ and $q = 4\omega^{'}+1$ or $p = 4\omega+1$ and $q = 4\omega^{'}+3$. Then*

1) If $\omega+\omega^{'}$ is even, which means $\frac{p+q}{4}$ is odd,
$(0,0) \equiv (1,0) \equiv (1,1) \equiv 0 \pmod 2$,
$(0,1) \equiv 1 \pmod 2$.

2) If $\omega+\omega^{'}$ is odd, which means $\frac{p+q}{4}$ is even,
$(0,0) \equiv (1,0) \equiv (1,1) \equiv 1 \pmod 2$,
$(0,1) \equiv 0 \pmod 2$.

Proof.
Let $p = 4\omega+3$ and $q = 4\omega^{'}+1$. Then

$$\begin{aligned}
(0,0) =&\frac{(p-2)(q-2)+1}{4}\\
=&\frac{(4\omega+1)(4\omega^{'}-1)+1}{4}\\
=&\frac{16\omega\omega^{'}+4\omega+4\omega^{'}}{4}\\
=&4\omega\omega^{'}+\omega+\omega^{'}\\
\equiv&\omega^{'}+\omega \pmod 2
\end{aligned}$$

$$\begin{aligned}
(0,1) =&\frac{(p-2)(q-2)-3}{4}\\
=&\frac{(4\omega+1)(4\omega^{'}-1)-3}{4}\\
=&\frac{16\omega\omega^{'}-4\omega+4\omega^{'}-4}{4}\\
=&4\omega\omega^{'}-\omega+\omega^{'}-1\\
\equiv&\omega^{'}+\omega+1 \pmod 2
\end{aligned}$$

The proof for the condition of $p = 4\omega+1$ and $q = 4\omega^{'}+3$ is similar. □

3.1 Self-Dual Codes Over GF(2)

Theorem 2. *Let $\frac{(p-1)(q-1)}{4}$ be even and $\frac{p+q}{4}$ be odd. Then the following holds:*

1) $p = 4\omega+1$ and $q = 4\omega^{'}+3$

The double circulant codes with generator matrices $G_n(1,0,1,0,1)$ and $G_n(1,0,1,1,0)$ over GF(2) are self-dual codes of length $2n$.
The double circulant codes with generator matrices $B_n(0,0,1,0,1,0)$ and $B_n(0,0,1,0,0,1)$ over GF(2) are self-dual codes of length $2n+2$.

2) $p = 4\omega+3$ and $q = 4\omega^{'}+1$

The double circulant codes with generator matrices $G_n(1,1,0,0,1)$ and $G_n(1,1,0,1,0)$ over GF(2) are self-dual codes of length $2n$.
The double circulant codes with generator matrices $B_n(0,0,0,1,1,0)$ and $B_n(0,0,0,1,0,1)$ over GF(2) are self-dual codes of length $2n+2$.

Proof.
Let $p = 4\omega+1$ and $q = 4\omega^{'}+3$, $\overrightarrow{m} = (0,1,0,1,0)$. Then

$$\begin{aligned}
\alpha + n =&4(4\omega\omega^{'} + \omega + 3\omega^{'}) + 3 \equiv 1 \pmod 2,\\
&-\alpha + m_0 + (p-1)m_1 + (q-1)m_2\\
&+ \frac{(p-1)(q-1)}{2}(m_3 + m_4) \equiv 0 \pmod 2,\\
a_0 \equiv&0 \pmod 2,\\
a_1 =&a_2 \equiv 1 \pmod 2,\\
a_3 =&(0,0)(m_3^2 + m_4^2) + ((1,0) + (0,1))m_3m_4\\
=&0 \times (0+1) + (1+0) \equiv 1 \pmod 2.
\end{aligned}$$

By Theorem 2, the double circulant code with generator matrix $B_n(0,0,1,0,1,0)$ over GF(2) is a self-dual code of length $2n+2$. The proof that the double circulant code with generator matrix $B_n(0,0,0,1,0,1)$ over GF(2) is self-dual is similar.

Under the same conditions, if $\overrightarrow{m} = (1,0,1,0,1)$, we have

$$\begin{aligned} &a_0 \equiv 1 \pmod 2,\\ &a_1 = a_2 \equiv 0 \pmod 2,\\ &a_3 = (0,0)(m_3^2+m_4^2) + ((1,0)+(0,1))m_3m_4\\ &\quad = 0\times(0+1)+0\times(1+0) \equiv 0 \pmod 2. \end{aligned}$$

Thus the code with generator matrix $G_n(1,0,1,0,1)$ over GF(2) is self-dual. The rest of this theorem can be proved similarly. □

Table 1: Some codes over GF(2)

Code	p	q	Construction	Comments
$[70,35,10]$	5	7	$G_{35}(1,0,1,0,1)$	almost optimal
$[72,36,12]$	5	7	$B_{36}(0,0,1,0,1,0)$	highest know

The Table 1 shows that [72, 36, 12] code meets the highest know minimum distance for its parameters. It also can be obtain by $B_{36}(0,0,0,1,0,1)$ with $p=7$ and $q=5$ over GF(2). The structures of those codes are similar.

3.2 Self-Dual Codes Over GF(4)

Let u be the fixed primitive element of GF(4) satisfying $u^2+u+1=0$. Then we have the following results.

Theorem 3. *Let $\frac{(p-1)(q-1)}{4}$ be even and $\frac{p+q}{4}$ be even. Then the following holds:*

1) $p=4\omega+1$ and $q=4\omega'+3$

The double circulant codes with generator matrices $G_n(1,0,1,u+1,u)$ and $G_n(1,0,1,u,u+1)$ over GF(4) are self-dual codes of length $2n$.

The double circulant codes with generator matrices $B_n(0,0,1,0,u,u+1)$ and $B_n(0,0,1,0,u+1,u)$ over GF(4) are self-dual codes of length $2n+2$.

2) $p=4\omega+3$ and $q=4\omega'+1$

The double circulant codes with generator matrices $G_n(1,1,0,u+1,u)$ and $G_n(1,1,0,u,u+1)$ over GF(4) are self-dual codes of length $2n$.

The double circulant codes with generator matrices $B_n(0,0,0,1,u+1,u)$ and $B_n(0,0,0,1,u,u+1)$ over GF(4) are self-dual codes of length $2n+2$.

Proof.
Let $p=4\omega+1$ and $q=4\omega'+3, \overrightarrow{m}=(0,1,0,u,u+1)$. Then

$$\begin{aligned} \alpha+n =& 4(4\omega\omega'+3\omega+\omega')+3 \equiv 1 \pmod 2,\\ & -\alpha+m_0+(p-1)m_1+(q-1)m_2\\ & +\frac{(p-1)(q-1)}{2}(m_3+m_4) \equiv 0 \pmod 2,\\ a_0 \equiv& 0 \pmod 2,\\ a_1 =& a_2 \equiv 1 \pmod 2,\\ a_3 =& (0,0)(m_3^2+m_4^2)+((1,0)+(0,1))m_3m_4\\ =& (u+1+u)+(1+0)\times(u+1+u) \equiv 1 \pmod 2. \end{aligned}$$

By Theorem 2, the code with generator matrix $B_n(0,0,1,0,u,u+1)$ over GF(4) is self-dual.

The rest of this theorem can be proved similarly. □

Table 2: Some codes over GF(4)

Code	p	q	Construction	Comments
$[30,15,6]$	3	5	$G_{15}(1,1,0,u+1,u)$	
$[32,16,8]$	3	5	$B_{16}(0,0,0,1,u+1,u)$	almost high known

The Table 2 shows two double circulant self-dual codes with better parameters.

4 Conclusions

In this paper, we construct several families of self-dual codes based on the generalized cyclotomic sets of order two. These can enrich the choices of methods to construct good self-dual codes. We believe that these constructions will lead to good self-dual codes. It seems that the construction method by generalized cyclotomic classes of higher order is a rich source to obtain double circulant self-dual codes with good parameters.

Acknowledgments

The work is supported by Fundamental Research Funds for the central Universities(No.17CX02030A), Shandong Provincial Natural science Foundation of China (N0.ZR2016FL01, No.ZR2017MA001, No.ZR2019MF070), Qingdao application research on special independent innovation plan project(o.16-5-1-5-jch), Key Laboratory of Applied Mathematics of Fujian Province University(Putian University) (No.SX201702,No.SX201806), Open Research Fund from Shandong provincial Key Laboratory of Computer Network (No.SDKLCN-2017-03) and The National Natural Science Foundation of China (No.2017010754).The authors are grateful to the anonymous reviewers for valuable comments.

References

[1] A. Alahmadi, C. uneri, B. Ozkaya, H. Shoaib, and P. Sole, "On self-dual double negacirculant codes," *Discrete Applied Mathematics*, vol. 222, no. C, pp. 205–212, 2017.

[2] A. Alahmadi, F. Ozdemir, and P. Sole, "On self-dual double circulant codes," *Designs Codes and Cryptography*, vol. 86, no. 6, pp. 1–9, 2016.

[3] K. T. Arasu and T. Gulliver, "Self-dual codes over fp and weighing matrices," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 2051–2055, 2001.

[4] W. Bosma, J. Cannon, and C. Playoust, "The magma algebra system i: The user language," *Journal of Symbolic Computation*, vol. 24, no. 3-4, pp. 235–265, 1997.

[5] Z. Chen, "Linear complexity and trace representation of quaternary sequences over Z_4 based on generalized cyclotomic classes modulo pq," *Cryptography and Communications*, vol. 9, no. 4, 2015.

[6] Z. Chen and V. Edemskiy, "Linear complexity of quaternary sequences over Z_4 derived from generalized cyclotomic classes modulo $2p$," *International Journal of Network Security*, vol. 19, no. 4, 2016.

[7] C. Ding, "Autocorrelation values of generalized cyclotomic sequences of order two," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1699–1702, 1998.

[8] C. Ding, "Cyclic codes from the two-prime sequences," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3881–3891, 2012.

[9] C. Ding, "Cyclic codes from cyclotomic sequences of order four," *Finite Fields and Their Applications*, vol. 23, no. 96, pp. 8–34, 2013.

[10] C. Ding, D. Pei, and A. Salomaa, "Chinese remainder theorem," *Applications in Computing, Coding, Cryptography*, pp. 224, 1996. (`https://doi.org/10.1142/3254`)

[11] T. Feng and X. Qing, "Cyclotomic constructions of skew hadamard difference sets," *Journal of Combinatorial Theory*, vol. 119, no. 1, pp. 245–256, 2012.

[12] P. Gaborit and A. Otmani, *Tables of Self-Dual Codes*, 2020. (`https://www.unilim.fr/pages\_perso/philippe.gaborit/SD/index.html`)

[13] M. Garcia-Rodriguez, Y. Yaez, M. Garcia-Hernandez, J. Salazar, A. Turo, and J. Chavez, "Application of golay codes to improve the dynamic range in ultrasonic lamb waves air-coupled systems," *NDT&E International*, vol. 43, no. 8, pp. 677–686, 2010.

[14] J. Gildea, A. Kaya, R. Taylor, and B. Yildiz, "Constructions for self-dual codes induced from group rings," *Finite Fields and Their Applications*, vol. 51, pp. 71–92, 2018.

[15] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes, Volume 16*, 1977. eBook ISBN: 9780080570877.

[16] G. Nebe, E. M. Rains, and N. J. A. Sloane, "Self-dual codes and invariant theory (algorithms and computation in mathematics)," in *Math Nachrichten*, vol. 17, 2006.

[17] E. M. Rains, "Shadow bounds for self-dual codes," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 134–139, 1998.

[18] R. G. Stanton and D. A. Sprott, "A family of difference sets," *Canadian Journal of Mathematics*, vol. 10, no. 1, pp. 73–77, 1958.

[19] Y. Sun, Q. Wang, and T. Yan, "A lower bound on the 2-adic complexity of the modified jacobi sequence," *Cryptography and Communications*, vol. 11, pp. 337–349, 2019.

[20] Z. Tao and G. Ge, "Fourth power residue double circulant self-dual codes," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4243–4252, 2015.

[21] Q. Wang and D. Lin, "Generalized cyclotomic numbers of order two and their applications," *Cryptography and Communications*, vol. 8, no. 4, pp. 605–616, 2016.

[22] A. Zhdanov, "New self-dual codes of length 72," *Information Theory*, 2017. (`arXiv:1705.05779v1`)

Biography

Wenpeng Gao was born in 1994 in Shandong Province of China. He was graduated from the Department of Mathematics, China University of Petroleum, China, in 2017.He is a graduate student of China University of Petroleum.

Tongjiang Yan was born in 1973 in Shandong Province of China. He was graduated from the Department of Mathematics, Huaibei Coal-industry Teachers College, China, in 1996. He received the M.S. degree in mathematics from the Northeast Normal University, Lanzhou, China, in 1999, and the Ph.D degree in cryptography from Xidian University, Xian, China, in 2007. Now he is a associate professor of China University of Petroleum. His research interests include cryptography and algebra.

Compromised Accounts Detection Based on Information Entropy

Yanpeng Cui, Kun Wang, Jianwei Hu, Wei Zhao, Luming Feng, and Junjie Cui
(Corresponding author: Kun Wang)

School of Cyber Engineering, Xidian University
No. 2 South Taibai Road Xi'an City, Shaanxi, China
Email: kunwang156@sina.com

(Received Nov. 15, 2019; Revised and Accepted Mar. 15, 2020; First Online Apr. 17, 2021)

Abstract

Since an ever-increasing part of the population uses online social networks (OSN) in their day-to-day lives, many cyber criminals turn their attention to these networking sites. Compared to using Sybil accounts, compromised accounts can provide greater maneuverability for attackers. As a result, an upsurge of such accounts has been observed. However, research aimed at detecting such accounts is scarce, and the existing algorithms are not well adapted to the user's change of behavior. In this work, we advance the principle of using information entropy to detect compromised accounts. Information entropy can measure the stability of the system, and the long-term behavior of users can be considered an orderly system. We exploit the well-established regularities in users' usage patterns and the influence of new behavior on those regularities to discriminate between genuine and malicious accounts. Furthermore, we demonstrate the validity and feasibility of the algorithm experimentally.

Keywords: Compromised Accounts; Information Entropy; Online Social Networks

1 Introduction

With the development of computer networks and the popularity of smartphones, online social networks had gradually become the commonplace to maintain contact between people. Facebook is the first social network with more than 1 billion registered users, and its monthly active users are estimated at 2.2 billion. Twitter was released in 2006, attracting more than 1.3 billion active users and posting more than 140 million tweets a day. As users use social networks, they will continue to expand their social circles and build trust relationships. It is this trust that facilitates cybercrimes. Previous research demonstrated the deceitful strategy of cyber criminals in exploiting this trustworthiness, as users are more prone to messages coming from accounts they trust [6]. Moreover, the hijacking of a high-profile account provides the attacker favorable conditions for the widespread of malicious behavior on the network.

The occurrence of such incidents is not uncommon. High-profile accounts, from @foxnewspolitics in 2011, @AP in 2013, @YahooNews in 2014, and @officialcafee in 2017, all have been broken and spread false information. For regular accounts, many Instagram accounts were stolen for publishing adult content. In 2016, a report from Pew Research found that 13 percent of online accounts had been compromised [16]. Previous surveys have also found that 10% of victims said friends or relatives stole their accounts [28]. In addition, Thomas, *et al.* [21] found that 57% of the Compromised accounts will be screened by friends and lost some friends, and 21% of users never get right to the use of the account.Zangerle and Specht [28] found that 27% of users of stolen accounts applied for new social network accounts.These phenomena have caused major concerns in using social networks. They also have many adverse on society, such as the fall of the stock markets and the panic of the masses. On July 4, 2011, Fox News' Twitter account was hacked, and the attackers published the News of Obama's assassination. On April 23, 2013, the Associated Press Twitter account was hacked, and the attackers reported that Obama had been injured by a White House explosion, which caused a perceived drop in the New York stock exchange's market index. These incidents show that it is critical for a social network to be able to reliably detect and block messages that have not been authored by an account's legitimate owner.

A wealth of research was proposed in the last few years to detect malicious accounts in online social networks. Many detection algorithms, based on crowdsourcing, graph theory and machine learning, have been proposed [7, 19, 23]. In 2012, Wang *et al.* proposed a crowdsourcing method for detecting malicious accounts [25]. The method uses a manual detection and outsources the identification process of malicious accounts to external experts. Modeling social networks as graphs offers new possibilities for identifying malicious behavior

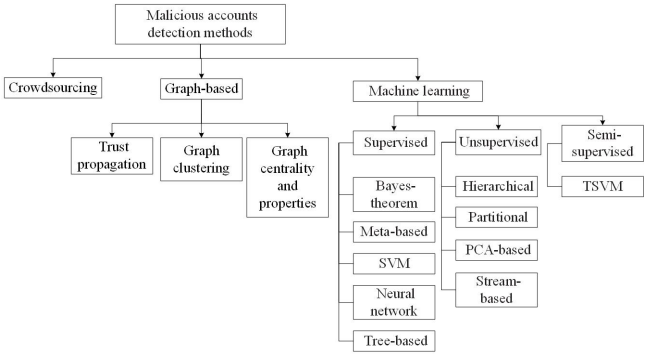


Figure 1: Detection methods

in OSN. The social graph is represented as $G = (V, E)$, where the vertices set V represents the users and the edges set E represents the users interactions within the network. Algorithms such as SybilGuard, Gatekeeper, SybilLimit, SybilRank, and SybilRadar are all based on graphs [1, 15, 27]. There are more detection methods based on machine learning algorithms. Yang *et al.* [26] used 18 features to train the NaiveBayes classifier in conjunction with the web and content. Fire *et al.* [8] developed a social privacy protector for Facebook users using the Rotating Forest Collection algorithm. Lee *et al.* [11] trained the SVM algorithm to cluster different account names to distinguish between benign accounts and suspicious accounts. Alsaleh *et al.* [2] introduced many content/behavior features extracted from tweet data and trained the multilayer perceptron (MLP) using the gradient descent method (GD). Lin and Huang [13] used the J48 to identify long-term surviving spam-sending accounts on Twitter. Bhat *et al.* [3] showed that the collection method can significantly improve the performance of a single classifier, and uses the J48 algorithm as a meta-classifier to identify malicious accounts. Gani *et al.* [9] used the K-means and Kohonen map algorithms to identify fake accounts on Twitter. Viswanath *et al.* [24] proposed a PCA-based detection system that captures normal user behavior in low-dimensional subspaces suitable for PCA algorithms. Behaviors that deviate from these patterns are considered abnormal behaviors. Miller *et al.* [14] adjusted two stream-based clustering algorithms, DenStream and StreamKM++, to detect spam accounts on Twitter. The entire existing detection algorithms can be partitioned into the categories shown in Figure 1.

But these detection algorithms are mainly aimed at sybil accounts and basically rely on macro-analysis of social networks, that is, using specific indicators to obtain global attributes. Therefore, these algorithms are inadequate to detect compromised accounts, since they have significantly different characteristics than sybil accounts. There are also techniques for classifying messages [10], detecting URLs in messages [4] and classifying account names [12], but these algorithms are not sufficient to detect compromising accounts.

In this paper, we present CPIE, an information entropy-based approach for detecting malicious accounts. CPIE is based on a simple assumption: as time goes by, the way a user uses social networks will appear to be regular, which is the user's usage pattern. Ruan *et al.* [17] and Egele *et al.* [6] showed evidences pointing to the existence of such a user's usage pattern. A social network user, for example, might consistently check his posts from 12 o'clock to 14 o'clock from his computer, and likes to focus on and publish posts related to basketball. Conversely, if the account falls under the control of an attacker, the account's stream may reveal some obvious differences. For example, many posts are sent in the morning, and their content is of political nature. Naturally, such unusual behavior raises strong suspicions that the posts are not from the account's legitimate owner.

Therefore, the user's usage pattern makes it possible, in principle, to detect compromised accounts. Based on this assumption, we propose the CPIE detection algorithm. CPIE builds a user profile for every social network account. The content of the user's portrait mainly includes the characteristics of the user's behavior and user's posts. When a new user behavior occurs, we compare it to the already constructed user's profile based on the principle of information entropy. Information entropy is used because it is a good way to assess the degree of confusion in a system. (A person is also a system in a sense.) And finally we evaluated the usability and applicability of the algorithm. In summary, this paper makes the following contributions:

- We present CPIE, an algorithm designed to detect compromised accounts.
- We use information entropy as a standard for assessing account behavior, more reflective of the user's current state, and can support real-time updates.
- We verify the applicability of CPIE through experiments.

The remainder in this paper is structured as follows. In Section 2, we provide some definitions about compromised accounts and information entropy. Section 3 details CPIE. Section 4 describes the experiments used for demonstration and validation. The conclusion is given in Section 5.

2 Preliminaries

In this section, we define the compromised accounts, information entropy and user portrait. These will help us to introduce CPIE subsequently.

2.1 Compromised Accounts

With the growing prominence of online social networks and the upsurge in the number of its users, more and more attackers focus their malicious behavior on such online networks. Zhang *et al.* divided abnormal accounts

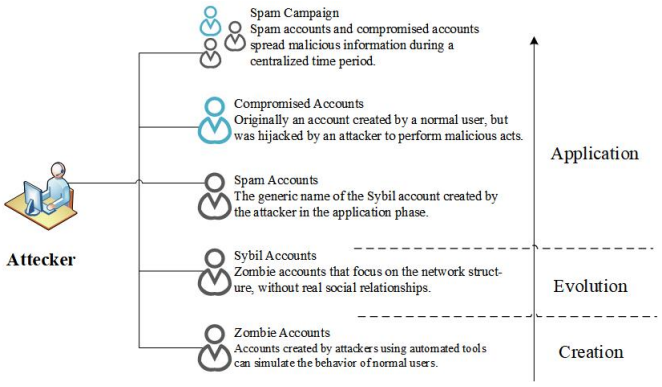


Figure 2: Abnormal accounts classification

into five categories according to the whole process from generation to application, as shown in Figure 2 [18]. From the figure we can see that the abnormal accounts can be divided into sybil accounts and compromised accounts, and whether the sybil account remains alive or malicious, the participation of the compromised account is required.

2.2 Information Entropy

Information is an abstract concept that is difficult to quantify. People often say that there is a lot of information, or that there is less information, but it is difficult to say how much information is there. For example, how much information is available in a 500,000-word book. To this end, C. E. Shannon [20] proposed the concept of information entropy in 1948.

Definition 1. *Information Entropy. Information entropy is the average rate at which information is produced by a stochastic source of data.*

The computation of information entropy is performed as follows: For discrete random variables,

$$entropy = -\sum_{x} p(x) \log p(x). \tag{1}$$

while for continuous random variables,

$$entropy = -\int_{x} p(x) \log p(x)\, dx. \tag{2}$$

where $p(x)$ is the probability of x. It is worth mentioning that the greater the uncertainty on x, the greater the entropy, and the greater the amount of information needed to figure it out. In other words, the more orderly a system is, the lower the information entropy is. Conversely, the more chaotic a system is, the higher the information entropy is. Therefore, information entropy can also be viewed as a measure of the degree of order within the system.

We regard people as a system, and the same person's use of social networks is regular. The use of an account by someone other than the legitimate owner is equivalent to the disturbance of the established rules, making the entire user behavior confusing. This is also the opportunity for us to detect compromised accounts through information entropy.

2.3 User Portrait

From the perspective of the English language, the three concepts of user portrait, user persona and user profile are different and can easily be confused. User persona tends to differentiate between user roles, such as distinguishing between user and administrators in online social networks. User portrait tends to portray different dimensions of the same class of users to achieve further subdivision and representation. User profile is more focused on the portrayal of attributes, such as gender, age and so on.

But in this paper, we do not distinguish between the three concepts. Therefore, we have redefined the user portrait as follows.

Definition 2. *User Portrait. User portraits are, as abstract as possible, representations of the user.*

By definition the user portrait is approximately equivalent to the user character, while the user attribute is a subset of the user portrait. In fact, practical applications using user portrait can be very costly. This is because a user portrait is a carrier that combines qualitative and quantitative methods. The qualitative methods mainly refer to the abstract generalization of the nature and features of the user, while the quantitative methods refer to the fine statistical analysis and calculation of the features.

3 CPIE

In the previous section, we introduced the concepts of compromised accounts, information entropy and user portrait. We also alluded to the rational of invoking information entropy to identify changes in behaviors. In this section, we will describe the CPIE detection algorithm in detail.

Considering the large users base, the extensive amount of information in online social networks and our intent to include user portraits for each user, our model should aim at lightening, as much as possible, the attributes associated with a user portrait. As the same time, we should aim at incorporating the most important discriminative features. Thus, the method of portraying the user portrait is fundamental to the success of our approach.

3.1 Build User Portrait

Before constructing the user portrait, we should stress that this study is aimed at twitter's users. Therefore, our research data are obtained through the Twitter API. The content returned by the API includes two parts, Tweet and User, which are used to provide information about the

tweet and the user. In Table 1, we show the main fields that list the return information to illustrate the displayed fields.

Table 1: The main fields in message

Structure	Fields				
Tweet	contributors	favorited	geo	retweeted	id(str)
	entities	coordinates	created at	lang	truncated
	in reply to screen name		text	retweet count	
	in reply to status id(str)		place	possibly sensitive	
	source	favorite count	in reply to user id(str)		
	quoted status	is quote status	quoted status id(str)		
User	created at	entities	id(str)	default profile	
	contributors enabled		lang	description	following
	default profile image		location	followers count	
	favourites count	notifications	url	verified	time zone
	follow request sent		utc offset	statuses count	
	friends count	protected	listed count	is translator	
	geo enabled	has extended profile		screen name	
	is translation enabled		translator type		

We shall follow the already existing tradition in sybil accounts of dividing features into two categories: social network analysis and content/behavior analysis. The main idea of our compromised accounts detection is to find out if there is a contradiction between the user's current and past behaviors. So we are concerned with the analysis of content/behavior, and thus we are going to extract features that embody the characteristics of the user's behavior to construct our user portrait. Moreover, we are limited in data resources and should full consider the information provided by the Twitter API. All facts considered, we chose 8 features that can be acquired through the API and reflect the user's behavior to construct the user portrait.

- **Time.** The time slot of using social networks each day. The time during which users use social networks is regular. For example, users are more likely to send Tweets during their break times, but are unlikely to pay attention to Twitter during working and sleeping hours. Therefore, a tweet appearing during a user's quite period raises suspicion and can be viewed as an exceptional occurrence.

- **Place.** Both time and place are indispensable. Needless to say that tweets emanating from the Japan from a user with posts from the USA just the day before are alarming. Although the account may not necessarily be stolen, the origin of the tweets is still a dimension of portrayal. Unfortunately, although the API provides a place field, it does not provide its content. (It requires a high-level API.)

- **Tweet Source.** Social networks offer different platforms and applications for users to send their tweets, such as web clients, iOS and Android, etc., but people tend to have their own habits. When a user, that generally uses an Apple application to post tweets, sends a tweet using an Android application, the change is perceptible and may reflects an anomaly in the account. (Of course, it does not rule out the possibility that the user's iPhone is broken, and the user has changed to an Android phone.)

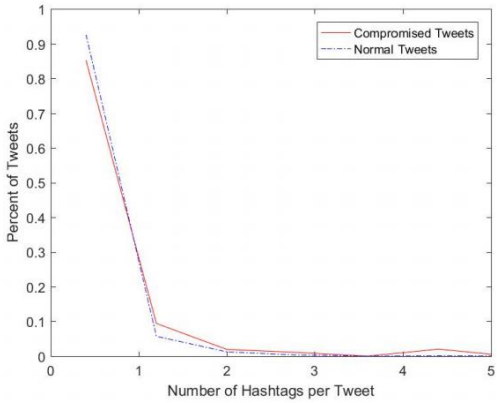


Figure 3: Analysis of the number of tweet topics

- **Tweet Language.** The return information of the Twitter API provides us with a lang tag, which makes it easy to get the language used in the tweet. The importance of language is self-evident. Is it suspicious that a user who has been using English suddenly sends a tweet in Japanese?

- **Tweet Topic.** Everyone has their own hobbies and interests, and so the range of topics broadcasted in online social networks is broad. We believe that users will not post content outside the topics of their own interest. Therefore, determining the major topics of interest of each user enriches the discriminatory strength of the user portrait and enhances the detection of compromised accounts. Fortunately, the hashtags field provided by the twitter API return information includes the topic of each tweet.

- **The Number of Tweet Topics.** Users tend to label Tweet topics when they send Tweets. According to the survey of literature [22], the number of Tweet Topics can also reflect the status of the account, as shown in Figure 3.

- **Links in Tweet.** Today's tweets tend to contain some external links which mirror, to a certain extent, the interest of the user. Similar to the Tweet Topic attribute, the links sent by a user present some regularities, such as similarities in the domain names of the links.

- **The Number of Links in Tweet.** There is also an underlying habit of the number of links attached to a user's Tweet. And the compromised account won't consider these problems, and only hope to spread malicious information as much as possible, Courtland *et al.* [22] provides the analysis as shown in Figure 4.

- **User Mentioned.** Users may mention some other users when sending their tweets. Users mention is

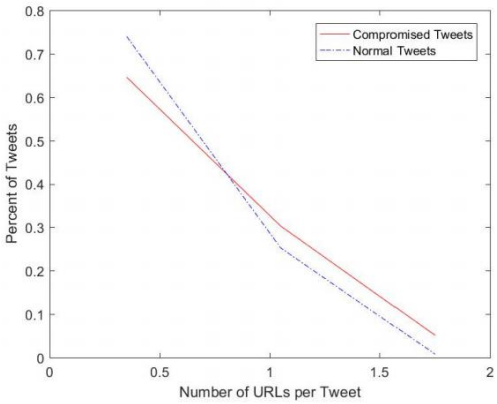


Figure 4: Analysis of the number of links in tweet

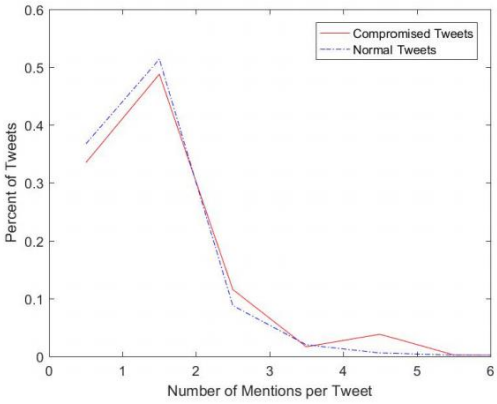


Figure 5: Analysis of user mentioned

usually based on acquaintances, and thus the mention of a user that has been so far ignored presents a noticeable change. In addition, changes in the number of users mentioned are also noteworthy. Figure 5 shows a research analysis of the number of Twitter accounts that mention users when sending messages.

- **Profile Changes.** User profile changes may be due to user behavior or to an exception, and so configuration file changes are worth our attention. We mainly focus on the protected field and also need to pay attention to changes from True to False.

Among the eight attributes proposed above, the place attribute is unavailable to us, and the configuration file will not change frequently, so we mainly consider the other six attributes, recorded as F. Below we describe in detail the construction process of the user portrait.

Algorithm 1. Build user portrait.

Input: The series of tweets (the tweet stream) that were extracted from a user account.

Output: User portrait file.

- Each feature model $f \in F$ is represented as a set M_f.
- Each element of M_f is a tuple $\langle m, c_m \rangle$, where c_m is the value of a feature (e.g., Web client for the Source of tweet). denotes the number of messages in which the specific feature value m was present.
- Record the total number (N) of values of each feature.
- Calculate and record the information entropy of each attribute at the present time, i.e.;

$$entropy_f = -\sum_{m \in M_f} p\left(\frac{c_m}{N}\right) \log p\left(\frac{c_m}{N}\right). \quad (3)$$

At this stage, some comments are in order. First of all, one thing is that the time feature itself is continuous and need to be discretized. Then, for each hour m_i we would store the number of tweets c'_i that were posted during this hour. This approach suffers from the rigidity of one-hour intervals. This simple subdivision is not faulty tolerant and may result in higher false positives. To avoid such scenarios, we calculate the average of the adjacent values for three consecutive time periods as a result. That is to say, for time m_i, its corresponding value $c_i = (c'_{i-1} + c'_i + c'_{i+1})/3$, where c'_{i-1} and c'_{i+1} are the statistics of the number of tweets in the previous hour and the next hour respectively.Then, in a tweet, some features may have only one value (such as the language used the tweet), and some may have multiple values (such as links in the tweet). And in a tweet, some features may have no value, such as no links in the tweet. When presented with such situations, we set the value of the feature to NULL. Note that a user who is reluctant to post links is also a behavioral habit of the user.

3.2 CIPE

In the previous section, we introduced a method for assigning user portraits to each user. In the following, we shall describe a method for assessing new tweets through CPIE. The core idea of CPIE is to use information entropy to quantify the orderliness of users behavior. Based on the user's past normal behavior, an increase in information entropy with new tweets reflects the breaking of order and an increase of chaos. Below we give a detailed algorithm of CPIE.

Algorithm 2. CIPE.

Input: User portrait file and a new tweet.

Output: The anomaly score of the tweet.

- Get the features values of the tweet T.
- Calculate the information entropy of each feature after adding the new tweet, recorded as $entropy'_f$.

- The abnormality score is computed using the tanh function as:

$$score_f = N * \tanh\left(\frac{entropy'_f - entropy_f}{entropy_f}\right). \quad (4)$$

where

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (5)$$

- Integrate all attribute anomaly scores to compute the overall anomaly score for the tweet:

$$score_T = \sum_{f \in F} \omega * score_f. \quad (6)$$

where ω is the weight of each feature.

- Determine if the tweet is abnormal, and update the user portrait if it is deemed to be normal.

In the following, using the language feature as an example, we illustrate the relevance of information entropy in quantifying behavioral changes of the users.

Example 1. *A user sent 300 tweets, 150 of which were written in English, 100 in German, and the remaining 50 in Arabic. The portrait of the language feature at this time is* $\{\langle English, 150\rangle, \langle German, 100\rangle, \langle Arabic, 50\rangle\}$, *and*

$$entropy_{language} = 1.0114042647073516 \quad (7)$$

When the user sends a tweet written in Japanese, the new information entropy is

$$entropy'_{language} = 1.030321351178594 \quad (8)$$

We have $entropy'_{language} - entropy_{language} = 0.01891708647124246 > 0$, that is to say, with the generation of this tweet, the user's behavior pattern becomes more confusing. If the tweet was written in Arabic, then

$$entropy'_{language} = 1.0139693345754932 \quad (9)$$

We would still have $entropy'_{language} - entropy_{language} = 0.002565069868141645 > 0$, but this value is much smaller than 0.01891708647124246. That is to say, this tweet will bring confusion, but in a much smaller scale than the previous one. If the tweet was written in English, then

$$entropy'_{language} = 1.0103414134888424 \quad (10)$$

Thus $entropy'_{language} - entropy_{language} = -0.001062851218509131 < 0$, that is to say, with the generation of this tweet, the user's behavior pattern becomes more orderly.

We should stress that when a tweet has multiple values in a feature, then we proceed first by updating all these values simultaneously and then we perform all the required computations. The following simple example illustrates this update procedure with the Tweet Link feature.

Example 2. *Suppose that the initial portrait of the user's Tweet Links feature is:*

$$\{\langle t.co, 100\rangle, \langle \text{youtube.com}, 50\rangle, \langle bing.com, 20\rangle\}$$

If a tweet, from the same user, contains two domain names t.co and bing.com, then the feature values become

$$\{\langle t.co, 101\rangle, \langle \text{youtube.com}, 50\rangle, \langle bing.com, 21\rangle\}$$

and the value of N increases by 2. It is only after this update that we proceed at quantifying the impact of introducing the tweet on the stability of the system.

3.3 Preferences

In this section, we discuss parameter settings in the algorithm. The weight between each feature is determined according to its own stability. For a more stable feature, the impact of the behavior that destroys its stability will be more serious. Therefore, the weight of a feature is calculated as follows:

$$\omega_f = \frac{\sum_{i \in F} entropy_i}{entropy_f} \quad (11)$$

Because changes in social network are very fast, feature values are different for different social networks or even for the same social network. Our weights do not necessarily represent the best option, as they were computed from a limited sample that is not representative of the complete social network.

There are many factors that affect the stability of the system, and which may lead to false positives. There are two main sources of false positives: a normal account that begins to change its behavior and characteristics; and false positives caused by Sybil accounts, because we only want to detect compromised accounts. To prevent false positives from the second source, we set a minimum requirement on the number of tweets in each user portrait to 20. Taking into account the already existing sybil accounts detection and penalty systems set by Twitter, such requirement not only reduce false positives, but also guarantee the quality of our user portraits.

For the first source of false positives, we use a text similarity algorithm to cluster the tweets in a certain period of time and count the proportion of violations to the user's portraits within the class. When this count is greater than a certain threshold θ, the account is considered to be stolen. In [5] the value of θ is chosen to be a linear function of the size of the class, more precisely $\theta = \max(0.1, kn + d)$, where $k = -0.005$ and $d = 0.82$. Moreover, one should realize that our CIPE is plastic and self-adaptive, namely as the number of tweets sent by a user increases, if his behavior changes, CPIE will gradually build a user portrait that is conform to the user's new behavior and eventually brings the system to a new stable state.

4 Model Analysis

4.1 Novelty of the CPIE

First of all, our main object of study is compromised accounts in social networks, and there are not many scholar works on this kind of research. From the micro perspective of social networks, we analyze the social behavior of each user in a given social network. Starting from the consistency of user behavior, we hypothesize that users are systems and use information entropy to measure the regularity of their behavior. This idea and the proposed method are novel.

4.2 Robustness of the CPIE

Numerous studies have shown that it is very difficult to imitate the behavior of others. At the same time, skilled imitations require costly resources, and go against the proper agenda of an attacker aiming to benefit the most from his intrusion effortlessly. This provides the conditions for us to use CPIE to detect compromised accounts.

In addition, our CPIE has greater applicability and robustness than other algorithms such as COMPA. Taking COMPA as an example, it only considers the probability of occurrence of various events, while we consider the stability of the system as a whole. Moreover, COMPA is not very adaptable to users because it only reconstructs the indicators after a certain period of time, and does not adapt well to the user's behavior and changes. CPIE can adjust the user's portrait based on users feedback, and the adaptability to users is stronger.

5 Experiment and Result Analysis

In this section, we evaluate the usefulness of our algorithm from three perspectives.

5.1 High-Profile Account

Studies have shown that popular corporate accounts tend to behave in a consistent manner [1]. Therefore, according to Alexa's statistical results, we selected the twitter's account of 40 companies as the target of our analysis. In the first set of experiments, we used the first 3,000 tweets from each account to build the user profile, and the last 200 as the test set to determine the false positives rate of the model. The experimental results are shown in Table 2.

From Table 2, we can find that in the experiments on the data of 40 official accounts, the average CPIE false alarm rate was 0.725%, and the method COMPA reached 11.3%. And from the perspective of a single account, the false positive rate of CPIE is less than COMPA. In addition, we can find that for these official accounts, their behavior is still very consistent. Figure 6 shows the changes in information entropy of the accounts @ap, @yahoo, @foxnews, @twitch and @bw in Table 2 during the detection process.

In the second set of experiments, we still used the first 3,000 Tweets of each account to build user portraits, but the test set was the last 200 Tweets of 10 accounts randomly obtained from the Twitter social network. The experimental results are shown in Table 3:

Through Table 3, we can find that when using the portraits of the 40 official accounts to detect information posted by other accounts, the average underreporting rate of CPIE is 5.71%, while the underreporting rate of COMPA has reached 0%. This is because COMPA The algorithm sacrifices the false alarm rate in exchange for a lower false alarm rate. In addition, this experiment also illustrates the usability of the model to a certain extent. Figure 7 shows the change of the information entropy of the @nih account in the detection process in Table 3.

5.2 Ordinary Account

In this section, we randomly obtained 15,000 users and 200 tweets from each of them through the Twitter Stream API. Of course this is only a small part of the entire social network. We evaluate the status of these accounts using CPIE and the results are shown in Table 4. Here, we manually tagged the account behavior of these 15,000 accounts. The CPIE is evaluated based on the results of the manual marking as a standard. We found that the false positive rate of our system is smaller than that of COMPA.

After our evaluation for this study was, we continued to track the relevant accounts and tweets. About 3 days or so, we found that more than 80% of the accounts have deleted the tweet or the account can no longer be accessed,as shown in Figure 8. (Of course we deleted the photos.) In addition, we also paid random attention to the normal accounts and found that basically all of them can be accessed normally.

6 Conclusions

In this paper, we presented a compromised accounts detection algorithm based on information entropy —— CIPE. We have the following contributions: first of all, we studied and proposed detection algorithms for compromised accounts. Compromised accounts have few detection algorithms, but their adverse effects in social networks are large. So it is necessary to study the algorithm for detecting compromised accounts.Then, we considered the stability of user behavior as a whole.Next, compared with methods such as COMPA, our algorithm can update its own mode with the increase of data and user feedback, and adapts better to user changes. And we again demonstrate the effectiveness of user behavior in detecting compromised accounts and verify the validity and usability of our algorithm through experiments.

Table 2: False positives

Account	False positives (rate)		Account	False positives (rate)		Account	False positives (rate)	
	CPIE	COMPA		CPIE	COMPA		CPIE	COMPA
@bw	0(0%)	0%	@bing	0(0%)	7%	@nfl	1(0.5%)	13%
@ap	0(0%)	0%	@cnn	0(0%)	9%	@google	1(0.5%)	14%
@msn	0(0%)	0%	@netflix	0(0%)	10%	@nih	1(0.5%)	12%
@reuters	0(0%)	0%	@yahooanswers	0(0%)	17%	@paypal	1(0.5%)	13%
@skype	0(0%)	0%	@instagram	0(0%)	23%	@imdb	2(1%)	10%
@latimes	0(0%)	0%	@wikipedia	0(0%)	25%	@ ebay	2(1%)	10%
@digg	0(0%)	1%	@bookingcom	0(0%)	44%	@weebly	2(1%)	16%
@facebook	0(0%)	1%	@twitter	0(0%)	46%	@youtube	4(2%)	10%
@yahoonews	0(0%)	2%	@guardian	0(0%)	47%	@yelp	6(3%)	19%
@amazon	0(0%)	2%	@yahoo	1(0.5%)	5%	@xe	8(4%)	12%
@ign	0(0%)	2%	@nytimes	1(0.5%)	7%	@yandexcom	8(4%)	15%
@9gag	0(0%)	3%	@stackfeed	1(0.5%)	10%	@microsoft	8(4%)	23%
@walmart	0(0%)	4%	@abcnews	1(0.5%)	2%	@linkedin	10(5%)	14%
@yahoosports	0(0%)	4%						

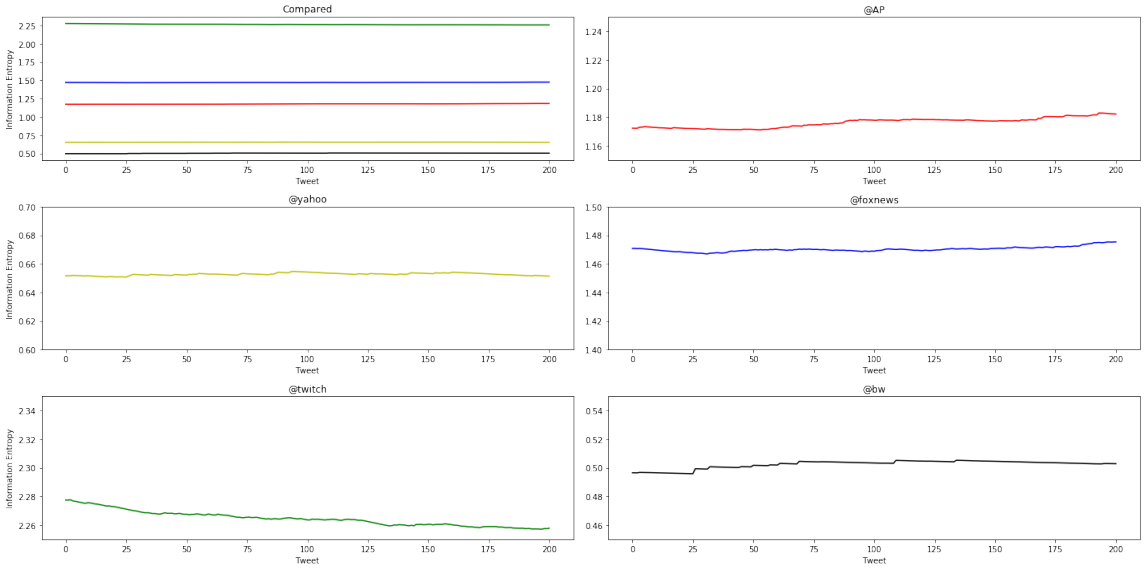


Figure 6: The graph of the tanh

Table 3: Underreport ratio

Account	Underreport ratio		Account	Underreport ratio		Account	Underreport ratio	
	CPIE	COMPA		CPIE	COMPA		CPIE	COMPA
@bw	0%	0%	@twitter	0%	0%	@nfl	1%	0%
@ap	0%	0%	@guardian	0%	0%	@yahoosports	4.50%	0%
@msn	0%	0%	@yahoo	0%	0%	@google	10%	0%
@reuters	0%	0%	@abcnews	0%	0%	@youtube	15%	0%
@skype	0%	0%	@nih	0%	0%	@yahooanswers	16%	0%
@latimes	0%	0%	@paypal	0%	0%	@cnn	17.50%	0%
@digg	0%	0%	@imdb	0%	0%	@facebook	19%	0%
@yahoonews	0%	0%	@weebly	0%	0%	@netflix	20%	0%
@amazon	0%	0%	@yelp	0%	0%	@nytimes	20%	0%
@walmart	0%	0%	@xe	0%	0%	@ ebay	20%	0%
@bing	0%	0%	@microsoft	0%	0%	@yandexcom	20%	0%
@instagram	0%	0%	@linkedin	0%	0%	@ign	27%	0%
@wikipedia	0%	0%	@stackfeed	1%	0%	@9gag	37.50%	0%
@bookingcom	0%	0%						

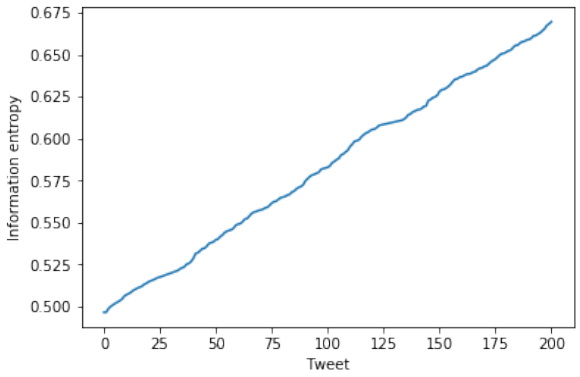


Figure 7: Bloomberg Businessweek's portrait and Twitch's tweets

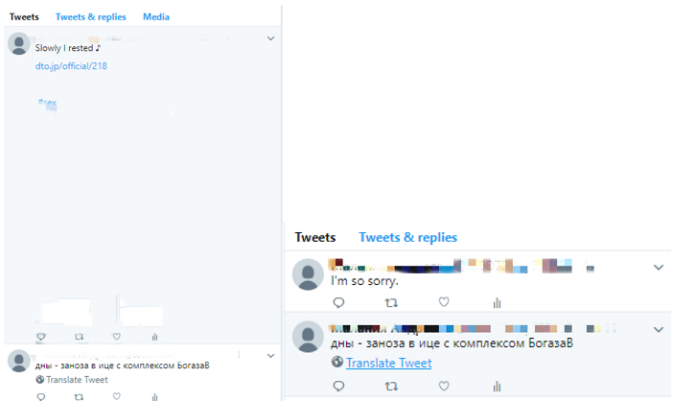


Figure 8: Account change

Table 4: Experimental result

	Result	
Accounts	15000	
Tweets	300000	
Similarity grouping	7328	
	Normal	Compromised
Manual marking	14814	186
CIPE detection	14808	192
CPIE False positives	3.13%	
COMPA detection	14806	194
COMPA False positives	4.12%	

However, there are some deficiencies in the algorithm, including the following: Nowadays, the content in the tweet may not match the theme provided; There is currently no public, standard, and excellent data set for Twitter which includes unusual accounts and suitable for scholar use. Existing data sets are intercepted according to different research needs; Our experiments are only done on a small piece of data on social networks, and do not show well the performance of the entire social network. We will further explore these issues in future research.

References

[1] K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan, and S. A. Razak, "Malicious accounts: Dark of the social networks," *Journal of Network and Computer Applications*, vol. 79, pp. 41–67, 2017.

[2] M. Alsaleh, A. Alarifi, A. M. Al-Salman, M. Alfayez, and A. Almuhaysin, "TSD: Detecting sybil accounts in twitter," in *The 13th International Conference*

on Machine Learning and Applications, pp. 463–469, 2014.

[3] S. Y. Bhat, M. Abulaish, and A. A. Mirza, "Spammer classification using ensemble methods over structural social network features," in *IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies (IAT'14)*, vol. 2, pp. 454–458, 2014.

[4] C. M. Chen, D. J. Guan, and Q. K. Su, "Feature set identification for detecting suspicious urls using bayesian classification in social networks," *Information Sciences*, vol. 289, pp. 133–147, 2014.

[5] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "COMPA: Detecting compromised accounts on social networks," in *NDSS*, 2013. (`http://www.people.vcu.edu/~cfung/bib/compromised_accounts_detection-ndss13.pdf`)

[6] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Towards detecting compromised accounts on social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 4, pp. 447–460, 2015.

[7] R. M. Eisa, M. Labib, and A. ElMougy, "SOS: Save our social network accounts," in *IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI'19)*, pp. 43–48, 2019.

[8] M. Fire, D. Kagan, A. Elyashar, and Y. Elovici, "Friend or foe? Fake profile identification in online social networks," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 194, 2014.

[9] K. Gani, H. Hacid, and R. Skraba, "Towards multiple identity detection in social networks," in *Proceedings of the 21st International Conference on World Wide Web*, pp. 503–504, 2012.

[10] S. R. Harsule and M. K. Nighot, "N-Gram classifier system to filter spam messages from OSN user wall," in *Innovations in Computer Science and Engineering*, pp. 21–28, 2016.

[11] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots+ machine learning," in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 435–442, 2010.

[12] S. Lee and J. Kim, "Early filtering of ephemeral malicious accounts on twitter," *Computer Communications*, vol. 54, pp. 48–57, 2014.

[13] P. C. Lin and P. M. Huang, "A study of effective features for detecting long-surviving twitter spam accounts," in *The 15th International Conference on Advanced Communications Technology (ICACT'13)*, pp. 841–846, 2013.

[14] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, "Twitter spammer detection using data stream clustering," *Information Sciences*, vol. 260, pp. 64–73, 2014.

[15] D. Mulamba, I. Ray, and I. Ray, "Sybilradar: A graph-structure based framework for sybil detection in on-line social networks," in *IFIP International Conference on ICT Systems Security and Privacy Protection*, pp. 179–193, 2016.

[16] K. Olmstead and A. Smith, "Americans and cybersecurity," *Pew Research Center*, vol. 26, pp. 311–327, 2017.

[17] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online social behaviors for compromised account detection," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 176–187, 2015.

[18] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," *Social Networks*, vol. 39, pp. 62–70, 2014.

[19] S. Shah, B. Shah, A. Amin, F. Al-Obeidat, F. Chow, F. J. L. Moreira, and S. Anwar, "Compromised user credentials detection in a digital enterprise using behavioral analytics," *Future Generation Computer Systems*, vol. 93, pp. 407–417, 2019.

[20] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[21] K. Thomas, F. Li, C. Grier, and V. Paxson, "Consequences of connectivity: Characterizing account hijacking on twitter," in *Proceedings of ACM SIGSAC Conference on Computer and Communications Security*, pp. 489–500, 2014.

[22] C. VanDam, J. Tang, and P. N. Tan, "Understanding compromised accounts on twitter," in *Proceedings of the International Conference on Web Intelligence*, pp. 737–744, 2017.

[23] A. Varpe and M. Mahajan, "Detecting twitter compromised accounts using anomalous user behavior," in *The 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT'19)*, vol. 1, pp. 22–25, 2019.

[24] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks," in *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, pp. 223–238, 2014.

[25] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, "Social turing tests: Crowdsourcing sybil detection," *Social and Information Networks*, 2012. (`arXiv:1205.3856v2`)

[26] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers," in *International Workshop on Recent Advances in Intrusion Detection*, pp. 318–337, 2011.

[27] Z. Yang, J. Xue, X. Yang, X. Wang, and Y. Dai, "Votetrust: Leveraging friend invitation graph to defend against social network sybils," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 4, pp. 488–501, 2015.

[28] E. Zangerle and G. Specht, ""sorry, i was hacked" a classification of compromised twitter accounts," in *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, pp. 587–593, 2014.

Biography

Yanpeng Cui biography. Yanpeng Cui is an associate professor in the Cyberspace Security at Xidian University. Her research interests are in network attack and defense, intelligent terminal security and protection, electronic warfare related equipment and the signal processing.

Kun Wang biography. Kun Wang is a Master Student in the Cyberspace Security at Xidian University. His research interests are in date privacy, data mining, machine learning and social networks.

Jianwei Hu biography. Jianwei Hu is an associate professor in the Cyberspace Security at Xidian University. His research interests are in network security, network confrontation, communications recon and communication countermeasure.

Wei Zhao biography. Wei Zhao is a Master Student in the Cyberspace Security at Xidian University. His research interests are in date privacy, data mining, machine learning and social networks.

Luming Feng biography. Luming Feng is a Master Student in the Cyberspace Security at Xidian University. His research interests are in date privacy, Web security, machine learning and malware detection.

Junjie Cui biography. Junjie Cui is a Master Student in the Cyberspace Security at Xidian University. Her research interests are in date privacy, Web security, machine learning.

Bound Estimation for Divisors of RSA Modulus with Small Divisor-ratio

Xingbo Wang
(Corresponding author: Xingbo Wang)
Department of Mechatronic Engineering, Foshan University
Guangdong Engineering Center of Information Security for Intelligent Manufacturing System, China
Email: xbwang@fosu.edu.cn; dr.xbwang@qq.com
(Received Nov. 16, 2019; Revised and Accepted Mar. 8, 2020; First Online Apr. 17, 2021)

Abstract

Through subtle analysis on relationships among ancestors, symmetric brothers, and the square root of a node on the T3 tree, the article puts a method forwards to calculate an interval that contains a divisor of a semiprime whose divisor-ratio is less than 3/2. Concrete mathematical reasonings to derive the method and programming procedure from realizing the calculations are shown in detail. Numerical experiments are made by applying the method on both ordinary small semiprimes and the RSA numbers. In the end, the paper makes predictions on the divisors' bounds of RSA232, RSA240, RSA250, and RSA260.

Keywords: Binary Tree; Bound Estimation; Cryptography; RSA Modulus

1 Introduction

Since the RSA numbers came into being, factoring an RSA number has been a challenge in the research of network security. In spite that reports of factoring certain RSA numbers have been made now and then, it is a fact that conventional and systematic methods to factorize RSA numbers in high efficiency are still under development [1, 4]. This is the reason why there are still researchers working on the issue of seeking new approaches to factorize large integers [5, 6]. In 2016, article [7] proposed an approach to study integers by putting the odd integers bigger than 1 on a full perfect binary tree from top to bottom and from left to right. Such a binary tree is called a $\mathbf{T}_3$ tree, as analyzed in article [8]. The $\mathbf{T}_3$ tree enables us to analyze an odd integer and its divisors' distribution on its layers (levels) and exhibits many new properties of semiprimes, especially in knowing a semiprime by means of analyzing its divisor-ratio that is calculated by the semiprime's big divisor divided by the small divisor. For example, article [9] proved that the small divisor of a semiprime would lie in an interval that is uniquely determined by the divisor-ratio and a bigger divisor-ratio could make it easier to find the small divisor; articles [2, 10] found out the distribution of an odd integer's square-root in the $\mathbf{T}_3$ tree; articles [11, 15, 17] investigated divisors' distribution of an RSA number, presenting in detail how two divisors distribute on the levels of the $\mathbf{T}_3$ tree in terms of their divisor-ratio.

This paper, following the studies in [11, 15, 17], and based on the inequalities proved in [12] as well as the theorems proved in [13, 16], gives in detail a bound estimation to the divisors of an RSA number. The results in this paper are helpful to design algorithm to search the divisors.

2 Preliminaries

2.1 Symbols & Notations

Symbol $\lfloor x \rfloor$ is the floor function, an integer function of real number x that satisfies inequality $x - 1 < \lfloor x \rfloor \leq x$, or equivalently $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$. Symbol $\{x\} = x - \lfloor x \rfloor$ is the fractional part of x.

A valuated binary tree T is such a binary tree that each of its nodes is assigned a value. An odd number N-rooted tree, denoted by T_N is a recursively constructed valuated binary tree whose root is the odd number N with $2N - 1$ and $2N + 1$ being the root's left and right sons, respectively. In this whole paper, symbol $T_{\mathbf{3}}$ is the $T_{\mathbf{3}}$ tree and symbol $N_{(k,j)}$ is by default the node at position j on level k of $T_{\mathbf{3}}$, where $k \geq 0$ and $0 \leq j \leq 2^k - 1$. Figure 1 shows the first 6 levels of the $T_{\mathbf{3}}$ tree. By using the asterisk wildcard *, symbol $N_{(k,*)}$ means a node lying on level k. Node $N_{(k,s)}$is called a brother of $N_{(k,j)}$ if $s \neq j$. $N_{(k,s)}$is called a symmetric brother of $N_{(k,j)}$ if $N_{(k,s)}$ and $N_{(k,j)}$ are at the symmetric positions in a subtree. An integer X is said to be clamped on level k of $T_{\mathbf{3}}$ if $2^{k+1} \leq X \leq 2^{k+2} - 1$ and use symbol $X \stackrel{\wedge}{=} k$ to indicate such a clamping relation. An odd integer O satisfying $2^{k+1} + 1 \leq O \leq 2^{k+2} - 1$ is said to be on level k of $T_{\mathbf{3}}$, and use symbol $O \stackrel{\Delta}{=} k$ to express it. Symbol $(p \stackrel{\circ}{=} q) = k$ means integers p and q are on the same level k or clamped

on the same level k. Symbol $p \doteqdot q$ means p and q separately lie on two different levels. Symbol $A \otimes B$ means A holds and simultaneously B holds, symbol $A \oplus B$ means A or B holds. Symbol $(a = b) > c$ means a takes the value of b and $a > c$. Symbol $A \Rightarrow B$ means conclusion B can be derived from condition A. Symbol $A <> B$ means the relation between A and B is uncertain. If a semiprime $N = pq$ satisfies $1 < p < q$ and $1 < \frac{q}{p} < \frac{3}{2}$, then N is said to be a *semiprime with small divisor-ratio*.

An odd interval $[a, b]$ is a set of consecutive odd numbers that take a as the lower bound and b as the upper bound, for example, $[3, 11] = \{3, 5, 7, 9, 11\}$. Intervals in this whole article are by default the odd ones unless particularly mentioned. Symbol *mod* is the modulo operation and expression $r = a \bmod b$ means $a \equiv r (\mathrm{mod} b)$.

2.2 Lemmas

Lemma 1. *(**See in [8]**). T_3 tree has the following fundamental properties.*

P1. *Every node is an odd integer and every odd integer bigger than 1 must be on T_3. Odd integer N with $N > 1$ lies on level $\lfloor \log_2 N \rfloor - 1$.*

P2. *$N_{(k,j)}$ is calculated by*

$$N_{(k,j)} = 2^{k+1} + 1 + 2j, j = 0, 1, ..., 2^k - 1$$

and thus

$$2^{k+1} + 1 \le N_{(k,j)} \le 2^{k+2} - 1.$$

P3. *Nodes $N_{(k+1,2j)}$ and $N_{(k+1,2j+1)}$ on level $k+1$ are respectively the left and right sons of node $N_{(k,j)}$. On level $(k+i)$ the descendants of $N_{(k,j)}$ are $N_{(k+i,2^i j+\omega)}$ with ω satisfying ($0 \le \omega \le 2^i - 1$),namely,*

$$N_{(k+i,2^i j)}, N_{(k+i,2^i j+1)}, N_{(k+i,2^i j+2)}, ..., N_{(k+i,2^i j+2^i-1)}$$

P4. *Multiplication of arbitrary two nodes of T_3, say $N_{(m,\alpha)}$ and $N_{(n,\beta)}$, is a third node of T_3. Let $J = 2^m(1+2\beta) + 2^n(1+2\alpha) + 2\alpha\beta + \alpha + \beta$; the multiplication $N_{(m,\alpha)} \times N_{(n,\beta)}$ is given by*

$$N_{(m,\alpha)} \times N_{(n,\beta)} = 2^{m+n+2} + 1 + 2J$$

If $0 \le J < 2^{m+n+1}$, then $N_{(m,\alpha)} \times N_{(n,\beta)} = N_{(m+n+1,J)}$ lies on level $m+n+1$; whereas, if $J \ge 2^{m+n+1}$,$N_{(m,\alpha)} \times N_{(n,\beta)} = N_{(m+n+2,\chi)}$ with $\chi = J - 2^{m+n+1}$ lying on level $m+n+2$.

Lemma 2. *(**See in [10]**).Let $N > 3$ be an odd integer and $k = \lfloor \log_2 N \rfloor - 1$; then $2^{\lfloor \frac{k+1}{2} \rfloor} \le \left\lfloor \sqrt{N} \right\rfloor < 2^{\lfloor \frac{k+1}{2} \rfloor + 1}$, namely, $\left\lfloor \sqrt{N} \right\rfloor \triangleq \lfloor \frac{k-1}{2} \rfloor$. Particularly, $(\left\lfloor \sqrt{N} \right\rfloor \le \left\lfloor 2^{\lfloor \frac{k+1}{2} \rfloor} \sqrt{2} \right\rfloor) \triangleq \lfloor \frac{k-1}{2} \rfloor$ when k is odd whereas $(\left\lfloor 2^{\lfloor \frac{k+1}{2} \rfloor} \sqrt{2} \right\rfloor \le \left\lfloor \sqrt{N} \right\rfloor) \triangleq \lfloor \frac{k-1}{2} \rfloor$ when k is even. If $N = pq$ with $1 < \frac{q}{p} < \alpha$ then $p \le \left\lfloor \sqrt{N} \right\rfloor < \alpha p \otimes \frac{q}{\alpha} < \left\lfloor \sqrt{N} \right\rfloor \le q$.*

Lemma 3. *(**See in [17]**).Let $N > 64$ be an odd integer and $k = \lfloor \log_2 N \rfloor - 1$, where divisors p and q satisfy $1 < p < q$ and $1 < \frac{q}{p} < 2$; then*

1) There are 3 possible cases in terms of the levels on which p and q lie; They are

$$\begin{cases} (p \triangleq \lfloor \frac{k+1}{2} \rfloor - 2) \otimes (q \triangleq \lfloor \frac{k+1}{2} \rfloor - 1) \\ (p \triangleq \lfloor \frac{k+1}{2} \rfloor - 1) \otimes (q \triangleq \lfloor \frac{k+1}{2} \rfloor - 1) \\ (p \triangleq \lfloor \frac{k+1}{2} \rfloor - 1) \otimes (q \triangleq \lfloor \frac{k+1}{2} \rfloor) \end{cases}$$

2) If $k > 2$ then

$$\left\lfloor \sqrt{N} \right\rfloor \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 1 \Rightarrow \left\lfloor \frac{\sqrt{N}}{2} \right\rfloor \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 2$$

In general, for arbitrary positive integers M and k it holds

$$M \triangleq k \Rightarrow \begin{cases} \lfloor \frac{M}{2} \rfloor \triangleq k - 1 \\ 2M \triangleq k + 1 \end{cases}$$

Lemma 4. *(**See in [11]**).For a semiprime $N = pq$ with $1 < p < q$ and $1 < \frac{q}{p} < \frac{3}{2}$, let $k = \lfloor \log_2 N \rfloor - 1$ then it holds for an even k*

$$2^{\lfloor \frac{k+1}{2} \rfloor} + 2^{\lfloor \frac{k+1}{2} \rfloor - 5} + 1 \le p \le \left\lfloor \sqrt{N} \right\rfloor$$

and

$$\left\lfloor \sqrt{N} \right\rfloor \le q \le 2^{\lfloor \frac{k+1}{2} \rfloor + 1} + 2^{\lfloor \frac{k+1}{2} \rfloor - 1} - 1$$

whereas it holds for an odd k

$$2^{\lfloor \frac{k+1}{2} \rfloor} - 2^{\lfloor \frac{k+1}{2} \rfloor - 2} + 1 \le p \le \left\lfloor \sqrt{N} \right\rfloor$$

and

$$\left\lfloor \sqrt{N} \right\rfloor \le q \le 2^{\lfloor \frac{k+1}{2} \rfloor + 1} - 1$$

Consequently,

$$((p \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 2) \oplus (p \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 1)) \otimes (q \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 1)$$

if k is odd, whereas

$$(p \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 1) \otimes ((q \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor - 1) \oplus (q \triangleq \left\lfloor \frac{k+1}{2} \right\rfloor))$$

if k is even.

Lemma 5. *(**See in [12]**). Let $N_{(m,\alpha)}$, $N_{(n,\beta)}$ be nodes of T_3 with $0 \le m \le n$, s be an integer with $0 \le s \le m$,$\Xi_1 = 2^{s+2} \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2+s}} \right\rfloor - 2 \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2}} \right\rfloor$ and $\Xi_2 = 2^{s+2} \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3+s}} \right\rfloor - 2 \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3}} \right\rfloor$; then*

it holds

$$\begin{aligned} N_{(m,\alpha)} &\leq \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+1}} \right\rfloor \\ &\leq 2N_{(m,\alpha)} - 1 \\ \frac{N_{(m,\alpha)} - 1}{2} - 1 &\leq \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2}} \right\rfloor \\ &\leq N_{(m,\alpha)} - 1 \\ \frac{N_{(m,\alpha)}}{2^2} - 2 &< \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3}} \right\rfloor \\ &\leq \frac{N_{(m,\alpha)} - 1}{2} \\ \frac{N_{(m,\alpha)} - 1}{2} - 2 &\leq 2\left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3}} \right\rfloor \\ &\leq N_{(m,\alpha)} - 1 \\ N_{(m,\alpha)} - 2^{s+2} + 1 &\leq \Xi_1 \leq 2N_{(m,\alpha)} - 1 \end{aligned}$$

and

$$\frac{N_{(m,\alpha)} - 1}{2} - 2^{s+2} \leq \Xi_2 \leq N_{(m,\alpha)} - 1.$$

Lemma 6. *(**See in [16]**).Let $N_{(k,j)}$ be a node in a perfect binary tree T with $k > 0$; then $N_{(k,j)}$ has k symmetric brothers on level k in T. If $A_1, A_2, ..., A_k$ are the father, the grandfather and the so-forth ancestors of $N_{(k,j)}$ respectively; then*

$$A_1 = N_{(k-1,\lfloor \frac{j}{2} \rfloor)}, ..., A_i = N_{(k-i,\lfloor \frac{j}{2^i} \rfloor)}, ..., A_k = N_{(0,\lfloor \frac{j}{2^k} \rfloor)}$$

And the k symmetric brothers of $N_{(k,j)}$ on level k of T are calculated by $N_{(k,2\lfloor \frac{j}{2} \rfloor + 1 - j \bmod 2)}$, $N_{(k,2^2\lfloor \frac{j}{2^2} \rfloor + 2^2 - 1 - j \bmod 2^2)}$, $\cdots$, $N_{(k,2^i\lfloor \frac{j}{2^i} \rfloor + 2^i - 1 - j \bmod 2^i)}$, $\cdots$, $N_{(k,2^k\lfloor \frac{j}{2^k} \rfloor + 2^k - 1 - j \bmod 2^k)}$.
Or equivalently, $N_{(k,2^{1+1}\lfloor \frac{j}{2} \rfloor + 1 - j)}$, $N_{(k,2^{2+1}\lfloor \frac{j}{2^2} \rfloor + 2^2 - 1 - j)}$, $\cdots$, $N_{(k,2^{i+1}\lfloor \frac{j}{2^i} \rfloor + 2^i - 1 - j)}$, $\cdots$, $N_{(k,2^{k+1}\lfloor \frac{j}{2^k} \rfloor + 2^k - 1 - j)}$.

Lemma 7. *(**See in [14]**).Properties of the floor functions with real numbers x and y, integers m, n and k.*

P1. $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor + 1$.

P2. $\lfloor x \rfloor - \lfloor y \rfloor - 1 \leq \lfloor x - y \rfloor \leq \lfloor x \rfloor - \lfloor y \rfloor < \lfloor x \rfloor - \lfloor y \rfloor + 1$.

P13. $x \leq y \Rightarrow \lfloor x \rfloor \leq \lfloor y \rfloor$.

P14. $\lfloor x \pm n \rfloor = \lfloor x \rfloor \pm n$.

P31. $i - 1 \leq 2\left\lfloor \frac{i}{2} \right\rfloor \leq i$ *with positive integer i.*

P32. *Let α and x be positive real numbers; Then it holds*

$$\alpha \lfloor x \rfloor - 1 < \lfloor \alpha x \rfloor < \alpha(\lfloor x \rfloor + 1).$$

Particularly, if α is a positive integer, say $\alpha = n$, then it yields

$$n \lfloor x \rfloor \leq \lfloor nx \rfloor \leq n(\lfloor x \rfloor + 1) - 1$$

Taking $n = 2$ yields

$$2 \lfloor x \rfloor \leq \lfloor 2x \rfloor \leq 2 \lfloor x \rfloor + 1$$

P34. *If k and x are positive, then*

$$0 \geq 2^k \left\lfloor \frac{x}{2^k} \right\rfloor - \lfloor x \rfloor \geq \begin{cases} 1 - 2^k, 0 \leq k \leq \lfloor \log_2 x \rfloor \\ -\lfloor x \rfloor, k > \lfloor \log_2 x \rfloor \end{cases}$$

3 Approach and Results

3.1 Approach: Train of Thought

Let $N = N_{(m,\alpha)} \times N_{(n,\beta)}$ with $0 \leq m \leq n$ be a composite odd integer whose divisors $N_{(m,\alpha)}$ and $N_{(m,\beta)}$ are to be found out. It can see with a simple observation on Figure 1 that, it is possible that $N_{(m,\alpha)}$ is N's ancestor on level m, e.g., $35 = 5 \times 7$, or $N_{(m,\alpha)}$ is a symmetric brother of N's ancestor on level m, e.g.,$77 = 7 \times 11$,where 5 is 77's ancestor on level 2 and 7 is the symmetric brother of 5. Accordingly, finding N's ancestor A_m on level m and then using A_m to evaluate $N_{(m,\alpha)}$ provide a clue to find out $N_{(m,\alpha)}$. By Lemmas 2,3 and 4, the three quantities $\left\lfloor \sqrt{N} \right\rfloor$, $N_{(m,\alpha)}$, and the number (of level) m, are closely related. This makes it possible to calculate $\left\lfloor \sqrt{N} \right\rfloor$ first by N, and then calculate m by $\left\lfloor \sqrt{N} \right\rfloor$, and then A_m by m under the situation that $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ are unknown. This forms the train of thought to the problem of evaluating $N_{(m,\alpha)}$ or finding out $N_{(m,\alpha)}$. Thereby, approaches related with calculating A_m, evaluating $N_{(m,\alpha)}$ are mandatory to investigate. The following subsections investigate symmetric paths, ancestors on path, symmetric brothers and the evaluation of $N_{(m,\alpha)}$ one by one. To meet the demands of mathematical reasonings in later sections, some necessary mathematical foundations (NMF) are also proved at the beginning.

3.2 Necessary Math Foundations

NMF 1. For positive integer k, it holds

(T1). $k \geq \left\lfloor \frac{k}{2} \right\rfloor + 1$ if $k > 2$.

(T2). $\left\lfloor \frac{k+1}{2} \right\rfloor - 1 \leq \left\lfloor \frac{k}{2} \right\rfloor \leq \left\lfloor \frac{k+1}{2} \right\rfloor$.

(T3). $k \leq 2\left\lfloor \frac{k+1}{2} \right\rfloor \leq k + 1$.

(T4). $\lfloor \alpha \lfloor x \rfloor \rfloor \leq \min(\lfloor \alpha x \rfloor, \alpha \lfloor x \rfloor)$ for positive numbers α and x; particularly, $\lfloor \alpha x \rfloor - 1 \leq \lfloor \alpha \lfloor x \rfloor \rfloor \leq \lfloor \alpha x \rfloor$ when $0 \leq \alpha < 1$.

Proof. First prove T1. By Lemma 7 (**P2**), direct calculation shows

$$k - \left(\left\lfloor \frac{k}{2} \right\rfloor + 1\right) = k - \left\lfloor \frac{k}{2} \right\rfloor - 1 \geq \left\lfloor k - \frac{k}{2} \right\rfloor - 1 = \left\lfloor \frac{k}{2} \right\rfloor - 1$$

Obviously, if $k > 2$, $\left\lfloor \frac{k}{2} \right\rfloor - 1 \geq 0$ and thus T1 holds. Now prove T2. Again by Lemma 7(**P2**), it yields

$$\left\lfloor \frac{k}{2} \right\rfloor - \left\lfloor \frac{k+1}{2} \right\rfloor \leq \left\lfloor \frac{k}{2} - \frac{k+1}{2} \right\rfloor + 1 = \left\lfloor -\frac{1}{2} \right\rfloor + 1 = 0$$

and

$$\left\lfloor \frac{k}{2} \right\rfloor - \left\lfloor \frac{k+1}{2} \right\rfloor \geq \left\lfloor \frac{k}{2} - \frac{k+1}{2} \right\rfloor = \left\lfloor -\frac{1}{2} \right\rfloor = -1$$

Hence T2 holds.
T3 is just a transformation of Lemma 7(**P**31). The proof of T4 can be obtained from the following deductions.

$$\begin{aligned} \lfloor \alpha \lfloor x \rfloor \rfloor &\leq \alpha \lfloor x \rfloor \\ \alpha \lfloor x \rfloor &\leq \alpha x \Rightarrow \lfloor \alpha \lfloor x \rfloor \rfloor \leq \lfloor \alpha x \rfloor \\ \lfloor \alpha \lfloor x \rfloor \rfloor &= \lfloor \alpha(x - \{x\}) \rfloor = \lfloor \alpha x - \alpha\{x\} \rfloor \\ &\geq \lfloor \alpha x \rfloor - \lfloor \alpha\{x\} \rfloor - 1 = \lfloor \alpha x \rfloor - 1. \end{aligned}$$

□

NMF 2. For integer $k \geq 0$ and integer $n > 1$, let a and b be given by

$$\begin{aligned} a &= 1 + 2\left\lfloor \frac{n}{2^k} \right\rfloor \\ b &= 1 + 2\left\lfloor \frac{n}{2^{k+1}} \right\rfloor \end{aligned}$$

then $b - 1 \leq \frac{a-1}{2} \leq b$ or $\frac{a-1}{2} \leq b \leq \frac{a+1}{2}$ and

$$\frac{3}{2} \leq \frac{a}{b} \leq \frac{5}{2}$$

under the condition $b \geq 2$.

Proof. By Lemma 7 (**P**32), direct calculation shows

$$\begin{aligned} \frac{a-1}{2} - b &= \left\lfloor \frac{n}{2^k} \right\rfloor - \left(2\left\lfloor \frac{n}{2^{k+1}} \right\rfloor + 1\right) \\ &= \left\lfloor \frac{n}{2^k} \right\rfloor - 2\left\lfloor \frac{n}{2^{k+1}} \right\rfloor - 1 \\ &\geq \left\lfloor \frac{n}{2^k} \right\rfloor - \left\lfloor \frac{2n}{2^{k+1}} \right\rfloor - 1 \\ &= -1 \end{aligned}$$

and

$$\frac{a-1}{2} - b = \left\lfloor \frac{n}{2^k} \right\rfloor - \left(2\left\lfloor \frac{n}{2^{k+1}} \right\rfloor + 1\right) \leq \left\lfloor \frac{n}{2^k} \right\rfloor - \left\lfloor \frac{2n}{2^{k+1}} \right\rfloor = 0$$

That is $b - 1 \leq \frac{a-1}{2} \leq b$, or $\frac{a-1}{2} \leq b \leq \frac{a+1}{2}$. Consequently it holds

$$2b - 1 \leq a \leq 2b + 1$$

Or equivalently,

$$2 - \frac{1}{b} \leq \frac{a}{b} \leq 2 + \frac{1}{b}$$

which immediately yields $\frac{3}{2} \leq \frac{a}{b} \leq \frac{5}{2}$ when $b \geq 2$. □

NMF 3. Suppose $N > 4$ is a positive integer; let $K = \lfloor \log_2 N \rfloor - 1$, $l = \left\lfloor \frac{K+1}{2} \right\rfloor - 1$, $A_1 = 1 + 2\left\lfloor \frac{N-1}{2^{l+2}} \right\rfloor$ and $A_2 = 1 + 2\left\lfloor \frac{N-1}{2^{l+3}} \right\rfloor$; then $\left\lfloor \sqrt{N} \right\rfloor - 1 \leq A_1 \leq \left\lfloor 2\sqrt{N} \right\rfloor + 1$, and $\left\lfloor \frac{\sqrt{N}}{2} \right\rfloor - 1 \leq A_2 \leq \left\lfloor \sqrt{N} \right\rfloor + 1$.

Proof. Since $N > 4$, $K > 1$ and $l > 0$. By definition of the floor function $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$ and Lemma 7 (**P**15), it holds

$$\begin{aligned} l &= \left\lfloor \frac{K+1}{2} \right\rfloor - 1 \Rightarrow l + 1 = \left\lfloor \frac{K+1}{2} \right\rfloor \\ &\Rightarrow 2^{l+1} = 2^{\left\lfloor \frac{K+1}{2} \right\rfloor} = 2^{\left\lfloor \frac{\lfloor log_2 N \rfloor}{2} \right\rfloor} = 2^{\lfloor \frac{log_2 N}{2} \rfloor} \\ &\Rightarrow \begin{cases} 2^{l+1} \leq 2^{\frac{\log_2 N}{2}} = \sqrt{N} \\ 2^{l+1} > 2^{\frac{\log_2 N}{2} - 1} = \frac{\sqrt{N}}{2} \end{cases} \end{aligned}$$

That is

$$l = \left\lfloor \frac{\lfloor \log_2 N \rfloor}{2} \right\rfloor - 1 \Rightarrow \begin{cases} \frac{\sqrt{N}}{2} < 2^{l+1} \leq \sqrt{N} \\ \sqrt{N} < 2^{l+2} \leq 2\sqrt{N} \end{cases}$$

Consequently, by Lemma 7 (**P**2), (**P**13) and (**P**32) it results in

$$\begin{aligned} &A_1 = 1 + 2\left\lfloor \tfrac{N-1}{2^{l+2}} \right\rfloor \geq \left\lfloor \tfrac{N-1}{2^{l+1}} \right\rfloor \geq \left\lfloor \tfrac{N-1}{\sqrt{N}} \right\rfloor \\ &\geq \left\lfloor \sqrt{N} - \tfrac{1}{\sqrt{N}} \right\rfloor \geq \left\lfloor \sqrt{N} \right\rfloor - 1 \end{aligned}$$

$$\begin{aligned} &A_1 = 1 + 2\left\lfloor \tfrac{N-1}{2^{l+2}} \right\rfloor \leq 1 + \left\lfloor \tfrac{N-1}{2^{l+1}} \right\rfloor \leq 1 + \left\lfloor \tfrac{2(N-1)}{\sqrt{N}} \right\rfloor \\ &\leq 1 + \left\lfloor 2\sqrt{N} - \tfrac{2}{\sqrt{N}} \right\rfloor \leq \left\lfloor 2\sqrt{N} \right\rfloor + 1 \end{aligned}$$

and

$$\begin{aligned} &A_2 = 1 + 2\left\lfloor \tfrac{N-1}{2^{l+3}} \right\rfloor \geq \left\lfloor \tfrac{N-1}{2^{l+2}} \right\rfloor \\ &\geq \left\lfloor \tfrac{(N-1)}{2\sqrt{N}} \right\rfloor \geq \left\lfloor \tfrac{\sqrt{N}}{2} \right\rfloor - 1 \end{aligned}$$

$$\begin{aligned} &A_2 = 1 + 2\left\lfloor \tfrac{N-1}{2^{l+3}} \right\rfloor \leq 1 + \left\lfloor \tfrac{N-1}{2^{l+2}} \right\rfloor \\ &\leq 1 + \left\lfloor \tfrac{N-1}{\sqrt{N}} \right\rfloor \leq 1 + \left\lfloor \sqrt{N} \right\rfloor \end{aligned}$$

□

3.3 Square Root Clamped on a Level

Proposition 1. Let $N_{(m,\alpha)}$, $N_{(m,\beta)}$ be two nodes of $\mathbf{T}_3$ with $0 \leq m \leq n$ and $0 \leq n - m \leq 1$; suppose $J = \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2} - 2^{m+n+1}$; then $\left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor \triangleq m$ if $0 \leq J \leq 2^{m+n+1} - 1$; $\left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor \triangleq n$ if $2^{m+n+1} \leq J \leq 2^{m+n+2} - 1$. Particularly, $\left\lfloor \sqrt{N_{(m,*)} \times N_{(m,*)}} \right\rfloor \triangleq m$ for an arbitrary $m \geq 0$.

Proof. The condition $0 \leq m \leq n$ and $0 \leq n - m \leq 1$ means

$$0 \leq n - 1 \leq m \leq n$$

Then by $0 \leq J \leq 2^{m+n+1} - 1$, it knows $N_{(m,\alpha)} \times N_{(n,\beta)}$ lies on level $m + n + 1$; Thus

$$\begin{aligned} 2^{m+n+2} + 1 &\leq N_{(m,\alpha)} \times N_{(n,\beta)} \leq 2^{m+n+3} - 1 \\ \Rightarrow 2^{2m+2} &< N_{(m,\alpha)} \times N_{(n,\beta)} < 2^{m+n+3} \leq 2^{m+(m+1)+3} \\ \Rightarrow 2^{m+1} &\leq \left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor < 2^{m+2} \\ &\Rightarrow \left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor \triangleq m \end{aligned}$$

Similarly, $2^{m+n+1} \le J \le 2^{m+n+2}-1$ yields $N_{(m,\alpha)} \times N_{(n,\beta)}$ lies on level $m+n+2$; Hence

$$\begin{aligned} 2^{m+n+3}+1 &\le N_{(m,\alpha)} \times N_{(n,\beta)} \le 2^{m+n+4}-1 \\ \Rightarrow 2^{2n+2} &< N_{(m,\alpha)} \times N_{(n,\beta)} \le 2^{2n+4}-1 < 2^{2n+4} \\ \Rightarrow 2^{n+1} &\le \left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor < 2^{n+2} \\ \Rightarrow & \left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor \triangleq n \end{aligned}$$

Now direct calculation shows

$$N_{(m,*)} \times N_{(m,*)} \ge (2^{m+1}+1)^2 > 2^{2(m+1)}$$

and

$$N_{(m,*)} \times N_{(m,*)} \le (2^{m+2}-1)^2 < 2^{2(m+2)}$$

Hence

$$\begin{aligned} 2^{m+1} &< \sqrt{N_{(m,*)} \times N_{(m,*)}} < 2^{m+2} \\ \Rightarrow 2^{m+1} &\le \left\lfloor \sqrt{N_{(m,*)} \times N_{(m,*)}} \right\rfloor < 2^{m+2} \\ \Rightarrow 2^{m+1} &\le \left\lfloor \sqrt{N_{(m,*)} \times N_{(m,*)}} \right\rfloor \le 2^{m+2}-1 \\ \Rightarrow & \left\lfloor \sqrt{N_{(m,*)} \times N_{(n,*)}} \right\rfloor \triangleq m \end{aligned}$$

□

Corollary 1. *Let $N_{(m,\alpha)}$ and $N_{(m,\beta)}$ be two nodes of $\boldsymbol{T}_3$ with $0 \le m \le n$ and $0 \le n-m \le 1$. If $2^{m+n+2}+1 \le N_{(m,\alpha)} \times N_{(n,\beta)} \le 2^{m+n+3}-1$ then $\left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor \triangleq m$; whereas if $2^{m+n+3}+1 \le N_{(m,\alpha)} \times N_{(n,\beta)} \le 2^{m+n+4}-1$, $\left\lfloor \sqrt{N_{(m,\alpha)} \times N_{(n,\beta)}} \right\rfloor \triangleq n$. Particularly, $\left\lfloor \sqrt{N_{(m,*)} \times N_{(m,*)}} \right\rfloor \triangleq m$ for an arbitrary $m \ge 0$.*

Proof. Let $J = \frac{N_{(m,\alpha)} \times N_{(n,\beta)}-1}{2} - 2^{m+n+1}$. Then direct calculations show

$$\begin{aligned} N_{(m,\alpha)} \times N_{(n,\beta)} &\ge 2^{m+n+2}+1 \\ \Rightarrow \frac{N_{(m,\alpha)} \times N_{(n,\beta)}-1}{2} - 2^{m+n+1} &\ge 0 \\ \Rightarrow J &\ge 0 \end{aligned}$$

$$\begin{aligned} N_{(m,\alpha)} \times N_{(n,\beta)} &\le 2^{m+n+3}-1 \\ \Rightarrow J - 2^{m+n+1} &\le \frac{2^{m+n+3}-2}{2} - 2^{m+n+2} = -1 \\ \Rightarrow J &\le 2^{m+n+1}-1 \end{aligned}$$

and

$$\begin{aligned} N_{(m,\alpha)} \times N_{(n,\beta)} &\ge 2^{m+n+3}+1 \\ \Rightarrow J - 2^{m+n+1} &\ge \frac{2^{m+n+3}}{2} - 2^{m+n+2} = 0 \\ \Rightarrow J &\ge 2^{m+n+1} \end{aligned}$$

$$\begin{aligned} N_{(m,\alpha)} \times N_{(n,\beta)} &\le 2^{m+n+4}-1 \\ \Rightarrow J - 2^{m+n+2} &= \frac{N_{(m,\alpha)} \times N_{(n,\beta)}-1}{2} - 2^{m+n+1} - 2^{m+n+2} \\ &\le \frac{2^{m+n+4}-2}{2} - 2^{m+n+1} - 2^{m+n+2} \\ &= -1 \\ \Rightarrow J &\le 2^{m+n+2}-1 \end{aligned}$$

□

By Proposition 1 it knows the corollary holds.

Corollary 1*. *For a semiprime $N = pq$ with $1 < p < q$ and $1 < \frac{q}{p} < \frac{3}{2}$, let $K = \lfloor \log_2 N \rfloor - 1$; then $K-1 \equiv 0(\mathrm{mod}2) \Rightarrow \left\lfloor \sqrt{N} \right\rfloor \triangleq q$ otherwise $\left\lfloor \sqrt{N} \right\rfloor \triangleq p$.*

Proof. Let $K_{sqrtN} = \left\lfloor \frac{K+1}{2} \right\rfloor - 1$; then by Lemma 2, $\left\lfloor \sqrt{N} \right\rfloor \triangleq K_{sqrtN}$; by Lemma 4 it holds

$$\begin{aligned} & K-1 \equiv 0(\mathrm{mod}2) \\ & \Rightarrow ((p \triangleq K_{sqrtN}-1) \oplus (p \triangleq K_{sqrtN})) \otimes (q \triangleq K_{sqrtN}) \end{aligned}$$

whereas

$$\begin{aligned} & K \equiv 0(\mathrm{mod}2) \\ & \Rightarrow (p \triangleq K_{sqrtN}) \otimes ((q \triangleq K_{sqrtN}) \oplus (q \triangleq K_{sqrtN}+1)) \end{aligned}$$

Hence, it is sure $K-1 \equiv 0(\mathrm{mod}2) \Rightarrow \left\lfloor \sqrt{N} \right\rfloor \triangleq q$ and $K \equiv 0(\mathrm{mod}2) \Rightarrow \left\lfloor \sqrt{N} \right\rfloor \triangleq p$. □

3.4 Quotient vs. Lying levels of Two Nodes

Proposition 2. *Let $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ be two nodes of $\boldsymbol{T}_3$ with $0 \le m \le n$; if $n-m > 1$, then $\frac{N_{(n,\beta)}}{N_{(m,\alpha)}} > 2$.*

Proof. Without loss of generality, assume $n = m+\alpha$ with integer $\alpha \ge 2$; then $N_{(m,2^m-1)} = 2^{m+2}-1$ is the biggest node on level m and $N_{(n,0)} = 2^{n+1}+1 = 2^{m+1+\alpha}+1$ is the smallest node on level n. Since $N_{(n,0)} \ge 2^{m+3}+1$, it immediately leads to $\frac{N_{(n,0)}}{N_{(m,2^m-1)}} \ge \frac{2^{m+3}+1}{2^{m+2}-1} > 2$. □

Corollary 2. *Let $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ be two nodes of $\boldsymbol{T}_3$ with $0 \le m \le n$; if $1 \le \frac{N_{(n,\beta)}}{N_{(m,\alpha)}} < 2$ then $0 \le n-m \le 1$.*

Proof. Since $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ are odd integers, $\frac{N_{(n,\beta)}}{N_{(m,\alpha)}} = 2$ is impossible. This fact and Proposition 2 plus using the proof by contradiction immediately lead to the claimed conclusion. □

3.5 Symmetric Path

Theorem 1. *Let $N_{(k,j)}$ be a node of $\boldsymbol{T}_3$ with $k > 0$; then there are $\lfloor \log_2 N_{(k,j)} \rfloor - 1$ symmetric paths connecting $N_{(k,j)}$ and its symmetric nodes; there are totally $\lfloor \log_2 N_{(k,j)} \rfloor^2 - 1$ nodes on all its symmetric paths.*

Proof. The first conclusion is proved by Lemma 1(**P1**) and $k = \lfloor \log_2 N_{(k,j)} \rfloor - 1$. Note that, there are 3 nodes on the symmetric path in its father's tree, 5 nodes on the symmetric path in its grandfather's tree, and so on. Hence the number of its total symmetric nodes is given by

$$\begin{aligned} S_{N_{(k,j)}} &= 3 + 5 + ... + 2(\lfloor \log_2 N_{(k,j)} \rfloor - 1) + 1 \\ &= \lfloor \log_2 N_{(k,j)} \rfloor^2 - 1. \end{aligned}$$

□

3.6 Ancestors on Path

By $N_{(m,\alpha)} \times N_{(n,\beta)} = 2^{m+n+2} + 1 + 2J$ it leads to Equation (1), namely,

$$J = \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2} - 2^{m+n+1} \tag{1}$$

By Lemma 1 (**P3**), it knows that, on level m, $N_{(m,\lfloor \frac{J}{2^{n+1}} \rfloor)}$ is the ancestor of $N_{(m,\alpha)} \times N_{(n,\beta)}$ if $J < 2^{m+n+1}$, whereas, $N_{(m,\lfloor \frac{J-2^{m+n+1}}{2^{n+2}} \rfloor)}$ is the ancestor if $J \geq 2^{m+n+1}$.

Example 1.

Take $N_{(m,\alpha)} = N_{(2,0)} = 9$ *and* $N_{(n,\beta)} = N_{(2,2)} = 13$*; then* $J = \frac{9\times 13 - 2^{2+2+2}-1}{2} = 26 < 2^{2+2+1}$*; hence* $\vartheta = \lfloor \frac{26}{2^{2+1}} \rfloor = 3$ *and* $N_{(2,3)} = 15$ *is surely the ancestor of* $N_{(2,0)} \times N_{(2,2)} = 9 \times 13 = 117$ *on level 2, as seen in Figure 1.*

Take $N_{(m,\alpha)} = N_{(1,1)} = 7$ *and* $N_{(n,\beta)} = N_{(2,1)} = 11$*; then* $J = \frac{7\times 11 - 2^{1+2+2}-1}{2} = 22 > 2^{1+2+1}$*; hence* $\vartheta = \left\lfloor \frac{22-2^{2+1+1}}{2^{2+2}} \right\rfloor = 0$ *and* $N_{(1,0)} = 5$ *is surely the ancestor of 77 on level 1, as seen in Figure 1.*

Now let

$$\vartheta = \begin{cases} \lfloor \frac{J}{2^{n+1}} \rfloor = \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2}} - 2^m \right\rfloor, J < 2^{m+n+1} \\ \lfloor \frac{J-2^{m+n+1}}{2^{n+2}} \rfloor = \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3}} - 2^m \right\rfloor, J \geq 2^{m+n+1} \end{cases} \tag{2}$$

By Lemma 2, it knows that,

$$N_{(m-1,\lfloor \frac{\vartheta}{2^1} \rfloor)}, N_{(m-2,\lfloor \frac{\vartheta}{2^2} \rfloor)}, \cdots, N_{(m-\sigma,\lfloor \frac{\vartheta}{2^\sigma} \rfloor)}, \cdots, N_{(0,0)} \tag{3}$$

are *direct ancestors* of $N_{(m,\alpha)} \times N_{(n,\beta)}$ on levels $m-1, m-2, \cdots, m-\sigma, \cdots$, and level 0 respectively. The ancestor A_m that lies on the same level as $N_{(m,\alpha)}$ lies on is particularly called a *homolayer ancestor* of $N_{(m,\alpha)} \times N_{(n,\beta)}$. Accordingly, the following Theorem 2 is true.

Theorem 2. *Let* $N_{(m,\alpha)}$ *and* $N_{(n,\beta)}$ *be two nodes of* $\boldsymbol{T}_3$ *with* $0 \leq m \leq n$*; then the ancestors of* $N_{(m,\alpha)} \times N_{(n,\beta)}$ *on level* $m, m-1, \cdots, m-\sigma$*, and so forth to level 0, denoted by* $A_m, A_{m-1}, \cdots, A_{m-\sigma}, \cdots, A_0$ *respectively, are calculated by*

$$A_{m-s} = 2^{m-s+1} + 1 + 2\Omega_s, s = 0, 1, 2, \cdots, m \tag{4}$$

where

$$\Omega_s = \left\lfloor \frac{\vartheta}{2^s} \right\rfloor = \begin{cases} \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2+s}} \right\rfloor - 2^{m-s}, J < 2^{m+n+1} \\ \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3+s}} \right\rfloor - 2^{m-s}, J \geq 2^{m+n+1} \end{cases} \tag{5}$$

Proof. By Equations (4) and (5) it knows that, the quantity $\lfloor \frac{\vartheta}{2^s} \rfloor_{(s=1,2,...,m)}$ is critical to determine the ancestors in sequence (3). By Equation (2) it holds when $J < 2^{m+n+1}$

$$\begin{aligned} \lfloor \tfrac{\vartheta}{2^s} \rfloor_{(s=1,2,...,m)} &= \lfloor \tfrac{J}{2^{n+1+s}} \rfloor_{(s=1,2,...,m)} \\ &= \left\lfloor \frac{\frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1 - 2^{m+n+2}}{2}}{2^{n+1+s}} \right\rfloor_{(s=1,2,...,m)} \\ &= \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2+s}} - 2^{m-s} \right\rfloor_{(s=1,2,...,m)} \end{aligned}$$

likewise, when $J \geq 2^{m+n+1}$

$$\begin{aligned} \lfloor \tfrac{\vartheta}{2^s} \rfloor_{(s=0,1,...,m)} &= \left\lfloor \tfrac{J-2^{m+n+1}}{2^{n+2+s}} \right\rfloor_{(s=0,1,...,m)} \\ &= \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3+s}} - 2^{m-s} \right\rfloor_{(s=0,1,...,m)} \end{aligned}$$

Letting $\Omega_s|_{(s=0,2,...,m)} = \lfloor \frac{\vartheta}{2^s} \rfloor_{(s=0,2,...,m)}$ immediately results in Equation (5). And then the ancestors of $N_{(m,\alpha)} \times N_{(n,\beta)}$ on level $m-s$ with $s = 0, 1, 2, \cdots, m$ respectively are given by Equation (4). □

3.7 Symmetric Brothers

By Lemma 6, it knows that, $N_{(m,2\Omega_1+1-\vartheta \bmod 2)}$, $N_{(m,2^2\Omega_2+2^2-1-\vartheta \bmod 2^2)}$, $\cdots$, $N_{(m,2^i\Omega_i+2^i-1-\vartheta \bmod 2^i)}$, $\cdots$, $N_{(m,2^k\Omega_k+2^k-1-\vartheta \bmod 2^k)}$ are symmetric brothers of A_m on level m corresponding to the ancestors $A_{m-s} (1 \leq s \leq m)$ respectively. For convenience, denote $N_{(m,2^i\Omega_i+2^i-1-\vartheta \bmod 2^i)}$ by $B_{(m,i)}$, namely

$$B_{(m,i)} = N_{(m,2^i\Omega_i+2^i-1-\vartheta \bmod 2^i)} = N_{(m,2^{i+1}\Omega_i+2^i-1-\vartheta)}$$

and let Ξ_1 and Ξ_2 as defined in Lemma 5. Then

$$\begin{aligned} B_{(m,s)} &= N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)} \\ &= 2^{m+1} + 1 + 2(2^{s+1}\Omega_s + 2^s - 1 - \vartheta) \\ &= 2^{m+1} + 2^{s+1} - 1 + 2^{s+2}\Omega_s - 2\vartheta. \end{aligned}$$

By Equation (5) it knows that, when $J < 2^{m+n+1}$,

$$\begin{aligned} &B_{(m,s)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)} \\ &= 2^{m+1} + 2^{s+1} - 1 + 2^{s+2}(\left\lfloor \tfrac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2+s}} \right\rfloor - 2^{m-s}) \\ &\quad - 2(\left\lfloor \tfrac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2}} \right\rfloor - 2^m) \\ &= 2^{m+1} + 2^{s+1} - 1 + 2^{s+2} \left\lfloor \tfrac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2+s}} \right\rfloor \\ &\quad - 2^{m+2} - 2 \left\lfloor \tfrac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2}} \right\rfloor + 2^{m+1} \\ &= 2^{s+1} - 1 + \Xi_1 \end{aligned}$$

whereas when $J \geq 2^{m+n+1}$

$$\begin{aligned} B_{(m,s)} &= N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)} \\ &= 2^{s+1} - 1 + \Xi_2 \end{aligned} \tag{6}$$

Then it is natural to lead to the following Theorem 3.

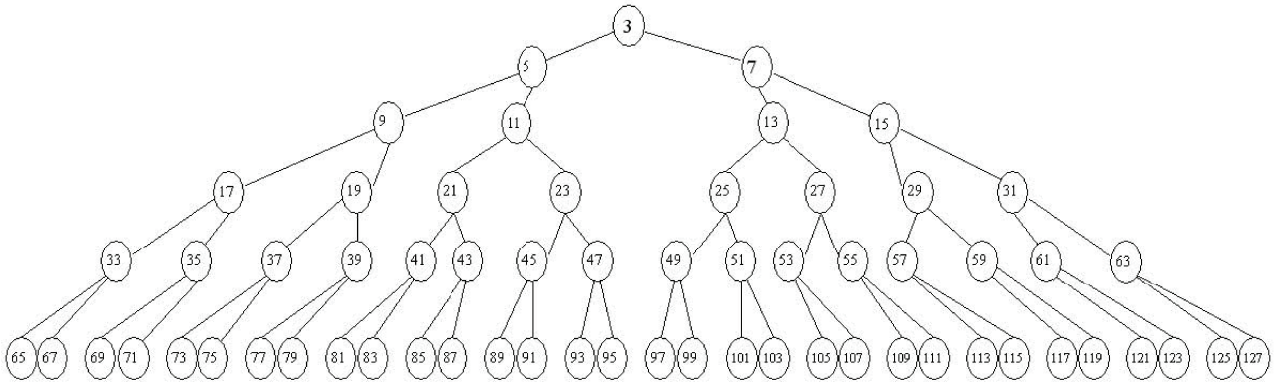


Figure 1: T_3 tree from level 0 to level 5

Proof. Let $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ be two nodes of $\mathbf{T}_3$ with $0 \le m \le n$ and A_m be the ancestors of $N_{(m,\alpha)} \times N_{(n,\beta)}$ on level m; then the symmetric brothers of A_m, denoted by $B_{(m,1)}, B_{(m,2)}, ..., B_{(m,s)}$, and so forth to $B_{(m,m)}$ respectively, are calculated by

$$B_{(m,s)} = 2^{s+1} - 1 + \begin{cases} \Xi_1, J < 2^{m+n+1} \\ \Xi_2, J \ge 2^{m+n+1} \end{cases}, s = 1, 2, ..., m \quad (7)$$

□

Example 2. *Take* $N_{(m,\alpha)} = N_{(2,1)} = 11$ *and* $N_{(n,\beta)} = N_{(2,2)} = 13$*; then* $m = n = 2$, $N_{(2,1)} \times N_{(2,2)} = 143 = N_{(6,7)}$ *and* $J = \frac{11\times13-2^{2+2+2}-1}{2} = 39 > 2^{2+2+1}$*; Hence* $\vartheta = \left\lfloor \frac{J-2^{m+n+1}}{2^{n+2}} \right\rfloor = \left\lfloor \frac{39-2^{2+2+1}}{2^{2+2}} \right\rfloor = 0$ *and* $N_{(2,0)} = 9$ *is surely the ancestor of* $N_{(6,7)}$ *on level 2; Consequently,*$N_{(1,0)} = 5$ *and* $N_{(0,0)} = 3$ *are ancestors on level 1 and level 0 respectively. Since*

$$\Omega_s = \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+3+s}} \right\rfloor - 2^{m-s} = \begin{cases} \left\lfloor \frac{143-1}{2^{2+3+1}} \right\rfloor - 2^{2-1} = 0, s = 1 \\ \left\lfloor \frac{143-1}{2^{2+3+2}} \right\rfloor - 2^{2-2} = 0, s = 2 \end{cases}$$

and $\vartheta = 0$*, it leads to*

$$2^{s+1}\Omega_s + 2^s - 1 - \vartheta = 2^{s+1}\Omega_s + 2^s - 1 = \begin{cases} 1, s = 1 \\ 3, s = 2 \end{cases}$$

Hence

$$B_{(2,1)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}|_{m=2,s=1} = N_{(2,1)} = 11$$

and

$$B_{(2,2)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}|_{m=2,s=2} = N_{(2,3)} = 15$$

are the two symmetric brothers of $N_{(2,0)} = 9$ corresponding to the ancestors $N_{(1,0)} = 5$ and $N_{(0,0)} = 3$, respectively.

Example 3. *Take* $N_{(n,\beta)} = N_{(1,1)} = 5$ *and* $N_{(m,\alpha)} = N_{(4,4)} = 41$*; then* $m = 1$, $n = 4$,$N_{(1,1)} \times N_{(4,4)} = 205 = N_{(6,38)}$ *and* $J = \frac{41\times5-1}{2} - 2^{4+1+1} = 38 < 2^{4+1+1}$*; hence* $\vartheta = \left\lfloor \frac{J}{2^{m+1}} \right\rfloor = \left\lfloor \frac{38}{2^{1+1}} \right\rfloor = 9$ *and* $N_{(4,9)} = 51$ *is the ancestor of* $N_{(6,38)}$ *on level 4. Consequently,* $N_{(3,4)} = 25$, $N_{(2,2)} = 13$, $N_{(1,1)} = 7$ *and* $N_{(0,0)} = 3$ *are ancestors on level 3 and level 0 respectively. Since*

$$\Omega_s = \left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{n+2+s}} \right\rfloor - 2^{m-s} = \begin{cases} = 4, s = 1 \\ = 2, s = 2 \\ = 1, s = 3 \\ = 0, s = 4 \end{cases}$$

and $\vartheta = 9$*, it leads to*

$$2^{s+1}\Omega_s + 2^s - 1 - \vartheta = 2^{s+1}\Omega_s + 2^s - 1 - 9 = \begin{cases} 2^{1+1} \times 4 + 2^1 - 1 - 9 = 8, s = 1 \\ 2^{2+1} \times 2 + 2^2 - 1 - 9 = 10, s = 2 \\ 2^{3+1} \times 1 + 2^3 - 1 - 9 = 14, s = 3 \\ 2^{4+1} \times 0 + 2^4 - 1 - 9 = 6, s = 4 \end{cases}$$

Hence

$$B_{(4,1)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}|_{m=4,s=1} = N_{(4,8)} = 49$$
$$B_{(4,2)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}|_{m=4,s=2} = N_{(4,10)} = 53$$
$$B_{(4,3)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}|_{m=4,s=3} = N_{(4,14)} = 61$$
$$B_{(4,4)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}|_{m=4,s=4} = N_{(4,6)} = 45$$

are four symmetric brothers of $N_{(4,9)} = 51$ corresponding to the ancestors $N_{(3,4)} = 25$, $N_{(2,2)} = 13$, $N_{(1,1)} = 7$ and $N_{(0,0)} = 3$ respectively.

Example 4. *Take* $N_{(n,\beta)} = N_{(1,1)} = 5$ *and* $N_{(m,\alpha)} = N_{(4,4)} = 41$*; then* $m = 1, n = 4$,$N_{(1,1)} \times N_{(4,4)} = 205 = N_{(6,38)}$ *and* $J = \frac{41\times5-1}{2} - 2^{4+1+1} = 38 < 2^{4+1+1}$*; hence* $\vartheta = \left\lfloor \frac{J}{2^{m+1}} \right\rfloor = \left\lfloor \frac{38}{2^{4+1}} \right\rfloor = 1$ *and* $N_{(2,1)} = 7$ *is the ancestor of* $N_{(6,38)}$ *on level 2. Consequently,* $N_{(0,0)} = 3$ *is the ancestor on level 0. Since*

$$\Omega_s|_{s=1} = \left(\left\lfloor \frac{N_{(m,\alpha)} \times N_{(n,\beta)} - 1}{2^{m+2+s}} \right\rfloor - 2^{n-s}\right)|_{s=1} = \left\lfloor \frac{205-1}{2^{4+2+1}} \right\rfloor - 2^{1-1} = 0$$

and $\vartheta = 1$*, it leads to*

$$2^{s+1}\Omega_s + 2^s - 1 - \vartheta|_{s=1} = 2^{s+1}\Omega_s + 2^s - 1 - 1|_{s=1} = 2^{1+1} \times 0 + 2^1 - 1 - 1 = 0$$

Hence

$$B_{(1,1)} = N_{(m,2^{s+1}\Omega_s+2^s-1-\vartheta)}\big|_{m=1,s=1} = N_{(1,0)} = 5$$

is the symmetric brother of $N_{(1,1)} = 7$ corresponding to the ancestors $N_{(0,0)} = 3$.

3.8 Bounds of Symmetric Brothers

Nodes $N_{(m,\alpha)}$,$A_m(= A_{m-0}$ defined by Equation (4)), $B_{(m,1)}$, $B_{(m,2)}$, $\cdots$, and $B_{(m,m)}$, which are defined by Equation (7), are all nodes on level m of $\mathbf{T}_3$. It is of course necessary to make clear their relative positions. This can surely be done by means of estimating their bounds.

Theorem 3. *Let $A_m = 2^{m+1}+1+2\Omega_0 = 2^{m+1}+1+2\vartheta$; Then*

$$\begin{cases} N_{(m,\alpha)} - 2 \le A_m \le 2N_{(m,\alpha)} - 1, J < 2^{m+n+1} \\ \frac{N_{(m,\alpha)}-1}{2} - 1 \le A_m \le N_{(m,\alpha)}, J \ge 2^{m+n+1} \end{cases}$$

or equivalently

$$\begin{cases} \frac{A_m-1}{2} \le N_{(m,\alpha)} \le A_m + 2, J < 2^{m+n+1} \\ A_m \le N_{(m,\alpha)} \le 2A_m + 3, J \ge 2^{m+n+1} \end{cases}$$

Proof. Direct calculation yields

$$A_m = 2^{m+1}+1+2\begin{cases} \left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+2}} \right\rfloor - 2^m, J < 2^{m+n+1} \\ \left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+3}} \right\rfloor - 2^m, J \ge 2^{m+n+1} \end{cases}$$

That is

$$A_m = 1+2\begin{cases} \left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+2}} \right\rfloor, J < 2^{m+n+1} \\ \left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+3}} \right\rfloor, J \ge 2^{m+n+1} \end{cases}$$

By Lemma 5, it knows

$$\begin{cases} N_{(m,\alpha)} - 2 \le A_m \le 2N_{(m,\alpha)} - 1, J < 2^{m+n+1} \\ \frac{N_{(m,\alpha)}-1}{2} - 1 \le A_m \le N_{(m,\alpha)}, J \ge 2^{m+n+1} \end{cases}$$

or equivalently

$$\begin{cases} \frac{A_m+1}{2} \le N_{(m,\alpha)} \le A_m + 2, J < 2^{m+n+1} \\ A_m \le N_{(m,\alpha)} \le 2A_m + 3, J \ge 2^{m+n+1} \end{cases}$$

□

Proposition 3. Let $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ be two nodes of $\mathbf{T}_3$ with $0 \le m \le n$, A_m be the ancestors of $N_{(m,\alpha)} \times N_{(n,\beta)}$ on level m and $B_{(m,1)}, B_{(m,2)}, ..., B_{(m,s)}$ be the symmetric brothers of A_m; Then

$$\begin{aligned} -2^{s+1} &\le B_{(m,s+1)} - B_{(m,s)} \le 2^{s+1}, \\ & \qquad s = 1, 2, ..., m-1 \\ -2^{s+1} &\le B_{(m,s)} - A_m \le 2^{s+1} - 4. \end{aligned}$$

Proof. Direct calculation by Equation (6) yields

$$B_{(m,s+1)} - B_{(m,s)} = 2^{s+1} + 2^{s+2}(2\Omega_{s+1} - \Omega_s)$$

Let $\Theta^i_{n+2} = \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+2+i}}$, $\Theta^i_{n+3} = \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+3+i}}$,$\Xi_3 = 2\left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+3+s}} \right\rfloor - \left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+2+s}} \right\rfloor$ and $\Xi_4 = 2\left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+4+s}} \right\rfloor - \left\lfloor \frac{N_{(m,\alpha)}\times N_{(n,\beta)}-1}{2^{n+3+s}} \right\rfloor$; then it holds for $s = 1, 2, ..., m-1$

$$2\Omega_{s+1} - \Omega_s = \begin{cases} \Xi_3, J < 2^{m+n+1} \\ \Xi_4, J \ge 2^{m+n+1} \end{cases}$$

By Lemma 7(**P**34), $-1 \le 2\Omega_{s+1} - \Omega_s \le 0$, it yields consequently

$$-2^{s+1} \le B_{(m,s+1)} - B_{(m,s)} \le 2^{s+1}$$

Direct calculation by Equations (4) and (6) shows

$$\begin{aligned} B_{(m,s)} - A_m &= 2^{s+1} - 2 + 2^{s+2}\Omega_s - 2\vartheta - 2\Omega_0 \\ &= 2^2(2^{s-1} - 1 + 2^s\Omega_s - \Omega_0). \end{aligned}$$

Thus when $J < 2^{m+n+1}$

$$\begin{aligned} &B_{(m,s)} - A_m = 2^{s+1} - 2 + 2^{s+2}\Omega_s - 2\vartheta - 2\Omega_0 \\ &= 2^2(2^{s-1} - 1 + 2^s(\left\lfloor \Theta^s_{n+2} \right\rfloor - 2^{m-s}) - \left\lfloor \Theta^0_{n+2} - 2^m \right\rfloor) \\ &= 2^2(2^{s-1} - 1 + 2^s \left\lfloor \Theta^s_{n+2} \right\rfloor - \left\lfloor \Theta^0_{n+2} \right\rfloor) \end{aligned}$$

and when $J \ge 2^{m+n+1}$

$$\begin{aligned} &B_{(m,s)} - A_m = 2^{s+1} - 2 + 2^{s+2}\Omega_s - 2\vartheta - 2\Omega_0 \\ &= 2^2(2^{s-1} - 1 + 2^s(\left\lfloor \Theta^s_{n+3} \right\rfloor - 2^{m-s}) - \left\lfloor \Theta^0_{n+3} - 2^m \right\rfloor) \\ &= 2^2(2^{s-1} - 1 + 2^s \left\lfloor \Theta^s_{n+3} \right\rfloor - \left\lfloor \Theta^0_{n+3} \right\rfloor). \end{aligned}$$

By Lemma 7 (**P**34), it knows

$$-2^{s+1} = 2^2(2^{s-1} - 1 + 1 - 2^s) \le B_{(m,s)} - A_m \le 2^{s+1} - 4.$$

□

Proposition 4. Let $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ be two nodes of $\mathbf{T}_3$ with $0 \le m \le n$, A_m be the ancestors of $N_{(m,\alpha)} \times N_{(n,\beta)}$ on level m and $B_{(m,1)}, B_{(m,2)}, \cdots, B_{(m,s)}$ be the symmetric brothers of A_m; Then it holds for $s = 1, 2, \cdots, m$

$$\begin{cases} N_{(m,\alpha)} - 2^{s+1} \le B_{(m,s)} \le 2N_{(m,\alpha)} + 2^{s+1} - 1, J < 2^{m+n+1} \\ \frac{N_{(m,\alpha)}-1}{2} - 2^{s+1} - 1 \le B_{(m,s)} \le N_{(m,\alpha)} + 2^{s+1} - 2 \\ J \ge 2^{m+n+1} \end{cases}$$

Or equivalently

$$\begin{cases} \frac{B_{(m,s)}-2^{s+1}+1}{2} \le N_{(m,\alpha)} \le B_{(m,s)} + 2^{s+1}, J < 2^{m+n+1} \\ B_{(m,s)} - 2^{s+1} + 2 \le N_{(m,\alpha)} \le 2B_{(m,s)} + 2^{s+2} + 3, J \ge 2^{m+n+1} \end{cases} \tag{8}$$

Proof. By Equation (7) and Lemma 3 it yields when $1 \le s \le m$

$$\begin{cases} N_{(m,\alpha)} - 2^{s+1} \le B_{(m,s)} \le 2^{s+1} - 2 + 2N_{(m,\alpha)}, J < 2^{m+n+1} \\ \frac{N_{(m,\alpha)}-1}{2} - 2^{s+1} - 1 \le B_{(m,s)} \le 2^{s+1} - 2 + N_{(m,\alpha)}, J \ge 2^{m+n+1} \end{cases}$$

Since $N_{(m,\alpha)}$ and $B_{(m,s)}$ are odd integers, it is sure

$$\begin{cases} N_{(m,\alpha)} - 2^{s+1} \le B_{(m,s)} \le 2N_{(m,\alpha)} + 2^{s+1} - 1, J < 2^{m+n+1} \\ \frac{N_{(m,\alpha)}-1}{2} - 2^{s+1} - 1 \le B_{(m,s)} \le N_{(m,\alpha)} + 2^{s+1} - 2, J \ge 2^{m+n+1} \end{cases}$$

That is equivalent to Equation (8). □

Corollary 3. Let $N_{(m,\alpha)}$ and $N_{(n,\beta)}$ be two nodes of $\mathbf{T}_3$ with $0 \le m \le n$, A_m be the ancestors of $N_{(m,\alpha)} \times N_{(n,\beta)}$ on level m and $B_{(m,1)}, B_{(m,2)}, \cdots, B_{(m,s)}$ be the symmetric brothers of A_m; then $N_{(m,\alpha)}$ lies near A_m and $B_{(m,1)}$.

Proof. Direct calculation by Equation (7) yields

$$\begin{cases} 1 \le \frac{A_m}{N_{(m,\alpha)}} \le 2 - \frac{1}{N_{(m,\alpha)}}, J < 2^{m+n+1} \\ \frac{1}{2} - \frac{3}{2N_{(m,\alpha)}} \le \frac{A_m}{N_{(m,\alpha)}} \le 1, J \ge 2^{m+n+1} \end{cases}$$

which yields

$$\begin{cases} 1 \le \frac{A_m}{N_{(m,\alpha)}} < 2 - o_1(m), J < 2^{m+n+1} \\ \frac{1}{2} - o_2(m) \le \frac{A_m}{N_{(m,\alpha)}} \le 1, J \ge 2^{m+n+1} \end{cases}$$

where $o_1(m)$ and $o_2(m)$ are small positive numbers with $o_1(m) < \frac{1}{3}$ and $o_2(m) < \frac{1}{2}$.
Similarly, by Equation (8) it yields when $1 \le s \le m$

$$\begin{cases} 1 - \frac{2^{s+1}}{N_{(m,\alpha)}} \le \frac{B_{(m,s)}}{N_{(m,\alpha)}} \le 2 + \frac{2^{s+1}-1}{N_{(m,\alpha)}}, J < 2^{m+n+1} \\ \frac{1}{2} - \frac{3}{2N_{(m,\alpha)}} - \frac{2^{s+1}}{N_{(m,\alpha)}} \le \frac{B_{(m,s)}}{N_{(m,\alpha)}} \le 1 + \frac{2^{s+1}-2}{N_{(m,\alpha)}} \\ J \ge 2^{m+n+1} \end{cases}$$

Obviously, when $J < 2^{m+n+1}$, the smaller s is, the closer $1 - \frac{2^{s+1}}{N_{(m,\alpha)}}$ and $2 + \frac{2^{s+1}-1}{N_{(m,\alpha)}}$ are to $\frac{B_{(m,s)}}{N_{(m,\alpha)}}$. Likewise, when $J \ge 2^{m+n+1}$, the smaller s is, the closer $\frac{1}{2} + \frac{1}{2N_{(m,\alpha)}} - \frac{2^{s+1}}{N_{(m,\alpha)}}$ and $1 + \frac{2^{s+1}-2}{N_{(m,\alpha)}}$ are to $\frac{B_{(m,s)}}{N_{(m,\alpha)}}$. □

Example 5. *Take* $N_{(m,\alpha)} = N_{(6,2)} = 2^{6+1} + 2 \times 2 + 1 = 133$ *and* $N_{(n,\beta)} = N_{(8,5)} = 2^{8+1} + 10 + 1 = 523$*; then* $m = 6, n = 8$, $N_{(6,2)} \times N_{(8,5)} = 69959 = N_{(15,2211)}$ *and* $J = \frac{69959-1}{2} - 2^{6+8+1} = 2211 < 2^{6+8+1} = 32768$*; hence* $\vartheta = \lfloor \frac{J}{2^{m+1}} \rfloor = \lfloor \frac{2211}{2^{6+1}} \rfloor = 17$ *and* $N_{(6,17)} = 2^{6+1} + 34 + 1 = 163$ *is the ancestor of* $N_{(15,2211)}$ *on level 6. Since*

$$A_m = A_6 = 1 + 2\left\lfloor \frac{69959-1}{2^{10}} \right\rfloor = 137$$

$$B_{(6,1)} = 2^{1+1} - 1 + 2^3 \left\lfloor \frac{69959-1}{2^{8+2+1}} \right\rfloor - 2\left\lfloor \frac{69959-1}{2^{8+2}} \right\rfloor = 139$$

$$B_{(6,2)} = 2^{2+1} - 1 + 2^4 \left\lfloor \frac{69959-1}{2^{8+2+2}} \right\rfloor - 2\left\lfloor \frac{69959-1}{2^{8+2}} \right\rfloor = 143$$

$$B_{(6,3)} = 2^{3+1} - 1 + 2^{3+2} \left\lfloor \frac{69959-1}{2^{8+2+3}} \right\rfloor - 2\left\lfloor \frac{69959-1}{2^{6+2}} \right\rfloor = 135$$

$$B_{(6,4)} = 2^{4+1} - 1 + 2^{4+2} \left\lfloor \frac{69959-1}{2^{8+2+4}} \right\rfloor - 2\left\lfloor \frac{69959-1}{2^{6+2}} \right\rfloor = 151$$

$$B_{(6,5)} = 2^{5+1} - 1 + 2^{5+2} \left\lfloor \frac{69959-1}{2^{8+2+5}} \right\rfloor - 2\left\lfloor \frac{69959-1}{2^{6+2}} \right\rfloor = 183$$

$$B_{(6,6)} = 2^{6+1} - 1 + 2^{6+2} \left\lfloor \frac{69959-1}{2^{8+2+6}} \right\rfloor - 2\left\lfloor \frac{69959-1}{2^{6+2}} \right\rfloor = 247$$

It can see that A_6 *and* $B_{(6,1)}$ *are very close to* $N_{(m,\alpha)} = N_{(6,2)} = 133$.

3.9 Ancestors and Square Root

Let $N = N_{(m,\alpha)} \times N_{(n,\beta)}$ with $2^{m+1} + 1 \le N_{(m,\alpha)} \le N_{(n,\beta)} \le 2^{n+2} - 1$ and $1 \le \frac{N_{(n,\beta)}}{N_{(m,\alpha)}} < 2$; let $k = \lfloor \log_2 N \rfloor - 1$ and $l = \lfloor \frac{k+1}{2} \rfloor - 1 = \lfloor \frac{k-1}{2} \rfloor$; Then by Lemma 2 $\left\lfloor \sqrt{N} \right\rfloor \overset{\wedge}{=} l$ and by Lemma 3 $N_{(m,\alpha)} \overset{\wedge}{=} l - 1 \oplus N_{(m,\alpha)} \overset{\wedge}{=} l$. On the other hand, by definition, N's homolayer ancestor A_m, which lies on level m, naturally lies on the same level as $N_{(m,\alpha)}$ lies. Thereby, it is mandatory to make clear the relations among the ancestor A_m, the divisor $N_{(m,\alpha)}$ and the square root $\left\lfloor \sqrt{N} \right\rfloor$. This subsection investigates the relationships.

Theorem 4. *Suppose* $N > 5$ *is an odd positive integer and* $K = \lfloor \log_2 N \rfloor - 1$*; Let*

$$\begin{aligned} l &= \left\lfloor \frac{K+1}{2} \right\rfloor - 1 \\ A_1 &= 1 + 2\left\lfloor \frac{N-1}{2^{l+2}} \right\rfloor \\ A_2 &= 1 + 2\left\lfloor \frac{N-1}{2^{l+3}} \right\rfloor \end{aligned}$$

then $((A_1 \overset{\circ}{=} \left\lfloor \sqrt{N} \right\rfloor) \overset{\wedge}{=} l) \otimes (A_2 \overset{\wedge}{=} l - 1)$ *if* K *is odd, whereas* $A_1 \overset{\Delta}{=} l + 1 \otimes (A_2 \overset{\circ}{=} \left\lfloor \sqrt{N} \right\rfloor) \overset{\wedge}{=} l$ *if* K *is even.*

Proof. By Lemma 2, $\left\lfloor \sqrt{N} \right\rfloor \overset{\wedge}{=} l$. Now it is to prove A_1 and A_2 lie on level $\lfloor \frac{K}{2} \rfloor$ and level $\lfloor \frac{K}{2} \rfloor - 1$ respectively. In fact,

$$\begin{aligned} & 2^{K+1} + 1 \le N \le 2^{K+2} - 1 \\ \Rightarrow\ & 2^{K+1} \le N - 1 \le 2^{K+2} - 2 < 2^{K+2} \\ \Rightarrow\ & \frac{2^{K+1}}{2^{l+2}} < \frac{N-1}{2^{l+2}} < \frac{2^{K+2}}{2^{l+2}} \\ \Rightarrow\ & 2^{K-l-1} < \frac{N-1}{2^{l+2}} < 2^{K-l} \\ \Rightarrow\ & 2^{K-\lfloor \frac{K+1}{2} \rfloor} < \frac{N-1}{2^{l+2}} < 2^{K-\lfloor \frac{K+1}{2} \rfloor + 1} \\ \Rightarrow\ & 2^{K-\lfloor \frac{K+1}{2} \rfloor - 1} < \frac{N-1}{2^{l+3}} < 2^{K-\lfloor \frac{K+1}{2} \rfloor} \end{aligned}$$

Since $\lfloor \frac{K}{2} \rfloor + \lfloor \frac{K+1}{2} \rfloor = K$, it yields

$$\begin{aligned} 2^{\lfloor \frac{K}{2} \rfloor} &< \frac{N-1}{2^{l+2}} < 2^{\lfloor \frac{K}{2} \rfloor + 1} \\ 2^{\lfloor \frac{K}{2} \rfloor - 1} &< \frac{N-1}{2^{l+3}} < 2^{\lfloor \frac{K}{2} \rfloor} \end{aligned}$$

By Lemma 7 (**P**13) it holds

$$\begin{aligned} 2^{\lfloor \frac{K}{2} \rfloor + 1} \le 2\left\lfloor \frac{N-1}{2^{l+2}} \right\rfloor &< 2^{\lfloor \frac{K}{2} \rfloor + 2} \\ 2^{\lfloor \frac{K}{2} \rfloor} \le 2\left\lfloor \frac{N-1}{2^{l+3}} \right\rfloor &< 2^{\lfloor \frac{K}{2} \rfloor + 1} \end{aligned}$$

which are

$$\begin{aligned} 2^{\lfloor \frac{K}{2} \rfloor + 1} + 1 \le 1 + 2\left\lfloor \frac{N-1}{2^{l+2}} \right\rfloor &< 2^{\lfloor \frac{K}{2} \rfloor + 2} + 1 \\ 2^{\lfloor \frac{K}{2} \rfloor} + 1 \le 1 + 2\left\lfloor \frac{N-1}{2^{l+3}} \right\rfloor &< 2^{\lfloor \frac{K}{2} \rfloor + 1} + 1 \end{aligned}$$

Since $1+2\left\lfloor\frac{N-1}{2^{l+2}}\right\rfloor$ and $1+2\left\lfloor\frac{N-1}{2^{l+3}}\right\rfloor$ are odd integers, it holds

$$\begin{aligned} 2^{\lfloor\frac{K}{2}\rfloor+1}+1 \le 1+2\left\lfloor\frac{N-1}{2^{l+2}}\right\rfloor &\le 2^{\lfloor\frac{K}{2}\rfloor+2}-1 \\ 2^{\lfloor\frac{K}{2}\rfloor}+1 \le 1+2\left\lfloor\frac{N-1}{2^{l+3}}\right\rfloor &\le 2^{\lfloor\frac{K}{2}\rfloor+1}-1 \end{aligned}$$

That is

$$\begin{aligned} 2^{\lfloor\frac{K}{2}\rfloor+1}+1 &\le A_1 \le 2^{\lfloor\frac{K}{2}\rfloor+2}-1 \\ 2^{\lfloor\frac{K}{2}\rfloor}+1 &\le A_2 \le 2^{\lfloor\frac{K}{2}\rfloor+1}-1 \end{aligned} \tag{9}$$

which says A_1 lies on level $\lfloor\frac{K}{2}\rfloor$ and A_2 lies on level $\lfloor\frac{K}{2}\rfloor-1$. Now consider the relationship between $\lfloor\frac{K}{2}\rfloor$ and $l=\lfloor\frac{K+1}{2}\rfloor-1$. By $\lfloor\frac{K+1}{2}\rfloor-1\le\lfloor\frac{K}{2}\rfloor\le\lfloor\frac{K+1}{2}\rfloor$ (NMF1 (T2)), it knows $\lfloor\frac{K}{2}\rfloor$ might take either of the two values, $\lfloor\frac{K+1}{2}\rfloor-1=l$ and $\lfloor\frac{K+1}{2}\rfloor=l+1$. Obviously, when $\lfloor\frac{K}{2}\rfloor=l$ then $(A_1\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)=l\otimes A_2\stackrel{\Delta}{=}l-1$; whereas $(A_2\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)=l\otimes A_1\stackrel{\Delta}{=}l+1$ when $\lfloor\frac{K}{2}\rfloor=l+1$. Meanwhile, it can easily prove that, $\lfloor\frac{K}{2}\rfloor=\lfloor\frac{K+1}{2}\rfloor$ when K is even whereas $\lfloor\frac{K}{2}\rfloor=\lfloor\frac{K+1}{2}\rfloor-1$ when K is odd. Actually, let $K=2s$ be even with integer $s>0$; then

$$K=2s\Rightarrow\left\lfloor\frac{K}{2}\right\rfloor=s\otimes\left\lfloor\frac{K+1}{2}\right\rfloor=s\Rightarrow\left\lfloor\frac{K}{2}\right\rfloor=\left\lfloor\frac{K+1}{2}\right\rfloor$$

If $K=2s-1$ then

$$\begin{aligned} &K=2s-1\Rightarrow(\lfloor\tfrac{K}{2}\rfloor=s-1)\otimes(\lfloor\tfrac{K+1}{2}\rfloor-1=s-1)\\ &\Rightarrow\lfloor\tfrac{K}{2}\rfloor=\lfloor\tfrac{K+1}{2}\rfloor-1 \end{aligned}$$

Hence $(A_1\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)=l\otimes A_2\stackrel{\Delta}{=}l-1$ when K is odd, whereas $(A_2\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)=l\otimes A_1\stackrel{\Delta}{=}l+1$ when K is even. □

Proposition 4. Suppose $N=N_{(m,\alpha)}\times N_{(n,\beta)}>8$ with $1<\frac{N_{(n,\beta)}}{N_{(m,\alpha)}}<\frac{3}{2}$. Let $K=\lfloor\log_2 N\rfloor-1$, $l=\lfloor\frac{K+1}{2}\rfloor-1$, $A_1=1+2\lfloor\frac{N-1}{2^{l+2}}\rfloor$ and $A_2=1+2\lfloor\frac{N-1}{2^{l+3}}\rfloor$; then it is impossible $N_{(m,\alpha)}<A_2\otimes N_{(n,\beta)}>A_1$.

Proof. Use proof by contradiction. Assume $N_{(m,\alpha)}<A_2\otimes N_{(n,\beta)}>A_1$ holds; then

$$\frac{N_{(n,\beta)}}{N_{(m,\alpha)}}>\frac{A_1}{A_2}$$

By $K=\lfloor\log_2 N\rfloor-1$, $N>8$ yields $K>2$ and by Equation (9) $A_2=1+2^{\lfloor\frac{K}{2}\rfloor}>2$; hence it yields by NMF2 $\frac{A_1}{A_2}>\frac{3}{2}$. That is the contradiction because of $1<\frac{N_{(n,\beta)}}{N_{(m,\alpha)}}<\frac{3}{2}$. □

3.10 Divisors, Ancestors and Square Root

Corollary 4. Suppose $N=pq$ with $1\le\frac{q}{p}<\frac{3}{2}$; let $K=\lfloor\log_2 N\rfloor-1$, $l=\lfloor\frac{K+1}{2}\rfloor-1$, $A_1=1+2\lfloor\frac{N-1}{2^{l+2}}\rfloor$ and $A_2=1+2\lfloor\frac{N-1}{2^{l+3}}\rfloor$; then p satisfies $A_2\le p\le\min(A_1,\lfloor\sqrt{N}\rfloor)$ when K is odd whereas q satisfies $\lfloor\sqrt{N}\rfloor\le q\le A_1$ when K is even.

Proof. The condition $1\le\frac{q}{p}<\frac{3}{2}$ naturally leads to $p\le\lfloor\sqrt{N}\rfloor\le q$ and that p and q lie on the same level or on two adjacent levels. By Lemma 2, $\lfloor\sqrt{N}\rfloor$ is permanently clamped on level $l=\lfloor\frac{K+1}{2}\rfloor-1$. By Corollary 1*, $\lfloor\sqrt{N}\rfloor\stackrel{\circ}{=}q$ for an odd K and $\lfloor\sqrt{N}\rfloor\stackrel{\circ}{=}p$ for an even K. Then by Theorem 5, the relationships among p, q, $\lfloor\sqrt{N}\rfloor$, l, A_1 and A_2 are sorted in Table 1.

It can see from Table 1, there are 4 cases to be made clear. Next is to investigate them one by one. Note that, during the deductions, Lemma 7 (**P**32), Lemma 5 and NMF1(T4) are referred to without declaration.

Case 1. $((A_1\stackrel{\circ}{=}p\stackrel{\circ}{=}q\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)\stackrel{\wedge}{=}l)\otimes(A_2\stackrel{\Delta}{=}l-1)$ makes it clear that $A_2<p\le\lfloor\sqrt{N}\rfloor\le q$. Note that, $(A_1\stackrel{\circ}{=}p\stackrel{\circ}{=}q\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)\stackrel{\wedge}{=}l$ this time leads to $2^{l+1}+1\le p<q\le 2^{l+2}-1$ and $p\le\lfloor\sqrt{N}\rfloor\le\lfloor 2^{l+1}\sqrt{2}\rfloor$ (Lemma 2). Consequently,

$$\begin{aligned} A_1 &= 1+2\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor\ge\left\lfloor\frac{pq-1}{2^{l+1}}\right\rfloor \\ &\ge \left\lfloor\frac{(2^{l+1}+1)q-1}{2^{l+1}}\right\rfloor=\left\lfloor q+\frac{q-1}{2^{l+1}}\right\rfloor \\ &\ge q+1>p>\max(p,q) \\ A_1 &= 1+2\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor\le 1+\frac{pq-1}{2^{l+1}} \\ &\le 1+\frac{\lfloor 2^{l+1}\sqrt{2}\rfloor\,q-1}{2^{l+1}}\le 1+\frac{2^{l+1}\sqrt{2}q-1}{2^{l+1}} \\ &= 1+\sqrt{2}q-\frac{1}{2^{l+1}}<1+\sqrt{2}q \\ \Rightarrow q &\ge \frac{A_1-1}{\sqrt{2}}\ge\left\lfloor\frac{A_1-1}{\sqrt{2}}\right\rfloor \end{aligned}$$

and thus

$$\begin{aligned} A_2 &< p<\min(A_1,\lfloor\sqrt{N}\rfloor)\otimes\max(\lfloor\sqrt{N}\rfloor,\left\lfloor\frac{A_1-1}{\sqrt{2}}\right\rfloor) \\ &\le q<A_1-1. \end{aligned}$$

Case 2. Since $((A_1\stackrel{\circ}{=}q\stackrel{\circ}{=}\lfloor\sqrt{N}\rfloor)\stackrel{\wedge}{=}l)\otimes(A_2\stackrel{\Delta}{=}l-1)\otimes(p\stackrel{\Delta}{=}l-1)$, it is obvious $p\le\min(A_1,\lfloor\sqrt{N}\rfloor)$. Because this time $2^l+1\le p\le 2^{l+1}-1$ and $2^{l+1}+1\le$

Table 1: Four cases among p, q, l, A_1,A_2 and $\left\lfloor\sqrt{N}\right\rfloor$

K's Odevity	$\left\lfloor\sqrt{N}\right\rfloor$'s level	p & q	Levels of p, q, l, A_1,A_2 and $\left\lfloor\sqrt{N}\right\rfloor$
K odd	$\left\lfloor\sqrt{N}\right\rfloor \stackrel{\circ}{=} q$	$p \stackrel{\circ}{=} q$	$((A_1 \stackrel{\circ}{=} p \stackrel{\circ}{=} q \stackrel{\circ}{=} \left\lfloor\sqrt{N}\right\rfloor) \stackrel{\wedge}{=} l) \otimes (A_2 \stackrel{\Delta}{=} l-1)$
		$p \doteq q$	$((A_1 \stackrel{\circ}{=} q \stackrel{\circ}{=} \left\lfloor\sqrt{N}\right\rfloor) \stackrel{\wedge}{=} l) \otimes (A_2 \stackrel{\Delta}{=} l-1) \otimes (p \stackrel{\Delta}{=} l-1)$
K even	$\left\lfloor\sqrt{N}\right\rfloor \stackrel{\circ}{=} p$	$p \stackrel{\circ}{=} q$	$(A_2 \stackrel{\circ}{=} p \stackrel{\circ}{=} q \stackrel{\circ}{=} \left\lfloor\sqrt{N}\right\rfloor) \stackrel{\wedge}{=} l \otimes A_1 \stackrel{\Delta}{=} l+1$
		$p \doteq q$	$((A_2 \stackrel{\circ}{=} p \stackrel{\circ}{=} \left\lfloor\sqrt{N}\right\rfloor) \stackrel{\wedge}{=} l) \otimes ((q \stackrel{\circ}{=} A_1) = l+1)$

$q \le 2^{l+2}-1$, it yields

$$\begin{aligned} A_1 &= 1+2\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor \ge \left\lfloor\frac{pq-1}{2^{l+1}}\right\rfloor \\ &\ge \left\lfloor\frac{q(2^l+1)-1}{2^{l+1}}\right\rfloor = \left\lfloor\frac{q}{2}+\frac{q-1}{2^{l+1}}\right\rfloor \ge \left\lfloor\frac{q}{2}\right\rfloor+1. \end{aligned}$$

$$\begin{aligned} A_1 &= 1+2\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor \le 1+\left\lfloor\frac{pq-1}{2^{l+1}}\right\rfloor \\ &\le 1+\left\lfloor\frac{(2^{l+1}-1)q-1}{2^{l+1}}\right\rfloor = 1+\left\lfloor q-\frac{q+1}{2^{l+1}}\right\rfloor \\ &\le 1+\left\lfloor q-\frac{2^{l+1}+2}{2^{l+1}}\right\rfloor < q. \end{aligned}$$

$$\begin{aligned} \left\lfloor\frac{3}{2}A_1\right\rfloor &= \left\lfloor\frac{3}{2}(1+2\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor)\right\rfloor \ge \left\lfloor\frac{3}{2}\left\lfloor\frac{pq-1}{2^{l+1}}\right\rfloor\right\rfloor \\ &\ge \left\lfloor\frac{3}{2}\cdot\frac{pq-1}{2^{l+1}}-\frac{3}{2}\right\rfloor \ge \left\lfloor\frac{3}{2}\cdot\frac{(\frac{2}{3}q)\cdot q-1}{2^{l+1}}-\frac{3}{2}\right\rfloor \\ &= \left\lfloor\frac{q^2-\frac{3}{2}}{2^{l+1}}-\frac{3}{2}\right\rfloor \\ &\ge \left\lfloor\frac{(2^{l+1}+1)q-\frac{3}{2}}{2^{l+1}}-\frac{3}{2}\right\rfloor \\ &= q-1+\left\lfloor\frac{2q-3}{2^{l+2}}-\frac{1}{2}\right\rfloor \\ &\ge q-1+\left\lfloor\frac{2(2^{l+1}+1)-3}{2^{l+2}}-\frac{1}{2}\right\rfloor \\ &= q-1+\left\lfloor 1-\frac{1}{2^{l+2}}-\frac{1}{2}\right\rfloor = q-1 \end{aligned}$$

and

$$\begin{aligned} A_2 &= 1+2\left\lfloor\frac{pq-1}{2^{l+3}}\right\rfloor \le 1+\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor \\ &\le 1+\left\lfloor\frac{p(2^{l+2}-1)-1}{2^{l+2}}\right\rfloor = 1+\left\lfloor p-\frac{p+1}{2^{l+2}}\right\rfloor \\ &\le p. \end{aligned}$$

Consequently

$$\begin{aligned} A_2 &\le p \le \min(A_1, \left\lfloor\sqrt{N}\right\rfloor)) \otimes (max(\left\lfloor\sqrt{N}\right\rfloor, A_1) \\ &\le q \le \left\lfloor\frac{3}{2}A_1\right\rfloor+1. \end{aligned}$$

Case 3. The condition $(A_2 \stackrel{\circ}{=} p \stackrel{\circ}{=} q \stackrel{\circ}{=} \left\lfloor\sqrt{N}\right\rfloor) \stackrel{\wedge}{=} l \otimes A_1 \stackrel{\Delta}{=} l+1$ yields $p \le \left\lfloor\sqrt{N}\right\rfloor \le q < A_1$ and $2^{l+1}+1 \le p < q \le 2^{l+2}-1$. Hence

$$\begin{aligned} A_2 &= 1+2\left\lfloor\frac{pq-1}{2^{l+3}}\right\rfloor \ge \left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor \\ &\ge \left\lfloor\frac{p(2^l+1)-1}{2^{l+2}}\right\rfloor = \left\lfloor\frac{p}{2}+\frac{p-1}{2^{l+2}}\right\rfloor \\ &\ge \left\lfloor\frac{p}{2}\right\rfloor+\left\lfloor\frac{p-1}{2^{l+2}}\right\rfloor = \left\lfloor\frac{p}{2}\right\rfloor \end{aligned}$$

$$\begin{aligned} A_2 &= 1+2\left\lfloor\frac{pq-1}{2^{l+3}}\right\rfloor \le 1+\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor \\ &\le 1+\left\lfloor\frac{p(2^{l+2}-1)-1}{2^{l+2}}\right\rfloor = 1+\left\lfloor p-\frac{p+1}{2^{l+2}}\right\rfloor \le p \end{aligned}$$

It holds

$$(A_2 \le p \le \min(A_1, \left\lfloor\sqrt{N}\right\rfloor)) \otimes (\left\lfloor\sqrt{N}\right\rfloor \le q < A_1$$

Case 4. The condition $(A_2 \stackrel{\circ}{=} p \stackrel{\circ}{=} \left\lfloor\sqrt{N}\right\rfloor) \stackrel{\wedge}{=} l \otimes (q \stackrel{\circ}{=} A_1) = l+1)$ yields $p \le \left\lfloor\sqrt{N}\right\rfloor \le q$, $2^{l+1}+1 \le p \le 2^{l+2}-1$ and $2^{l+2}+1 \le q \le 2^{l+3}-1$. Thus

$$\begin{aligned} A_2 &= 1+2\left\lfloor\frac{pq-1}{2^{l+3}}\right\rfloor \ge \left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor \\ &\ge \left\lfloor\frac{p(2^{l+2}+1)-1}{2^{l+2}}\right\rfloor = \left\lfloor p+\frac{p-1}{2^{l+2}}\right\rfloor \ge p \end{aligned}$$

$$\begin{aligned} \left\lfloor\frac{2}{3}A_2\right\rfloor &= \left\lfloor\frac{2}{3}(1+2\left\lfloor\frac{pq-1}{2^{l+3}}\right\rfloor)\right\rfloor \\ &\le \left\lfloor\frac{2}{3}(1+\left\lfloor\frac{pq-1}{2^{l+2}}\right\rfloor)\right\rfloor \le \left\lfloor\frac{2}{3}+\frac{2}{3}\cdot\frac{pq-1}{2^{l+2}}\right\rfloor \\ &\le \left\lfloor\frac{2}{3}+\frac{2}{3}\cdot\frac{p\cdot(\frac{3}{2}p)-1}{2^{l+2}}\right\rfloor = \left\lfloor\frac{2}{3}+\frac{p^2-\frac{2}{3}}{2^{l+2}}\right\rfloor \\ &\le \left\lfloor\frac{2}{3}+\frac{(2^{l+2}-1)p-\frac{2}{3}}{2^{l+2}}\right\rfloor \\ &= \left\lfloor p+\frac{2}{3}-\frac{p+\frac{2}{3}}{2^{l+2}}\right\rfloor \le p \end{aligned}$$

$$\begin{aligned}
\left\lfloor \frac{3}{2}A_2 \right\rfloor &= \left\lfloor \frac{3}{2}(1+2\left\lfloor \frac{pq-1}{2^{l+3}} \right\rfloor) \right\rfloor \\
&\geq \left\lfloor \frac{3}{2}\left\lfloor \frac{pq-1}{2^{l+2}} \right\rfloor \right\rfloor \geq \left\lfloor \frac{3}{2}\frac{pq-1}{2^{l+2}} - \frac{3}{2} \right\rfloor \\
&\geq \left\lfloor \frac{3}{2}\frac{(\frac{2}{3}q)q-1}{2^{l+2}} - \frac{3}{2} \right\rfloor = \left\lfloor \frac{q^2-\frac{3}{2}}{2^{l+2}} - \frac{3}{2} \right\rfloor \\
&\geq \left\lfloor \frac{(2^{l+2}+1)q-\frac{3}{2}}{2^{l+2}} - \frac{3}{2} \right\rfloor \\
&= \left\lfloor q + \frac{q-\frac{3}{2}}{2^{l+2}} - \frac{3}{2} \right\rfloor \\
&\geq q + \left\lfloor \frac{2^{l+2}+1-\frac{3}{2}}{2^{l+2}} - \frac{3}{2} \right\rfloor \\
&= q + \left\lfloor 1 - \frac{1}{2^{l+3}} - \frac{3}{2} \right\rfloor = q-1 \\
A_1 &= 1+2\left\lfloor \frac{pq-1}{2^{l+2}} \right\rfloor \geq \left\lfloor \frac{pq-1}{2^{l+1}} \right\rfloor \\
&\geq \left\lfloor \frac{(2^{l+1}+1)q-1}{2^{l+1}} \right\rfloor = q + \left\lfloor \frac{q-1}{2^{l+1}} \right\rfloor \\
&\geq q+2 > p > \max(p,q).
\end{aligned}$$

Therefore

$$\begin{aligned}
\left\lfloor \frac{2}{3}A_2 \right\rfloor &\leq p \leq \min(A_2, \left\lfloor \sqrt{N} \right\rfloor) \otimes \left\lfloor \sqrt{N} \right\rfloor \\
&\leq q < \min(A_1, \left\lfloor \frac{3}{2}A_2 \right\rfloor + 1).
\end{aligned}$$

Summarizing all the cases as well as their results immediately leads to Corollary 4. □

4 Divisors' Bounds of Semiprimes with Small Divisor-Ratio

Let $N = pq$ be a semiprime with $1 < p < q$ and $1 < \frac{q}{p} < \frac{3}{2}$. Corollary 4 together with its proving process surely provides a way to locate p or q in an interval. Meanwhile, by Lemma 4, a more accurate bounds of p and q can be given by Theorem 6.

Theorem 5. *Suppose $N = pq$ is an odd integer satisfying $1 < \frac{q}{p} < \frac{3}{2}$ and $K = \lfloor \log_2 N \rfloor - 1$. Let*

$$\begin{aligned}
l &= \left\lfloor \frac{K+1}{2} \right\rfloor - 1 \\
A_1 &= 1+2\left\lfloor \frac{N-1}{2^{l+2}} \right\rfloor \\
A_2 &= 1+2\left\lfloor \frac{N-1}{2^{l+3}} \right\rfloor \\
p_{ol} &= 2^{l+1} - 2^{l-1} + 1 \\
q_{er} &= 2^{l+2} + 2^l - 1
\end{aligned}$$

then p is bounded by $\max(A_2, p_{ol}) \leq p \leq \min(A_1, \left\lfloor \sqrt{N} \right\rfloor)$ when K is odd whereas q is bounded by $\left\lfloor \sqrt{N} \right\rfloor \leq q \leq \min(A_1, q_{er})$ when K is even.

Proof. The thereom is a combination of Corollary 4 and Lemma 4. □

Corollary 5. Let $N, p, q, K, l, A_1, A_2, p_{ol}$ and q_{er} be defined as those in Theorem 6; denote $p_s = \max(A_2, p_{ol})$, $p_b = \min(A_1, \left\lfloor \sqrt{N} \right\rfloor)$, $q_s = \left\lfloor \sqrt{N} \right\rfloor$ and $q_b = \min(A_1, q_{er})$; then $p_s \leq p \leq p_b \otimes \frac{N}{p_b} \leq q \leq \frac{N}{p_s}$ when K is odd whereas $\frac{N}{q_b} \leq p \leq \frac{N}{q_s} \otimes q_s \leq q \leq q_b$ when K is even.

Proof. (Omitted) □

4.1 Procedure to Find an Interval Containing a Divisor

Based on Corollary 5, a procedure to find an interval (divisor's bound) that contains a divisor of a semiprime with small divisor-ratio can be designed as follows.

Algorithm 1 Divisor's Bounds Detecting Algorithm

1: Begin
2: Input: Large semiprime N;
3: Step 1. Calculate K,l by $K = \lfloor \log_2 N \rfloor - 1$ and $l = \lfloor \frac{K-1}{2} \rfloor$;
4: Step 2. Calculate A_1, A_2,M, p_{ol} and q_{er} by $A_1 = 1 + 2\lfloor \frac{N-1}{2^{l+2}} \rfloor$, $A_2 = 1 + 2\lfloor \frac{N-1}{2^{l+3}} \rfloor$,$M = \left\lfloor \sqrt{N} \right\rfloor$,$p_{ol} = 2^{l+1} - 2^{l-1} + 1$ and $q_{er} = 2^{l+2} + 2^l - 1$;
5: Step 3. If K is odd then $p_s = \max(A_2, p_{ol})$,$p_b = \min(A_1, M)$, $I_p = [p_s, p_b]$ and $I_q = [\left\lfloor \frac{N}{p_b} \right\rfloor, \left\lfloor \frac{N}{p_s} \right\rfloor]$; .
6: Else calculate $Q_s = M$,$q_b = \min(A_1, q_{er})$,

$$I_q = [q_s, q_b]$$

and $I_p = [\left\lfloor \frac{N}{q_b} \right\rfloor, \left\lfloor \frac{N}{q_s} \right\rfloor]$.
7: End

Programming with Maple 18.0 on Windows 7.0, taking N = 3551, 16637, 72593, 489779, 753041, 2350553, 4538873, 8772041, 14351501, 23552813, 1035918371, 2512642129, 5783560579, 9048212729, 80735174503 and 211041144109, the running results are list in Table 2. In the table, k_1, k_2 and k_m are levels of A_1, A_2 and M respectively, k_p and k_q are levels of p and q respectively, and q/p is the divisor-ratio. Referring to the process in proving Corollary 4, it can see that, $N = 3551$ and $N = 14351501$ match to the Case 4 that says $(A_2 \stackrel{\circ}{=} p \stackrel{\circ}{=} \left\lfloor \sqrt{N} \right\rfloor) \stackrel{\wedge}{=} l \otimes (q \stackrel{\circ}{=} A_1) = l + 1)$ and for which it is proved $\lfloor \frac{2}{3}A_2 \rfloor \leq p \leq \min(A_2, \left\lfloor \sqrt{N} \right\rfloor) \otimes \left\lfloor \sqrt{N} \right\rfloor \leq q < \min(A_1, \lfloor \frac{3}{2}A_2 \rfloor + 1)$, while all the other Ns match exactly to the other cases enumerated in the proof of Corollary 4. The Maple programs are shown in the end of this paper, readers can test them with Maple software.

4.2 Estimation on Divisors of RSA Numbers

Applying the procedure introduced in Section 4.1, numerical experiments also were made to the known factorized RSA numbers with small divisor-ratio, for example, the RSA100, RSA120, RSA130, and so on. Experiments showed that the calculated bounds of the divisors exactly matched to the known facts. Due to limitation of the space, here merely list RSA100, RSA120, RSA130, and RSA150 in Table 3. Meanwhile, considering each of the remained unfactorized RSA modulus must have their divisor-ratios less than $\sqrt{2}$, as specified in American Digital Signature Standard (DSS) [3], divisors' bounds for the unfactorized RSA numbers are predicted and list in Table 3.

5 Conclusions

Finding an interval that contains a divisor of a semiprime is certainly helpful for factoring the semiprime, and the interval is of course a bound to the divisors. The method and procedure developed in this paper might be a valuable attempt to know the RSA numbers. However, readers can see that, there are still work to do. For example, the way to make the interval smaller will be continuously seeking. Hope there are followers to continue the work and find better results.

Acknowledgments

The research is supported by the Open Project Program of the State Key Lab of CAD & CG Grant No. A2002 and by Foshan University and Foshan Bureau of Science and Technology under project that constructs Guangdong Engineering Center of Information Security for Intelligent Manufacturing System.

References

[1] F. Bao, C. C. Lee, M. S. Hwang, "Cryptanalysis and improvement on batch verifying multiple RSA digital signatures", *Applied Mathematics and Computation*, vol. 172, no. 2, pp. 1195-1200, Jan. 2006.

[2] G. Chen, J. Li, "Brief investigation on square root of a node of $\mathbf{T}_3$ tree," *Advances in Pure Mathematics*, vol. 8, no. 7, pp. 666-671, 2018.

[3] Federal Information Processing Standards Publication, "Digital signature standard (DSS)," *Information Technology Laboratory*, FIPS publication 186-3, June 2009. (`https://csrc.nist.gov/csrc/media/publications/fips/186/3/archive/2009-06-25/documents/fips_186-3.pdf`)

[4] M. S. Hwang, C. C. Lee, Y. C. Lai, "Traceability on RSA-based partially signature with low computation", *Applied Mathematics and Computation*, vol. 145, no. 2-3, pp. 465–468, Dec. 2003.

[5] M. S. Hwang, Chao-Chen Yang, S. F. Tzeng, "Improved digital signature scheme based on factoring and discrete logarithms", Journal of Discrete Mathematical Sciences & Cryptography, vol. 5, no. 2, pp. 151-155, Aug. 2002.

[6] S. F. Tzeng, C. Y. Yang, M. S. Hwang, "A new digital signature scheme based on factoring and discrete logarithms", *International Journal of Computer Mathematics*, vol. 81, no. 1, pp. 9-14, 2004.

[7] X. Wang, "Valuated binary tree: A new approach in study of integers," *International Journal of Scientific and Innovative Mathematical Research*, vol. 4, no. 3, pp. 63-67, 2016.

[8] X. Wang, "$\mathbf{T}_3$ tree and its traits in understanding integers," *Advances in Pure Mathematics*, vol. 8, no. 5, pp. 494-507, 2018.

[9] X. Wang, "Influence of divisor-ratio to distribution of semiprime's divisor," *Journal of Mathematics Research*, vol. 10, no. 4, pp. 54-61, 2018.

[10] X. Wand, "More on square and square root of a node on $\mathbf{T}_3$ tree," *International Journal of Mathematics and Statistics Study*, vol. 6, no. 3, pp. 1-7, 2018.

[11] X. Wang, "Divisors' distribution of an RSA modulus on $\mathbf{T}_3$ tree," *International Journal of Mathematics and Statistics Study*, vol. 6, no. 4, pp. 15-32, 2018.

[12] X. Wang, "Some inequalities on $\mathbf{T}_3$ tree," *Advances in Pure Mathematics*, vol. 8, no. 8, pp. 711-719, 2018. (`https://doi.org/10.4236/apm.2018.88043`)

[13] X. Wang, "Number of digits in two integers and their multiplication," *Journal of Advances in Applied Mathematics*, vol. 4, no. 2, pp. 69-74, 2019.

[14] X. Wang, "Brief summary of frequently-used properties of the floor function: Updated 2018," *IOSR Journal of Mathematics*, vol. 15, no. 1, pp. 30-33, 2019. doi:10.9790/5728-1501023033

[15] X. Wang, Y. Miao, "Relationship between divisors' distribution and square root of an RSA modulus," *International Journal of Information and Electronics Engineering*, vol. 9, no. 1, pp. 7-11, 2019.

[16] X. Wang, Z. Shen, "Analytic formulas to calculate symmetric brothers of a node in a perfect binary tree," *Journal of Mathematics Research*, vol. 10, no. 5, pp. 45-48, 2018.

[17] X. Wang, Z. Shen, "Traits of an RSA modulus on $\mathbf{T}_3$ tree," *Journal of Mathematics Research*, vol. 10, no. 6, pp. 15-29, 2018.

Biography

Dr. & Prof. Xingbo Wang got his Master and Doctor's degrees at National University of Defense Technology (NUDT) of China. Since 1994, he had worked at NUDT on CAD/CAM/CNC technologies till 2010. Since 2010, he has been a professor in Foshan University with research interests in intelligent manufacturing system and computer applications. Prof. Wang is now in charge of Guangdong Engineering Center of Information Security for Intelligent Manufacturing System.

Table 2: Test with ordinary semiprimes in maple

N	A_1	M	A_2	K	l	k_1	k_2	km	$[Ps, Pb]$	$[Qs, Qb]$	p	q	kp	kq	q/p
3551	111	59	55	10	4	5	4	4	[43,61]	[59, 79]	53	67	4	5	1.2642
16637	129	128	65	13	6	6	5	6	[97,129]	[127,171]	127	131	5	6	1.0315
72593	283	269	141	15	7	7	6	7	[193,269]	[269,377]	229	317	6	7	1.3843
489779	957	699	479	17	8	8	7	8	[479,699]	[699,1023]	647	757	8	8	1.1700
753041	1471	867	735	18	8	9	8	8	[587,869]	[867,1279]	739	1019	8	8	1.3789
2350553	2295	1533	1147	20	9	10	9	9	[1023,1533]	[1533,2295]	1259	1867	9	9	1.4829
4538873	2217	2130	1109	21	10	10	9	10	[1537,2131]	[2129,2953]	2029	2237	9	10	1.1025
8772041	4283	2961	2141	22	10	11	10	10	[2047,2963]	[2961,4283]	2659	3299	10	10	1.2407
14351501	7007	3788	3503	22	10	11	10	10	[2803,3789]	[3787,5119]	3491	4111	10	11	1.1776
23552813	5751	4853	2875	23	11	11	10	11	[3073,4853]	[4853,7665]	4663	5051	11	11	1.0832
1035918371	63227	32185	31613	28	13	14	13	13	[25291,32187]	[32185,40959]	31663	32717	13	13	1.0333
2512642129	76679	50126	38339	30	14	15	14	14	[32767,50127]	[50125,76679]	49199	51071	14	14	1.0380
5783560579	88251	76049	44125	31	15	15	14	15	[49153,76049]	[76049,117665]	71387	81017	15	15	1.1349
9048212729	138065	95122	69033	32	15	16	15	15	[65535,95123]	[95121,138065]	90599	99871	15	15	1.1023
80735174503	307981	284139	153991	35	17	17	16	17	[196609,284139]	[284139,410639]	259271	311393	16	17	1.2010
211041144109	805059	459392	402529	36	17	18	17	17	[322023,459393]	[459391,655359]	412771	511279	17	17	1.2387

Table 3: Divisors' bounds of some RSA numbers

RSA Modulus	K	q/p	Bounds of p &q
RSA100	328	1.056	qs:39020571855401265512289573339484371018905006900194 qb:58460065493236116728147393308651320786237301719039 ps:26045215910657851738937410052610101605450381262848 pb:39020571855401265512296365669150564414504357068800 q=40094690950920881030683735292761468389214899724061 ∈ $[qs, qb]$
RSA110	362	1.047	qs:59828282759683040041003178541182303136857938437236090 73 qb:76624777043294442917917351357515459180936956 10918010879 ps:46713655244310142619235613244214189471410835709619077 12 pb:59828282759683040041003178587831603911481633405675765 76 q=61224210904935475769370373175614188412257585542531069 99 ∈ $[qs, qb]$
RSA130	428	1.147	qs:42509788151523465407452697662505294703186899165086434856734890373 qb:65820182292848241686198767302294020199309434625343194533944360959 ps:27454832632448590673149681387526649708671363703288226741678505984 pb:42509788151523465407452697672385147726002667100788997061307531264 q=45534498646735972188403686897274408864356301263205069600999044599 ∈ $[qs, qb]$
RSA150	494	1.281	qs:39381443914405733799015713304153570178670710260362452639411239967444526155 4 qb:56539106072908298546665520023773392506479484700019806659891398441363832831 9 ps:27430538480455819663423717399656593360063676754450126638994590024334940569 6 pb:39381443914405733799015713318154707602083044846311236917615295586713049497 6 q=44564774490364074153324112578708617600544253629776615349341972453246029619 9 ∈ $[qs, qb]$
RSA232	766	?	qs:31778631151639045179686388493300715963579687858851199671688113190565970121 13 2383850648339 14560665325846904492282993 qb:49252507745493099015348800125179517256349674088081808334935366755307152214371513263 26782 8328186 06144551008284987883519 ps:20504162003087726530524231266151890028838809030421134074165904222129784360726792359243279 4851 14020053144233840738304 pb:31778631151639045179686388493460258559790492608506204494863444524725674502533969522805105233 4 84775800504617795584000
RSA240	793	?	ps:24208592607064768028024242237528236321840991807773970432787431467568571456407886199652925 18700129216971159223724450775 05 pb:35301609989024407676277184230754669625591247945328832161339587199109937626315697180300630222 14786064772315627183366893 98 qs:35301609989024407676277376201215623494809100345835202482007420959779051400431763598705283397 43802630270818573939072040 96 qb:51477741314607010225948129476132851481976294299155271062661198751714361529219087624214875028 34335690459478792782961704 96
RSA250	827	?	ps:31730686501931932749691934785573009911763424782285498525663142173171477979342944599609082 10110633372684378177201121198604289 pb:46263642855278322335316947335501789288773209888512730218438986286143041213336360936304994055 64967740269443836891832736830327 qs:46263642855278322335316947335946623061458775676639813213532599622958589938825942881009811620 53299122087098548208209686953984 0 qb:67452831507771841841969181568437402850342166336387702891215741529944576744800108078873515077 33579980027700688337927513781043 2
RSA260	860	?	qs:47024276208709120795061378748873252735571493245640595713707096298486086720087632250363079 99162383887826987432678940935201933498833 qb:69316742353020371489460354603577092585910926884395414379261989515365532695140640575999360152 60348945243478027403508929572435394559 ps:31901132077027179765050649029228297709014926458423900186946803716772326021845665874908676203 10106277907616162689106735859975061504 pb:47024276208709120795061457104860034792688322806356770024545913304764822020508266950362669087 09976391144559837102489549982891245568

Analysis and Improvement of Otway-Rees Based on Enhanced Authentication Tests

Lei Yu[1,2,4], Yu-Yan Guo[1], Ze-Peng Zhuo[3], and Shi-Min Wei[1]
(Corresponding author: Lei Yu)
School of Computer Science and Technology, Huaibei Normal Universtiy[1]
School of Information and Control Engineering, China University of Mining and Technology[2]
School of Mathematical Sciences, Huaibei Normal University[3]
Anhui Big-Data Research Center on University Management[4]
Huaibei, Anhui 235000, China
Email: yulei@chnu.edu.cn
(Received Feb. 11, 2020; Revised and Accepted Oct. 5, 2020; First Online Apr. 25, 2021)

Abstract

To improve the formalization and accuracy of enhanced authentication tests and the effectiveness of security protocol analysis and improvement, the theory and methods of authentication tests are extended and optimized from principal knowability, principal identity, parameter types, and roles challenge-response structure. Then, based on the optimized enhanced authentication tests method, the correspondence of the original Otway-Rees protocol is analyzed, and the main factors that cause principals to fail to reach consistency on the session key are found. According to the correspondence degree of principals, by the reconstruction and construction of the test components on the existing messages, three new, improved schemes of Otway-Rees protocol are given without changing the original cryptosystem, and fresh value mechanism, the order, and the number of messages exchanged among principals. The practice of protocol analysis shows that optimized enhanced authentication tests can improve the formalization of protocol analysis and reduce the complexity of protocol analysis and accurately locate protocol defects and provide correction schemes.

Keywords: Authentication Tests; Formal Analysis; Otway-Rees; Security Protocol

1 Introduction

Security protocol is an interactive protocol based on cryptosystem. It uses cryptographic algorithms and protocol logics to achieve security objectives such as identity authentication and key distribution. The analysis of protocol flaw is not only an important part of protocol design, but also the main means for designing the security protocol [5, 9, 18]. The current formal methods have been extensively recognized as the most scientific, rigorous and effective method for analyzing the flaws in security protocol [2–4, 11].

Authentication tests [7] is a formal analysis method of security protocol based on challenge and response model, which is developed from strand spaces model theory [13]. The basic idea of authentication tests is to judge the legitimacy of principals according to the operation capability of principals to an encrypted message, so as to realize the identity authentication of principals. Using authentication test can not only realize the authentication analysis of a protocol, but also the confidentiality analysis, and can also be used in the design of a security protocol. At present, authentication tests theory has been widely used in the design and analysis of security protocols [6,8,12,20]. However, in the original authentication testing theory, the correctness analysis of a protocol is based on the consistency of the strand parameters, the proving method used in the parameter consistency inference is not only complex but also low formalized. For this reason, the original authentication tests theory is further improved in [19], and enhanced authentication tests (EAT) method based on correspondence of principals are proposed. This new method makes the correctness analysis of protocol more concise and intuitive. Because authentication tests is based on the encryption components and uses the challenge and response mode to analyze the security properties of a protocol, but in the original theory, there is no research on the knowability and confidentiality of test components, nor challenge and response roles of test component and principals in the definition of test components.

The inadequate researches of these basic theories directly leads to the problems of loose logic and reasoning in the formal expression of related concept definition and property theorem, as well as in protocol analysis. Therefore, we will further optimize and perfect the enhanced authentication tests theory based on the researches of the cryptographic and structural properties of test compo-

nents.

Otway-Rees protocol is an authentication and secret key distributing protocol based on symmetric cryptosystem. The protocol has three principals, trusted third-party server distributes session key for other two communication clients. The security properties of the protocol include confidentiality and authentication. Because of the diversity and generality in protocol principal, security properties, protocol targets and attack types, it is often used to verify the effectiveness of a new formal analysis method. At the same time, many improved protocols are proposed, including message component reconstruction, message component addition, interactive message addition and other methods. Based on BAN logic [17], strand spaces model (SSM) [15], original authentication tests (OAT) [7], protocol composition logic(PCL) [10] and other methods [1], the correctness of the corresponding improved schemes is proved.

Most of the above improvement schemes are put forward on the basis of known defects, and then their correctness is verified by using the existing formal analysis method. The improvement scheme is lack of reasoning and analysis, which fully reflects the shortcomings of most formal methods in protocol defect location and scheme improvement analysis. In this paper, the original Otway-Rees protocol is used to verify the feasibility and preciseness of the optimized enhanced authentication tests method compared with other analysis methods in the localization of protocol defects and the output of the improved schemes.

In short, there are three contributions of our work:

1) Extend the cryptographic and structural properties of authentication tests components from the aspects of principal knowability, parameter types and roles in challenge response structure, and give the parameter consistency judgment theorem.

2) Optimize the enhanced authentication test method, improve the formalization of protocol analysis, and improve the intuitiveness and simplicity of protocol analysis.

3) Use the optimized enhanced authentication tests method to analyze and locate the defects of Otway-Rees protocol, and give three improvement schemes.

The rest of this paper is organized as follows. Section 2 further expand the theoretical system of the original authentication tests, then formalize and optimize the method system of the enhanced authentication tests. In Section 3, we use the optimized enhanced authentication tests method to analyze the correspondence of Otway-Rees protocol, and locate the defects in Otway-Rees protocol. In Section 4, we deduce and output the improved schemes of Otway-Rees protocol based on the parameter consistency judgment theorem. The advantages of the optimized enhanced authentication test method over other formal methods are analyzed in Section 5, and our conclusions are made in Section 6.

2 Authentication Tests

2.1 Symbols and Semantics

The symbols used in this paper and their semantics are shown in Table 1.

2.2 Extension of Authentication Tests

The basic concepts and theorems of authentication tests can be found in [14, 16].

Definition 1. *Suppose X is a principal, n is a node on the strand of X, a is a atomic term, t is a component of n and satisfies $a \sqsubset t \sqsubset term(n)$. If the X can obtain a through t, then t is said to be knowable to X on a, otherwise t is said to be unknowable to X on a.*

The following two properties can be derived from Definition 1.

Property 1. *Suppose X is a principal, n is a node on the strand of X, a and b are terms, t is a component of n and satisfies $a \sqsubset t \wedge b \sqsubset t$. If t to be knowable to X on a, and t to be knowable to X on b, then t to be knowable to X on ab.*

Property 2. *Suppose X is a principal, n is a node on the strand of X, a is a atomic term, t is a component of n.*

1) *If t is an atomic component, then t is knowable to X;*

2) *If t is an encryption component, let $t = \{h\}_k$ where a is the component of h, and K_X is the key set of X.*

   a. *If $k^{-1} \in K_X$, t is knowable to X on a;*

   b. *If $k^{-1} \notin K_X$ and t does not source from X, t is unknowable to X on a.*

Definition 2. *Suppose X, Y are principals, s_X is a strand of X, s_Y is a strand of Y, $n, n' \in s_X$, $sign(n) = +$, $sign(n') = -$, a is a atomic term originated from n uniquely, t is a component of $term(n)$ and uniquely originates at n, t' is a component of $term(n')$ and uniquely originates in s_Y, $a \sqsubset t \wedge a \sqsubset t'$. If $n \Rightarrow^+ n'$ or n' is a test of $T(a)$, then:*

1) *t is called challenge component of a in $T(a)$, and t' is called response component of a in $T(a)$;*

2) *X is called test initiator of $T(a)$, and Y is called test responder of $T(a)$.*

Definition 3. *Suppose t is an atomic term, X, Y are principal identifiers. If one of the following conditions is met, then X can determine that t is the identity of X to Y, which is denoted as $t \in id(X, X, Y)$.*

1) *$t = X$;*

2) *$t = k_X{-}1 \wedge t \notin K_I$;*

Table 1: The semantics of symbols in the paper

Symbols	Semantics of symbols
Σ	The strand spaces of a protocol.
N	The set of nodes.
C	A bundle of a protocol.
K	The set of keys.
K_X	The key set of principal X.
K_I	The key set of infiltrators.
s_X	The strand of principal X.
k_X	The public key of principal X.
k_X^{-1}	The private key of principal X.
k_{XY}	The shared key of principal X and Y, $k_{XY}^{-1} = k_{XY}$.
gh	A concatenated term consisting of g and h.
$\{t\}_k$	Term t is encrypted with k.
$a \sqsubset b$	Term a is a subterm of term b.
$n_1 \Rightarrow n_2$	A edge expresses that n_1 is an immediate causal predecessor of n_2 on the same strand.
$n_1 \Rightarrow^+ n_2$	A edge expresses that n_1 is an predecessor of n_2 (not necessarily immediately) on the same strand.
$sign(n)$	The sign of a node n, $'+'$ indicates that the node sends a message, and $'-'$ indicates that the node accepts a message.
$term(n)$	The signed term of a node n.
$un\_term(n)$	The unsigned part of a node n.
$< s, i >$	The ith node on the strand s.
$height(x)$	The height of a strand or a node in a protocol bundle.
$P(x)$	The parameter set of x, x may be a term or a strand.
$T(x)$	An authentication test of term x.
$OT(x)$	An outgoing authentication test of term x.
$IT(x)$	An incoming authentication test of term x.
$UT(x)$	An unsolicited authentication test of term x.

3) $t = k_{XY} \wedge t \notin K_I$;

4) $t = k(XY)$.

Definition 4. *Suppose t is an atomic term, X, Y, Z are principal identifiers. If one of the following conditions is met, then X can determine that t is the identity of Y to Z, which is denoted as $t \in id(X, Y, Z)$.*

1) $t = Y$;

2) $t = k(XY)$, *and X can confirm that Z knows that a is a shared secret between X and Y.*

Definition 5. *Suppose a is an atomic term, X, Y are principals of $T(a)$, X is the test initiator of $T(a)$, Y is the test responder of $T(a)$, $t = \{h\}_k$ is a test component of $T(a)$, g is an atomic term with $g \sqsubset t$.*

1) If $g = k$, then g is called key parameter of t. The key parameters set of t is denoted as $P_{key}(t)$.

2) If $g = a$, then g is called fresh value parameter of t. The fresh value parameters set of t is denoted as $P_{fresh}(t)$.

3) If $g = X$ or $g = Y$, g is called principal identifier parameter of t. The principal identifier parameters set of t is denoted as $P_{id}(t)$.

4) If $g \notin \{a, k, X, Y\}$, g is called negotiation data parameter of t. The negotiation data parameters set of t is denoted as $P_{data}(t)$.

The following two properties can be derived from Definition 5.

Property 3. *Suppose P is a protocol, X, Y, Z are principals of P, a is an atomic term, t is a test component of $T(a)$, X is the test initiator of $T(a)$, Y is the test responder of $T(a)$, $P(t)$ is the parameter set of t. If $Z \in P(t)$, then $Z \in P_{data}(t)$.*

Property 4. *Suppose a is an atomic term, t is a test component of $T(a)$, $P(t)$ is the parameter set of t, then $P(t) = P_{key}(t) \cup P_{fresh}(t) \cup P_{id}(t) \cup P_{data}(t)$.*

Theorem 1. *Suppose a is an atomic term, X, Y are principals of a $T(a)$, X is the test initiator of $T(a)$, Y is the test responder of $T(a)$, t is a test component of $T(a)$, $P(t)$ is the parameter set of t, g is an atomic term and $g \in P(t)$. Then:*

1) X can reach consistency on $\{k^{-1}\}$ with Y;

2) X can reach consistency on $\{a, Y\}$ with Y;

3) If $g \in id(X, X, Y)$ or $g \in id(Y, X, X)$, and g is a subterm of the test component listed in Table 2, X can reach consistency on $P_{id}(t)$ with Y;

4) If X can reach consistency on $P_{id}(t)$ with Y, X can reach consistency on $P_{data}(t)$ with Y.

Table 2: Conditions for authentication tests based on the keys of principals

K_X	K_Y			
	$null$	k_X	k_Y^{-1}	k_{XY}
$null$	$\times$	$\times$	$IT(a)/UT(a):R$	$IT(a)/UT(a):OK[R]$
k_X^{-1}	$\times$	$\times$	$IT(a)/UT(a):R$	$IT(a)/UT(a):OK[C/R]$
k_Y	$OT(a):C$	$OT(a):C$	$OT(a)/IT(a):C/R$	$OT(a)/IT(a):OK[C/R]$
k_{XY}	$OT(a):OK[C]$	$OT(a):OK[C/R]$	$OT(a)/IT(a):OK[C/R]$	$OT(a)/IT(a):OK[C/R]$

In Table 2, X and Y are the test initiator and test responder of $T(a)$ respectively; $\times$ indicates X cannot confirm identity of Y through the test components composed of secret keys, the part before the colon indicates the type of authentication tests, and the part after the colon represents the type of test components. C is for the challenge component, R is for the response component; OK indicates that X can reach consistency on $\{X,Y\}$ with Y directly; the value in square brackets indicates the type of test component which the negotiation data should be in.

Definition 6. *Suppose* $\boldsymbol{C}$ *is a bundle,* X,Y *are principals,* s_X, s_Y *are strands of* X *and* Y *respectively,* $P(s_X)$ *is the parameter set of* s_X*,* $P(s_Y)$ *is the parameter set of* s_Y*. Let* X *be initiator,* Y *be responder,* $P(s) = P(s_A) \cap P(s_B)$*. IF* X *and* Y *can reach consistency on* $P(s)$*, then for* X*,* s_X *and* s_Y *satisfy correspondence.*

2.3 Optimization of Enhanced Authentication Tests

Theorem 2. *Suppose* $\boldsymbol{C}$ *is a bundle,* X,Y *are principals,* s_X *is a strand of* X*,* s_Y *is a strand of* Y*,* $P(s_X)$ *is the parameter set of* s_X*,* $P(s_Y)$ *is the parameter set of* s_Y*,* $n, n^{'} \in s_X$ *are nodes,* $sign(n) = +$*,* $sign(n^{'}) = -$*,* a *is a term,* t *is a test component of* $term(n)$ *on* a*,* $P(t)$ *is the parameter set of* t*, and* $n \Rightarrow^{+} n^{'}$ *is an* $T(a)$*. Let* X *be test initiator of* $T(a)$*,* Y *be test responder of* $OT(a)$*, then there must exist regular nodes* $m, m^{'} \in s_Y$ *that satisfy* t *is a component of* $term(m)$*, and* $m \Rightarrow^{+} m^{'}$ *is a transforming edge of* a*. Furthermore, if* $P(s_X) \cap P(s_Y) = P(t)$ *and* X *canreach consistency on* $P(t)$ *with* Y*,* $n \Rightarrow^{+} n^{'}$ *and* $m \Rightarrow^{+} m^{'}$ *satisfy correspondence on* t*.*

Theorem 3. *Suppose* $\boldsymbol{C}$ *is a bundle,* X,Y *are principals,* s_X *is a strand of* X*,* s_Y *is a strand of* Y*,* $P(s_X)$ *is the parameter set of* s_X*,* $P(s_Y)$ *is the parameter set of* s_Y*,* $n, n^{'} \in s_X$ *are nodes,* $sign(n) = +$*,* $sign(n^{'}) = -$*,* a *is a term,* t *is a test component of* $term(n^{'})$ *on* a*,* $P(t)$ *is the parameter set of* t*, and* $n \Rightarrow^{+} n^{'}$ *is an* $IT(a)$*. Let* X *be test initiator of* $IT(a)$*,* Y *be test responder of* $IT(a)$*, then there must exist regular nodes* $m, m^{'} \in s_Y$ *that satisfy* t *is a component of* $term(m^{'})$*, and* $m \Rightarrow^{+} m^{'}$ *is a transforming edge of* a*. Furthermore, if* $P(s_X) \cap P(s_Y) = P(t)$ *and* X *reach consistency on* $P(t)$ *with* Y*,* $n \Rightarrow^{+} n^{'}$ *and* $m \Rightarrow^{+} m^{'}$ *satisfy correspondence on* t*.*

Theorem 4. *Suppose* $\boldsymbol{C}$ *is a bundle,* X,Y *are principals,* s_X *is a strand of* X*,* s_Y *is a strand of* Y*,* $P(s_X)$ *is the parameter set of* s_X*,* $P(s_Y)$ *is the parameter set of* s_Y*,* $n \in s_X$*,* $sign(n) = -$*,* a *is a term,* t *is a test component* $term(n)$ *on* a*, and* $P(t)$ *is the parameter set of* t*,* n *is an* $UT(a)$*. Let* X *be test initiator of* $UT(a)$*,* Y *be test responder of* $UT(a)$*, then there must exist a regular node* $m \in s_B$ *that satisfies* t *is a component of* $term(m)$*. Furthermore, if* $P(s_X) \cap P(s_Y) = P(t)$ *and* X *can reach consistency on* $P(t)$ *with* Y*,* n *and* m *satisfy correspondence on* t*.*

Definition 7. *Suppose* X,Y *are principals,* s_X *is a strand of* X*,* s_Y *is a strand of* Y*, let* Con_{XY} *be the degree of correspondence of* Y *to* X*.*

1) *If* X *cannot confirm the correspondence with* Y*, then* $Con_{XY} = 0$*;*

2) *If* X *can confirm the correspondence with* Y *on* $< s_Y, i > \Rightarrow^{+} < s_Y, j > (j > i)$*, then* $Con_{XY} = j$*.*

Definition 8. *Suppose* X,Y *are principals,* s_X *is a strand of* X*,* s_Y *is a strand of* Y*,* $T(s_X) = \{T_1(a_1), T_2(a_2), \cdots, T_n(a_n)\}(n < length(s_X))$ *is a ordered set of authentication tests whose test responder is* Y *on* s_X*. Let* Con_{XY} *be the degree of correspondence of* Y *to* X*, and* Con_{XY} *is initialized to* 0*. Traversing all elements of* $T(s_X)$*, if there exist an edge* $< s_Y, i > \Rightarrow^{+} < s_Y, j >$ *or a node* $< s_Y, j >$ *that satisfy correspondence with* $T_k(a_k)(1 \leqslant k \leqslant n)$*, then* $Con_{XY} = (j > Con_{XY})?j : Con_{XY}$*.*

Definition 9. *Suppose* P *is a protocol,* X,Y *are principals of* P*,* s_X *is a strand of* X*. Let* M_p *be the correspondence matrix of* P*, the elements in* M_P *are denoted as* m_{XY}*.*

1) *If* $X = Y$*, then* $m_{XY} = height(s_X)$*;*

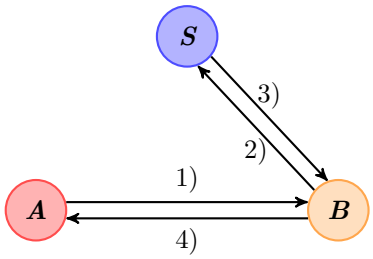
2) *If* $X \neq Y$*, then* $m_{XY} = Con_{XY}$*.*

3 Analysis of Otway-Rees

3.1 Formal Description of Otway-Rees

Otway-Rees is a key distribution protocol based on the third party. The principals of Otway-Rees include two communication clients and one authentication server. Messages exchanging in Otway-Rees is shown in Figure 1. A is a protocol initiator, B is a protocol responder, S is a credible server responsible for generating and distributing the session secret key for A and B, k_{AS} is a shared secret key between A and S, k_{BS} is a shared secret key between B and S, N_a and N_b are random values generated by A and B respectively. k_{AB} is a session key distributed by S for A and B in the same round. M is a value selected randomly by A as a unique identifier of a round of Otway-Rees.

The design goals of Otway-Rees is that S distributes a shared secret key k_{AB} for A and B, and A can confirm that k_{AB} is the shared secret key distributed by S for A and B in the same round of protocol.



1) $A \rightarrow B$: $MAB\{N_aMAB\}_{k_{AS}}$

2) $B \rightarrow S$: $MAB\{N_aMAB\}_{k_{AS}}\{N_bMAB\}_{k_{BS}}$

3) $S \rightarrow B$: $M\{N_ak_{AB}\}_{k_{AS}}\{N_bMk_{AB}\}_{k_{BS}}$

4) $B \rightarrow A$: $M\{N_ak_{AB}\}_{k_{AS}}\{M\}_{k_{AB}}$

Figure 1: Message exchange in Otway-Rees

Definition 10. *Define Σ as a strand space of Otway-Rees, then the regular strands in Σ is defined as follows:*

1) *Let $Init[*]$ be the set of initiator strands whose trace is*

$$< +MAB\{N_aMAB\}_{k_{AS}}, -M\{N_ak_{AB}\}_{k_{AS}} > .$$

The principal associated with $s \in Init[*]$ is A, $P(s) = \{A, B, S, N_a, k_{AB}, k_{AS}, M\}$.

2) *Let $Resp[*]$ be the set of responder strands whose trace is*

$$< -MABH_1, +MABH_1\{N_bMAB\}_{k_{BS}}, \\ -MH_2\{N_bk_{AB}\}_{k_{BS}}, +MH_2 > .$$

where $H_1 = \{N_aMAB\}_{k_{AS}}$, $H_2 = \{N_ak_{AB}\}_{k_{AB}}$, H_1 and H_2 are unknowable to B. The principal associated with $s \in Resp[*]$ is B, $P(s) = \{A, B, S, N_b, k_{AB}, k_{BS}, H_1, H_2, M\}$.

3) *Let $Serv[*]$ be the set of server strands whose trace is*

$$< -MAB\{N_aMAB\}_{k_{AS}}\{N_bMAB\}_{k_{BS}}, \\ +M\{N_ak_{AB}\}_{k_{AS}}\{N_bk_{AB}\}_{k_{BS}} > .$$

The principal associated with $s \in Serv[*]$ is S, $P(s) = \{A, B, S, N_a, N_b, k_{AB}, k_{AS}, k_{BS}, M\}$.

3.2 Analysis of Correspondence

Proposition 1. *Suppose Σ is an Otway-Rees space, $\boldsymbol{C}$ is a bundle in Σ, A, B, S are principals, s_A, s_B, s_S are strands of A, B and S respectively. If:*

1) *$s_A \in Init[*]$ with $height(s_A) = 2$;*

2) *N_a is uniquely originating in Σ with $N_a \notin \{M, A, B, S, N_b\} \cup \boldsymbol{K}$;*

3) *$k_{AS}, k_{BS} \notin K_I$.*

Then $Con_{AB} = 0$, $Con_{AS} = 2$.

Proof. 1) Prove $Con_{AS} = 2$.

Since N_a is uniquely originating in Σ and originates from s_A, $k_{AS} \notin K_I$, $\{N_aMAB\}_{k_{AS}}$ is a test component of $term(< s_A, 1 >)$ on N_a, $< s_A, 1 > \Rightarrow < s_A, 2 >$ is an outgoing test of N_a, labeled as $OT(N_a)$. Since $k_{AS} \notin K_I$, there exist a unique stand $s_S \in Serv[*]$ in Σ, and the corresponding principal is S. By Definition 10, $< s_S, 1 > \Rightarrow < s_S, 2 >$ is the transforming edge of N_a. By Definition 2, A is the test initiator of $OT(N_a)$, S is the test responder of $OT(N_a)$. By Definition 3, $A \in id(A, A, S)$, $B \in id(A, B, S)$, $k_{AS} \in id(A, S, S)$. By Definition 5, $P_{key}(\{N_aMAB\}_{k_{AS}}) = \{k_{AS}\}$, $P_{fresh}(\{N_aMAB\}_{k_{AS}}) = \{N_a\}$, $P_{id}(\{N_aMAB\}_{k_{AS}}) = \{A, S\}$, $P_{data}(\{N_aMAB\}_{k_{AS}}) = \{M, B\}$. By Property 4 and Theorem 1, A can reach consistency on $P(\{N_aMAB\}_{k_{AS}}) = \{A, B, S, N_a, k_{AS}, M\}$ with S.

$\{N_ak_{AB}\}_{k_{AS}}$ is a test component of $term(< s_A, 2 >)$ on N_a, $< s_A, 1 > \Rightarrow < s_A, 2 >$ is also an incoming test of N_a, labeled as $IT(N_a)$. Since $k_{AS} \notin K_I$, there exist a unique stand $s_S \in Serv[*]$ in Σ, and the corresponding principal is S. By Definition 10, $< s_S, 1 > \Rightarrow < s_S, 2 >$ is the transforming edge of N_a. By Definition 2, A is the test initiator of $IT(N_a)$, S is the test responder of $IT(N_a)$. By Definition 3, $k_{AS} \in id(S, S, A)$. By Definition 5, $P_{key}(\{N_ak_{AB}\}_{k_{AS}}) = \{k_{AS}\}$, $P_{fresh}(\{N_ak_{AB}\}_{k_{AS}}) = \{N_a\}$, $P_{id}(\{N_ak_{AB}\}_{k_{AS}}) = \{A, S\}$, $P_{data}(\{N_ak_{AB}\}_{k_{AS}}) = \{k_{AB}\}$. By Property 4 and Theorem 1, A can reach consistency on $P(\{N_ak_{AB}\}_{k_{AS}}) = \{A, B, S, N_a, k_{AS}, k_{AB}\}$ with S.

$P(\{N_ak_{AB}\}_{k_{AS}}) \cup P(\{N_aMAB\}_{k_{AS}}) = \{A, B, S, N_a, k_{AS}, k_{AB}, M\}$. Therefore, A can reach consistency on

$\{A, B, S, N_a, k_{AS}, k_{AB}, M\}$ with S. $P(s_A) = \{A, B, S, N_a, k_{AS}, k_{AB}, M\}$, $P(s_S) = \{A, B, S, N_a, N_b, k_{AS}, k_{AB}, M\}$, $P(s_A) \cap P(s_S) = \{A, B, S, N_a, k_{AS}, k_{AB}, M\}$. By Theorem 5 and Theorem 6, $< s_A, 1 > \Rightarrow < s_A, 2 >$ and $< s_S, 1 > \Rightarrow < s_S, 2 >$ satisfy correspondence on $\{A, B, S, N_a, k_{AS}, k_{AB}, M\}$. Since $height(< s_S, 2 >) = 2$, by Definition 8, $Con_{AS} = 2$.

2) Prove $Con_{AB} = 0$.

Since $k_{AS} \notin K_B$, by Definition 1, $\{N_a MAB\}_{k_{AS}}$ and $\{N_a k_{AB}\}_{k_{AS}}$ are unknowable to B on subterm. By Table 1, Neither $\{N_a MAB\}_{k_{AS}}$ nor $\{N_a k_{AB}\}_{k_{AS}}$ can constitute a authentication test for B. Therefore, A cannot confirm that there exist a s_B with the corresponding principal is B, nor can reach consistency on $P(s_A) = \{A, B, S, N_a, k_{AS}, M\}$ with B. By Definition 10, Theorem 2 and Theorem 2, there is no edges that have correspondence with $< s_A, 1 > \Rightarrow < s_A, 2 >$. By Definition 8, $Con_{AB} = 0$.

$Con_{AS} = 2$ and $Con_{AB} = 0$, the conclusion of Proposition 1 is proved. □

Proposition 2. *Suppose Σ is an Otway-Rees space, $\boldsymbol{C}$ is a bundle in Σ, A, B, S are principals, s_A, s_B, s_S are strands of A, B and S respectively. If:*

1) $s_B \in Resp[*]$ with $height(s_B) = 4$;

2) N_b is uniquely originating in Σ with $N_b \notin \{M, A, B, S, N_a\} \cup \boldsymbol{K}$;

3) $k_{AS}, k_{BS} \notin K_I$.

Then $Con_{BA} = 0$, $Con_{BS} = 2$.

Proposition 3. *Suppose Σ is an Otway-Rees space, $\boldsymbol{C}$ is a bundle in Σ, A, B, S are principals, s_A, s_B, s_S are strands of A, B and S respectably, If:*

1) $s_S \in Serv[*]$ with $height(s_S) = 2$;

2) $k_{AS}, k_{BS} \notin K_I$.

Then $Con_{SA} = 0$, $Con_{SB} = 0$.

The same principle can prove the correctness of the conclusion of proposition 2 and proposition 3, which is omitted here.

3.3 Discussion of Correspondence Matrix

According to the conclusions of Proposition 1, Proposition 2 and Proposition 3, by Definition 9, the correspondence matrix of Otway-Rees on $\{A, B, S, k_{AB}, M\}$ is as follows.

$$M_{Otway-Rees} = \begin{array}{c} \\ A \\ B \\ S \end{array} \begin{array}{c} \begin{array}{ccc} A & B & S \end{array} \\ \begin{bmatrix} 2 & 0 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 2 \end{bmatrix} \end{array}$$

From the correspondence matrix of Otway-Rees protocol, the following conclusions can be drawn.

1) $Con_{AS} = 2$. A can reach consistency on $\{A, B, S, M, N_a, k_{AS}, k_{AB}\}$ with S, A can confirm that k_{AB} is the shared secret key distributed by S for A and B in the round of N_a.

2) $Con_{BS} = 2$. B can reach consistency on $\{A, B, S, M, N_b, k_{BS}, \underline{k_{AB}}\}$ with S, B can confirm that $\underline{k_{AB}}$ is the shared secret key distributed by S for A and B in the round of N_b.

3) $Con_{AB} = 0$, $Con_{BA} = 0$. There is no correspondence between A and B. That is, neither A nor B can confirm that k_{AB} and $\underline{k_{AB}}$ distributed by S are the same secret key in the same round identified as M.

4) $Con_{SA} = 0$, $Con_{SB} = 0$. S can neither reach consistency on $\{A, B, S, M, N_a, k_{AS}, k_{AB}\}$ with A, nor reach consistency on $\{A, B, S, M, N_b, k_{BS}, k_{AB}\}$ with B. That is, S cannot confirm that k_{AB} was distributed successfully for A and B in the same round identified as M.

In Otway-Rees protocol, A is the initiator of distribution of k_{AB}, S is the generator and distributor of k_{AB}, and B is both the receiver and the forwarder of k_{AB}. whether k_{AB} is successfully distributed is finally confirmed by A. When a round of protocol is completed, both A and B should confirm that k_{AB} they receive is distributed by S for A and B, and A can confirm that the k_{AB} received by A is the same as that received by B. Therefore, in order to meet design goals of Otway-Rees, not only $Con_{AS} > 0$ and $Con_{BS} > 0$, but also $Con_{AB} > 0$ is required.

A is the initiator of Otway-Rees, whose main purpose is to obtain a session secret key k_{AB} from S for communicating with B, so the validity of k_{AB} can be guaranteed only if A can reach consistency on k_{AB} with B. Since B is the forwarder of k_{AB}, not the terminator of the protocol, it is unnecessary for design goals of Otway-Rees that B can reach consistency on k_{AB} with A. Therefore, $Con_{BA} = 0$ does not affect the correctness of Otway-Rees.

By Definition 10, S exits a round of running when k_{AB} is distributed at $< s_S, 2 >$. Therefore, whether k_{AB} is successfully distributed or not is not the design goal of Otway-Rees. So $Con_{SA} = 0$ and $Con_{SB} = 0$ do not affect the correctness of Otway-Rees.

According to the correspondence matrix of Otway-Rees, the original Otway-Rees meets the goals of $Con_{AS} > 0$ and $Con_{BS} > 0$, but not $Con_{AB} > 0$. Therefore, $Con_{AB} = 0$ is the main factor that causes the attacks of Otway-Rees, which is pointed out in [7, 10, 16]. The specific attack examples of Otway-Rees are not given here again.

4 Improvements of Otway-Rees

4.1 Analysis of Improvement Schemes

The main improvement goal of the original Otway-Rees is to make $Con_{AB} > 0$, so A should reach consistency on $P(s_A) \cap P(s_B) = \{A, B, S, k_{AB}, M\}$ with B. By Theorem 5, Theorem 6 and Theorem 7, then there should exist a test component t of $term(< s_A, i >)(1 \leq i \leq 2)$ with $P(t) = \{A, B, S, k_{AB}, M\}$. Without changing the order and the number of messages exchanged between principals, there are two ways to improve the original Otway-Rees: to reconstruct the existing test components of $term(< s_A, i >)$, or to add new test components on $term(< s_A, i >)$.

Since $k_{AS} \notin K_B$, by Table 2, Neither $\{N_a, M, A, B\}_{k_{AS}}$ nor $\{N_a, k_{AB}\}_{k_{AS}}$ can constitute a authentication test to B. Therefore, it is unfeasible to improve the original Otway-Rees just by reconstructing the existing test components of $term(< s_A, i >)$.

Suppose $t = \{h\}_k$ is a new test components to be added to $term(< s_A, i >)$, then $P(t) = \{A, B, S, k_{AB}, M\}$. By Definition 10, $< s_S, 1 >$ where k_{AB} uniquely originates at occurs after $< s_A, 1 >$ in the same round of Otway-Rees, k_{AB} cannot be a subterm of $term(< s_A, 1 >)$, so t can only be a subterm of $term(< s_A, 2 >)$. By Definition 10, t is a components of $term(< s_B, 4 >)$ and uniquely originates at $< s_B, 4 >$.

Lemma 1. *Suppose P is the Otway-Rees protocol, Σ is a space of P, A, B, S are principals of P, $s_A \in Init[*]$, $s_B \in Resp[*]$, $s_S \in Serv[*]$, k_{AB} and $\underline{k_{AB}}$ are session keys distributed by S to A and B respectively in the same round of P, A can confirm a is a term uniquely originating in Σ, $t = \{a\}_{\underline{k_{AB}}}$ is a components of $term(< s_B, 4 >)$ and uniquely originates at $< s_B, 4 >$. If $\underline{k_{AB}} = k_{AB}$, then A can reach consistency on k_{AB} with B.*

Proof. By Definition 10, $un\_term(< s_A, 2 >) = un\_term(< s_A, 2 >)$, $t = \{a\}_{\underline{k_{AB}}}$ is a components of $term(< s_B, 4 >)$, then $t = \{a\}_{\underline{k_{AB}}}$ is also a components of $term(< s_A, 2 >)$. $\underline{k_{AB}} = k_{AB}$, then $\underline{k_{AB}} \in K_A$ and $\underline{k_{AB}} \notin K_I$. A can confirm a is uniquely originating, t is uniquely originates at $< s_B, 4 >$.

If a is uniquely originates at $< s_A, 1 >$, by Definition 10, $< s_A, 1 > \Rightarrow < s_A, 2 >$ is an incoming test of a. By Definition 3, $\underline{k_{AB}} \in id(B, A, A)$, $\underline{k_{AB}} \in id(B, B, A)$, so A can reach consistency on $\{A, B\}$ with B. $\underline{k_{AB}} \sqsubset \{a\}_{\underline{k_{AB}}}$. By Theorem 1, A can reach consistency on k_{AB} with B.

If a is uniquely originates at other node in Σ, $sign(< s_A, 2 >) = -$, $< s_A, 2 >$ is an unsolicited test of a. Similarly, A can also reach consistency on k_{AB} with B. □

Lemma 2. *Suppose P is the Otway-Rees protocol, Σ is a space of P, A, B, S are principals of P, $s_A \in Init[*]$, $s_B \in Resp[*]$, $s_S \in Serv[*]$, k_{AB} is distributed by S for A and B in the same round of P. If A can reach consistency on k_{AB} with B, then A can reach consistency on $\{A, B, S, k_{AB}\}$ with B.*

Proof. A can reach consistency on k_{AB} with B, then $k_{AB} \in K_B$ and $k_{AB} \notin K_I$. By Definition 3, $k_{AB} \in id(B, B, A)$ and $k_{AB} \in id(B, A, A)$, so A can reach consistency on $\{A, B\}$ with B.

According to Lemma 1, $Con_{AS} = 2$ and $Con_{AS} = 2$, both A and B can confirm that k_{AB} is the shared secret key distributed by S for A and B, so k_{AB} is also a shared secret among A, B, and S. By Definition 4, $k_{AB} \in id(B, S, A)$, then A can reach consistency on $\{S\}$ with B.

$\{k_{AB}\} \cup \{A, B\} \cup \{S\} = \{A, B, S, k_{AB}\}$. Therefore, A can reach consistency on $\{A, B, S, k_{AB}\}$ with B. □

Proposition 4. *Suppose P is an improvement of Otway-Rees, Σ is a space of P, A, B, S are principals of P, $s_A \in Init[*]$, $s_B \in Resp[*]$, $s_S \in Serv[*]$, k_{AB} is distributed by S for A and B in the same round of P, $t = \{h\}_{k_{AB}}$ is a new test component added to $< s_B, 4 >$, $a \sqsubset h$ is a term. If:*

1) *A can confirm that A can reach consistency on k_{AB} with B;*
2) *A can confirm that a is uniquely originating in Σ;*
3) *$M \sqsubset h$.*

Then $Con_{AB} > 0$.

Proof. Since $t = \{h\}_{k_{AB}}$ is a new test component added to $< s_B, 4 >$ and $a \sqsubset h$, by Definition 10, $un\_term(< s_A, 2 >) = un\_term(< s_B, 4 >)$, then $a \sqsubset term(< s_A, 2 >)$.

By Definition 10, k_{AB} is just a proper subterm of $\{N_a k_{AB}\}_{k_{AS}}$ and $\{N_b k_{AB}\}_{k_{BS}}$ in Σ, since $k_{AS}, k_{BS} \notin K_I$, then $k_{AB} \notin K_I$. If A can confirm that A can reach consistency on k_{AB} with B, then A can confirm that $k_{AB} \in K_A$ and $k_{AB} \in K_B$.

If a uniquely originates at $< s_A, 1 >$, $sign(< s_A, 1 >) = +$, and $sign(< s_A, 2 >) = -$, by Definition 10, $< s_A, 1 > \Rightarrow < s_A, 2 >$ is an incoming test of a to B.

If a uniquely originates at other node, $sign(< s_A, 2 >) = -$, $< s_A, 2 >$ is an unsolicited test of a to B.

By Lemma2, A can reach consistency on $\{A, B, S, k_{AB}\}$ with B. If $M \sqsubset h$, by Theorem 4, then A can reach consistency on $\{A, B, S, M, a, k_{AB}\}$ with B. By Definition 10, $P(s_A) \cap P(s_B) = \{A, B, S, M, a, k_{AB}\}$, $height(< s_B, 4 >) = 4$, by Theorem 6, Theorem 7 and Definition 19, $Con_{AB} = 4 > 0$. □

4.2 Improvement based on $IT(N_a)$

Suppose $t = \{h\}_{k_{AB}}$ is a new test component added to $< s_B, 4 >$, $a \sqsubset h$. By Definition 10, N_a originates at $< s_A, 1 >$, if N_a is uniquely originating in Σ, let $a = N_a$ and $h = MN_a$, then $t = \{MN_a\}_{k_{AB}}$, $< s_A, 1 > \Rightarrow < s_A, 2 >$ is an $IT(N_a)$ to B, by Proposition 4, $Con_{AB} = 4$.

Since $sign(< s_B, 4 >) = -$, there must exist a node $< s_B, i > \Rightarrow^+ < s_B, 4 >$ with $sign(< s_B, i >) = +$ and $N_a \sqsubset term(< s_B, i >)$. By Definition 10, only $< s_B, 1 >$ and $< s_B, 3 >$ are satisfied. $k_{AS} \notin K_B$, by Property 2,

$\{N_aMAB\}_{k_{AS}}$ and $\{N_ak_{AB}\}_{k_{AS}}$ is unknowable to B on N_a. $un\_term(< s_B, 3 >) = un\_term(< s_S, 2 >)$. Therefore, N_a should be added to $\{N_bk_{AB}\}_{k_{BS}}$ at $< s_S, 2 >$. At the same time, this makes the length of $\{N_ak_{AB}\}_{k_{AS}}$ and $\{N_aN_bk_{AB}\}_{k_{BS}}$ different, which can effectively avoid the type defect attacks pointed out in [10].

The improved Otway-Rees based on $IT(N_a)$ is shown in Figure 2.

1) $A \rightarrow B$: $MAB\{N_aMAB\}_{k_{AS}}$
2) $B \rightarrow S$: $MAB\{N_aMAB\}_{k_{AS}}\{N_bMAB\}_{k_{BS}}$
3) $S \rightarrow B$: $M\{N_ak_{AB}\}_{k_{AS}}\{N_aN_bk_{AB}\}_{k_{BS}}$
4) $B \rightarrow A$: $M\{N_ak_{AB}\}_{k_{AS}}\{MN_a\}_{k_{AB}}$

Figure 2: Improved Otway-Rees based on $IT(N_a)$

At present, This improved scheme of Otway-Rees based on $IT(N_a)$ has not been seen in other literatures.

4.3 Improvement based on $UT(N_b)$

In the original Otway-Rees protocol, N_b originates at $< s_B, 2 >$, $k_{BS} \notin K_B$, by Property 2, $\{N_bMAB\}_{k_{BS}}$ and $\{N_bk_{AB}\}_{k_{BS}}$ is unknowable to A on N_b, so A cannot confirm the freshness of N_b, that is, A cannot confirm that N_b and N_a are created in the same round.

According to Proposition 4, $Con_{AS} = 2$. If N_b is added to $\{N_ak_{AB}\}_{k_{AS}}$ at $< s_S, 2 >$, then A can reach consistency on $\{N_b\}$ with B, so A can confirm through $\{N_aN_bk_{AB}\}_{k_{AS}}$ that N_b and N_a are created in the same round. The length of $\{N_aN_bk_{AB}\}_{k_{AS}}$ added with N_b is different from that of $\{N_bk_{AB}\}_{k_{BS}}$, which can effectively prevent type defect attack pointed out in [10].

Suppose $t = \{h\}_{k_{AB}}$ is a new test component added to $< s_B, 4 >$, $a \sqsubset t$. According to Proposition 4, if A can confirm N_b is uniquely originating in Σ, let $a = N_b$ and $h = MN_b$, then $< s_A, 2 >$ is an $UT(N_b)$ to B and $Con_{AB} = 4$.

The improved Otway-Rees based on $UT(N_b)$ is shown in Figure 3.

1) $A \rightarrow B$: $MAB\{N_aMAB\}_{k_{AS}}$
2) $B \rightarrow S$: $MAB\{N_aMAB\}_{k_{AS}}\{N_bMAB\}_{k_{BS}}$
3) $S \rightarrow B$: $M\{N_aN_bk_{AB}\}_{k_{AS}}\{N_bk_{AB}\}_{k_{BS}}$
4) $B \rightarrow A$: $M\{N_aN_bk_{AB}\}_{k_{AS}}\{MN_b\}_{k_{AB}}$

Figure 3: Improved Otway-Rees based on $UT(N_b)$

This improved scheme of Otway-Rees based on $IT(N_b)$ is similar to the improvement scheme in [10] shown in Figure 4.

Compared with the improvement based on $UT(N_b)$, the improvement in [10] still exist parameter inconsistencies and data redundancy.

1) $A \rightarrow B$: $MAB\{N_aMAB\}_{k_{AS}}$
2) $B \rightarrow S$: $MAB\{N_aMAB\}_{k_{AS}}\{N_bMAB\}_{k_{BS}}$
3) $S \rightarrow B$: $M\{AN_aN_bk_{AB}\}_{k_{AS}}\{BN_bk_{AB}\}_{k_{BS}}$
4) $B \rightarrow A$: $M\{N_aN_bk_{AB}\}_{k_{AS}}\{N_b\}_{k_{AB}}$

Figure 4: Improved Otway-Rees in [10]

1) $\{AN_aN_bk_{AB}\}_{k_{AS}}$ and $\{BN_bk_{AB}\}_{k_{BS}}$ exists redundancy of principal identifier.

By Definition 3, $k_{AS} \in id(S, A, A)$ and $k_{BS} \in id(S, B, B)$, so A in $\{AN_aN_bk_{AB}\}_{k_{AS}}$ and B in $\{BN_bk_{AB}\}_{k_{BS}}$ are redundant.

2) A cannot confirm reach consistency on M with B based on the new test component $\{N_b\}_{k_{AB}}$.

M is a random value created by A as a unique identifier of a round of Otway-Rees. In cooperation with N_a and N_b, M can enable the principals to determine the uniqueness of the protocol round and further guarantee the correctness of Otway-Rees.

4.4 Improvement based on $IT(M)$

If A can confirm that M is uniquely originating in Σ, the improvements scheme based on $IT(N_a)$ and $UT(N_b)$ can be further optimized to be based on $IT(M)$.

Suppose $t = \{h\}_{k_{AB}}$ is a new test component added to $< s_B, 4 >$, $a \sqsubset h$. Let $h = a = M$. By Definition 10, M originates at $< s_A, 1 >$, then $t = \{M\}_{k_{AB}}$, $< s_A, 1 > \Rightarrow < s_A, 2 >$ is an $IT(M)$ to B, by Proposition 4, $Con_{AB} = 4$.

In order to avoid the type defect attacks, M is added to $\{N_bk_{AB}\}_{k_{BS}}$ to make the length of $\{N_ak_{AB}\}_{k_{AS}}$ and $\{N_bMk_{AB}\}_{k_{BS}}$ different.

The improved Otway-Rees based on $IT(M)$ is shown in Figure 5.

1) $A \rightarrow B$: $MAB\{N_aMAB\}_{k_{AS}}$
2) $B \rightarrow S$: $MAB\{N_aMAB\}_{k_{AS}}\{N_bMAB\}_{k_{BS}}$
3) $S \rightarrow B$: $M\{N_ak_{AB}\}_{k_{AS}}\{N_bMk_{AB}\}_{k_{BS}}$
4) $B \rightarrow A$: $M\{N_ak_{AB}\}_{k_{AS}}\{M\}_{k_{AB}}$

Figure 5: Improved Otway-Rees based on $IT(M)$

r Like the improvement scheme based on $IT(N_a)$, the improved scheme of Otway-Rees based on $IT(M)$ has not been seen in other literatures.

5 Comparison with Other Methods

Table 3 shows the comparison of different security protocol formalization methods in protocol defect location,

Table 3: Compared with other methods in defect location and improvement scheme output

Methods	Locating defects	Formalization theory	Improvement schemes
BAN [17]	N	N	$Y(1)$
SSM [15]	N	N	$N(2)$
OAT [7]	N	N	$N(2)$
PCL [10]	Y	N	$N(1)$
EAT	Y	Y	$N(3)$

formalization theoretical basis and improvement scheme. In locating defects column, Y means that the corresponding method can realize the precise location of protocol defects, otherwise, it means that it cannot. In formalization theory column, Y indicates that the theoretical basis of the corresponding method is complete, N indicates that it is missing or incomplete. Y indicates that the number or order of protocol interaction information has been changed, N indicates no change, and the number after Y or N indicates the number of improvement schemes in improvement schemes column.

From table 3, we can draw the conclusion that the optimized enhanced authentication tests has the following advantages over other formal methods.

1) The parameter consistency inference theorem based on test components not only realizes the precise positioning of protocol defects, but also points out the causes of protocol defects and gives the correction scheme.

2) The correspondence degree based on the consistency of parameters can show the authentication relationship among different principals and accurately reflect the defects of protocol design goals.

3) The expansion of test components in cryptography and structure improves the theoretical basis of enhanced authentication tests, which makes protocol analysis process more accurate and logical reasoning more scientific and rigorous.

4) The functionalization of concepts, the quantification of correspondence degree, and the application of parameter consistency judgment table and correspondence degree matrix not only make protocol analysis more clear and concise, but also have a higher degree of formalization, which is convenient for automation.

6 Conclusion

Through the analysis and improvement of the original Otway-Rees protocol, it is found that the expansion of test components in cryptography and structure properties makes enhanced authentication tests further optimized and improved both in theory and method, which makes the protocol analysis process more clear, concise, rigorous and efficient than other formal methods [7, 15] based on theorem proving.

The three new improved schemes of Otway-Rees proposed in this paper have not changed the original cryptosystem and fresh value mechanism, nor changed the number and order of message exchanged among principals. They not only fix the defect that principals can not be inconsistent on the session key, but also avoid the attack of type defect. Compared with other improved schemes [10, 15, 17], they have the advantage of low redundancy.

Acknowledgments

This study was supported by the Natural Science Foundation of Anhui University (KJ2020A0034,KJ2020A0036,KJ2018A0678 and KJ2018A396), the National Natural Science Foundation of China (61902140 and 61300048), Science Foundation for The Excellent Youth Scholars of Anhui University (gxyq2017154), and in part by Anhui Provincial Natural Science Foundation (1608085MF159 and 1908085QF288). The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] M. Backes, "Real-or-random key secrecy of the otway-rees protocol via a symbolic security proof," in *Proceedings of the 21st Annual Conference on Mathematical Foundations of Programming Semantics (MFPS'05)*, pp. 111–145, 2005.

[2] T. Y. Chang, M. S. Hwang, and C. C. Yang, "Password authenticated key exchange and protected password change protocols," *Symmetry*, vol. 9, no. 8, p. 134, 2017.

[3] X. Chen and H. Deng, "Analysis of cryptographic protocol by dynamic epistemic logic," *IEEE Access*, vol. 7, pp. 29981–29988, 2019.

[4] V. Cheval, V. Cortier, and B. Warinschi, "Secure composition of pkis with public key protocols," in *IEEE 30th Computer Security Foundations Symposium (CSF'17)*, pp. 144–158, Aug. 2017.

[5] S. F. Chiou, H. T. Pan, E. F. Cahyadi, and M. S. Hwang, "Cryptanalysis of the mutual authentication and key agreement protocol with smart cards for wireless communications," *International Journal Network Security*, vol. 21, no. 1, pp. 100–104, 2019.

[6] J. D. Guttman, "Cryptographic protocol composition via the authentication tests," in *International Conference on Foundations of Software Science and Computational Structures*, pp. 303–317, 2009.

[7] J. D. Guttman and F. J. Thayer, "Authentication tests," in *Proceeding of IEEE Symposium on Security and Privacy (S&P'00)*, pp. 96–109, 2000.

[8] R. Jiang, A. Hu, and J. Li, "Formal protocol design of ESIKE based on authentication tests," *International Journal Network Security*, vol. 6, no. 3, pp. 246–254, 2008.

[9] C. H. Ling, S. M. Chen, and M. S. Hwang, "Cryptanalysis of tseng-wu group key exchange protocol," *International Journal Network Security*, vol. 18, no. 3, pp. 590–593, 2016.

[10] L. Lu, X. Duan, and J. Ma, "Improvement and security analysis of the otway-rees protocol," *Journal on Communications*, vol. 33, no. Z1, pp. 250–254, 2012.

[11] B. Meng, J. T. Lu, D. J. Wang, and X. D. He, "Survey of security analysis of security protocol implementations," *Journal of Shandong University (Natural Science)*, vol. 53, no. 1, pp. 1–18, 2018.

[12] S. Muhammad, "Applying authentication tests to discover man-in-the-middle attack in security protocols," in *Eighth International Conference on Digital Information Management (ICDIM'13)*, pp. 35–40, 2013.

[13] F. J. Thayer, J. C. Herzog, and J. D. Guttman, "Honest ideals on strand spaces," in *Proceedings of the 11th IEEE Computer Security Foundations Workshop*, pp. 66–77, 1998.

[14] F. J. Thayer, J. C. Herzog, and J. D. Guttman, "Strand spaces: Why is a security protocol correct?," in *Security and Privacy - IEEE Symposium on Security and Privacy*, pp. 160–171, 1998.

[15] F. J. Thayer, J. C. Herzog, and J. D. Guttman, "Mixed strand spaces," in *Proceedings of the 12th IEEE Computer Security Foundations Workshop, (CSFW'99)*, pp. 72–82, 1999.

[16] F. J. Thayer, J. C. Herzog, and J. D. Guttman, "Strand spaces: Proving security protocols correct," *Journal of Computer Security*, vol. 7, no. 1, pp. 191–230, 1999.

[17] J. Wang, Y. Zan, A. Liu, and M. Qu, "Improvement and security analysis of the otway-rees protocol," *Journal of information engineering university*, vol. 15, no. 5, pp. 525–530, 2014.

[18] C. H. Wei, M. S. Hwang, and A. Y. H. Chin, "A secure privacy and authentication protocol for passive RFID tags," *International Journal of Mobile Communications*, vol. 15, no. 3, pp. 266–277, 2017.

[19] M. Yang and J. Luo, "Analysis of security protocols based on authentication test," *Journal of Software*, vol. 1.17, no. 1, pp. 148–156, 2006.

[20] L. Zhang, P. Jiang, T. Jiang, and J. Li, "Security analysis and improvement of lte-r authentication and key agreement protocol in security protocols," *Journal of East China Jiaotong University*, vol. 36, no. 05, pp. 120–128, 2019.

Biography

Lei Yu was born in 1978. He received the MS an BS degree in computer science and technology from Huaibei Normal University of China. Currently, he is an assistant professor and MS supervisor in the school of computer science and technology, Huaibei Normal University, China. His major research interests include cryptography and information security. He has published many papers in related journals (yulei@chnu.edu.cn).

Yu-Yan Guo was born in 1984. She received the PhD degree in cryptography from Hohai University of China. Currently, She is an assistant professor in the school of computer science and technology, Huaibei Normal University, China. Her research interests include cryptography and information security (guoyuyan428@163.com).

Ze-Peng Zhuo was born in 1978. He received the M.S.degree from Huaibei Normal University in 2007, and the Ph.D. degree from Xidian University in 2012. Since 2002, he has been with the School of Mathematical Science, Huaibei Normal Universtiy, where he is currently a professor. His research interests include cryptography and information theory (zepengzhuo@chnu.edu.cn).

Shi-Ming Wei was born in 1962. In 1986 and 1993, he received his bachelor's and master's degrees in basic mathematics from Huaibei Coal Normal University and Northwest University respectively. In 2001, he received his doctoral degree in cryptography from Xidian University. From April 2001 to July 2003, he was engaged in postdoctoral research at the Software Institute of School of Information Science and Technology, Peking University. His research interests include cryptography andinformation theory (weism02@aliyun.com).

A Novel Privacy-preserving User Authentication Protocol for Big Data Environment

Jiabing Liu, Xudong He, Huoye Tang, Dejun Wang, and Bo Meng
(Corresponding author: Bo Meng)

School of Computer Science, South-Central University for Nationalities
Wuhan 430074, China
Email: mengscuec@gmail.com

(Received Dec. 4, 2019; Revised and Accepted July 23, 2020; First Online Apr. 24, 2021)

Abstract

Many authentication protocols use private information as a factor to implement user authentication. The authentication server can obtain user private information. Therefore, there exists a risk of privacy leakage with the rapid development of data mining technology. Hence, it's important to protect privacy in the authentication protocol. In this paper, we first model the PPMUAS (privacy-preserving remote user authentication protocol) protocol using Applied PI calculus and analyze it with ProVerif(Protocol Verifier). We found that it has three vulnerabilities. And then, we propose the PPUAPBDE protocol (privacy-preserving user authentication protocol for big data environment), which uses signcryption, homomorphic encryption, and fuzzy hash to protect user privacy of multi-behavior features. After that, PPUAPBDE is modeled using Applied PI calculus and analyzed with ProVerif. The result shows that it achieves confidentiality, authentication, and privacy. Finally, we develop an authentication system based on PPUAPBDE to evaluate Recall(recall) and FPR (false positive rate). The Recall is about 94.8%, and the FPR is 5.1%, which is better than PPMUAS.

Keywords: Multi-behavior; Authentication; Privacy; Fuzzy Hash; Homomorphic Encryption

1 Introduction

Many authentication protocols used private information, such as biometrics, behavior characteristics and hardware information, as factors to implement identity authentication [2, 12, 13, 16, 18, 19, 21, 27]. But they do not consider the privacy of private information. When private information are sent to authentication server, the authentication server can obtain user privacy. Therefore, there exists a risk of private information leakage with big data mining tools and techniques [26]. Hence it's very important to protect the privacy of the private information in authentication protocol.

In general, there are two methods to protect privacy. On the one hand, authentication server applies privacy protection technology to protect privacy, such as data anonymity, data distortion or cryptography [6, 25]. On the other hand, authentication server adopts authentication protocol provide authentication and privacy. The second method is adopted mostly as it solves the problem fundamentally. Privacy is protected before being sent to authentication server. Therefore, privacy-preserving authentication protocol [7, 11, 14, 15, 17, 23, 24] has attracted attention. The main contributions of this paper are summarized as follows.

1) Apply Applied PI calculus in the symbolic model to formalize PPMUAS protocol and analyze its confidentiality, authentication and privacy with ProVerif. The result shows that it achieves confidentiality and privacy, but AS(Authentication Server) cannot authenticate DB(Database), and AS cannot authenticate User mutually;

2) Present PPUAPBDE protocol, which protects privacy of user private information, such as user keyboard, mouse usage habits, system processes and network behavior, with signcryption, homomorphic encryption and fuzzy hash. PPUAPBDE protocol provides the mutual authentication between User and AS, the authentication from DB to AS and the authentication from User to DB;

3) Apply Applied PI calculus in the symbolic model to formalize PPUAPBDE protocol, and then analyze its confidentiality, authentication, and privacy with ProVerif. The results indicate that it achieves confidentiality, authentication and privacy;

4) Develop an authentication system based on PPUAPBDE protocol to evaluate the Recall and FPR. The Recall is about 94.8%, and the FPR is 5.1%, which are better than PPMUAS protocol.

2 Related Work

Nowadays, lots of private information, such as biometric, behavior characteristic and hardware information, are used for authentication protocols [2, 13, 16, 18, 19, 21, 27], but the privacy of private information has not been implemented. Hence privacy-preserving authentication protocols that provides privacy are introduced [7, 11, 17, 23, 24].

Without privacy preservation: Based on biometrics, S. X. Fang *et al.* [27] expounded the advantages of biometrics identification technology compared with traditional identification technology. The characteristics of face recognition methods are analyzed and compared. A. Rassan and H. Alshaher [21] used fingerprint recognition as authentication factor and proposed an authentication method with user handwritten response code. The method consists of registration and login phase. In the registration phase, the user fingerprint is collected with mobile device and is stored in the database after the pre-processing. In the login phase, the fingerprint is collected again. After pre-processing, the corresponding entry in the database is matched to judge whether user is valid or not. R. H. Li and Y. Z. Li [16] proposed a dynamic continuous authentication system. After user login successfully, the character combination of Keystroke behavior and mouse behavior is used to monitor user operation behavior. It is used to judge whether operation object is genuine. In general, the security of the system is guaranteed. G. Kambourakis *et al.* [19] explored the potential of keystroke dynamics for touchscreen-equipped smartphones. A touchstroke system is implemented in the Android platform. And different methodologies, such as every pair of pressed keys and the average value of pressed keys, are executed under different scenarios to estimate the effectiveness in authenticating the end-user. The result shows that there still need advanced privacy protection. M. Babaeizadeh [2] used user keystroke duration as authentication factor and proposed a method for authenticating mobile users. When the user registers, the information of user keystroke duration is stored in the database after user registers successfully. And the session expiration time is set to force the user to log in before the end of the session. Unauthorized personnel cannot log into the system for session hijacking. S. Kang *et al.* [13] designed a practical method for biometric authentication based on electrocardiogram signals which is collected from mobile or wearable devices. The proposed approach can reduce the time required to achieve the target FAR (false acceptance rate) and FRR (false rejection rate). The proposed method are implemented in a wearable watch to verify its feasibility. In the experiment results, the FAR and FRR are 5.2% and 1.9%, respectively. H. Jeong and E. Choi [18] proposed a security authentication using profiling techniques for access control and user authentication, in which the profile consists of user information (name, ID, personal preference, hobby, etc) and service information (service name, provider name, context, frequency value, etc). However, details and requirements are not presented.

With privacy preservation: Based on homomorphic encryption, Hatin [7] proposed a privacy-preserving biometric authentication protocol. The protocol relies on the homomorphic Goldwasser-Micali cryptosystem [10]. And, they proved its security against malicious, but cannot resist insider attacks. Ren *et al.* [17] proposed a privacy-preserving authentication and access control scheme to insure the interaction security between mobile user and service in PCEs. The proposed scheme integrates underlying cryptographic primitives: Blind signature and hash chain, into highly flexible and lightweight authentication protocol. Based on the ring signature, Gamage *et al.* [11] proposed an identity-based ring signature scheme to create privacy-preserving authenticatable messages. Ruj *et al.* [23] proposed a privacy-preserving authenticated access control scheme to insure the security of private information in clouds. In the proposed scheme, the cloud verifies the authentication of the user without knowing the user identity before storing information. The cloud, however, does not know the identity of the user who stores information, but only verifies the user credential, which protect the user privacy effectively. Vorugunti [24] claimed that they proposed the first privacy-preserving remote user authentication protocol(PPMUAS), which uses the user-specific information and behavioral features. They claimed PPMUAS is the first authentication protocol to consider multi-factors and hybrid profile for privacy-preserving remote user authentication. Because of Proverif's support for cryptographic primitives [5], Proverif is used to analyze and validate security protocols described in Horn clause or Applied PI calculus. Based on Applied PI calculus [1], we use ProVerif [4] to analyze the PPMUAS. And we found that it has three security vulnerabilities.

3 PPMUAS Protocol

In this section, the messages in PPMUAS are given. And then, we model the PPMUAS protocol using Applied PI calculus and analyze the protocol with ProVerif. We found that it has protection mechanism for user privacy, but it has three authentication vulnerabilities.

3.1 Message of PPMUAS

In 2017, C. Vorugunti [24] proposed PPMUAS, which uses user password and keystroke dynamics to generate user profile. The messages in PPMUAS protocol are shown in Figure 1.

Registration request: User sends message 1 to AS to launch the registration. The message 1 is composed of User ID Id_i. User number i, User profile Cup_i and Rpw_i. The inputs to function Fuzzy-Hash Pw_i include U_OlInfo and U_BrInfo , where U_OlInfo is User online information, and U_BrInfo is User browser information. The inputs to function FH_Enc consists of U_KeySD, U_AMM and

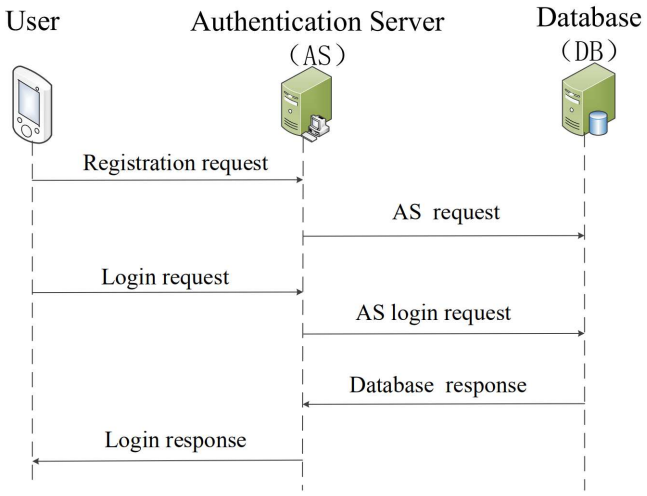


Figure 1: PPMUAS messages

U_GM, where *U_KeySD* is User keystroke dynamics information, *U_AMM* is accelerometer measurement information, *U_GM* is gyroscope measurement information. At last, message 1 is sent to AS through public channel c.

Authentication server request: After receiving message 1, Function Hash uses ID_i, RPW_i and i to generate hash value V_i. And AS stores it locally. At the same time, it creates message 2. The message contains α, $Rpup_i$ and i. $\alpha = E_{PUK}\ (r_i,\ k_n)$, which means α is obtained by encrypting the random number r_i which are generated by AS, and the User specific secret key k_n with the public key $Pu(k)$ of DB. $Rpup_i$ is received by the following method: Firstly, permute the user profile Cup_i with $Pup_i = \pi\ (Cup_i)$ to produce Pup_i, and then Pup_i is XORed with the random number r_i to get $Rpup_i$ with $Rpup\_i = Pup_i \bigoplus r_i$.After that, message 2 is generated and sent to DB through public channel c. After DB receives message 2, it decrypts α with the private key $Pr(k)$ to get r_i and k_n and obtains Pup_i by $Rpup_i$ and r_i through XOR. Finally, r_i, k_n, and Pup_i are saved in the database, and User registration is completed.

Login request: When User logs in, User enters the user id Id_i, the password Pw_i, and uses related information to generate Cup_i'. And then, message 3 is created in which it contains Id_i, Rpw_i, Cup_i' and i. Finally, the message 3 is sent to AS through public channel c.

AS login request: After AS receives message 3, V_i' is calculated firstly by the hash function and is compared with V_i saved in AS. If V_i and V_i' are equal, message 4 is generated. Message 4 contains α', $Rpup_i$' and i, where α' is created by the random number r_i' and the User session key k_{nn} by the public key $Pu(k_1)$ of DB. $Rpup_i$' is obtained by the following method: Firstly, permute the user profile Cup_i with Pup_i'$= \pi(Cup_i$') to produce Pup_i', and then Pup_i' is XORed with the random number r_i' to get $Rpup_i$' with $Rpup_i$' = Pup_i' *bigoplus* r_i'. Finally, message 4 is sent to DB through public channel c.

Database response: After receiving message 4, DB first decrypts with the private key to obtain r_i' and k_{nn}. Then $Rpup_i$' is XORed with r_i' to obtain Pup_i'. Finally, Pup_i' is compared with Pup_i stored in DB. If $HW((Pup_i' \bigoplus Pup_i)) \leq \triangle t$, it is legitimate user. Otherwise, AS denies access. HW is the Hamming weight. After that, message 5 is produced and sent to AS through public channel c, in which it contains "yes" or "no" only.

Login response: On receiving the message "yes" from DB, AS sends Content Service Ticket to User to access the content server resources, otherwise AS rejects and ends the connection.

3.2 Formal Modeling of PPMUAS

3.2.1 Function and Equational Theory

The functions and equations used in the modeling process are described in this section. We use Applied PI calculus to formalize PPMUAS protocol. The message x is encrypted by function *aenc(x, Pu)* with public key *Pu*, and message y is decrypted by function *adec(y, Pr)* with private key *Pr*. *XOR(x, y)* performs *XOR* calculation on message x and y. The Hamming weight algorithm *HW(x, y)* performs a Hamming weight measurement on the message x and y. The permutation function P permutes the message x. Figure 2 depicts the PPMUAS protocol function and equation theory.

$$\begin{bmatrix} fun\ FuzzyHash/1 \\ fun\ Hash/2 \\ fun\ FH\_Enc/1 \\ fun\ Hashone/3 \\ fun\ P/1 \\ fun\ XOR/2 \\ fun\ adec/2 \\ fun\ aenc/2 \\ fun\ Pu/1 \\ fun\ HW/2 \\ equation\ adec(aenc(x, Pu(y)), \Pr(y)) = x \\ equation\ HW(x, y) = (x, y) \\ equation\ XOR(XOR(x, y), y) = x \end{bmatrix}$$

Figure 2: Function and equation theory

3.2.2 Process

The whole PPMUAS protocol process mainprocess consists of three processes: *processUser*, *processAS* and *processDB*. They constitute the main process together, as shown in Figure 3.

$$mainprocess$$
$$(!\,processUser\,|!\,processAS\,|!\,processDB)$$

Figure 3: PPMUAS mainprocess

The all processes are shown in Figures 4, 5 and 6, respectively.

$$\begin{bmatrix} let\ processUser = \\ new\ U\_OlInfo \\ new\ U\_BrInfo \\ new\ U\_KeySD \\ new\ U\_AMM \\ new\ U\_GM \\ let\ Cup_i = \begin{bmatrix} (FuzzyHash(Pw_i),\ FuzzyHash(U\_OlInfo), \\ FuzzyHash(U\_BrInfo),\ FH\_Enc(U\_KeySD), \\ FH\_Enc(U\_AMM),\ FH\_Enc(U\_GM)) \end{bmatrix} in \\ let\ Rpw_i = Hash(Pw_i, b_i)\ in \\ let\ e = (ID_i, Rpw_i, Cup_i, i)\ in \\ out(c,e) \quad //registration\ phase \\ \\ new\ U\_OlInfo' \\ new\ U\_BrInfo' \\ new\ U\_KeySD' \\ new\ U\_AMM' \\ new\ U\_GM' \\ let\ Cup_i' = \begin{bmatrix} (FuzzyHash(Pw_i),\ FuzzyHash(U\_OlInfo'), \\ FuzzyHash(U\_BrInfo'),\ FH\_Enc(U\_KeySD'), \\ FH\_Enc(U\_AMM'),\ FH\_Enc(U\_GM')) \end{bmatrix} in \\ let\ Rpw_i = Hash(Pw_i, b_i)\ in \\ let\ e' = (ID_i, Rpw_i, Cup_i', i)\ in \\ out(c,e') \\ in(c,ticket) \quad //\log in\ phase \end{bmatrix}$$

Figure 4: PPMAUS processUser

$$\begin{bmatrix} let\ processAS = \\ in(c,m_1) \\ new\ r_i \\ new\ k_n \\ let\ (ID_i, Rpw_i, Cup_i, i) = m_1\ in \\ let\ V_i = Hashone(ID_i, Rpw_i, i)\ in \\ let\ Pup_i = P(Cup_i)\ in \\ let\ Rpup_i = XOR(Pup_i, r_i)\ in \\ let\ a = aenc((r_i, k_n), Pu(k_1))\ in \\ let\ DB = (a, Rpup_i, i)\ in \\ out(c,DB) \quad //registration\ phase \end{bmatrix} \begin{bmatrix} in(c,m_3) \\ new\ yes \\ let\ (ID_i, Rpw_i, Cup_i', i) = m_3\ in \\ let\ V_i' = Hashone(ID_i, Rpw_i, i)\ in \\ if\ V_i' = V_i\ then \\ new\ r_i' \\ new\ k_m \\ let\ Pup_i' = P(Cup_i')\ in \\ let\ Rpup_i' = XOR(Pup_i', r_i')\ in \\ let\ a' = aenc((r_i', k_m), Pu(k_n))\ in \\ let\ DB' = (a', Rpup_i', i)\ in \\ out(c,DB') \\ in(c,yes) \\ out(c,ticket) \quad //\log in\ phase \end{bmatrix}$$

Figure 5: PPMAUS processAS

$$\begin{bmatrix} let\ processDB = \\ in(c,m_2) \\ let\ (a, Rpup_i, i) = m_2\ in \\ let\ (r_i, k_n) = adec(a, Pu))\ in \\ let\ Pup_i = XOR(Rpup_i, r_i)\ in \\ //registration\ phase \end{bmatrix} \begin{bmatrix} in(c,m_4) \\ let\ (a', Rpup'_i, i) = m_4\ in \\ let\ (r_i, k_m) = adec(a', Pu)\ in \\ let\ Pup_i' = XOR(Rpup_i', r_i')\ in \\ if\ HW(Pup_i', Pup_i) = (Pup_i', Pup_i)\ then \\ new\ yes \\ out(c,yes) \quad //\log in\ phase \end{bmatrix}$$

Figure 6: PPMAUS processDB

3.3 Security Analysis of PPMUAS

In this section, the security analysis results of PPMUAS are given. The results indicate that PPMUAS has achieved confidentiality and privacy. But in the aspect of authentication, it has three security vulnerabilities that User cannot authenticate AS mutually and AS cannot authenticate DB.

3.3.1 Confidentiality

Here we use query attacker: Pw_i to model confidentiality of user password Pw_i. The analyzed result is shown in Figure 7. Figure 7 is the result of the confidentiality of Pw_i. The result is true, which proves that the user password Pw_i is secret. Because Pw_i is hashed by user to obtain a hash value. And then, the hash value is sent to AS. Owning to the one-way property of the hash function, and the attacker cannot obtain the password Pw_i through the hash value.

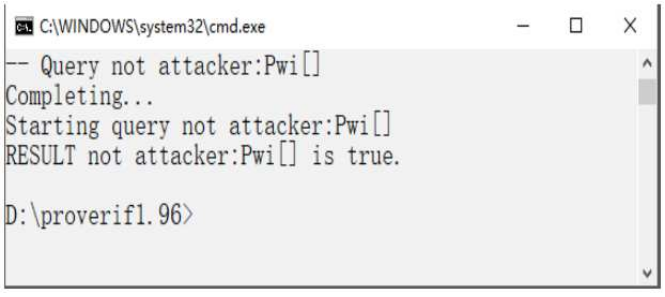


Figure 7: The confidentiality of Pw_i

3.3.2 Authentication

Here we use non-injective agreements to model authentication among User, AS and DB. The non-injective agreement is used to model authentication in ProVerif. PPMUAS protocol authentications are shown in Table 1.

Table 1: Authentication model

Non-injective agreement	Authentication
ev:endauthAs_User(x)⟶ ev:beginaauthAs_User(x)	AS authenticates User
ev:endauthUser_As(x)⟶ ev:beginaauthUser_As(x)	User authenticates AS
ev:endauthDB_As(x)⟶ ev:beginaauthDB_As(x)	DB authenticates AS
ev:endauthAs_DB(x)⟶ ev:beginaauthAs_DB(x)	AS authenticates DB

Figures 8 and 9 are mutual authentication results between AS and DB. The authentication result of DB to AS in Figure 8 is "true", it indicates that the authentication of DB to AS is achieved. AS sends Authentication Server login request message to DB, in which it contains α', $Rpup_i$' and i, where $Rpup_i$' is generated by AS. After receiving Authentication Server login request message, DB first decrypts with the private key to get r_i' and $Rpup_i$', and then creates Pup_i', which is compared to Pup_i saved in DB. If they are equal, it indicates that Authentication Server login request message is produced by the AS. Therefore, DB can authenticate AS. The authentication result of AS to DB in Figure 9 is "Can Not Be Proved", indicating that AS cannot authenticate DB. Because the DB response message is without any security mechanism. Thus the attacker can launch a impersonate attack. Therefore, AS cannot authenticate DB.

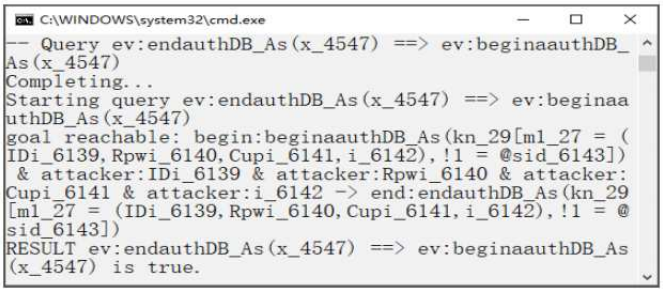


Figure 8: The authentication result of DB to AS

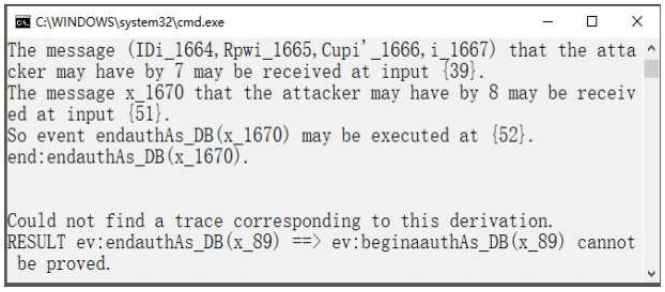


Figure 9: The authentication result of AS to DB

Figures 10 and 11 are mutual authentication results between AS and User. The results of both are "Can Not Be Proved", indicating that AS and User Cannot authenticate mutually. Because the attacker can pretend to be User to launch an impersonate attack. The Login response message sent by AS to User contains only one parameter ticket/reject, and the message does not have any security mechanisms. The attacker can initiate an impersonate attack, so User cannot authenticate AS.

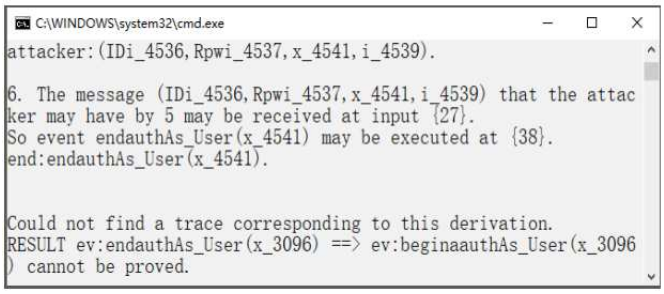


Figure 10: The authentication result of AS to User

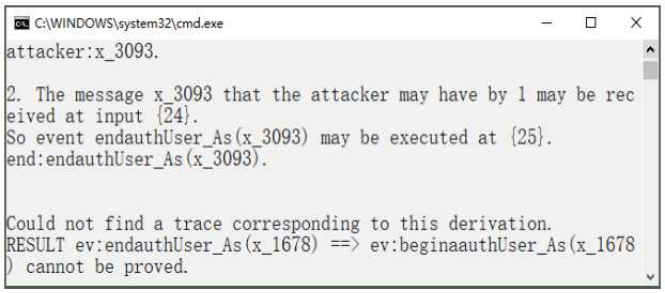


Figure 11: The authentication result of User to AS

3.3.3 Privacy

Here the privacy is modeled as the confidentiality of profile. We use query attacker: Cup_i and Cup_i' to model the privacy of Cup_i and Cup_i'. The results of user profile privacy analysis are shown in Figures 12 and 13. The results of Cup_i and Cup_i' are true, indicating that the privacy of PPMUAS has been implemented. Because Cup_i and Cup_i' are ciphertext containing private information, and they have confidentiality.

4 PPUAPBDE Protocol

The PPMUAS protocol claimed to be privacy protection while achieving remote user authentication. But with the analysis of section III, we find that it has not supported the one-way authentication from AS to DB and

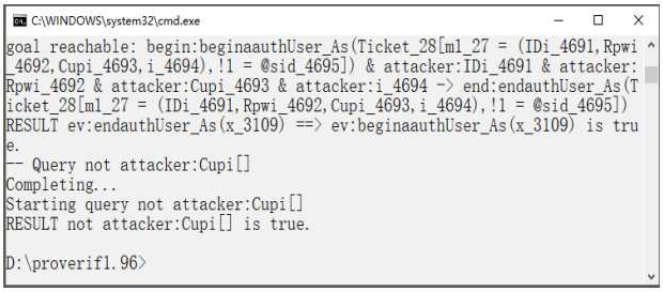


Figure 12: Privacy analysis of Cup_i

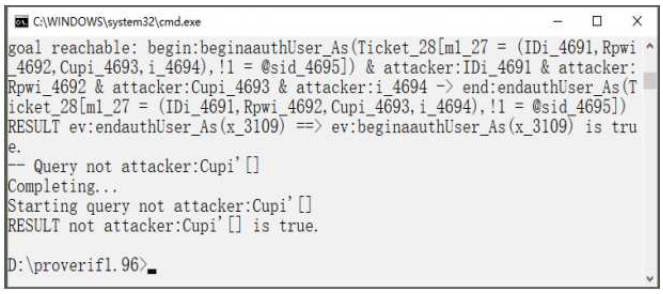


Figure 13: Privacy analysis of Cup_i'

the mutual authentication between User and AS. In order to achieve remote user authentication and privacy, the PPUAPBDE protocol is proposed based on user profile which contains the behavioral features of user mouse movement and keystroke, system processes, and network as user authentication factors. At the same time, it applies homomorphic encryption and fuzzy hash to protect the security and privacy of authentication factors. The fuzzy hash scheme can solve the avalanche effect of the normal hash algorithm. PPUAPBDE protocol provides the mutual authentication between User and AS, the authentication from DB to AS and the authentication from User to DB.

4.1 Message of PPUAPBDE

The PPUAPBDE protocol includes three roles: User, Authentication Server (AS) and Database (DB) and is composed of the registration phase and login phase. In the registration phase, User generates the encrypted user profiles using homomorphic signcryption and fuzzy hash with the behavioral patterns of user mouse movement and keystroke, system processes, and network behavior. And then the encrypted user profiles are sent to DB. In the login phase, AS verifies password and regenerates profiles, and then the regenerated profiles are sent to DB. DB compares the user profiles produced in the Login phase and in the registration phase and produces a result. If the result is higher than the preset threshold value, the User is allowed to log in, otherwise denied. The PPUAPBDE protocol mainly uses signcryption [3], fuzzy hash [22] and homomorphic encryption(HE) [8]. Notations in PPUAPBDE protocol are explained in Table 2.

The messages of PPUAPBDE protocol are shown in Figure 14. It has two phases, including eight messages. Registration phase includes Registration request, Authentication Server request, and Registration response messages. Login phase includes Login request, Authentication Server login request, Database response, User grant, and Login response messages.

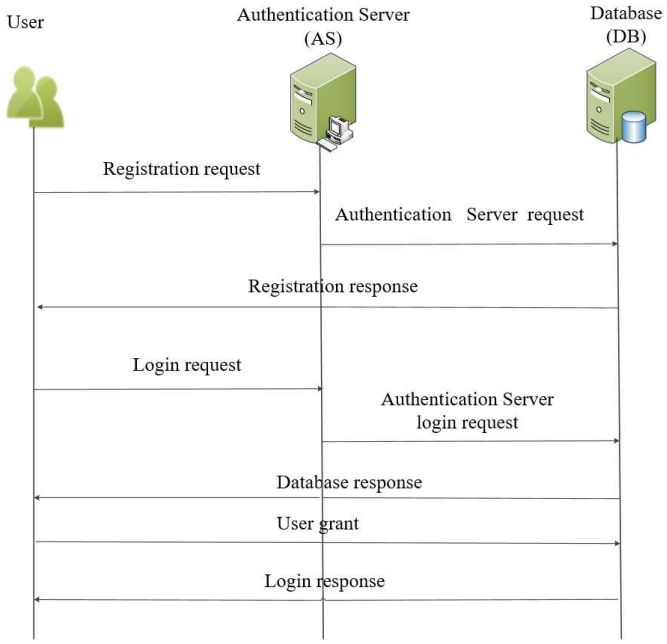


Figure 14: PPUAPBDE messages

Registration request: When User registers, it sends message 1 to AS:

——***Registration request***: $t_1, c_{reg}, R_{reg}, s_{reg}$ ——

Message 1 is composed of timestamp t_1, signcryption parameters c_{reg}, R_{reg} and s_{reg}. The calculation steps of signcryption parameters are shown in Table 3.

Authentication Server request: After AS receives message 1, it verifies whether the timestamp t_1 is in the time skew T or not. If t_1 exceeds T, the message 1 is a replay message, and is ignored, otherwise, $r_{reg} \cdot D$ will be compared with R_{reg}, which is shown in Table 4.

If $r_{reg} \cdot D$ and R_{reg} are equal, User password hash value h_{Pw} will be saved locally. And then, message 2 will be generated:

——**Authentication Server request**:t_2,c_{reg}',R_{reg}',s_{reg}'——

It includes timestamps t_2, signcryption parameters C_{reg}', R_{reg}' and s_{reg}'. C_{reg}', R_{reg}' and s_{reg}' are calculated , which is shown in Table 5.

Finally, message 2 is sent to DB.

Table 2: Notations and definitions

Symbol	Definition	Symbol	Definition
UP_i	Encrypted User Profile	Dis_{FHE}	Metric-data Characteristic deviation
C	Signcryption parameter	$\mathrm{Dis}_{FuzzyHash}$	String Characteristic deviation
R	Signcryption parameter	t_i	Timestamp
S	Signcryption parameter	T	Effective Time Period
PU_i,PR_i	Public/Private Key of HE	c	Public Channel
$FuzzyHash$	Fuzzy Hash Function	$\triangle t_1$	String Characteristic deviation
P_i,d_i	Public/Private Key of Signcryption	$\triangle t_2$	Metric-data Characteristic deviation

Table 3: Registration request calculation

Step1: $(PU_{UID}, PR_{UID} = THE_{keygen}$.
Step2: $d_{UID} \in (0,1,2,\cdots,p-1)$, $PU_{UID} = d_{UID} \cdot D$, $D \in n$, n is a large prime number.
Step3: $UP_{reg} = \{FuzzyHash(pw_i), FuzzyHash(U\_OlInfo), FuzzyHash(U\_BrInfo), FH\_Enc(U\_KeySD), FH\_Enc(U\_AMM), FH\_Enc(U\_GM)$
Step4: $k_1 = h(k \cdots D), k_2 = h(k \cdot P_{AS}), k \in (1,2,\cdots,n-1)$.
Step5: h_{Pw}=hash(Password).
Step6: c_{reg}=E((h_{Pw},UID,UP_{reg}),k_2), UID is user id.
Step7: r_{reg}=hash(k_1,h_{Pw}).
Step8: $s_{reg} = \frac{k}{r_{reg}+d_{UID}} \bmod n$.
Step9: $R_{reg} = r_{reg}D$.

Table 4: AS request comparison

step1: choose $\mathrm{d}_{AS} \in$ (0,1,2...p-1), P_{AS}=d_{AS} D.
step2: k_1=hash(s_{reg}(P_{UID}+R_{reg})).
step3: k_2=hash(s_{reg}(d_{AS}(P_{UID}+R_{reg}))).
step4: (h_{PW},UID,UP_{reg})=D(c_{reg},k_2).
step5: r_{reg}=hash(k_1,$\mathrm{h}_p w$).
step6: r_{reg} D?=R_{reg}.

Table 5: AS request calculation

step1: choose k, k $\in$ (0,1,2...n-1).
step2: $k_3 = h(kD), k_4 = h(kP_{DB})$.
step3: c_{reg}'=E((UID,UP_{reg}),k_4).
step4: r_{reg}'=hash(k_3,UID).
step5: $s'_{reg} = (\frac{k}{r'_{reg}+d_{AS}}) \bmod n$.
step6: $R'_{reg} = r'_{reg}D$.

Table 6: Registration responce calculation

step1: $d_{DB} \in (0,1,\cdots,n-1), P_{DB} = d_{DB}D$.
step2: k_3=hash(s_{reg}'(P_{AS}+R_{reg}')).
step3: k_4=hash(s_{reg}'(d_{DB}(P_{AS}+R_{reg}'))).
step4: (UID,UP_{reg})=D(s_{reg}',k_4).
step5: r_{reg}'=hash(k_3,UID).
step6: $r'_{reg}D = R'_{reg}$.

Registration response: After DB receives message 2, and then it verifies whether the timestamp t_2 is in the time skew T or not. If t_2 exceeds T, message 2 is a replay message, and is ignored, otherwise, r_{reg}' $\cdot D$ and R_{reg}' are calculated, which is shown in Table 6.

If r_{reg}' $\cdot D$ and R_{reg}' are equal, the User profile UP_{reg} will be saved locally. And then, message 3 will be generated:

——Registration responce: *Yes*——

Where the message "Yes" means that the User registers successfully. And then, message 3 is sent to AS.

Login request: When User enters id and password to log in, UP_{login}, c_{login}, R_{login}, s_{login} will be generated just like the calculation in the registration phase. And then, message 4 will be created:

——Login request: t_3, c_{login}, R_{login}, s_{login}——

where it contains a timestamp t_3, signcryption parameter c_{login}, R_{login}, s_{login}. Finally, message 4 is sent to AS.

Authentication Server login request: After AS receives message 4, and then it verifies whether the timestamp t_3 is in the time skew T or not. If t_3 exceeds T, message 4 is a replay message and is ignored. Otherwise, $r_{login} \cdot D$ is compared with R_{login}. If they are equal, we will continue to compare the User password hash value h_{Pw}' saved in login phase and the

User password hash value h_{Pw} saved locally. If h_{Pw}' and h_{Pw} are equal, message 5 is generated, too.

——AS Login request: t_4,c_{login}', R_{login}', s_{login}'——

Where it contains timestamp t_4, Signcryption parameters c_{login}', R'_{login}, s'_{login}. Signcryption parameters c'_{login}, R'_{login}, s'_{login} are produced by the same method in Registration phrase. Finally, message 5 is sent to DB. If h_{Pw}' and h_{Pw} are not equal, User will be requested to login again.

Database response: After DB receives message 5, and then it verifies whether the timestamp t_4 is in the time skew T or not. If t_4 exceeds T, Authentication Server login request message is ignored, otherwise, r_{login}'D is compared with R_{login}'. If they are equal, the following computation will be executed.

$$
\begin{aligned}
&Dis_{FuzzyHash} = \\
&\{Fuzzycmp(FuzzyHash(U\_SFP), \\
&FuzzyHash(U'\_SFP)), \\
&Fuzzycmp(FuzzyHash(U\_SP), \\
&FuzzyHash(U'\_SP)), \\
&Fuzzycmp(FuzzyHash(U\_NS), \\
&FuzzyHash(U'\_NS)).\}
\end{aligned}
$$

TThus, we get the String Characteristic Deviation $Dis_{FuzzyHash}$ of the login phase profile UP_{reg} and the registration phase profile UP_{login}. If $Dis_{FuzzyHash}$ is in the range of deviation Δt_1, we continue to calculate the Metric-data Characteristic deviation Dis_{FHE} of the User profile.

$$
\begin{aligned}
&Dis_{FHE} = \\
&\{FHE_{sub}(FHE(U\_AFI, PU_{UID}), \\
&FHE(U'\_AFI, PU_{UID}), PU_{UID}), \\
&FHE_{sub}(FHE(U\_MI, PU_{UID}), \\
&FHE(U'\_MI, PU_{UID}), PU_{UID}), \\
&FHE_{sub}(FHE(U\_KI, PU_{UID}), \\
&FHE(U'\_KI, PU_{UID}), PU_{UID}).\}
\end{aligned}
$$

Finally, DB will sign the Dis_{FHE} with the private key of DB. Thus, the result is sent to User:

——Database responce: Sign(Dis_{FHE})——

User grant: When User receives the response message 6, the signature is verified first. And then, Dis_{FHE} is decrypted by the private key of Homomorphic Encryption to get the plaintext of User profile Metric-data Characteristic deviation Dis_{FHE-pt}. After that the deviation is sent to DB after signing with the private key of User:

——User grant: Sign(Dis_{FHE-pt})——

Finally, the message 7 is sent to DB.

Login response: When DB receives message 7, it verifies the signature to get Dis_{FHE-pt}. Then, it compares Dis_{FHE}' with Dis_{FHE-pt}, PU_{UID}. If they are equal, we continue to compare whether Dis_{FHE-pt} is in the range of deviation Δt_2. If this condition is met, it means that the user is legitimate. And message 8 will be generated. Otherwise, the User is rejected.

——Login responce:Ticket)——

4.2 Formal Modeling of PPUAPBDE

In this section, the Applied PI calculus is used to describe the PPUAPBDE protocol formally, and the non-injective and Query are used to model the authentication and confidentiality. Then the software tool ProVerif is used to formalize and prove the confidentiality and authentication of the PPUAPBDE. Finally, the confidentiality of user privacy information is analyzed, and the privacy analysis results of the PPUAPBDE protocol is given.

4.2.1 Function and Equational Theory

Function and equation theory mainly contains public-key encryption algorithm *E(x, Pu)* and decryption algorithm *D(y, Pu)*. The public key encryption algorithm *E(x, Pu)* encrypts the message x using the public key Pu to generate the ciphertext. The decryption algorithm *D(y, Pu)* decrypts the message y using the public key Pu to obtain the plaintext. The sign algorithm Sign*(x,Pr(y))* signs the message x with private key *Pr(y)*, while the signature is verified with public key *Pu(y)*. Its formal modeling is shown in Figure 15.

```
fun FuzzyHash/1.
fun hash/1.
fun FHE/2.
fun E/2.
fun D/2.
fun mod/2.
fun FHE_keygen/0.
fun Fuzzycmp/2.
fun h/1.
fun Pu/1.
fun Pr/1.
fun sign/2.
fun versign/2.
fun FHE_Decrypt/2.
fun FHE_sub/2.
equation D(E(x,Pu(y)),Pu(y)) = x.
equation versign(sign(x,Pr(y)),Pu(y)) = x.
equation FHE_Decrypt(FHE_sub((FHE(x,PU(y)),
FHE(y,PU(y))),Pu(y)),Pr(y)) = x - y.
```

Figure 15: Function and equation theory

4.2.2 Process

The PPUAPBDE protocol is composed of *processUser*, *processAS* and *processDB*. The three processes constitute the main process, as shown in Figure 16.

$$\begin{bmatrix} Mainprocess \\ (!\,processUser|!\,processAS|!\,processDB) \end{bmatrix}$$

Figure 16: PPUAPBDE mainprocess

We model the three processes using Applied PI calculus. The *processUser* all processes are shown in Figures 17, 18 and 19, respectively.

4.3 Security Analysis of PPUAPBDE

In this section, the security analysis results of PPUAPBDE are given. The results indicate that PPUAPBDE has implemented confidentiality, authentication, and privacy.

4.3.1 Confidentiality

In this section, the confidentiality of user password Psw_i is analyzed. And we use query attacker: Password in ProVerif to verify the confidentiality of the user password Psw_i. The result of query attacker: Password is shown in Figure 20. The result is "true" to prove that the confidentiality of Password has been satisfied. This is because the User uses the hash function to process the Password to get the hash value, and the hash value is sent to the AS. Owning to the one-way property of the hash function, the attacker cannot obtain the User password plaintext.

4.3.2 Authentication

The non-injective agreement is used to model authentication in ProVerif. The authentication model of PPUAPBDE is the same as shown in Table 1.

The authentication results between User and AS are shown in Figures 21 and 22. The results of both are true, indicating that the mutual authentications between AS and User have been achieved. This is because the Login request message and Registration request message sent by User to AS, are signed with their private keys. Only the signcryption public key can verify the signature. Therefore, User can authenticate AS mutually.

In Figure 23, User authenticates DB. In Figure 24, DB authenticates AS. The results of both are true, indicating that User can authenticate DB and DB can authenticate AS. The Registration response message is handled by the signcryption function. And the Authentication server Login request message and Authentication server request message sent by AS to DB, are signed with their private keys, too.

$$\begin{bmatrix}
new\ t_1; \\
new\ UID; \\
new\ Password; \\
new\ U_{_SFP}; \\
new\ U_{_SP}; \\
new\ U_{_NS}; \\
new\ U_{_AFI}; \\
new\ U_{_MT}; \\
new\ U_{_KI}; \\
let\ UP_{reg} = \begin{bmatrix} FuzzyHash(U_{_SFP}), FuzzyHash(U_{_SP}), FuzzyHash(U_{_NS}) \\ FHE(U_{_AFI}, PU_{UID}), FHE(U_{_MT}, PU_{UID}), FHE(U_{_KI}, PU_{UID}) \end{bmatrix} in \\
let(PU_{UID}, PR_{UID}) = FHE_keygen\ in \\
in(c, P_{AS}); \\
\cdots \\
let\ c_{reg} = E((h_{pw}, UID, UP_{reg}), k_1)\ \ in \\
\cdots \\
let\ s_{reg} = (\frac{k}{r_{reg} + d_{UID}}) \bmod\ n\ in \\
\cdots \\
out(c, m_1); //\ registration\ phase \\
\\
new\ t_3; \\
new\ Password'; \\
new\ U_{_SFP}'; \\
new\ U_{_SP}'; \\
new\ U_{_NS}'; \\
new\ U_{_AFI}'; \\
new\ U_{_MT}'; \\
let\ UP_{login} = \begin{bmatrix} FuzzyHash(U_{_SFP}'), FuzzyHash(U_{_SP}'), FuzzyHash(U_{_NS}') \\ FHE(U_{_AFI}', PU_{UID}), FHE(U_{_MT}', PU_{UID}), FHE(U_{_KI}', PU_{UID}) \end{bmatrix} in \\
\cdots \\
out(c, m_3); \\
in(c, Ticket). //\ \log in\ phase
\end{bmatrix}$$

Figure 17: PPUAPBDE processUser

4.3.3 Privacy

Here the privacy is modeled as the confidentiality of profile. We use query attacker: UP_{reg} and UP_{login} to model the privacy of UP_{reg} and UP_{login}.The results of user profile privacy analysis are shown in Figures 25 and 26. The results of UP_{reg} and UP_{login} are true, indicating that it has User profile privacy of UP_{reg} and UP_{login}. This is because the homomorphic signcryption adopted by the PPUAPBDE is to provide privacy protection of user profile UP_{reg} and UP_{login}.

5 Evaluation and Analysis

In order to evaluate the performance of PPUAPBDE protocol, we develop an authentication system with Visual Studio Community 2015 and MySQL Community 6.3. The signcryption scheme is the open-source software library [9]. The fuzzy hash is the sdhash proposed by Roussev [22]. Homomorphic encryption is Microsoft's Open Encrypted Arithmetic Library project [20], which

$$
\begin{aligned}
&in(c, m_1);\\
&in(c, P_{UID});\\
&in(c, P_{DB});\\
&let(t_1, c_{reg}, R_{reg}, s_{reg}) = m_1 \ in\\
&let\ k_1 = hash(s_{reg}(P_{UID} + R_{reg}))\ in\\
&let\ k_2 = hash(s_{reg}(d_{AS}(P_{UID} + R_{reg})))\ in\\
&let\ (h_{pw}, UID, UP_{reg}) = D(c_{reg}, k_2)\ in\\
&if\ R_{reg} = hash(k_1, h_{pw}) \times D\ then\\
&new\ t_2\\
&let\ k_3 = h(k \times D)\ in\\
&let\ k_4 = h(k \times P_{DB})\ in\\
&let\ c_{reg}' = E((UID, UP_{reg}), k_4)\ in\\
&let\ r_{reg}' = hash(k_3, UID)\ in\\
&let\ s_{reg}' = (\frac{k}{r_{reg}' + d_{AS}}) \bmod n\ in\\
&let\ R_{reg}' = r_{reg}' \times D\\
&let\ m_2 = (t_2, c_{reg}', R_{reg}', s_{reg}')\ in\\
&out(c, m_2);//registration\ phase
\end{aligned}
\qquad
\begin{aligned}
&in(c, m_3);\\
&let\ (t_1, c_{\log in}, R_{\log in}, s_{\log in}) = m_3\ in\\
&\cdots\\
&if\ R_{\log in} = r_{\log in} \times D\ then\\
&let\ c_{\log in}' = E(UID, k_4)\ in\\
&let\ r_{\log in}' = hash(k_3, UID)\ in\\
&let\ s_{\log in}' = (\frac{k}{r_{\log in}' + d_{AS}}) \bmod n\ in\\
&let\ R_{\log in}' = r_{\log in}' \times D\\
&let\ m_4 = (t_4, c_{\log in}', R_{\log in}', s_{\log in}')\ in\\
&\cdots\\
&out(c, m_4).//\log in\ phase
\end{aligned}
$$

Figure 18: PPUAPBDE processAS

$$
\begin{aligned}
&new\ t;\\
&in(c, m_2);\\
&in(c, P_{AS});\\
&let\ (t_2, c_{reg}', R_{reg}', s_{reg}') = m_2\ in\\
&let\ k_3 = hash(s_{reg}'(P_{AS} + R_{reg}'))\ in\\
&let(UID, UP_{reg}) = D(c_{reg}', k_4)\ in\\
&\cdots\\
&if\ R_{reg}' = hash(k_3, UID) \times D\ then\\
&out(c, yes);//registration\ phase\\
&\\
&in(c, m_4);\\
&let\ (t_4, c_{\log in}', R_{\log in}', s_{\log in}') = m_4\ in\\
&\cdots\\
&if\ R_{\log in}' = c_{\log in}' \times D\ then\\
&let\ Dis_{FuzzyHash} = \begin{bmatrix} Fuzzycmp(FuzzyHash(U_{\_SFP}), FuzzyHash(U_{\_SFP}')) \\ Fuzzycmp(FuzzyHash(U_{\_SP}), FuzzyHash(U_{\_SP}')) \\ Fuzzycmp(FuzzyHash(U_{\_NS}), FuzzyHash(U_{\_NS}')) \end{bmatrix} in\\
&let\ Dis_{FHE} = \begin{bmatrix} FHE_{stib}(FHE(U_{\_AFI}, PU_{UID}), FHE(U_{\_AFI}', PU_{UID}'), PU_{UID}) \\ FHE_{stib}(FHE(U_{\_MT}, PU_{UID}), FHE(U_{\_MT}', PU_{UID}), PU_{UID}) \\ FHE_{stib}(FHE(U_{\_KI}, PU_{UID}), FHE(U_{\_KI}', PU_{UID}), PU_{UID}) \end{bmatrix} in\\
&\cdots\\
&newTicket;\\
&out(c, Ticket);//\log in\ phase
\end{aligned}
$$

Figure 19: PPUAPBDE processDB

Figure 20: The confidentiality of Psw_i

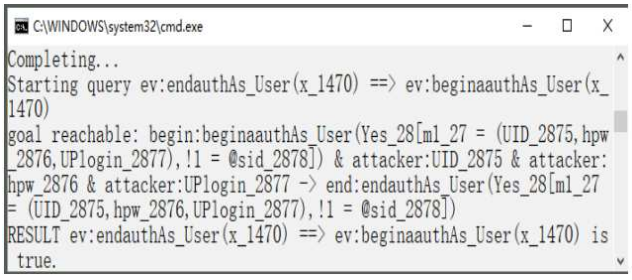


Figure 21: AS authenticates User

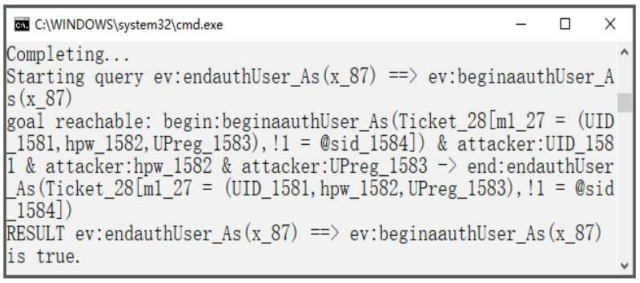


Figure 22: User authenticates AS

is an easy-to-use but powerful homomorphic encryption library that is called via API. We choose two closely related important parameters: Recall and FPR to evaluate the PPUAPBDE authentication system. The Recall R=TP/(TP+FN) represents the probability that a legitimate user will successfully log in to the system, in which TP is the number of times a legitimate user attempts to log in, and FN represents the number of times an illegal user attempts to log in. FPR that is, the probability that the system will misjudge the attacker as a legitimate user. We select 663 students in a university as a user to evaluate the Recall and FPR in 31 days. The related evaluation data is shown in Table 7, and the data in the table is taken from the database log. The experimental results are shown in Figure 27. The Recall is 94.8%. The last FPR is 5.1%. The cloud-based privacy protection mobile user authentication system PPMUAS based on big data characteristics described in [24] has a Recall of 84.9% and an FPR of 12.6%. The Recall value continues to increase, and the FPR value continues to decrease, indicating that the PPUAPBDE can be used in the authentication protocol effectively. If we want to put the experimental authentication system into practice, there is a work on using

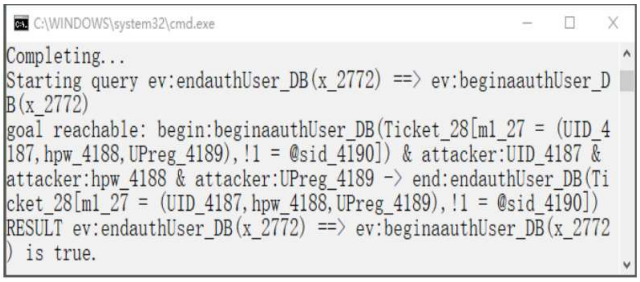


Figure 23: User authenticates DB

Figure 25: Privacy analysis of UP_{reg}

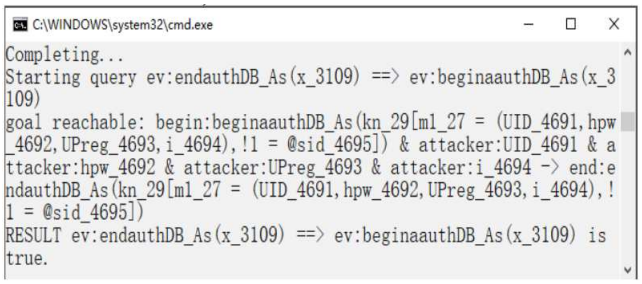


Figure 24: DB authenticates AS

Figure 26: Privacy analysis of UP_{login}

better algorithms for collecting dynamic behavioral features to improve the FPR.

6 Conclusion

Many authentication protocols used private information as authentication factors to implement user authentication. The authentication server can get private data. Therefore, with the rapid development of data mining technology, there exists a risk of private information leakage in the authentication server. In general, there are two methods to protect privacy. One is that the authentication server uses privacy protection technology, such as data anonymity, data distortion, or cryptography, to protect privacy. This method depends on authentication server cooperation. The other is to develop the privacy-preserving authentication protocol that provides not only user authentication but also privacy preservation of private data. The second method is adopted mostly as it solves the problem fundamentally. The privacy information is protected before being sent to the authentication server. Hence, we analyzed the authentication and confidentiality of PPMUAS protocol using Applied PI calculus and found it has some security vulnerabilities. In order to implement remote authentication and privacy of user information, the PPUAPBDE protocol is proposed based on user profile which contains behavioral patterns of user mouse movement and keystroke, system processes, and network as user authentication factors. At the same time, it applies homomorphic encryption and fuzzy hash to protect the security and privacy of authentication factors. And then, the confidentiality, authentication and privacy of the PPUAPBDE protocol are analyzed using Applied PI calculus. The results show that the confidentiality, authentication, and privacy of PPUAPBDE have been implemented. Finally, based on the PPUAPBDE protocol, we design, implement and test the PPUAPBDE authentication system. The Recall is about 94.8%, and FPR is 5.1%, which are better than PPMUAS protocol. The scheme has not been put into practical use yet, only for testing the actual effect of multi-behavior features to protect user privacy in the authentication protocol. PPMUAS was proposed by Vorupunti *et al.* in 2017, and they claimed that it is the first privacy-preserving remote user authentication. Hence, compared to PPMUAS protocol from the key three aspects:

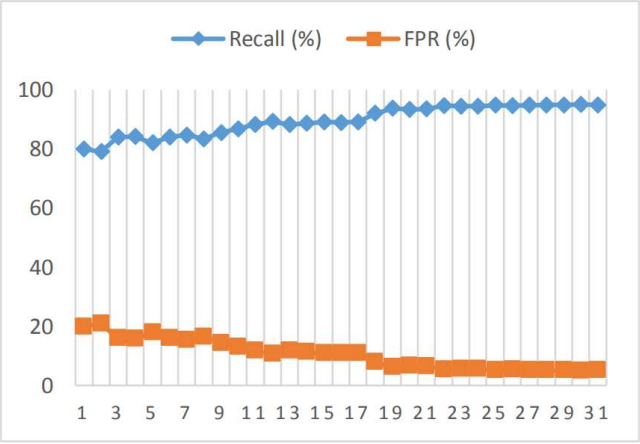


Figure 27: Experimental results

- Authentication factor. PPUAPBDE uses multi-factor mechanisms: User mouse movement and keystroke, system processes, and network as authentication factors, while PPMUAS only use a single factor;;

Table 7: Related evaluation data

Day	TP	FN	Recall(%)	FPR(%)
1	531	132	80.0	19.9
2	524	139	79.0	20.9
3	557	106	84.0	15.9
4	558	105	84.1	15.8
5	544	119	82.0	17.9
6	557	106	84.0	15.9
7	561	102	84.6	15.3
8	553	110	83.4	16.5
9	567	96	85.5	14.4
10	576	87	86.8	13.1
11	585	78	88.2	11.7
12	592	71	89.2	10.7
13	585	78	88.2	11.7
14	588	75	88.6	11.3
15	591	72	89.1	10.8
16	590	73	88.9	11.0
17	591	72	89.1	10.8
18	610	53	92.0	7.9
19	621	42	93.7	6.2
20	619	44	93.3	6.6
21	619	44	93.4	6.5
22	627	36	94.6	5.3
23	626	37	94.4	5.5
24	626	37	94.4	5.5
25	628	35	94.7	5.2
26	627	36	94.6	5.3
27	628	35	94.7	5.2
28	629	34	94.8	5.1
29	629	34	94.8	5.1
30	630	33	95.0	4.9
31	629	34	94.8	5.1

- Authentication. PPUAPBDE protocol provides the mutual authentication between User and AS, the authentication from DB to AS and the authentication from User to DB, while PPMUAS does not provide authentication from AS to DB and the mutual authentication between User and AS;

- Recall and FPR. Recall and the FPR of the PPUAPBDE authentication system are tested, and the results show that PPUAPBDE is better than PPMUAS.

There are open issues in the PPUAPBDE protocol. For example, the Recall of the PPUAPBDE authentication system is not high enough. In the future, we can establish a contour model for the normal behavior of legitimate users, and then use the similarity curve of the behavior profile as the credential to authenticate the legitimate user to improve the Recall.

Acknowledgments

The research is supported in part by the Fundamental Research Funds for the Central Universities, South-Central University for Nationalities No. CZZ21001 and QSZ17007, and in part by the natural science foundation of Hubei Province under the grants No.2018ADC150.

References

[1] M. Abadi and C. Fournet, "Mobile values, new names, and secure communication," in *Proceedings of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, vol. 36, no. 3, pp. 104–115, 2001.

[2] M. Babaeizadeh, "Keystroke dynamic authentication in mobile cloud computing," *International Journal of Computer Applications*, vol. 90, no. 1, pp. 29–36, 2014.

[3] S. M. Bellovin, P. Gutmann and M. Blaze, "An IBE-based signcryption scheme for group key management," *arXiv: Cryptography and Security*, 2016. arXiv:1603.09526.

[4] B. Blanchet, "An efficient cryptographic protocol verifier based on prolog rules," in *Proceedings of 14th IEEE Computer Security Foundations Workshop*, pages 82–96, 2001.

[5] Z. F. Cao, "New development of cryptography (in Chinese)," *Advanced Engineering Sciences*, vol. 47, no. 1, pp. 1–12, 2015.

[6] Y. C. Chen, W. L. Wang, M. S. Hwang, "RFID authentication protocol for anti-counterfeiting and privacy protection", in *The 9th International Conference on Advanced Communication Technology*, pp. 255–259, 2007.

[7] E. Cherrier, J. Hatin and J. Schwartzmann, "Privacy preserving transparent mobile authentication," in *The 3rd International Conference on Information Systems Security and Privacy*, pp. 354–361, 2017.

[8] C. Gentry, M. Dijk and S. Halevi, "Fully homomorphic encryption over the integers," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 24–43, 2010.

[9] A. Ghani, *et al.*, S. Ashraf, A. Irshad, "An efficient signcryption scheme with forwarding confidentiality and public verifiability based on hyper elliptic curve cryptography," *Multimedia Tools and Applications*, vol. 74, no. 5, pp. 171–173, 2015.

[10] S. Goldwasser and S. Micali, "Probabilistic encryption & how to play mental poker keeping secret all partial information," in *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, pp. 365–377, 1982.

[11] B. Gras, C. Gamage and B. Crispo, "An identity-based ring signature scheme with enhanced privacy," *Second International Conference on Security and Privacy in Communication Networks and the Workshops*, pp. 1–5, 2006.

[12] C. C. Lee, C. H. Liu, M. S. Hwang, "Guessing attacks on strong-password authentication protocol", *International Journal of Network Security*, vol. 15, no. 1, pp. 64–67, 2013.

[13] Y. Lee, S. Kang and I. Cho, "ECG authentication system design based on signal analysis in mobile and wearable devices," *IEEE Signal Processing Letters*, vol. 23, no. 6, pp. 805–808, 2016.

[14] C. T. Li, M. S. Hwang, Y. P. Chu, "Further improvement on a novel privacy preserving authentication and access control scheme for pervasive computing environments", *Computer Communications*, vol. 31, no. 18, pp. 4255–4258, Dec. 2008.

[15] C. T. Li, M. S. Hwang, Y. P. Chu, "A secure and efficient communication scheme with authenticated key establishment and privacy preserving for vehicular ad hoc networks", *Computer Communications*, vol. 31, no. 12, pp. 2803-2814, July 2008.

[16] R. H. Li and Y. Z. Li, "Research on behavior characteristics in dynamic continuous identity authentication system," (in Chinese), *Computer and Digital Engineering*, vol. 46, no. 1, pp. 138–143, 2018.

[17] W. Lou, K. Ren and K. Kim, "A novel privacy preserving authentication and access control scheme for pervasive computing environments," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1373–1384, 2016.

[18] H. Jeong and E. Choi, "User authentication using profiling in mobile cloud computing," *AASRI Procedia*, vol. 2, pp. 262–267, 2012.

[19] G. Kambourakis and D. Damopoulos, "Introducing touchstroke: keystroke-based authentication system for smartphones," *Security & Communication Networks*, vol. 9, no. 6, pp. 542–554, 2016.

[20] Microsoft/SEAL. (`https://github.com/microsoft/seal`)

[21] A. Rassan and H. Alshaher, "Securing mobile cloud computing using biometric authentication (SMCBA)," in *International Conference on Computational Science and Computational Intelligence*, vol. 1, pp. 157–161, 2014.

[22] V. Roussev, "An evaluation of forensic similarity hashes," *Digital Investigation*, vol. 8, pp. S34-S41, 2011.

[23] S. Ruj and M. Stojmenovic, "Privacy preserving access control with authentication for securing data in clouds," in *The 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pp. 556–563, 2012.

[24] C. Vorugunti, "PPMUAS: A privacy preserving mobile user authentication system for cloud environment utilizing big data features," in *IEEE International Conference on Advanced Networks and Telecommunications Systems*, pp. 1–6, 2016.

[25] C. H. Wei, M. S. Hwang, Augustin Y. H. Chin, "A secure privacy and authentication protocol for passive RFID tags," *International Journal of Mobile Communications*, vol. 15, no. 3, pp. 266–277, 2017.

[26] L. Yao, J. Lu and X. He, "A security analysis method for security protocol implementations based on message construction," *Applied Sciences*, vol. 8, no. 12, pp. 2543, 2018.

[27] L. W. Zhang, S. X. Fang and H. L. Wang, "Research on the development trend of identity authentication technology based on face biometrics," *Information Security Research*, vol. 003, no. 006, pp. 533–537, 2017.

Biography

Jiabing Liu. was born in 1994 and is now a postgraduate at the School of Computer Science, South-Central University for Nationalities. His research interests include Blockchain and Smart contract security.

Xudong He. was born in 1991 and is now a postgraduate at school of Computer Science, South-Central University for Nationalities. His research interests include: Security protocol implementations and reverse engineering.

Huoye Tang. was born in 1991 and is now a postgraduate at school of Computer Science, South-Central University for Nationalities. His research interests include security protocol implementations and reverse engineering.

Dejun Wang. was born in 1974 and received his Ph.D. in information security at Wuhan University in China. Currently, he is an associate professor in the school of computer, South-Center University for Nationalities, China. He has authored/coauthored over 20 papers in international/national journals and conferences. His current research interests include security protocols and formal methods.

Bo Meng. was born in 1974 in China. He received his M.S. degree in computer science and technology in 2000 and his Ph.D. degree in traffic information engineering and control from Wuhan University of Technology at Wuhan, China in 2003. From 2004 to 2006, he worked at Wuhan University as a postdoctoral researcher in information security. Currently, he is a full Professor at the school of computer, South-Center University for Nationalities, China. He has authored/coauthored over 50 papers in International/National journals and conferences. In addition, he has also published a book "secure remote voting protocol" in the science press in China. His current research interests include Cyberspace security.

On Security of Privacy-Preserving Remote User Authentication with k-Times Untraceability

Qijia Zhang[1], Jianhong Zhang[1,2], Linhan Liu[1], Jing Wang[3], and Pei Liu[3]
(Corresponding author: Jianhong Zhang)
School of Information Sciences and Technology, North China University of Technology[1]
Beijing 100144, China
GuiZhou University,Guizhou Provincial Key Laboratory of Public Big Data[2]
Guizhou Guiyang 550025, China
Beijing Jingdong Century Information Technology Company, Limited[3]
Beijing 100100, China
Email: zjhncut@163.com
(Received Dec. 31, 2019; Revised and Accepted July 23, 2020; First Online Apr. 17, 2021)

Abstract

As an important access control technique, k-times anonymous authentication (k-TAA) plays a vital role in e-coupon and e-bill. It allows a user to anonymously authenticate himself to a remote server a bounded number of times. However, most of the existing k-TAA schemes require heavy computation, which brings a challenge to resource-limited devices. In 2018, Tian *et al.* proposed a privacy-preserving remote user authentication with k-times untraceability. Unlike the traditional k-TAA schemes, Tian *et al.*'s is more suitable for mobile devices due to avoiding expensive pairing operations. And they claim that their scheme provides user authenticity and k-times untraceability. Unfortunately, in this paper, we find that their scheme is insecure by analyzing it. Their scheme can neither prevent a malicious user from passing the authentication nor trace the identity of a dishonest user authenticating for more than k times. Finally, the corresponding attacks are given.

Keywords: Anonymity; Attack; Authentication; User Privacy

1 Introduction

The development of internet makes it possible that people enjoy the service of remote providers. In the past few decades, various online applications and services emerged. In the most typical online service, users need to interact with a server. The first step of interaction is user authentication. Before opting for services, a user has to authenticate himself to an authentication server, and the server saves this user's identity into its database. However, in some cases, identity is a part of the users' privacy information that should be protected. In this situation, anonymity is one of the fundamental security properties that a secure system should provide. In order to preserve the privacy of users, anonymous remote user authentication is proposed and applied in a lot of fields [6, 7, 11, 14]. It allows users to anonymously authenticate themselves to a remote authentication server. To improve anonymous remote user authentication, researchers have exploited many cryptographic techniques in it, such as group signature [1], blind signature [8, 10], and ring signature [19].

However, sometimes the anonymity of users may bring harm to the system. For example, in the e-coupon system [5, 9, 20], for any customer who purchasing goods in a shop, the merchant can grant an electronic coupon as a bonus to the customer. To benefit from the service, a dishonest user may redeem this electronic coupon for more than the predetermined number of times. In this case, the service provider should be able to trace this dishonest user and obtain his identity. However, most of existing anonymous remote user authentication schemes neglect traceability against the dishonest users.

To address this issue, Teranishi, Furukawa and Sako [15] proposed *k-times anonymous remote user authentication* (k-TAA) scheme. It is a fine-grained method of privacy-preserving which enables users to anonymously authenticate themselves for a bounded number of times. If a dishonest user authenticate himself beyond the predetermined number of times, he will not remain anonymous, which means the service provider can immediately trace the identity of him. However, k-TAA can not support that a service provider control over giving users access permission to his service, so it is not flexible enough. Shortly after that, in order to make service providers have better control over their users, Nguyen and Safavi Naini [13] proposed dynamic k-TAA scheme, which allows service providers to grant and revoke the access of registered users. Dynamic k-TAA allows service providers to restrict

access to their services based on not only the number of times, but also other factors such as expiry date. So it can be used in a much wider range of realistic scenarios.

Later on, Nguyen [12] proposed an efficient dynamic k-TAA scheme, where computation and communication costs are constant and do not depend on the limited number k. After that, a lot of schemes [3, 4, 16, 18] that focus on reducing the computation and communication costs of k-TAA are proposed. In 2006 , Au *et al.* [2] optimized Nguyen's dynamic k-TAA scheme and reduced time complexity from $O(k)$ to $O(log(k))$.

Although k-TAA and dynamic k-TAA realize the anonymity in authentication and k-times untraceability against dishonest users, how to apply them to mobile devices is still a challenge. With the breakthrough of some key technologies in communication (*e.g.* 5G), people tend to rely on mobile devices, such as smartphones. These devices are low-power and limited resources, which means they can not handle complex cryptographic operations. However, most of existing k-TAA schemes require certain computation-intensive operations such as pairing and proof of knowledge. Therefore, it is infeasible to simply apply k-TAA schemes to these weak devices with low performance.

Recently, tian *et al.* proposed a privacy-preserving remote user authentication with k-times untraceability scheme [17], which avoids expensive pairing operations and makes it possible to apply to mobile devices. They claimed that their scheme supports user authenticity and k-times untraceability. In other words, a malicious third party can not impersonate an authorized user. Besides, the authority can trace the real identities of dishonest users who have authenticated themselves for more than k times. However, in this work, we find that their scheme is not as secure as they claimed. Their scheme does not satisfy user authenticity or k-time untraceability. That is to say, by forging a credential, a dishonest user who do not enroll to the server can successfully pass the authentication. Moreover, the dishonest user can still keep anonymous after authenticating for more than k times.

The rest of this paper is organized as follows. In Section 2, we review the system model and security goals of Tian *et al.*'s scheme. Then, we briefly review Tian *et al.*'s scheme in Section 3. In Section 4, we give two concrete attacks and analyze the corresponding reasons. Finally, we draw our conclusion in Section 5.

2 Preliminary

2.1 System Model

There are two entities in Tian *et al.*'s privacy-preserving remote user authentication with k-times untraceability scheme: Enrolled users and an authentication server.

Users: The users should generate their key pairs, and enroll themselves. After enrolling themselves, they will generate a valid credential and submit it to the authentication server in the authentication phase. What's more, they need to generate k-size commitments and send them to the authentication server.

Authentication Server: The authentication server is an honest-but-curious entity. It is in charge of generating master key pair and authenticate users. If there is a dishonest user who authenticates himself for more than k times, the authentication server will detect his misbehavior and trace his real identity.

2.2 Security Goals

A k-times anonymous remote user authentication scheme should satisfy the following properties:

User Authenticity: An authorized user can generate a valid and legal credential to anonymously authenticate himself. During a session, a third party cannot impersonate an authorized user and forge a credential to pass the authentication.

k-time Untraceability: Users are only allowed to anonymously authenticate for k times. If there is a dishonest user who authenticate for more than k times, he will not remain anonymous, and the authentication server will be able to trace his real identity.

3 Reviews of Tian *et al.*'s Scheme

In this section, we briefly review Tian *et al.*'s scheme. Their scheme includes 5 phases: System initialization, key generation, enrollment, authentication, trace. These phases are given as follows.

3.1 System Initialization

Authentication server $\mathbb{S}$ takes a secure parameter as input, and output a multiplicative cyclic group $\mathbb{G}$ with order q. Let g denote a generator of $\mathbb{G}$. $\mathbb{S}$ chooses two hash function: $H_1 : \{0,1\}^* \rightarrow \mathbb{G}$, $H_2 : \mathbb{G} \rightarrow \mathbb{Z}_q$. $\mathbb{S}$ chooses $y, e, f \in \mathbb{Z}_q$, then computes $g^y, g_1 = g^e, g_2 = g^f, g_1^y, g_2^y$. Finally secretly keep master secret key: $msk = (y, e, f)$, and publish the public parameters:

$$PK = (\mathbb{G}, q, g, g^y, g_1, g_2, g_1^y, g_2^y).$$

3.2 Key Generation

The user i randomly chooses the secret key $x \in \mathbb{Z}_q$ as his secert key sk_i. Then uses system public key to computes g_1^x as his public key pk_i.

3.3 Enrollment

In this part, a user gets his credential by interacting with authentication server $\mathbb{S}$.

1) Firstly user i requests authentication server $\mathbb{S}$ for enrollment. Upon receiving i's request, $\mathbb{S}$ randomly chooses $\mathcal{K} \in \mathbb{Z}_q$, and computes:

$$\delta_1 = g^{\mathcal{K}}, \delta_2 = (g_1^x g_2)^{\mathcal{K}},$$

and sends them to user i.

2) User i randomly chooses $x_1, a, b \in \mathbb{Z}_q$, and computes:

$$\begin{aligned} \alpha &= (g_1^x g_2)^{y \cdot x_1}, \\ \beta &= (g_1^x g_2)^{x_1}, \\ m &= H_1(\alpha||\beta), \\ r &= m \cdot \beta^a \cdot \delta_2^{b \cdot x_1}, \\ m' &= H_2(m||r)/b, \end{aligned}$$

and sends m' to $\mathbb{S}$.

3) Upon receiving m', the authentication server $\mathbb{S}$ computes blinded signature:

$$s' = \mathcal{K} + y \cdot m',$$

and sends s' to user i.

4) Upon receiving s', user i checks

$$g^{s'} \stackrel{?}{=} g^{y \cdot m'}.$$

If it does not hold, abort. Otherwise, user computes

$$s = s' \cdot b + a,$$

and keeps (α, β, r, s) as a valid credential.

3.4 Authentication

In this part, $\mathbb{S}$ authenticates valid users.

1) Before making request for authentication, user i randomly chooses $s_1, s_2, ..., s_k \in \mathbb{Z}_q$ and computes two k-size sets:

$$\begin{aligned} (S_1, S_2, ..., S_k) &= (g_1^{x \cdot s_1}, g_1^{x \cdot s_2}, ..., g_1^{x \cdot s_k}) \\ (\bar{S}_1, \bar{S}_2, ..., \bar{S}_k,) &= (g_2^{s_1}, g_2^{s_2}, ..., g_2^{s_k}). \end{aligned}$$

Then encrypts them and generates ciphertext $C_i = Enc(S_i, \bar{S}_i)$ by using master public key, and sends it to the authentication server $\mathbb{S}$.

2) Upon receiving the request from the user i, $\mathbb{S}$ randomly chooses $c_i \in \mathbb{Z}_q$ and sends it to user i.

3) User i computes

$$\begin{aligned} R_1 &= x_1 + s_1 \cdot c_i + s_2 \cdot c_i^2 + ... + s_k \cdot c_i^k, \\ R_2 &= x \cdot R_1, \end{aligned}$$

and sends $(R_1, R_2, \alpha, \beta, r, s)$ to $\mathbb{S}$.

4) $\mathbb{S}$ checks whether user i's credential is valid:

$$H_1(\alpha||\beta) \stackrel{?}{=} \beta^{-s} \cdot \alpha^{H_2(H_1(\alpha||\beta)||r)} \cdot r. \quad (1)$$

If it does not hold, aborts. Otherwise, checks whether the following equation holds:

$$g_1^{R_2} \cdot g_2^{R_1} \stackrel{?}{=} \beta \cdot S_1^{c_i} \cdot S_2^{c_i^2} \cdots S_k^{c_i^k} \cdot \bar{S_1}^{c_i} \cdot \bar{S_2}^{c_i^2} \cdots \bar{S_k}^{c_i^k}.$$

If it does not hold, aborts. Otherwise, $\mathbb{S}$ authenticates user i.

3.5 Trace

For each authentication request, $\mathbb{S}$ randomly chooses a c_i and sends it to the user i, which is used to compute R_1 and R_2. Therefore, the user i generates different R_1 and R_2 every time. If there is a dishonest user i who maliciously uses a valid credential more than k times, the authentication server $\mathbb{S}$ will be able to get at least $k+1$ different R_1 and R_2. Then $\mathbb{S}$ has the following equations:

$$\begin{cases} R_{1_1} = x_1 + s_1 \cdot c_{i_1} + s_2 \cdot c_{i_1}^2 + ... + s_k \cdot c_{i'_1}^k \\ R_{1_2} = x_1 + s_1 \cdot c_{i_2} + s_2 \cdot c_{i_2}^2 + ... + s_k \cdot c_{i'_2}^k \\ \quad ... \\ R_{1_{k+1}} = x_1 + s_1 \cdot c_{i_{k+1}} + s_2 \cdot c_{i_{k+1}}^2 + ... + s_k \cdot c_{i_{k+1}}^k \end{cases}, \quad (2)$$

and

$$\begin{cases} R_{2_1} = x \cdot (x_1 + s_1 \cdot c_{i_1} + s_2 \cdot c_{i_1}^2 + ... + s_{k_1} \cdot c_{i_1}^k) \\ R_{2_2} = x \cdot (x_1 + s_1 \cdot c_{i_2} + s_2 \cdot c_{i_2}^2 + ... + s_k \cdot c_{i_2}^k) \\ \quad ... \\ R_{2_{k+1}} = x \cdot (x_1 + s_1 \cdot c_{i_{k+1}} + s_2 \cdot c_{i_{k+1}}^2 + ... + s_k \cdot c_{i_{k+1}}^k) \end{cases}. \quad (3)$$

By calculating (2) and (3), $\mathbb{S}$ can get x_1 and $x_1 \cdot x$ respectively. Obviously, the secret key sk_i of the user i can be successfully obtained by $\mathbb{S}$. After getting the secret key of the user i, the authentication server $\mathbb{S}$ can easily trace the real identity of this user.

4 Attacks on Tian *et al.*'s Scheme

In this section, we show the detail attacks on Tian *et al.*'s scheme. By analyzing it, we show that their scheme satisfies neither user authenticity nor k-times untraceability. This means that a dishonest user can authenticate himself without enrollment by forging a credential, and the dishonest user can still keep anonymous after authenticating himself more than k times. The detail attacks are given as follows.

4.1 Attack on User Authenticity

We assume a misbehaving user i', who does not enroll himself to the authentication server $\mathbb{S}$ or generate a valid credential. However, i' can forge a credential and pass

the verifying equations. Since the server $\mathbb{S}$ cannot distinguish him from other honest users, it will allow him to be authenticated. The description of this attack is given as follows.

In order to forge a credential without enrollment, user i' randomly chooses $x', x'_1 \in \mathbb{Z}_q$ and computes:

$$\begin{aligned}\alpha' &= (g_1^{x'} g_2)^{x'_1},\\ \beta' &= \alpha',\\ r' &= H_1(\alpha'||\beta'),\\ s' &= H_2(H_1(\alpha'||\beta')||r'),\end{aligned}$$

and saves $(\alpha', \beta', r', s')$ as his forged credential.

To authenticate himself by using this forged credential, user i' needs to interact with $\mathbb{S}$ as below.

1) User i' randomly chooses $s_1, s_2, ..., s_k \in \mathbb{Z}_q$ and computes two k-size sets:

$$\begin{aligned}(S_1, S_2, ..., S_k) &= (g_1^{x'\cdot s_1}, g_1^{x'\cdot s_2}, ..., g_1^{x'\cdot s_k})\\ (\bar{S_1}, \bar{S_2}, ..., \bar{S_k},) &= (g_2^{s_1}, g_2^{s_2}, ..., g_2^{s_k}).\end{aligned}$$

Then i' encrypts them and generates ciphertext $C'_i = Enc(S'_i, \bar{S'_i})$ by using master public key, and sends it to the authentication server $\mathbb{S}$ as his authentication request.

2) Upon receiving the request from the user i', $\mathbb{S}$ randomly chooses $c_{i'} \in \mathbb{Z}_q$ and sends it to user i'.

3) Upon receiving c', user i' computes:

$$\begin{aligned}R'_1 &= x'_1 + s_1 \cdot c_{i'} + s_2 \cdot c_{i'}^2 + ... + s_k \cdot c_{i'}^k,\\ R'_2 &= x' \cdot R_1,\end{aligned}$$

and sends $(R'_1, R'_2, \alpha', \beta', r', s')$ to $\mathbb{S}$.

4) $\mathbb{S}$ checks whether user i's credential $(\alpha', \beta', r', s')$ is valid:

$$H_1(\alpha'||\beta') \stackrel{?}{=} \beta'^{-s'} \cdot \alpha'^{H_2(H_1(\alpha'||\beta')||r')} \cdot r'. \quad (4)$$

The Equation (4) can hold correctly. Because we have:

$$\begin{aligned}&\beta'^{-s'} \cdot \alpha'^{H_2(H_1(\alpha'||\beta')||r')} \cdot r'\\ =&\beta'^{-H_2(H_1(\alpha'||\beta')||r')} \cdot \beta'^{H_2(H_1(\alpha'||\beta')||r')} \cdot r'\\ =&r'\\ =&H_1(\alpha'||\beta').\end{aligned}$$

Then $\mathbb{S}$ checks whether the following equation holds:

$$g_1^{R'_2} \cdot g_2^{R'_1} \stackrel{?}{=} \beta' \cdot S_1^{c_{i'}} \cdot S_2^{c_{i'}^2} \cdots S_k^{c_{i'}^k} \cdot \bar{S_1}^{c_{i'}} \cdot \bar{S_2}^{c_{i'}^2} \cdots \bar{S_k}^{c_{i'}^k}. \quad (5)$$

The Equation (5) can hold correctly. Because we have:

$$\begin{aligned}&\beta' \cdot S_1^{c_{i'}} \cdot S_2^{c_{i'}^2} \cdots S_k^{c_{i'}^k} \cdot \bar{S_1}^{c_{i'}} \cdot \bar{S_2}^{c_{i'}^2} \cdots \bar{S_k}^{c_{i'}^k}\\ =&(g_1^x g_2)^{x'_1} \cdot g_1^{x' s_1 c_{i'}} \cdot g_1^{x' s_2 c_{i'}^2} \cdots g_1^{x' s_k c_{i'}^k} \cdot g_2^{s_1 c_{i'}}\\ &\cdot g_2^{s_2 c_{i'}^2} \cdots g_2^{s_k c_{i'}^k}\\ =&g_1^{x'(x'_1 + s_1 c_{i'} + s_2 c_i^2 + \cdots + s_k c_{i'}^k)} \cdot g_2^{(x'_1 + s_1 c_{i'} + s_2 c_{i'}^2 + \cdots + s_k c_{i'}^k)}\\ =&g_1^{R'_2} \cdot g_2^{R'_1}.\end{aligned}$$

Both of the above two equations hold, so $\mathbb{S}$ authenticates the misbehaving user i'.

This attack indicates that Tian *et al.*'s scheme can not satisfy unforgeability. We show that a misbehaving user i' can successfully authenticate himself to the authentication server $\mathbb{S}$ without enrollment by forging a credential. What's more, the misbehaving user i' can use the same $(C_{i'}, R'_1, R'_2, \alpha', \beta', r', s')$ up to k times to authenticate himself to $\mathbb{S}$ without being detected.

4.2 Attack on k-Time Untraceability

Tian *et al.* claims their scheme can achieve k-time untraceability, which allows an honest user be authenticated anonymously only up to k times. If there is a dishonest user who authenticated himself more than k times, he will not remain anonymous and the authentication server will trace his real identity. We give an attack on it and show that in Tian *et al*'s scheme, a dishonest user can keep anonymous after authenticating for $k+1$ times. The detail of attack is described as follows:

User i' randomly chooses $x', x'_1 \in \mathbb{Z}_q$ and forges a credential:

$$\begin{aligned}\alpha' &= (g_1^{x'} g_2)^{x'_1},\\ \beta' &= \alpha',\\ r' &= H_1(\alpha'||\beta'),\\ s' &= H_2(H_1(\alpha'||\beta')||r').\end{aligned}$$

User i' uses this forged credential to authenticate himself as below.

1) User i' randomly chooses $s_1, s_2, ..., s_k \in \mathbb{Z}_q$ and computes two k-size sets:

$$\begin{aligned}(S_1, S_2, ..., S_k) &= (g_1^{x'\cdot s_1}, g_1^{x'\cdot s_2}, ..., g_1^{x'\cdot s_k}),\\ (\bar{S_1}, \bar{S_2}, ..., \bar{S_k},) &= (g_2^{s_1}, g_2^{s_2}, ..., g_2^{s_k}).\end{aligned}$$

Then i' encrypts them and generates ciphertext $C'_i = Enc(S'_i, \bar{S'_i})$ by using master public key, and sends it to the authentication server $\mathbb{S}$ as his authentication request.

2) Upon receiving the request from the user i', $\mathbb{S}$ randomly chooses $c_{i'} \in \mathbb{Z}_q$ and sends it to user i'.

3) Upon receiving c', user i' computes:

$$R'_1 = x'_1 + s_1 \cdot c_{i'} + s_2 \cdot c_{i'}^2 + ... + s_k \cdot c_{i'}^k,$$
$$R'_2 = x' \cdot R_1,$$

and sends $(R'_1, R'_2, \alpha', \beta', r', s')$ to $\mathbb{S}$.

4) $\mathbb{S}$ checks whether the following two equations are hold:

$$H_1(\alpha'||\beta') \stackrel{?}{=} \beta'^{-s'} \cdot \alpha'^{H_2(H_1(\alpha'||\beta')||r')} \cdot r', \quad (6)$$

$$g_1^{R'_2} \cdot g_2^{R'_1} \stackrel{?}{=} \beta' \cdot S_1^{c_{i'}} \cdot S_2^{c_{i'}^2} \cdots S_k^{c_{i'}^k} \cdot \bar{S_1}^{c_{i'}} \cdot \bar{S_2}^{c_{i'}^2} \cdots \bar{S_k}^{c_{i'}^k}. \quad (7)$$

According to the attack on authentication, we can know that Equation (6) and Equation (7) are both hold, and $\mathbb{S}$ authenticates the misbehaving user i'.

5) If this misbehaving user i' uses his forged credential to authenticate himself for more than k times, the authentication server $\mathbb{S}$ will detect his misbehavior and try to trace him. $\mathbb{S}$ gets at least $k+1$ different R'_1 and R'_2 with different $c_{i'}$. Then $\mathbb{S}$ computes the following equations:

$$\begin{cases} R'_{1_1} = x'_1 + s_1 \cdot c_{i'_1} + s_2 \cdot c_{i_1^*}^2 + ... + s_k \cdot c_{i'_1}^k \\ R'_{1_2} = x'_1 + s_1 \cdot c_{i'_2} + s_2 \cdot c_{i'_2}^2 + ... + s_k \cdot c_{i'_2}^k \\ \quad ... \\ R'_{1_{k+1}} = x'_1 + s_1 \cdot c_{i'_{k+1}} + s_2 \cdot c_{i'_{k+1}}^2 + ... \\ \qquad + s_k \cdot c_{i'_{k+1}}^k \end{cases}, \quad (8)$$

and

$$\begin{cases} R'_{2_1} = x' \cdot (x'_1 + s_1 \cdot c_{i'_1} + s_2 \cdot c_{i'_1}^2 + ... + s'_k \cdot c_{i'_1}^k) \\ R'_{2_2} = x' \cdot (x'_1 + s_1 \cdot c_{i'_2} + s_2 \cdot c_{i'_2}^2 + ... + s_k \cdot c_{i'_2}^k) \\ \quad ... \\ R'_{2_{k+1}} = x' \cdot (x'_1 + s_1 \cdot c_{i'_{k+1}} + s_2 \cdot c_{i'_{k+1}}^2 + ... \\ \qquad + s_k \cdot c_{i'_{k+1}}^k) \end{cases}, \quad (9)$$

to get x'_1 and $x'_1 \cdot x'$ respectively. Then $\mathbb{S}$ can easily obtain the secret key of i', which is x'.

After getting x', $\mathbb{S}$ computes $g_1^{x'}$ and try to find out the identity whose public key matches this value. However, $\mathbb{S}$ could not determine the identity of dishonest user i' in its database since there is no public key of i'.

This attack indicates that Tian *et al.*'s k-RUA scheme does not satisfy k-time untraceability. We show that by forging a credential, a misbehaving user can not only successfully authenticate himself to the authentication server, but also keep anonymous after authenticating for more than k times.

4.3 Discussion

In the following, we analyze the reason to lead to the above attacks and provide our suggestions. In authentication phase of Tian *et al.*'s scheme, we can find that any one can forge a credential to impersonate an authorized user and pass authentication. We can see that in the Equation (1) hash value $H_2(H_1(\alpha||\beta)||r)$ is irrelevant to s. This makes the verifying Equation (1) vulnerable. In this situation, attackers can randomly choose s. When setting $s' = H_2(H_1(\alpha||\beta)||r)$, attackers can forge a credential $(R'_1, R'_2, \alpha', \beta', r', s')$ to pass authentication. To improve Tian *et al.*'s scheme, we suggest applying group signature technique to associate hash value $H_2(H_1(\alpha||\beta)||r)$ with s. With anonymity, group signature can protect users' privacy and help to get rid of the threats that attackers set $s = H_2(H_1(\alpha||\beta)||r)$. Then attackers are not able to forge a credential to pass authentication in this way.

5 Conclusion

In this paper, we analyze the security of Tian *et al.*'s privacy-preserving remote user authentication with k-times untraceability scheme. We found that their scheme is insecure although it is proven to be secure. It can not satisfy user authenticity and k-time untraceability. This is to say, in their k-RUA scheme, any third party can forge a credential and impersonate an authorized user to pass authentication. Furthermore, the authentication server can not trace the real identity of a dishonest user, even though he authenticated for more than k times. Our analysis is confirmed thought two concrete attacks. In the last, we show the reason to produce such attacks and give the corresponding suggestions. Our future work is improving privacy-preserving remote user authentication with k-times untraceability scheme.

Acknowledgments

This research was was supported by Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS201808, GCIS201710), the Engineering Program Project of CUC (3132015XNG1541), National Key *R&D* Program of China(2018YFB0803900) and the Foundation of Guizhou Provincial Key Laboratory of Public Big Data (No. 2019BDKFJJ012). The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] G. Ateniese, J. Camenisch, M. Joye, and G. Tsudik, "A practical and provably secure coalition-resistant group signature scheme," in *Annual International Cryptology Conference*, pp. 255–270, 2000.

[2] M. H. Au, W. Susilo, and Y. Mu, "Constant-size dynamic k-TAA," in *International Conference on Security and Cryptography for Networks*, pp. 111–125, 2006.

[3] J. Camenisch, S. Hohenberger, M. Kohlweiss, A. Lysyanskaya, and M. Meyerovich, "How to win the clonewars: Efficient periodic n-Times anonymous authentication," in *Proceedings of the 13th ACM conference on Computer and Communications Security*, pp. 201–210, 2006.

[4] J. Camenisch, M. Kohlweiss, and C. Soriente, "An accumulator based on bilinear maps and efficient revocation for anonymous credentials," in *International Workshop on Public Key Cryptography*, pp. 481–500, 2009.

[5] L. Chen, M. Enzmann, A. R. Sadeghi, M. Schneider, and M. Steiner, "A privacy-protecting coupon system," in *International Conference on Financial Cryptography and Data Security*, pp. 93–108, 2005.

[6] D. He, N. Kumar, J. Chen, C. C. Lee, N. Chilamkurti, and S. S. Yeo, "Robust anonymous authentication protocol for health-care applications using wireless medical sensor networks," *Multimedia Systems*, vol. 21, no. 1, pp. 49–60, 2015.

[7] D. He, S. Zeadally, N. Kumar, and J. H. Lee, "Anonymous authentication for wireless body area networks with provable security," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2590–2601, 2016.

[8] M. S. Hwang, C. C. Lee, Y. C. Lai, "An untraceable blind signature scheme", *IEICE Transactions on Foundations*, vol. E86-A, no. 7, pp. 1902–1906, July 2003.

[9] J. V. Jokinen, L. Blants, R. Pitkänen, S. Pienimäki, J. Mattila, and R. Suomela, *Real-Time Wireless E-Coupon (promotion) Definition based on Available Segment*, US20080120186A1, 2008.

[10] C. C. Lee, M. S. Hwang, and W. P. Yang, "A new blind signature based on the discrete logarithm problem for untraceability", *Applied Mathematics and Computation*, vol. 164, no. 3, pp. 837–841, May 2005.

[11] H. Mun, K. Han, Y. S. Lee, C. Y. Yeun, and H. H. Choi, "Enhanced secure anonymous authentication scheme for roaming service in global mobility networks," *Mathematical and Computer Modelling*, vol. 55, no. 1-2, pp. 214–222, 2012.

[12] L. Nguyen, "Efficient dynamic k-times anonymous authentication," in *International Conference on Cryptology in Vietnam*, pp. 81–98, 2006.

[13] L. Nguyen and R. S. Naini, "Dynamic k-times anonymous authentication," in *International Conference on Applied Cryptography and Network Security*, pp. 318–333, 2005.

[14] J. Shao, X. Lin, R. Lu, and C. Zuo, "A threshold anonymous authentication protocol for vanets," *IEEE Transactions on vehicular technology*, vol. 65, no. 3, pp. 1711–1720, 2015.

[15] I. Teranishi, J. Furukawa, and K. Sako, "K-times anonymous authentication," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 308–322, 2004.

[16] I. Teranishi and K. Sako, "K-times anonymous authentication with a constant proving cost," in *International Workshop on Public Key Cryptography*, pp. 525–542, 2006.

[17] Y. Tian, Y. Li, B. Sengupta, R. H. Deng, A. Ching, and W. Liu, "Privacy-preserving remote user authentication with k-times untraceability," in *International Conference on Information Security and Cryptology*, pp. 647–657, 2018.

[18] Y. Yang, H. Cai, Z. Wei, H. Lu, and K. K. R. Choo, "Towards lightweight anonymous entity authentication for iot applications," in *Australasian Conference on Information Security and Privacy*, pp. 265–280, 2016.

[19] F. Zhang and K. Kim, "Id-based blind signature and ring signature from pairings," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 533–547, 2002.

[20] H. Zhu, Y. Zhang, and X. Wang, "A novel one-time identity-password authenticated scheme based on biometrics for e-coupon system," *International Journal of Network Security*, vol. 18, no. 3, pp. 401–409, 2016.

Biography

Qijia Zhang received the B.E. degree from Northeastern University at Qinhuangdao, Qinhuangdao, China, in 2018. He is currently working towards a M.E. degree, North China University of Technology, Beijing. His research interests include applied cryptography and information security.

Jianhong Zhang received his Ph.D. degrees in Cryptography from Xidian University, Xian, Shanxi, in 2004 and his M.S. degree in Computer Software from Guizhou University, Guiyang, Guizhou, in 2001. He was engaging in postdoctoral research at Peking University from October 2005 to December 2007. He now is a Professor of School of Electronic and Information Engineering, North China University of Technology, Beijing China. His research interests include computer networks, cryptography, electronic commerce security, computer software.

Linhan Liu is currently working towards a B.E. degree in North China University of Technology, Beijing. Her research interests are information security and computer networks.

Jing Wang is an employee of Beijing Jingdong Century Information Technology Company, Limited. His research interests include applied cryptography and information security.

Pei Liu is an employee of Beijing Jingdong Century Information Technology Company, Limited. His research interests include applied cryptography and information security.

The Linear Complexity of the Interleaved Polynomial Quotient Sequences

Chun-e Zhao[1,2], Tongjiang Yan[1,2], Xubo Zhao[1], and Qihua Niu[1]
(Corresponding author: Chun-e Zhao)

College of Sciences, China University of Petroleum[1]
Shandong Qingdao, 266555, China
Key Laboratory of Applied Mathematics (Putian University)[2]
Fujian Province University, Fujian Putian, 351100, China
Email:zhaochune1981@163.com

(Received Dec. 24, 2019; Revised and Accepted July 23, 2020; First Online Apr. 11, 2021)

Abstract

Pseudorandom sequences have a wide range of applications in many fields, such as radar navigation systems, spread-spectrum communication systems, stream ciphers, etc. When used in stream ciphers, linear complexity is a significant index. This paper constructs a new class of interleaved sequences derived from binary polynomial quotient sequences, which have larger linear complexity, longer period, and better balance character than some known ones. The result shows that the linear complexity is large enough to resist BM-algorithm attacks.

Keywords: Euler Quotient; Fermat Quotient; Interleaved Sequences; Linear Complexity; Polynomial Quotient

1 Introduction

Let p be an odd prime and $u \in Z$, where Z is the set of all nonnegative integers. The *polynomial quotient* [6] modulo p is defined by

$$F_w(u) \equiv \frac{u^w - u^{wp}}{p} \pmod{p}, \tag{1}$$

where $w \in Z_p$ and $Z_p = \{0, 1, 2, \cdots, p-1\}$. In particular, when $w = p-1$, $F_w(u)$ is referred to *Fermat quotient*.

As far as we know, the Fermat quotient has first been studied from the viewpoint of stream ciphers by Ostafe and Shpalinsk [9]. Later, three families of binary sequences have been introduced from polynomial quotient. One is the binary threshold sequence $(e(u))$ [11] defined by

$$e(u) = \begin{cases} 0, & \text{if } 0 \le F_w(u) \le \frac{p-1}{2}, \\ 1, & \text{if } \frac{p+1}{2} \le F_w(u) < p. \end{cases}$$

The second one is the Legendre polynomial quotient sequence $(f(u))$ [1] defined by

$$f(u) = \begin{cases} 0, & \text{if } \left(\frac{F_w(u)}{p}\right) = 1 \text{ or } F_w(u) = 0, \\ 1, & \text{otherwise}, \end{cases}$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol.

The third one is the least significant bit of polynomial sequence $(s(u))$ [14] defined by

$$s(u) = \begin{cases} 0, & \text{if } F_w(u) \equiv 0 \pmod 2, \\ 1, & \text{if } F_w(u) \equiv 1 \pmod 2, \end{cases} \quad u \ge 0.$$

Chen, Ostafe and Winterhof studied the pseudorandomness of $(e(u))$ [11]. Gomez and Winterhof discussed the pseudo-randomness of $(f(u))$ [12] for $w = p-1$. For both $(e(u))$ and $(f(u))$, Chen, Du and other coauthors considered the linear complexity when 2 is a primitive element modulo p^2 and later extended to the case $2^{p-1} \not\equiv 1 \pmod{p^2}$ [2–5, 7]. In particular, Du, Klapper and Chen gave a conjecture of the linear complexity [3]. Chen gave the exact linear complexity and its trace representation [2]. They go deeply into the k-error linear complexity [13]. Zhao, Ma, etc. deliberated the linear complexity of $(s(u))$ [14]. The results show that these sequences have large linear complexity which can resist BM algorithm attacks.

To improve the balance property of these three families of sequences, Zhao,Yan and Niu constructed a class of balanced polynomial quotient sequences [8],

$$t(u) = \begin{cases} 1, & \text{if } F_w(u) \in I, \\ 0, & otherwise, \end{cases}$$

where $|I| = \frac{p+1}{2}$ and $|I|$ is odd.

In order to expand the period from p^2 to p^3, we use the interleaving technique to polynomial quotient sequences. Then we construct a new class of interleaved polynomial sequences and discuss the linear complexity. The result shows that the linear complexity is large enough to resist the BM algorithm attacks. And it has better balance character than the p^{r+1}-period Euler quotient sequences for $r = 2$ [7].

2 Preliminaries

Let $\mathbb{F}_q$ be a finite field. For a T-periodic sequence (a_u) over $\mathbb{F}_q$, the polynomial

$$a(x) = a_0 + a_1 x + \cdots + a_{T-1}x^{T-1} \in \mathbb{F}_q[x]$$

is called the *generating polynomial* of (a_u). If $a_{u+L} = c_0 a_u + c_1 a_{u+1} + \cdots + c_{L-1}a_{u+L-1}$ for all $u \geq 0$, the polynomial

$$c(x) = x^L + c_{L-1}x^{L-1} + \cdots + c_0 \in \mathbb{F}_q[x]$$

is called the *characteristic polynomial* of (a_u). Among all the characteristic polynomials, the one with the smallest degree is called the *minimal polynomial*. Then the linear complexity is the degree of the minimal characteristic polynomial. It is well known that the minimal polynomial can be obtained by computing

$$\frac{x^T - 1}{\gcd(x^T - 1, a(x))},$$

and the linear complexity of (a_u) is

$$L((a_u)) = T - \deg(\gcd(x^T - 1, a(x))).$$

So the number of the common roots of $a(x)$ and $x^T - 1$ will lead to the values of the linear complexity.

Let p be an odd prime and $F_w(u)$ be defined in Equation (1). We construct a set of p^2-periodic binary sequences $\{t_i(u)\}$ as follows:

$$t_i(u) = \begin{cases} 1, & \text{if } F_w(u) \in I + i, \\ 0, & otherwise, \end{cases} \quad u \geq 0. \tag{2}$$

where $I \subset Z_p, |I| = 2\left\lfloor \frac{p-1}{4} \right\rfloor + 1$ and $I + i = \{a + i \pmod p | a \in I\}$ for $i \in Z_p$. Especially, $t_0(u) = t(u)$ for $u \geq 0$.

Let $\{(t_0(u)), (t_1(u)), \cdots, (t_{p-1}(u))\}$ be a set of p^2- periodic sequences defined in Equation (2). An $p^2 \times p$ matrix A is formed by placing the sequence $(t_j(u))$ on the j-th column, where $0 \leq j \leq p-1$. Then one can obtain an interleaved sequence $(s(u))$ of period p^3 by concatenating the successive rows of the matrix A. For simplicity, the interleaved sequence $(s(u))$ can be written as

$$s(u) = t_i(j), \ where \ u = i + jp \tag{3}$$

for $i = 0, 1, \cdots, p-1$, *and* $j = 0, 1, \cdots, p^2 - 1$. Let

$$\begin{aligned} H_w(u) &= u^{-w}F_w(u) \pmod p, \\ Z^*_{p^2} &= \{u \in Z_{p^2} | gcd(u, p) = 1\}, \\ D_l &= \{u \in Z^*_{p^2} : H_w(u) = l\}, \\ aD_l &= \{au \pmod{p^2} | u \in D_l\}, \\ D^p_l &= \{u \pmod p | u \in D_l\}, \\ D_l(x) &= \sum_{u \in D_l} x^u \ T_i(x) = \sum_{u=0}^{p^2-1} t_i(u)x^u, \end{aligned}$$

where $l, i \in Z_p = \{0, 1, \cdots, p-1\}$.

3 Linear Complexity of the Interleaved Polynomial Quotient Sequences

In this section, we will determine the linear complexity for the cases $w = p - 1$ and $w = \frac{p-1}{2}$, respectively. We note here that almost primes satisfy $2^{p-1} \neq \pmod{p^2}$. When $2^{p-1} \equiv 1 \pmod{p^2}$, such primes are rare and the only known such primes are $p = 1093$ and $p = 3511$ and it was reported that there are no new such primes when $p < 4 \times 10^{12}$. We always consider $2^{p-1} \neq 1 \pmod{p^2}$ in this section..

3.1 Linear Complexity of Interleaved Polynomial Sequences for $w = p - 1$

In this section, we will focus on the condition $w = p-1$. In order to determine the linear complexity of the interleaved sequence $(s(u))$, we need the following Lemmas 1-9.

Lemma 1. *[6] Let D_l, aD_l, D^p_l be defined in Section 2.*

1) $|D_l| = p - 1$ for $l = 0, 1, \cdots, p-1$,

2) $aD_l = D_{l+l' \pmod p}$ if $a \in D_{l'}$,

3) $D^p_l = Z^*_p$ for $l = 0, 1, \cdots, p-1$.

Lemma 2. *[5] Let $\beta \in \overline{\mathbb{F}}_2$ be a primitive p^2-th root of unity. Then for any $n \in Z^*_{p^2}$, we have*

$$\sum_{l=0}^{p-1} D_l(\beta^n) = 0.$$

Lemma 3. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3). Then*

$$s(x) = T_0(x^p) + xT_1(x^p) + \cdots + x^{p-1}T_{p-1}(x^p).$$

Proof. Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3). By the definition of $(s(u))$, we have

$$\begin{aligned} s(x) &= \sum_{u=0}^{p^3-1} s(u)x^u \\ &= \sum_{j=0}^{p^2-1} \sum_{i=0}^{p-1} s(i + jp)x^{i+jp} \\ &= \sum_{i=0}^{p-1} x^i \sum_{j=0}^{p^2-1} t_i(j)x^{jp} \\ &= \sum_{i=0}^{p-1} x^i T_i(x^p). \end{aligned}$$

□

Lemma 4. *[10] Let p be an odd prime and N be an positive integer, Z_N and Z_N^* be defined as in Section 2, then*

$$Z_{p^3} = p^2 Z_p \cup pZ_{p^2}^* \cup Z_{p^3}^*,$$

where $p^2 Z_p = \{p^2 u \pmod{p^3} | u \in Z_p\}$ and $pZ_{p^2}^* = \{pu \pmod{p^3} | u \in Z_{p^2}^*\}$.

Lemma 5. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3) and $\theta \in \overline{\mathbb{F}}_2$ be a primitive p^3-th root of unity. Then for every $j \in Z_p$, we have*

$$s(\theta^{jp^2}) = 0.$$

Proof. Let $s(x)$ be the generating polynomial of the interleaved sequence defined in Equation (3). Then by Lemma 3, $s(x) = \sum_{i=0}^{p-1} x^i T_i(x^p)$. For $j = 0$, we have

$$s(\theta^{jp^2}) = s(1) = p(p-1)|I| = 0 \pmod 2.$$

For $j = 1, 2, \cdots, p-1$, we have

$$s(\theta^{jp^2}) = \sum_{i=0}^{p-1} \theta^{jp^2} T_i(1) = p(p-1)|I|\theta^{jp^2} = 0 \pmod 2.$$

Then $s(\theta^{jp^2}) = 0$ for every $jp^2 \in P^2 Z_p$. □

Lemma 6. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3) and $\theta \in \overline{\mathbb{F}}_2$ be a primitive p^3-th root of unity. Then*

$$s(\theta^{jp}) \neq 0.$$

for any $j \in Z_{p^2}^*$.

Proof. Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3). By Lemma 3, $s(x) = \sum_{i=0}^{p-1} x^i T_i(x^p)$. Then

$$s(\theta^{jp}) = T_0(\theta^{jp^2}) + \theta^{jp} T_1(\theta^{jp^2}) + \cdots + \theta^{jp(p-1)} T_{p-1}(\theta^{jp^2}).$$

For every $j \in Z_{p^2}^*$, let $\theta^{jp} = \beta$ and $\theta^{jp^2} = \gamma$. Then we have

$$T_i(\theta^{jp^2}) = T_i(\gamma) = \sum_{a \in I+i} D_a(\gamma)$$

for each $i \in Z_p$. And for every $a \in I + i$, where $i \in Z_p$, by the definition of $D(x)$ defined in Section 2,

$$D_a(\gamma) = \sum_{u \in D_a} \gamma^u.$$

For γ is a primitive p-th root of unity, then $\gamma^u = \gamma^{u \pmod p}$. By Lemma 1,

$$D_a(\gamma) = \sum_{u \in D_a} \gamma^u = \sum_{u \in D_a^p} \gamma^u = 1.$$

Then $T_i(\gamma) = |I + i| = |I|$ and $s(\theta^{jp}) = |I|(1 + \beta + \beta^2 + \cdots + \beta^{p-1})$. Because β is primitive p^2-th root of unity, then $1 + \beta + \beta^2 + \cdots + \beta^{p-1} \neq 0$. By Equation (2), $|I| = 2\left\lfloor \frac{p-1}{4} \right\rfloor + 1 \neq 0$. Therefore, $s(\theta^{jp}) \neq 0$ for any $j \in Z_{p^2}^*$. □

Lemma 7. *[6] Let $\beta \in \overline{\mathbb{F}}_2$ be a primitive p^2-th root of unity. If $2 \in D_{l_0}$, where $1 \leq l_0 \leq p-1$, then for any $0 \leq l \leq p-1$ and $n \in Z_{p^2}^*$, we have*

$$D_l(\beta^n) \neq 0.$$

Let the cyclic matrix

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & c_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix}$$

Then C is inversible if and only if $gcd(c(x), x^n - 1) = 1$, where $c(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$.

Lemma 8. *Let $\beta \in \overline{\mathbb{F}}_2$ be a primitive p^2-th root of unity. If $2^{p-1} \neq 1 \pmod{p^2}$, then*

$$T_i(\beta^n) \neq 0$$

for every $n \in Z_{p^2}^*$ and $i \in Z_p$.

Proof. For the reason that $2^{p-1} \not\equiv 1 \pmod{p^2}$, we get $H_{p-1}(2) \neq 0$. Let $H_{p-1}(2) = l_0$ $(1 \leq l_0 < p)$ and $2 \in D_{l_0}$. We obtain $2^j \in D_{jl_0 \pmod p}$. Suppose $T_i(\beta^{n_0}) = 0$ for some $n_0 \in D_{i_0} \subset \mathbb{Z}_{p^2}^*$. Then

$$0 = T_i(\beta^{n_0})^{2^j} = T_i(\beta^{n_0 2^j}) = T_{i+i_0+jl_0 \pmod p}(\beta).$$

Because $\gcd(l_0, p) = 1$, then $i_0 + jl_0 \pmod p$ runs through $\mathbb{Z}_p$ once when j turns over $\{0, 1, \cdots, p-1\}$. This means that for every $j \in \mathbb{Z}_p$ and every $n \in \mathbb{Z}_{p^2}^*$, $T_j(\beta^n) = 0$ always holds. Then we have

$$\begin{aligned} T_0(\beta^n) &= \sum_{i \in I} \sum_{u \in D_i} (\beta^{nu}) = 0, \\ T_1(\beta^n) &= \sum_{i \in I+1} \sum_{u \in D_i} (\beta^{nu}) = 0, \\ T_2(\beta^n) &= \sum_{i \in I+2} \sum_{u \in D_i} (\beta^{nu}) = 0, \\ &\vdots \qquad\quad \vdots \qquad\quad \vdots \\ T_{p-1}(\beta^n) &= \sum_{i \in I+(p-1)} \sum_{u \in D_i} (\beta^{nu}) = 0. \end{aligned}$$

Define c_i as follows

$$c_i = \begin{cases} 0, & \text{if } i \notin I, \\ 1, & \text{if } i \in I. \end{cases}$$

where $i = 0, 1, \ldots, p-1$. Thus the above system can be expressed as

$$\begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{p-1} \\ c_{p-1} & c_0 & c_1 & \cdots & c_{p-2} \\ c_{p-2} & c_{p-1} & c_0 & \cdots & c_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix} \begin{pmatrix} D_0(\beta^n) \\ D_1(\beta^n) \\ D_2(\beta^n) \\ \vdots \\ D_{p-1}(\beta^n) \end{pmatrix} = 0.$$

The coefficient matrix C is a circular matrix. Let $c(x) = \sum_{i=0}^{p-1} c_i x^i$. Because of the fact that $|I| = 2\left\lfloor \frac{p-1}{4} \right\rfloor + 1$ and $c(1) = 2\left\lfloor \frac{p-1}{4} \right\rfloor + 1 \neq 0 \pmod 2$, then $gcd(c(x), x^p - 1) = 1$. Therefore, the matrix C is invertible and thus the system of linear equation $CX = 0$ has only zero solution. We get $D_l(\beta^n) = 0$ for all $0 \leq l \leq p-1$. This contracts with Lemma 7. Then for every $n \in Z_{p^2}^*$ and $i \in Z_p$, we have $T_i(\beta^n) \neq 0$. □

Lemma 9. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Eq(3) and $\theta \in \overline{\mathbb{F}}_2$ be a primitive p^3-th root of unity, then for every $j \in Z_{p^3}^*$,*

$$s(\theta^j) \neq 0.$$

Proof. For each $j \in Z_{p^3}^*$, by Lemma 3 we have

$$s(\theta^j) = T_0(\theta^{jp}) + \theta T_1(\theta^{jp}) + \cdots + \theta^{p-1} T_{p-1}(\theta^{jp}).$$

Because θ^{jp} is p^2-primitive unit root, by Lemma 8, $T_i(\theta^{jp}) \neq 0$. Let $\mathbb{F}_{2^d}$ and $\mathbb{F}_{2^{d_1}}$ be the splitting fields of θ and θ^p,respectively. Then $\mathbb{F}_{2^d}$ is a linear space over $\mathbb{F}_{2^{d_1}}$ with $\{1, \theta, \theta^2, \cdots, \theta^{p-1}\}$ as a set of basis. Then $\{1, \theta, \theta^2, \cdots, \theta^{p-1}\}$ is linear independent. Because of the fact that $T_i(\theta^{jp}) \neq 0$ for $i = 0, 1, 2, \cdots, p-1$. Therefore $s(\theta^j) = T_0(\theta^{jp}) + \theta T_1(\theta^{jp}) + \cdots + \theta^{p-1} T_{p-1}(\theta^{jp}) \neq 0$. □

By Lemmas 5, 6 and 9, we have the following result.

Theorem 1. *Let p be an odd prime and $2^{p-1} \neq 1 \pmod{p^2}$, $(s(u))$ be the interleaved polynomial quotient sequence defined in Equation (3). When $w = p-1$, the linear complexity of $(s(u))$ is*

$$L(s(u)) = p^3 - p.$$

3.2 Linear Complexity of the Interleaved Polynomial Sequences for $w = \frac{p-1}{2}$

In this section, we will determine the linear complexity of the interleaved sequence $(s(u))$ for $w = \frac{p-1}{2}$. For the sake of simplicity, let

$$Q_l = \{u \in D_l, \left(\frac{u}{p}\right) = 1\}, N_l = \{u \in D_l : \left(\frac{u}{p}\right) = -1\},$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol as introduced in Section 1. Let

$$Q_l(x) = \sum_{u \in Q_l} x^u \in F_2[x], \qquad N_l(x) = \sum_{u \in N_l} x^u \in F_2[x],$$

$$Q_l^p = \{u \pmod p : u \in Q_l\}, \ N_l^p = \{u \pmod p : u \in N_l\}.$$

For $i \in Z_p$, define the binary sequence $(h_i(u))$ by

$$h_i(u) = \begin{cases} 1, & \text{if } H_w(u) \in I + i, \\ 0, & \text{if } H_w(u) \notin I + i, \end{cases} \tag{4}$$

where $I+i$ is the set defined in Equation (2). Then $(t_i(u))$ can be written as

$$t_i(u) = \begin{cases} h_i(u), & if\ u \in P \cup D_0 \\ h_i(u), & if\ u \in Q_1 \cup Q_2 \cup \cdots \cup Q_{p-1} \\ h_i(u) + 1, & if\ u \in N_1 \cup N_2 \cup \cdots \cup N_{p-1}. \end{cases}$$

This implies a relationship between the generating polynomials of $(t_i(u))$ and $(h_i(u))$. That is

$$T_i(x) = \sum_{u=0}^{p^2-1} t_i(u) x^u = H_i(x) + \sum_{l=1}^{p-1} N_l(x) \in F_2[x], \tag{5}$$

where $H_i(x) = \sum_{u=0}^{p^2-1} h_i(u) x^u = \sum_{l \in I+i} D_l(x) \in F_2[x]$. In order to determine the linear complexity of $(s(u))$, we need the following Lemmas.

Lemma 10. *[6] For $0 \leq l, l' < p$, we have*

(1) $aQ_l = \begin{cases} Q_{l+l' \pmod p}, & \text{if } a \in Q_{l'}, \\ N_{l+l' \pmod p}, & \text{if } a \in N_{l'}, \end{cases}$

(2) $aN_l = \begin{cases} N_{l+l' \pmod p}, & \text{if } a \in Q_{l'}, \\ Q_{l+l' \pmod p}, & \text{if } a \in N_{l'}. \end{cases}$

(3) $N_l^p = N_0^p$, $|Q_l| = |N_l| = (p-1)/2$.

Lemma 11. *[6] Let $\beta \in \overline{\mathbb{F}}_2$ be a primitive p^2-th root of unity. For any $n \in Z_{p^2}^*$, we have*

$$\sum_{l=0}^{p-1} N_l(\beta^n) = 0.$$

Lemma 12. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3) with $w = \frac{p-1}{2}$, then for any $j \in Z_p$,*

$$s(\theta^{jp^2}) = 0.$$

Proof. By Lemma 3, we have $s(\theta^{jp^2}) = \sum_{i=0}^{p-1} \theta^{ijp^2} T_i(1)$. Then by Equation (5) and Lemma 10, we have

$$T_i(1) = h_i(1) + \sum_{l=1}^{p-1} N_l(1) = (p-1)|I| + \frac{(p-1)^2}{2} = 0.$$

So $s(\theta^{jp^2}) = \sum_{i=0}^{p-1} \theta^{ijp^2} T_i(1) = 0$. □

Lemma 13. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3) and $\theta \in \overline{\mathbb{F}}_2$ be a p^3-th root of unity, then for every $j \in Z_{p^2}^*$, we have*

$$s(\theta^{jp}) \neq 0.$$

Proof. By Lemma 3, we have $s(\theta^{jp}) = \sum_{i=0}^{p-1} \theta^{ijp} T_i(\theta^{jp^2})$. For every $i \in Z_p$,

$$T_i(\theta^{jp^2}) = h_i(\theta^{jp^2}) + \sum_{l=1}^{p-1} N_l(\theta^{jp^2}).$$

Since θ^{jp^2} is a primitive p-th root of unity, we have $\theta^{jp^2 u} = \theta^{jp^2 a}$, where $a \equiv u \pmod p$. By Lemmas 1 and 10, we have

$$h_i(\theta^{jp^2}) = \sum_{l \in I+i} \sum_{u \in D_l} \theta^{jp^2 u} = \sum_{l \in I+i} \sum_{a \in D_l^p} \theta^{jp^2 a}$$

$$= \sum_{l \in I+i} \sum_{a \in \mathbb{Z}_p^*} \theta^{jp^2 a} = |I|,$$

and

$$\sum_{l=1}^{p-1} (N_l(\theta^{jp^2})) = \sum_{l=1}^{p-1} \sum_{u \in N_l} \theta^{jp^2 u} = \sum_{l=1}^{p-1} \sum_{a \in N_l^p} \theta^{jp^2 a}$$

$$= \sum_{l=1}^{p-1} \sum_{a \in N_0} \theta^{jp^2 a} = (p-1) \sum_{a \in N_0} \theta^{jp^2 a} = 0.$$

So $s(\theta^{jp}) = |I| = 2\left\lfloor \frac{p-1}{4} \right\rfloor + 1 \neq 0 \pmod 2$.

□

Lemma 14. *Let $s(x)$ be the generating polynomial of the interleaved sequence $(s(u))$ defined in Equation (3) and $\theta \in \overline{\mathbb{F}}_2$ be a primitive p^3-th root of unity, then for every $j \in Z_{p^3}^*$, we have*

$$s(\theta^j) \neq 0.$$

Proof. Let θ be a primitive p^3-th root of unity, then θ^{jp} is a primitive p^2-th root of unity for every $j \in Z_{p^3}^*$. Denote θ^p by β, then for every $j \in Z_{p^3}^*$, there always exists $n \in Z_{p^2}^*$, such that $\theta^{jp} = \beta^n$. Next we will show that for any $n \in Z_{p^2}^*$, $(T_0(\beta^n), T_1(\beta^n), \cdots, T_{p-1}(\beta^n))$ can not be zero vector.

Suppose that there exists $n_0 \in \mathbb{Z}_{p^2}^*$ such that $T_i(\beta^{n_0}) = 0$ for all $i \in Z_p$. For the reason that $2^{p-1} \neq 1 \pmod{p^2}$, then $2 \in Q_{\tau_0}$ or $2 \in N_{\tau_0}$ for some $1 \leq \tau_0 < p$. We will discuss $(T_0(\beta^n), T_1(\beta^n), \cdots, T_{p-1}(\beta^n))$ according to these two conditions, respectively. And in order to discuss conveniently, we list the following fact first.

For any $n \in \mathbb{Z}_{p^2}^*$ and every $i \in Z_p$, by Equation (5) and Lemma 11, we have

$$T_i(\beta^n) = h_i(\beta^n) + N_0(\beta^n). \tag{6}$$

If $n \in D_\tau$, where $0 \leq \tau < p$, we have

$$h_i(\beta^n) = \sum_{l \in I+i} \sum_{u \in D_l} \beta^{un} = \sum_{l \in I+i} \sum_{u \in D_{l+\tau}} \beta^u = h_{i+\tau}(\beta). \tag{7}$$

By Lemma 10, we have

$$N_0(\beta^n) = \begin{cases} N_\tau(\beta), & \text{if } n \in Q_\tau, \\ Q_\tau(\beta), & \text{if } n \in N_\tau. \end{cases} \tag{8}$$

Together with Equations (6), (7), and (8), we have

$$T_i(\beta^n) = \begin{cases} h_{i+\tau}(\beta) + N_\tau(\beta), & \text{if } n \in Q_\tau, \\ h_{i+\tau}(\beta) + Q_\tau(\beta), & \text{if } n \in N_\tau. \end{cases} \tag{9}$$

(1) $2 \in Q_{\tau_0}$

By Equation (6), we have $T_i(\beta^{n_0}) = h_i(\beta^{n_0}) + N_0(\beta^{n_0}) = 0$ for $i \in Z_p$. Therefore, $\sum_{i=0}^{p-1} h_i(\beta^{n_0}) + pN_0(\beta^{n_0}) = |I| \sum_{l=0}^{p-1} D_l(\beta^{n_0}) + N_0(\beta^{n_0}) = 0$. By Lemma 2, $\sum_{l=0}^{p-1} D_l(\beta^{n_0}) = 0$, so $N_0(\beta^{n_0}) = 0$. Then $h_i(\beta^{n_0}) = 0$ for $i \in Z_p$. By the proof of Lemma 8, we have $D_i(\beta^{n_0}) = 0$. This contracts with Lemma 7. So for any $n \in Z_{p^2}^*$

$$(T_0(\beta^n), T_1(\beta^n), \cdots, T_{p-1}(\beta^n)) \neq 0.$$

(2) $2 \in N_{\tau_0}$

By Lemma 10, we get

$$2^2 \in Q_{2\tau_0}, 2^3 \in N_{3\tau_0}, \cdots, 2^p \in N_0.$$

If $n_0 \in N_{i_0}$, then by Equation (9), we have $T_i(\beta^n) = 0$ for all $n \in N_{i_0}$. And the equation $T_i(\beta^{n_0})^{2^p} = T_i(\beta^{n_0 2^p}) = 0$ always holds. By Lemma 10 again, $n_0 2^p \in Q_{i_0}$. Then by Equation (9), $T_i(\beta^n) = 0$ holds for all $n \in Q_{i_0}$. Therefore $T_i(\beta^n) = 0$ for all $n \in D_{i_0}$.

If $n_0 \in Q_{i_0}$, then $n_0 2^p \in N_{i_0}$. Using similar method as discussed above, we can also get the result that $T_i(\beta^n) = 0$ for $n \in D_{i_0}$ holds when $n_0 \in Q_{i_0}$.

Therefore, $T_i(\beta^n) = 0$ always holds for any $n \in D_{i_0}$. By Equation (9), we get $N_{i_0}(\beta) = Q_{i_0}(\beta)$ and hence $D_{i_0}(\beta) = 0$ which contradicts to Lemma 7. Therefore, for any $n \in \mathbb{Z}_{p^2}^*$, $(T_0(\beta^n), T_1(\beta^n), \cdots, T_{p-1}(\beta^n)) \neq 0$.

By the proof of Lemma 8, $\{1, \theta, \theta^2, \cdots, \theta^{p-1}\}$ is linear independent. Therefore

$$s(\theta^j) = T_0(\beta^n) + \theta T_1(\beta^n) + \cdots + \theta^{p-1} T_{p-1}(\beta^n) \neq 0.$$

□

By Lemmas 12-14, we have the following result.

Theorem 2. *Let $(s(u))$ be the p^3-periodic interleaved binary sequence defined in Equation (3) with $w = \frac{p-1}{2}$. Assume that $2^{p-1} \neq 1 \pmod{p^2}$, then*

$$L((s_u)) = p^3 - p.$$

4 Conclusion

We have defined a new class of p^3-th periodic interleaved sequences using polynomial quotient modulo an prime p. Furthermore, we analyze their linear complexity for the cases of $w = p-1$ and $w = \frac{p-1}{2}$, respectively. The results show that when $2^{p-1} \not\equiv 1 \pmod{p^2}$, the linear complexity is $p^3 - p$ which is large enough to resist BM algorithm attacks. These interleaved sequences have longer period than the one constructed in [14] and better balance character than the one mentioned in [7] for $r = 2$.

Acknowledgments

The work is financially supported by the Key Laboratory of Applied Mathematics of Fujian Province University (Putian University)(No.SX201806), the National Natural Science Foundation of China (No. 61902429, No.11775306), the Fundamental Research Funds for the Central Universities (No. 19CX02058A), Shandong Provincial Natural Science Foundation of China (ZR2019MF070), the Open Research Fund from Shandong provincial Key Laboratory of Computer Networks, Grant No. SDKLCN-2018-02, This work was supported by Fundamental Research Funds for the Central Universities (No. ZD2019-183-008), the Major Scientific and Technological Projects of CNPC under Grant ZD2019-18 (No. ZD2019-183-001)

References

[1] Z. Chen, "Linear complexity of legendre-polynomial quotients," *IET Information Security*, vol. 12, no. 5, pp. 414–418, 2018.

[2] Z. Chen, "Trace representation and linear complexity of binary sequences derived from fermat quotients.," *Science China Information Sciences*, vol. 57, pp. 112101–112109, 2014.

[3] Z. Chen, X. Du, A. Klapper, "Linear complexity of pseudorandom sequences generated by fermat quotients and their generalizations," *Information Processing Letters*, vol. 112, pp. 233–237, 2012.

[4] X. Du, Z. Chen, "Linear complexity of some binary sequences derived from fermat quotients," *China Communications*, vol. 9, pp. 105–108, 2012.

[5] X. Du, Z. Chen, "On the linear complexity of binary threshold sequences derived from fermat quotients.," *Design, Codes and Cryptography.*, vol. 67, pp. 317–323, 2013.

[6] D. Gómez, Z.Chen, "Linear complexity of binary sequences derived from polynomial quotients," in *Sequences and Their Applications (SETA'12)*, pp. 181–189, June 2012.

[7] L. Hu, X. Du, Z. Chen, "Linear complexity of binary sequences derived from euler quotients with prime-power modulus," *Information Processing Letters*, vol. 112, pp. 604–609, 2012.

[8] Q. Niu, C. Zhao, T. Yan, "Linear complexity of the balanced polynomial quotients sequences," in *The 3rd International Conference on Circuits and Systems*, pp. 573–578, Sep. 2018.

[9] I. E. Shparlinski, A. Ostafe, "Pseudorandomness and dynamics of fermat quotients," *SIAM Journal on Discrete Mathematics (SIDMA'11)*, vol. 1, no. 25, pp. 50–71, 2011.

[10] H. Tang, X. Zeng, H. Cai, "Optimal frequency hopping sequences of odd length," *IEEE Transactions on Informations on Information Theory*, vol. 59, pp. 3237–3248, 2013.

[11] A. Winterhof, Z. Chen, A. Ostafe, "Structure of pseudorandom numbers derived from fermat quotients," in *Proceedings of the 3rd International Conference on Arithmetic of Finite Fields*, pp. 73–85, 2010.

[12] A. Winterhof, D. Gomez, "Multiplicative character sums of fermat quotients and pseudorandom sequences," *Periodica Mathematica Hungarica*, vol. 64, pp. 161–168, 2012.

[13] C. Wu, Z. Chen, V. Edemskiy, P. Ke, "On k-error linear complexity of pseudorandom binary sequences derived from euler quotients," *Advances in Mathematics of Communications*, vol. 12, no. 4, pp. 805–816, 2018.

[14] T. Yan, Y. Sun, C. Zhao, W. Ma, "Linear complexity of least significant bit of polynomial quotients," *Chinese Journal of Electronics*, vol. 26, pp. 573–578, 2017.

Biography

Chun-e Zhao was born in Shandong Province, China, in 1981. She received the Ph.D. degree at Xidian university in 2015. She is now a lecture of China University of Petroleum. Her research interests include cryptogrophy coding and sequence design. (Email: zhaochune1981@163.com).

Tongjiang Yan was born in Shandong province, China, in 1973. He received the Ph.D. degree at Xidian University in 2007. He is now a professor of China University of Petroleum. His research interests include cryptography and algebra.

Xubo Zhao was born in Shandong province, China, in 1979. She received the Ph.D. degree at Xidian University in 2013. She is now a lecture of China University of Petroleum. Her interests include Coding.

Qihua Niu was born in Shandong province, China, in 1979. He received the Ph.D. degree at Chinese Academy of Science in 2015. He is now a lecture of China University of Petroleum. His interests include protocol.

Expressive Ciphertext Policy Attribute-based Searchable Encryption for Medical Records in Cloud

Qing Wu[1], Xujin Ma[1], Leyou Zhang[2], and Yanru Chen[3]
(Corresponding author: Xujin Ma)
School of Automation, Xi'an University of Posts and Telecommunications[1]
Xi'an, Shaanxi 710121, China
School of Mathematics and Statistics, Xidian University[2]
Xi'an, Shaanxi 710071, China
School of Humanity and Foreign Languages, Xi'an University of Posts and Telecommunications[3]
Xi'an, Shaanxi 710121, China
Email: mxj419@126.com
(Received Dec. 24, 2019; Revised and Accepted July 23, 2020; First Online Apr. 11, 2021)

Abstract

As the medical technique develops, the sharing of electronic medical records has become convenient and economical. But the encrypted electronic medical records (EMRs) are difficult to be popularized in practical applications because of the massive data storage and privacy leakage. It is important to adopt a method of searchable encryption with some access policy. Attribute-based searchable encryption is considered secure and practical and has been a hot topic recently. However, the current works show some shortcomings, such as high computational costs, restricted access structure, *etc.* Additionally, access policy is related to sensitive information, which may enhance the chances of privacy leakage. A new method, ciphertext policy attribute-based searchable encryption, is proposed in this paper to eliminate these disadvantages. The new scheme supports large attributes universe and multi-keyword search in the sharing of electronic medical records. It also achieves low computational costs because the number of public keys and master keys is constant. Furthermore, the proposed scheme can support a hiding policy that can protect data owners and users' privacy.

Keywords: Attribute-based Searchable Encryption; Cloud Storage; Electronic Medical Record; Hiding Policy

1 Introduction

With the development of the Internet and the continuous deepening of informational medical treatment, electronic medical records have been used in many hospitals or institutions. In the current situation of medical associations, it is inevitably required to establish interconnections between the patient information. In order to save computing and storage resources, some hospitals or medical institutions store electronic medical records in the cloud, which not only reduces storage costs but also facilitates management. However, as the main carrier of medical information, the electronic medical records cover a lot of valuable information, such as the patient's identity, contact, health information, treatment information and so on. Once the information are leaked out to criminals, patients and hospitals will suffer a lot [29], which also destroy the trust between doctors and patients. Therefore, how to protect the security and privacy of electronic medical records from various links is an important part of hospital information construction.

Encryption technology is an effective method to protect sensitive information of patients' electronic medical records in cloud servers. After records are encrypted, only the person with the correct key can use the data. Even if it is stolen by the network hacker, the information cannot be decrypted. However, the encrypted storage of electronic medical records is difficult to be popularized in practical applications. One of the reasons is that there is a contradiction between the confidentiality and availability of data [32], which makes it impossible to retrieve encrypted data. In order to prevent unauthorized users accessing data, it is important to adopt searchable encryption with access policy and privacy protection mechanism. With this technique, each patient has a medical record with the same attributes, such as ID number, name, gender, age, and diagnostic information. These attributes are multi-valued numeric or non-numeric attributes. An attribute contains a keyword field and the attribute value is a keyword value. Medical data information is stored in form of ciphertext in the database of all field attribute domains, which protects the privacy of users and data security. The medical information including the ID number can be searched according to a specified key attribute

or attribute-related access structure, and transmitted to ciphertext, and the user decrypts locally to prevent the information from being stolen in the process of transmission.

Although there are many encryption technologies, like symmetric encryption, asymmetric encryption,*etc.*, these schemes possess different advantages and disadvantages. In this paper, an expressive ciphertext policy attribute-based searchable encryption (CPABSE) is proposed, which can solve key escrow and authenticate identity, and it can efficiently conduct multi-keyword search. Our structure is based on prime order groups and on a large universe, which will improve the computing performance obviously.

1.1 Related Work

In 2005, Sahai and Waters *et al.* [18] proposed first Attribute-Based Encryption (ABE). In this system, the encryptor does not need to know the specific identity information of the decrypter, but only needs to obtain the attributes of the decrypter's series of descriptions. In 2006, Goyal *et al.* [8] divided the Attribute-Based Encryption (ABE) into a ciphertext policy Attribute-Based Encryption (CP-ABE) and a key policy Attribute-Based Encryption (KP-ABE). The CP-ABE [27] refers to the combination of ciphertext and an access structure, and the key corresponds to a set of attributes, which can be decrypted if and only if the attributes can satisfy the access structure. While KP-ABE means that the key corresponds to an access structure, and the ciphertext is combined with a set of attributes, and can be decrypted if and only if the set of the attribute ciphertext satisfies the key of the access structure. Considering that CP-ABE can formulate flexible access policies and is more suitable for access control, thus this paper focuses on ciphertext policy attribute-based encryption (CP-ABE).

In the existing works, searchable encryption schemes generally fall into two types, that is symmetric key searchable encryption, and public key searchable encryption. Song *et al.* [22] proposed the first scheme to achieve searchability in symmetric encryption, and Boneh *et al.* [2] proposed the scheme called public key encryption with keyword search (PEKS). For networks with too many users, public key searchable encryption is more adaptive than symmetric key searchable encryption. In order to realize the retrieval of multiple keywords, Golle *et al.* [7] first proposed the concept of public-key encryption with conjunctive keyword search (PECKS) and established a security model. Subsequently, some improvements were proposed in the literature [12,14,15,20,23,33] to improve the efficiency of the algorithm and the privacy protection of users. Baek *et al.* [1] proposed a new public key searchable encryption scheme in 2008, removing the assumption of a secure channel. Later, Byun *et al.* [3] pointed out that the public key encryption scheme proposed by Boneh *et al.* [2] suffered from offline keyword guessing attacks. Yau *et al.* [30] also illustrated stealing into the common channel. In 2013, Xu *et al.* [28] proposed a public key fuzzy keyword search scheme. Each keyword corresponds to a precise and fuzzy keyword search trapdoor, and two or more keywords share a fuzzy keyword. Trapdoors have achieved resistance to internal keyword guessing attacks. Wang *et al.* [24] proposed a novel ciphertext-policy attribute-based encryption with equality test cryptosystem. It shows that the cloud server is unable to obtain any knowledge of the message encrypted under either access policy during the delegated equivalence test. In 2017, Huang *et al.* [9] introduced the concept of keyword search for public key authentication encryption. By encrypting and authenticating keywords, the sender achieved resistance against internal keyword guessing attacks. In 2019, Liu *et al.* [13] introduced a novel concept of searchable attribute-based authenticated encryption (SAAE), which can achieve expressive fine-grained access control, efficient data retrieval and authentication, simultaneously.

Access control policy can be used to protect resources from unauthorized access to specific information or unauthorized access to over-privileged data. In [9], a ciphertext policy attribute-based encryption (CP-ABE) is used to design a hybrid cloud re-encryption (HCRE) to implement ciphertext access control. However, the program does not support searching the ciphertext keywords and consumes a lot of resources when revoking permissions in a large-universe system. In [31], the scheme is called searchable ciphertext-policy attribute-based encryption with multi-keywords. In this scenario, CP-ABE and keywords are grouped together in a way that keywords are treated as file attributes. To solve the problems in cloud storage, the access structure is hidden so that the receiver cannot get sensitive information from the ciphertext. In [21], authors propose a P3 structure and an effective privacy protection phrase search scheme for intelligent encrypted data processing in cloud-based Internet of Things. The scheme utilizes homomorphic encryption and bilinear map to determine the positional relationship of multiple query keywords on the encrypted data. A probabilistic trapdoor generation algorithm is also used to protect the user's search patterns. Cui [5] put forth a key policy attribute-based encryption (KP-ABE) based on the prime-order groups in the cloud, which is effective and has expressiveness implements keyword searchable data encryption. Feng [6] introduced the PEKS into the multiple authority CP-ABE cloud storage schemes which supports the direct revocation of users. The access control of users is achieved by the central authority, which avoids the security risks caused by submitting the private keys and access structure to cloud server. However the complexity of the encryption and decryption algorithm increases with the increasing number of attributes and the efficiency of search mechanism needs to be improved.

1.2 Main Contributions

Our solution is based on ciphertext policy attribute-based encryption (CP-ABE), which supports hidden access pol-

icy [17]. So the proposed scheme realizes privacy protection. In addition, the proposed scheme is based on the large-universe which leads to the effective and expressive keyword search. The scheme is based on the prime order groups, which is much more efficient than the composite order group.

The keyword information retrieved by the data is submitted in the form of a trapdoor. The key value related index information exists in ciphertext, and the server cannot obtain any related keyword information. A group of users does not have access to private data, and legal identity can only access data correctly, which is a good resistance to collusion attacks. Moreover, the proposed scheme realizes multi-keyword search, which improves the search efficiency and meets the search demands for electronic medical records.

In addition, the proposed CPABSE scheme is proved to be chosen-plaintext attack security in the standard model. Under the decisional (q-1) assumption, the proposed scheme achieves selective indistinguishability security against chosen keyword-set attack (IND-CKA). Performance comparisons and experimental analyses show the proposed scheme is efficient and practical.

1.3 Organization

The main structure of this paper is organized as follows. In Section 2, we mainly introduce the basic knowledge and the complexity assumption used in this paper. Section 3 gives system framework and security model. The details of the proposed scheme are arranged in Section 4. The security of the proposed scheme is proved in Section 5. In Section 6, the comparison and experimental analysis with other structures are given. Section 7 is the conclusion.

2 Preliminaries

In this section, we will introduce the notions and definitions adopted in this paper.

2.1 Bilinear Map

Definition 1. *(Bilinear Maps). Let G be the cyclic group of prime order p whose generator is g, and G_1 is the multiplicative cyclic group with the same order. It is assumed that the discrete logarithm problem in G and G_1 is a difficult problem. Bilinear mapping $e: G \times G \to G_1$ that satisfies the following properties.*

1) *Bilinearity: For all $g \in G$ and $a, b \in \mathbb{Z}_p$, it has $e(g^a, g^b) = e(g, g)^{ab}$,*

2) *Non-degeneracy: There exists $e(g, g) \neq 1$, where $g \in G$.*

3) *Computability: For any $g \in G$, there exists an efficient algorithm to compute $e(g, g)$.*

2.2 Access Structures and Linear Secret-Sharing Scheme

Definition 2. *(Access Structure). The access structure defines the concept of an authorized access subset and an unauthorized access subset, which is a description of the access control policy. Let $U = \{A_1, A_2, \cdots, A_n\}$ is a set of attribute universe. An access structure on U is a collection $\mathbb{A} \subseteq 2^U$, which is monotone if $\forall B, C \in \mathbb{A}$: if $B \in \mathbb{A}$ and $B \subseteq C$, then $C \subseteq \mathbb{A}$. The structure on U is non-empty set,* i.e., *$\mathbb{A} \subseteq 2^U \setminus \{\emptyset\}$. The authorized set is the set in $\mathbb{A}$, the set not in $\mathbb{A}$ is called unauthorized set. In our scheme, non monotonic access structures are beyond our scope. Hence we only consider monotonic structures.*

Definition 3. *(Linear Secret-Sharing Scheme (LSSS)). Suppose there is a sharing scheme Π, including attributes. The scheme Π meets the following conditions, which is called a linear secret-sharing scheme if*

1) *The vector in the domain $\mathbb{Z}_p$ can represent the secret share owned by each attributes.*

2) *For each shared scheme Π there is a shared generation matrix $\boldsymbol{M}$ of l rows and n columns, that is $\boldsymbol{M} \in \mathbb{Z}_p^{l \times n}$. And $r(i)$ indicates that the ith row of the matrix is mapped to an entity, where $i = 1, 2, \cdots, l$. Considering a column vector $\boldsymbol{v} = (\varphi, y_1, \cdots, y_n)$, where φ is the secret to be shared, $y_1, \cdots, y_n \in \mathbb{Z}_p$ are chosen randomly, then the secret possessed by the entity user is defined as $\lambda_i = (\boldsymbol{M_i v})$. The share $(\boldsymbol{M_i v})$ "belongs" to the attributes $r(i)$.*

The linear secret-sharing scheme satisfies the linear reconstruction property. There is an LSSS scheme Π. U is a legal authorization set, and define $I \subseteq \{1, \cdots, l\}$ as $I = \{i : r(i) \in U\}$, that is the row related to the attribute in $\boldsymbol{M}$. Suppose λ_i can effectively share φ, then we have a constant set $\{w_i\}$ which satisfy the equation $\sum_{i \in I} w_i v_i = \varphi$.

2.3 Complexity Assumptions

Definition 4. *(Decisional (q-1) assumption). Let G and G_1 be cyclic groups of the order prime p with the generator g and a bilinear map $e: G \times G \to G_1$. $a, \varphi, b_1, \cdots, b_q \in \mathbb{Z}_p$ are chosen randomly. We have $\boldsymbol{y} = \{g, g^{\varphi}, g^a, \cdots, g^{a^q}, g^{a^{q+1}}, \cdots, g^{a^{2q}}, \forall_{1 \leqslant j \leqslant q} g^{\varphi \cdot b_j}, g^{a/b_j}, \cdots, g^{a^q/b_j}, g^{a^{q+1}/b_j}, \cdots, g^{a^{2q}/b_j} \forall_{1 \leqslant j \leqslant q, k \neq j} g^{a \cdot \varphi \cdot b_k/b_j}, \cdots, g^{a^q \cdot \varphi \cdot b_k/b_j}\}$. It is difficult to distinguish $\left\{(\boldsymbol{y}, e(g, g)^{a^{q+1}\varphi})\right\}$ from $\{(\boldsymbol{y}, Z)\}$, where $Z \in G$ is randomly chosen.*

Definition 5. *(Generic Group Model). The definition [19] follows here: we consider two random encodings $\varphi_0, \varphi_1 : \mathbb{Z}_p \to \{0,1\}^m$, where $m > 3\log(p)$. For $i = 0, 1$, we let $G_i = \{\varphi_i(x) | x \in \mathbb{Z}_p\}$. It is given the oracles to calculate the induced group action on G, G_1 and compute a non-degenerate bilinear map $e: G \times G \to G_1$. It is also given a random oracle to represent the hash function H.*

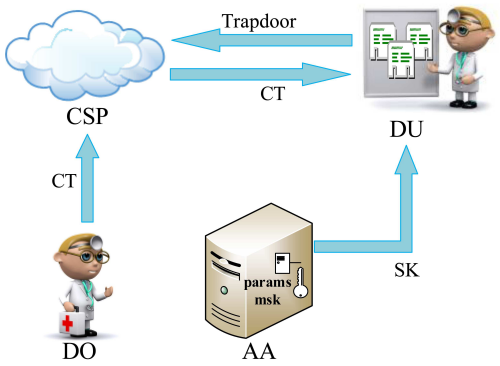


Figure 1: The proposed scheme

3 System Framework and Security Model

In this section, we will provide the system framework of CPABSE scheme, and give the security model of the scheme.

3.1 Scheme Model

As depicted in Figure 1, our CPABSE scheme includes four entities, Attributes Authority (AA), Data Owner (DO), Data User (DU), and Cloud Provide Server (CPS), which are showed as below.

Attributes Authority (AA) performs initialization processing and user authorization. AA is assumed to be in charge of generating and distributing the public parameters and master secret key.

Data Owner (DO) collects the EMRs information (*e.g.* patient's personal information, examination records, treatment records, care records,*etc.*). It is responsible for encrypting the original massage and keywords in order to protect the data privacy. It executes encryption algorithm, including massage encryption and keywords encryption, and generates massage ciphertext and index ciphertext. Then it uploads the massage ciphertext and index ciphertext to CPS.

Data User (DU). A data user, for example a doctor or a medical staff, needs to search an EMR. Firstly a search user receives a secret key from AA. Next, according to the search keywords, the search user generates a trapdoor and sends it to CSP. Finally, if the searching is successful, CSP sends the results to the search user, and he/she decrypts the massage and retrieves the original massage.

Cloud Server Provider(CSP) has a large amount of storage spaces. It stores the encrypted massage and index ciphertext which gets from DO. CSP performs the search operation for DO. It would check if the keywords in the trapdoor match the keywords in the index, then it returns the corresponding results.

3.2 Threat and Adversary Model

In CPABSE, AA and DO are completely trusted by third entities. However, we assume that CSP is honest-but-curious. It can only honestly implement the algorithm according to the protocol. But at the same time, it is curious to analyze and may guess an extra sensitive information. Besides, it requires that the malicious DU sohuld not collude with the CSP.

Then we propose the following adversary model and Note that the adversary can be a malicious DU. As for data security, adversaries can eavesdrop on encrypted EMRs transmitted over public channels and attempt to access unauthorized EMRs. Regarding to attributes privacy protection, the adversary aims to extract information about sensitive attribute values from the encrypted EMRs.

3.3 Scheme Definitions

A CPABSE for a general access structure over the monotone attribute universe space is composed of six PPT algorithms, Setup, KeyGen, Encrypt, TrapdoorGen, Search, and Decrypt. The details of these algorithm are as follows:

Setup$((1^\kappa) \rightarrow (params, msk))$**:** The setup algorithm is carried out by AA. It takes a security parameter κ and attributes universal U as input and outputs the system public parameters *params* and the master secret key *msk*. AA takes the public parameters *params* for public, and keeps the master private key *msk* secretly.

KeyGen$((params, msk, U) \rightarrow SK)$**:** The key generation algorithm is executed by AA. This algorithm takes the public parameters *params*, master private key *msk* and an attribute set U as input and outputs a secret key SK.

Encrypt: The encryption algorithm is performed by DO. This phase is divided into two parts, with one for encrypting the message and the other for encrypting the keyword.

1) **Massage Encryption**$((params, \mathcal{M}, (\mathbf{M}, \mathbf{r}, \{\mathbf{A_{r(i)}}\})) \rightarrow CT')$**:** This algorithm takes the parameter *params*, a massage $\mathcal{M}$ and the access structure encoded in LSSS structure $(\mathbf{M}, r, \{A_{r(i)}\})$ as the input and outputs the partial ciphertext CT'.

2) **Keyword Encryption**$((params, W) \rightarrow Index)$**:** The algorithm inputs public parameter *params*, keyword set W, in which it contains m keywords. Then, it outputs the index *Index*. Subsequently, the algorithm outputs the ciphertext $CT = \{CT', Index\}$.

TrapdoorGen$((params, w_{j'}, SK) \rightarrow TK)$**:** The trapdoor generation algorithm is carried out by DU.

Inputting public parameter *params*, the ciphertext CT and the keyword set $w_{j'}$ which the user wants to be queried, it outputs the trapdoor key TK and send it to CSP.

Search$((Index, TK) \rightarrow (0\ or\ 1))$**:** The CSP performs the search algorithm. It takes the the index $Index$ and the trapdoor key TK as input, if the keywords in the trapdoor match the keywords in the index successfully, the algorithm outputs 1, otherwise outputs 0.

Decrypt$((params, CT', SK) \rightarrow \mathcal{M})$**:** The decryption algorithm is performed by DU. It inputs the public parameters *params* and ciphertext CT' related to an access structure $(\mathbf{M}, r)$ and a private key SK, and outputs $\mathcal{M}$ if and only if the attribute set U satisfies the monotone access structure.

3.4 Security Model

We define security game for expressive CPABSE in the sence of semantic-security.

First we consider the indistinguishable keyword index against the adaptive chosen keyword attacks. In this game, the adversary can get the keywords set in the trapdoor which he wants to inquire. But there is a restriction that he cannot distinguish the encrypted cipertext of the keywords set between W_0 and W_1. The security game of adversary $\mathcal{A}$ and challenge $\mathcal{C}$ as follows:

1) Challenger $\mathcal{C}$ performs the setup algorithm to get *params* and send it to adversary $\mathcal{A}$;

2) Adversary $\mathcal{A}$ can adaptively ask the challenge $\mathcal{C}$ for the trapdoor to conduct trapdoor inquiry. Challenger $\mathcal{C}$ calculates the secret key SK and then generates a keyword set trapdoor TK and sends it to $\mathcal{A}$;

3) Adversary $\mathcal{A}$ commits two keyword sets W_0 and W_1 as the challenge keyword sets, then $\mathcal{A}$ gives a challenge access policy, and the constraint condition is that the attribute set U cannot satisfy $\mathbb{A}$. $\mathcal{C}$ selects 0 or 1 to generate the challenge index of W_c and send it to $\mathcal{A}$;

4) $\mathcal{A}$ repeats the inquiry in (2) until W_0 and W_1 cannot be queried;

5) Finally, $\mathcal{A}$ output guess c' of c, where $c' \in \{0, 1\}$.

The advantage of $\mathcal{A}$ is defined as $Adv = |\Pr[c = c'] - 1/2|$ in the above game.

Definition 6. *If the probabilities of all PPT adversary in the above game are negligible, the CPABSE scheme has an indistinguishable of keyword index against the adaptive chosen keyword attacks.*

Now we present the definitions of chosen-plaintext attack security of the scheme. If the adversary $\mathcal{A}$ submits a challenge access policy $\mathbb{A}^*$ before the setup game, it is called selective security. The security interactive game between the adversary $\mathcal{A}$ and challenger $\mathcal{C}$ as follows:

Int: The challenger $\mathcal{C}$ receives an access structure $(\mathbf{M}^*, r^*)$ from an adversary $\mathcal{A}$.

Setup: The adversary $\mathcal{A}$ runs the Setup algorithm, produces public parameter *params* and master secret key *msk*. It makes the *params* public and keeps *msk* security.

Phase 1: The adversary $\mathcal{A}$ sends the attributes set U to $\mathcal{C}$ and issues the adaptive query. The limitation of each query is that the attributes set U cannot satisfy the access policy $(\mathbf{M}^*, r^*)$. Then $\mathcal{C}$ returns the corresponding the secret key SK to $\mathcal{A}$.

Challenge: The adversary $\mathcal{A}$ submits the two equal-length massages $\mathcal{M}_0$ and $\mathcal{M}_1$ to C. In addition $\mathcal{A}$ gives the challenge access structure $(\mathbf{M}^*, r^*)$. The only construction is that the attributes set U cannot satisfy the access structure. $\mathcal{A}$ tosses a random coin and forwards a guess c' for c, where $c \in \{0, 1\}$ and performS the Massage Encryption$((params, m_c, (\mathbf{M}^*, r^*)))$ algorithm to get the ciphertext CT', then sends it to $\mathcal{A}$.

Phase 2: $\mathcal{A}$ continues the query of Phase 1, but the restriction is that none of the attributes set satisfies the challenge access policy.

Guess: $\mathcal{A}$ outputs a guess c', If $c' = c$, it outputs 1. That means $\mathcal{A}$ wins the game.

The advantage of the adversary is defined as $Adv = |\Pr[c = c'] - 1/2|$ in this game.

Definition 7. *If any PPT adversary wins the above game at most with negligible adversary, our CPABSE scheme proves to be chosen-plaintext attack security.*

4 Expressive Ciphertext Policy Attribute-Based Searchable Encryption

4.1 Construction

Let G and G_1 be groups of the prime order p with the generator g and a bilinear map $e : G \times G \rightarrow G_1$. Based on the above basic preliminaries and formal structure, our algorithm in prime order groups is as follows:

Setup$((1^\kappa) \rightarrow (params, msk))$**:** Take the security parameter 1^κ as input. The algorithm chooses $u, \beta, \omega, y \in G$, $\alpha, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{Z}_p$ randomly, and computes $g_1 = g^{\lambda_1}$, $g_2 = g^{\lambda_2}$, $g_3 = g^{\lambda_3}$, $g_4 = g^{\lambda_4}$. Suppose $H : \{0, 1\} \rightarrow \mathbb{Z}_p$ is a one way

hash function. Finally, it outputs public parameters $params = (g, u, \beta, \omega, y, g_1, g_2, g_3, g_4, H, e(g,g)^{\alpha})$ for public, and keeps the master private key $msk = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, g^{\alpha})$ security.

KeyGen($(params, msk, U) \rightarrow SK$): This algorithm takes the public parameters $params$, master private key msk and an attribute set U as input, where $U = \{A_1, A_2, \cdots, A_k\}$, and the size of U is k, and let $A_1, A_2, \cdots, A_k \subseteq \mathbb{Z}_p$ be the attribute value. It randomly chooses $\tau, \tau' \subseteq \mathbb{Z}_p$, and computes $K = g^{y}$, $K_1 = g^{\alpha}\omega^{\lambda_1\lambda_2\tau+\lambda_3\lambda_4\tau'}$, $K_2 = g^{\tau\lambda_1\lambda_2+\tau'\lambda_3\lambda_4}$, $K_{i,1} = ((u^{A_i}\beta)^{\tau})^{-\lambda_2}$, $K_{i,2} = ((u^{A_i}\beta)^{\tau})^{-\lambda_1}$, $K_{i,3} = ((u^{A_i}\beta)^{-\tau'})^{\lambda_4}$, $K_{i,4} = ((u^{A_i}\beta)^{-\tau'})^{\lambda_3}$. Then the algorithm outputs the secret key $SK = \left\{K, K_1, K_2, \{K_{i,1}, K_{i,2}, K_{i,3}, K_{i,4}\}_{i\in[1,k]}\right\}$.

Encrypt: The encryption phase is divided into two parts, one for encrypting the message and the other for encrypting the keyword.

1) **Massage Encryption($(params, \mathcal{M}, (\mathbf{M}, r, \{A_{r(i)}\})) \rightarrow CT'$):]** This algorithm takes the parameter $params$, a massage $\mathcal{M} \in \mathbb{Z}_p$ and the access structure encoded in LSSS structure $(\mathbf{M}, r, \{A_{r(i)}\})$ as the input, among which the size of the sharing matrix $\mathbf{M}$ is $l \times n$, and r is a map that relates each row of matrix $\mathbf{M}$ to an attribute, $r(i)$ is an attribute value. It chooses a vector $\mathbf{v} = (\varphi, y_2, \cdots, y_n) \in \mathbb{Z}_p^n$, where φ is the random secret exponent to be shared among the shares, and $y_2, \cdots, y_n$ are randomly chosen. The algorithm defines $I = \{i : r(i) \in A\}$, for $i \in I$, it calculates $v_i = \mathbf{M}_i \cdot \mathbf{v}$, where $\mathbf{M}_i$ represents the ith rows of $\mathbf{M}$. Then, it chooses $a_1, \cdots, a_l, d_{1,1}, \cdots, d_{i,1} \in \mathbb{Z}_p$ randomly, and calculates:

$$\begin{aligned} C &= \mathcal{M} \cdot e(g,g)^{\alpha\varphi}, \\ D &= g^{\varphi}, \\ C_i &= \omega^{v_i}(u^{A_{r(i)}}\beta)^{a_i}, \\ D_{i,1} &= g_1^{a_i - d_{i,1}}, \\ D_{i,2} &= g_3^{a_i - d_{i,2}}, \\ D_{i,3} &= g_2^{d_{i,1}}, \\ D_{i,4} &= g_4^{d_{i,2}}. \end{aligned}$$

Finally, it outputs a ciphertext $CT' = \left\{C, D, \{C_i, D_{i,1}, D_{i,2}, D_{i,3}, D_{i,4}\}_{i\in[1,l]}\right\}$.

2) **Keyword Encryption($(params, W) \rightarrow Index$):]** The algorithm inputs public parameter $params$, keyword set W, in which it contains m keywords. Then it picks $z, r, r_1 \in \mathbb{Z}_p$ randomly, for any keywords $\forall w_j \in W\,(j = \{1, \cdots, m\})$, computes $I_j = g^{zrH(w_j)}$, $I_1 = g^{r_1}$, $I_2 = g^{r_1 z}$, $I_3 = g_1^{r}$. Finally, it outputs the index $Index = \left\{\{I_j\}_{j\in[1,m]}, I_1, I_2, I_3\right\}$.

Subsequently, the algorithm outputs the ciphertext $CT = \{CT', Index\}$.

TrapdoorGen($(params, w_{j'}, SK) \rightarrow TK$): Based on [25], the trapdoor key is generated as follows. The user enters public parameter $params$, the secret key SK and a set of keywords $w_{j'} \in W\,(j' = 1, \cdots m')$, where $w_{j'}$ is what the user wants to be queried. For any $w_{j'}$, the user selects $f \in \mathbb{Z}_p^*$ randomly and calculates $T_1 = K^{zf}$, $T_2 = K^{f}$, $T_3 = \prod_{j'=1}^{m'} g^{H(w_{j'})z}$, $T_4 = g_1$. Then it outputs the trapdoor key $TK = \{T_1, T_2, T_3, T_4\}$ and sends it to the cloud server provider.

Search($(Index, TK) \rightarrow (0\ or\ 1)$): The cloud server provider uses trapdoor key TK and $Index$ as input and checks whether the following equation is true.

$$e(T_2, I_2)\, e\left(T_4, \prod_{k=1}^{m'} I_{j_k}\right) \stackrel{?}{=} e(T_1, I_1)\, e(T_3, I_3).$$

The above formula checks whether the keyword in the index matches the keyword in the trapdoor. In this process, the algorithm encrypts m keywords to generate the index, and the user generates m' trapdoor keywords to be queried, $m' \leq m$, and selects m' keywords from m keywords for a total of $C_m^{m'} = \frac{m\times(m-1)\times\cdots\times(m-m'+1)}{m'!}$, therefore the keywords in the trapdoor match the keywords in the index at least $C_m^{m'}$ times. If there is a match, return 1, that is the equation is true, otherwise return 0.

Decrypt($(params, CT', SK) \rightarrow \mathcal{M}$): Suppose there is an access policy $(\mathbf{M}, r, \{A_{r(i)}\})$ to decrypt the ciphertext CT'. It inputs the public parameters $params$, secret key SK and CT' related to an access structure $(\mathbf{M}, r, \{A_{r(i)}\})$. If the users attribute set U is not an authority set of the access policy, the algorithm outputs $\perp$. Otherwise the attribute set meet with the access policy, then it calculates the constant $\{\omega_i \in \mathbb{Z}_p\}_{i\in I}$ such that $\sum_{i\in I}\omega_i \mathbf{M}_i = (1, 0, \cdots, 0)$, where $\mathbf{M}_i$ is the ith row of the matrix $\mathbf{M}$. It performs the following operation

$$B = \frac{e(D, K_1)}{\prod_{i\in I}[e(C_i, K_2)e(D_{i,1}, K_{i,1})e(D_{i,3}, K_{i,2})e(D_{i,2}, K_{i,3})e(D_{i,4}, K_{i,4})]^{\omega_i}}$$

Finally, the algorithm outputs the partially decrypted ciphertext B, and from $\mathcal{M} = C/B$ it can be calculated to obtain $\mathcal{M}$.

Correctness. Following equation will show that the CPABSE index can be searched correctly if it has corrected trapdoor.

$$
\begin{aligned}
& e\left(T_2, I_2\right) e\left(T_4, \prod_{k=1}^{m^{\prime}} I_{j_k}\right) \\
= \; & e\left(K^f, g^{r_1 z}\right) e\left(g_1, \prod_{k=1}^{m^{\prime}} g^{z r H\left(w_{j_k}\right)}\right) \\
= \; & e(g, g)^{y f r_1 z} e\left(g, \prod_{k=1}^{m^{\prime}} g^{H\left(w_{j_k}\right)}\right)^{z r \lambda_1}
\end{aligned}
$$

$$
\begin{aligned}
& e\left(T_1, I_1\right) e\left(T_3, I_3\right) \\
= \; & e\left(K^{z f}, g^{r_1}\right) e\left(\prod_{j^{\prime}=1}^{m^{\prime}} g^{H\left(w_{j^{\prime}}\right) z}, g_1^r\right) \\
= \; & e(g, g)^{y z f r_1} e\left(\prod_{j^{\prime}=1}^{m^{\prime}} g^{H\left(w_{j^{\prime}}\right)}, g\right)^{z r \lambda_1}
\end{aligned}
$$

$$
\begin{aligned}
& \prod_{i \in I}[e(C_i, K_2) e(D_{i,1}, K_{i,1}) e(D_{i,3}, K_{i,2}) e(D_{i,2}, K_{i,3}) \\
& \qquad \cdot e(D_{i,4}, K_{i,4})]^{\omega_i} \\
= \; & \prod_{i \in I} e\left(\omega^{v_i}\left(u^{A_{r(i)}} \beta\right)^{a_i}, g^{\tau \lambda_1 \lambda_2 + \tau^{\prime} \lambda_3 \lambda_4}\right)^{\omega_i} \\
& \cdot e\left(g_1^{a_i - d_{i,1}},\left(\left(u^{A_{r(i)}} \beta\right)^{\tau}\right)^{-\lambda_2}\right)^{\omega_i} \\
& \cdot e\left(g_2^{d_{i,1}},\left(\left(u^{A_{r(i)}} \beta\right)^{\tau}\right)^{-\lambda_1}\right)^{\omega_i} \\
& \cdot e\left(g_3^{a_i - d_{i,2}},\left(\left(u^{A_{r(i)}} \beta\right)^{-\tau^{\prime}}\right)^{\lambda_4}\right)^{\omega_i} \\
& \cdot e\left(g_4^{d_{i,2}},\left(\left(u^{A_{r(i)}} \beta\right)^{-\tau^{\prime}}\right)^{\lambda_3}\right)^{\omega_i} \\
= \; & e(g, \omega)^{\left(\tau \lambda_1 \lambda_2 + \tau^{\prime} \lambda_3 \lambda_4\right) \sum\limits_{i \in I} \omega_i v_i} \\
B = \; & \frac{e\left(g^{\varphi}, g^{\alpha} \omega^{\lambda_1 \lambda_2 \tau + \lambda_3 \lambda_4 \tau^{\prime}}\right)}{e(g, \omega)^{\left(\tau \lambda_1 \lambda_2 + \tau^{\prime} \lambda_3 \lambda_4\right) \sum\limits_{i \in I} \omega_i v_i}}.
\end{aligned}
$$

If the attribute authorized set has been given, we then have $\sum\limits_{i \in I} \omega_i v_i = \varphi$. Therefore

$$
\begin{aligned}
B &= \frac{e\left(g^{\varphi}, g^{\alpha} \omega^{\lambda_1 \lambda_2 \tau + \lambda_3 \lambda_4 \tau^{\prime}}\right)}{e(g, \omega)^{\left(\tau \lambda_1 \lambda_2 + \tau^{\prime} \lambda_3 \lambda_4\right) \sum\limits_{i \in I} \omega_i v_i}} \\
&= \frac{e\left(g^{\varphi}, g^{\alpha} \omega^{\lambda_1 \lambda_2 \tau + \lambda_3 \lambda_4 \tau^{\prime}}\right)}{e(g, \omega)^{\left(\tau \lambda_1 \lambda_2 + \tau^{\prime} \lambda_3 \lambda_4\right) \varphi}} \\
&= e(g, g)^{\alpha \varphi}
\end{aligned}
$$

5 Security Proof

Theorem 1. *If the decisional $(q-1)$ assumption in G holds, all probabilistic polynomial time (PPT) adversaries break the proposed scheme with a challenge matrix of size $m \times n$ in selective condition, which have a negligible advantage.*

Proof. Considering the proof, we will assume that a PPT adversary $\mathcal{A}$ with a challenge matrix that meets the constraints, which has a non-negligible advantage to break the proposed model selectively. Then we can build a PPT challenger $\mathcal{C}$ that attracks the $(q-1)$ assumption with a non-negligible advantage. □

Init: A challenger $\mathcal{C}$ receives an access structure $(\mathbf{M}^*, r^*, \{r(i)^*\})$ from an adversary $\mathcal{A}$ (for simplicity ,we use $\{r(i)^*\}$ to take the place of $\left\{A^*_{r(i)}\right\}$ in rest of the proof), where $\mathbf{M}^*$ is an $m \times n$ matrix, and $r^* : [m] \to \mathbb{Z}_p$, $m = \{1, 2, \cdots, m\}$.

Setup: The challenger picks $\widetilde{\alpha}, \widetilde{u}, \widetilde{\beta}, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{Z}_p^*$, and calculates $g_1 = g^{\lambda_1}$, $g_2 = g^{\lambda_2}$, $g_3 = g^{\lambda_3}$, $g_4 = g^{\lambda_4}$. Besides it implicitly sets $\alpha = a^{q+1} + \widetilde{\alpha}$, where a, q are set in assumption and $\widetilde{\alpha} \leftarrow \mathbb{Z}_p$ is a known to $\mathcal{C}$ random exponent. In this way, α is correctly distributed and a is hidden theoretically from $\mathcal{A}$. The rest of the public parameters are implicitly set as

$$
\begin{aligned}
\omega &= g^a, \\
u &= g^{\widetilde{u}} \cdot \prod_{j, j^{\prime} \in [m,n]} \left(g^{a^{j^{\prime}} / b_j^2}\right)^{M^*_{j,j^{\prime}}}, \\
\beta &= g^{\widetilde{\beta}} \prod_{j, j^{\prime} \in [m,n]} \left(g^{a^{j^{\prime}} / b_j^2}\right)^{-r(j)^* M^*_{j,j^{\prime}}}, \\
e(g,g)^{\alpha} &= e(g,g)^{a^{q+1}} \cdot e(g,g)^{\widetilde{\alpha}}
\end{aligned}
$$

Phase 1: The challenger $\mathcal{C}$ outputs the attributes set $U = \{A_1, A_2, \cdots, A_k\}$ of private key. Since U is non authorized for $(\mathbf{M}^*, r^*, \{r(i)^*\})$, there is available a vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_n)^{\perp} \in \mathbb{Z}_p^n$, in which $\omega_1 = -1$ and $(M_i^*, \boldsymbol{\omega}) = 0$ for all $i \in I = \left\{i | i \in [m] \wedge r(i)^* \in U\right\}$. The simulator computes $\boldsymbol{\omega}$ with linear algebra. In addition, it picks $\widetilde{\tau}, \widetilde{\tau^{\prime}} \in \mathbb{Z}_p^*$ and implicitly sets

$$
\begin{aligned}
\tau &= \widetilde{\tau} + \omega_1 a^q + \cdots + \omega_n a^{q+1-n} \\
&= \widetilde{\tau} + \sum_{i \in [n]} \omega_i a^{q+1-i}, \\
\tau^{\prime} &= \widetilde{\tau^{\prime}} + \omega_1 a^q + \cdots + \omega_n a^{q+1-n} \\
&= \widetilde{\tau^{\prime}} + \sum_{i \in [n]} \omega_i a^{q+1-i},
\end{aligned}
$$

and calculates $K = g^y$,

$$
\begin{aligned}
K_1 &= g^{\alpha} \omega^{\lambda_1 \lambda_2 \tau + \lambda_3 \lambda_4 \tau^{\prime}} \\
&= \left(g^{a^{q+1}} \cdot g^{\widetilde{\alpha}}\right)\left(g^{a \widetilde{\tau}} \prod_{i \in [n]} g^{a^{q+2-i} \omega_i}\right)^{\lambda_1 \lambda_2} \\
&\quad \cdot \left(g^{a \widetilde{\tau^{\prime}}} \prod_{i \in [n]} g^{a^{q+2-i} \omega_i}\right)^{\lambda_3 \lambda_4}
\end{aligned}
$$

$$
\begin{aligned}
K_2 &= g^{\lambda_1\lambda_2\tau+\lambda_3\lambda_4\tau'} \\
&= \left(g^{\widetilde{\tau}}\prod_{i\in[n]} g^{a^{q+1-i}\omega_i}\right)^{\lambda_1\lambda_2} \\
&\quad \cdot\left(g^{\widetilde{\tau'}}\prod_{i\in[n]} g^{a^{q+1-i}\omega_i}\right)^{\lambda_3\lambda_4} \\
K_{i,1} &= \left(\left(u^{r(i)^*}\beta\right)^{\tau}\right)^{-\lambda_2} \\
&= \left(u^{r(i)^*}\beta\right)^{-\widetilde{\tau}\lambda_2}\cdot\prod_{i\in[n]}\left(u^{r(i)^*}\beta\right)^{-\lambda_2\omega_i a^{q+1-i}} \\
K_{i,2} &= \left(\left(u^{r(i)^*}\beta\right)^{\tau}\right)^{-\lambda_1} \\
&= \left(u^{r(i)^*}\beta\right)^{-\widetilde{\tau}\lambda_1}\cdot\prod_{i\in[n]}\left(u^{r(i)^*}\beta\right)^{-\lambda_1\omega_i a^{q+1-i}} \\
K_{i,3} &= \left(\left(u^{r(i)^*}\beta\right)^{-\tau'}\right)^{\lambda_4} \\
&= \left(u^{r(i)^*}\beta\right)^{-\widetilde{\tau'}\lambda_4}\cdot\prod_{i\in[n]}\left(u^{r(i)^*}\beta\right)^{-\lambda_4\omega_i a^{q+1-i}} \\
K_{i,4} &= \left(\left(u^{r(i)^*}\beta\right)^{-\tau'}\right)^{\lambda_3} \\
&= \left(u^{r(i)^*}\beta\right)^{-\widetilde{\tau'}\lambda_3}\cdot\prod_{i\in[n]}\left(u^{r(i)^*}\beta\right)^{-\lambda_3\omega_i a^{q+1-i}}
\end{aligned}
$$

then challenger $\mathcal{C}$ can output the secret key $SK = \left\{K, K_1, K_2, \{K_{i,1}, K_{i,2}, K_{i,3}, K_{i,4}\}_{i\in[1,k]}\right\}$.

Challenge: The adversary $\mathcal{A}$ submits the two equal-length massages m_0 and m_1 to $\mathcal{C}$, and implicitly sets $a_i = -\varphi b_i$, then calculates

$$
\begin{aligned}
C^* &= \mathcal{M}_c\cdot Y\cdot e(g,g^{\varphi})^{\widetilde{\alpha}} \\
D^* &= g^{\varphi} \\
C_i^* &= \omega^{v_i}(u^{r(i)^*}\beta)^{a_i} \\
v_i &= \sum_{j\in[n]} M_i^*\cdot\varphi a^{j-1}+\widetilde{v}_i \\
C_i^* &= \omega^{v_i}\left(u^{r(i)^*}\beta\right)^{a_i} \\
&= (g^a)^{v_i}\left(g^{\varphi b_i}\right)^{-\left(\widetilde{\alpha}r(i)^*+\widetilde{\beta}\right)} \\
&\quad\cdot\prod_{\substack{i',j'\in[m,n]\\ i'\neq j'}}\left(g^{\frac{\varphi a^{j'}b_i}{b_{i'}^2}}\right)^{-\left(r(i)^*-r(i')^*\right)M^*_{j,j'}} \\
D_{i,1}^* &= g_1^{a_i-d_{i,1}} = \left(g^{-\varphi b_i}\right)^{\lambda_1}\cdot g^{-\lambda_1 d_{i,1}} \\
D_{i,2}^* &= g_3^{a_i-d_{i,2}} = \left(g^{-\varphi b_i}\right)^{\lambda_3}\cdot g^{-\lambda_3 d_{i,2}} \\
D_{i,3}^* &= g_2^{d_{i,1}} = g^{\lambda_2 d_{i,1}} \\
D_{i,4}^* &= g_4^{d_{i,2}} = g^{\lambda_4 d_{i,2}},
\end{aligned}
$$

where the term $d_{i,1}, d_{i,2} \in \mathbb{Z}_p^*$. Then $\mathcal{C}$ outputs the challenge ciphertext $CT^* = \left(C^*, D^*, \left\{\left(C_i^*, D_{i,1}^*, D_{i,2}^*, D_{i,3}^*, D_{i,4}^*\right)\right\}_{i\in[1,m]}\right)$

Phase 2: $\mathcal{A}$ continues the query of Phase 1, but the restriction is that none of the attributes set satisfies the challenge access policy.

Guess: $\mathcal{A}$ outputs a guess c' for c. If $c' = c$, it outputs 1, that is the challenge term is $Y = e(g,g)^{\varphi a^{q+1}}$. From the view of $\mathcal{A}$, if $Y = e(g,g)^{\varphi a^{q+1}}$, the simulation process represents the security game since $C = \mathcal{M}_c\cdot Y\cdot e(g,g^{\varphi})^{\widetilde{\alpha}} = \mathcal{M}_c\cdot e(g,g)^{\alpha\varphi}$. Otherwise, if Y is a random term, then all the information of the massage is hidden. Hence the advantage of algorithm $\mathcal{A}$ is 0. As the consequence, if algorithm $\mathcal{A}$ has a non-negligible advantage in breaking the security game, then algorithm $\mathcal{C}$ has a non-negligible advantage to break the $(q-1)$ assumption.

Then we consider the security of the keyword index and adopt the indistinguishable keyword index against witch is the chosen keyword attacks through the security game.

Theorem 2. *Under the generic group model, For any adversary $\mathcal{A}$, let q be a bound on the total number of group elements, it receives from the queries it makes to the oracles for the hash function groups G and G_T, and the bilinear mapping e, and from its interaction with the IND-CPA security game. And we have that the advantage of the adversary in this security game is $O(q^2/p)$.*

Proof. A challenge $\mathcal{C}$ and a adversary $\mathcal{A}$ conduct the following game. $\mathcal{A}$ generates two sets of series: $L_G = \{\langle F_{0,l}, \varphi_{0,l}\rangle : l = 1,\cdots,\Upsilon_0\}, L_{G_T} = \{\langle F_{1,l}, \varphi_{1,l}\rangle : l = 1,\cdots,\Upsilon_1\}$, where $F_{0,l}$ and $F_{1,l}$ are the queries of two variable polynomials for queries of $\mathcal{A}$, $\varphi_{0,l}$, $\varphi_{1,l}$ are the query of random string in $\{0,1\}^*$ for each query result, where $\varphi_{0,l} = \varphi_0(F_{0,l})$, $\varphi_{1,l} = \varphi_1(F_{1,l})$. We set $F_{0,l} = 1, F_{1,l} = 1$, thus $g = \varphi_0(1)$, $g^x = \varphi_0(x)$, $e(g,g)^y = \varphi_1(y)$. Now, we show the query results to random oracle for $\mathcal{A}$ as follows:

Group action. Given two operands $\varphi_i(x)$ and $\varphi_i(y)$, $x, y \in \mathbb{Z}_p$, $i \in \{1,2\}$. If $\varphi_i(x)$ and $\varphi_i(y)$ are not in the list L_G and L_{G_T}, return $\perp$. Otherwise, $\mathcal{C}$ computes $F = x + y \,(\mathrm{mod} p)$ and check whether F is in the list L_G and L_{G_T}. If it exists, $\mathcal{C}$ returns $\varphi_i(F)$, otherwise $\mathcal{C}$ sets $\varphi_i(F)$ to a random string in $\{0,1\}^*$. At last $\mathcal{C}$ adds $\langle F, \varphi_i(F)\rangle$ to the list and return to $\mathcal{A}$ the string $\varphi_i(F)$.

Bilinear mapping. This step is the same as the group action step except that $\mathcal{C}$ calculate $F = xy \,(\mathrm{mod} p)$.

□

Through the above basic operations, the interactive game between challenger $\mathcal{C}$ and adversary $\mathcal{A}$ is as follows:

1) Challenger $\mathcal{C}$ randomly selected $\lambda_1, \alpha, y \in \mathbb{Z}_p^*$, where $g_1 = g^{\lambda_1}$, H is a one way function $H : \{0,1\}^* \to \mathbb{Z}_p^*$. Then C sets $params = (y, g_1, H, e(g,g)^{\alpha})$ and send it to $\mathcal{A}$;

2) Adversary $\mathcal{A}$ uses the representative keyword set $W' = \{w_{j'}\}_{j' \in [1,d']}$ and randomly selectes $y, z, \sigma \in \mathbb{Z}_p^*$ to conduct trapdoor inquiry, $\mathcal{C}$ calculates $K^* = g^y$ and gets the secret key $SK^* = \{K^*\}$. $\mathcal{C}$ then generates a keyword set trapdoor $TK^* = \{T_1^*, T_2^*, T_3^*, T_4^*\}$ and sends it to $\mathcal{A}$, where $T_1^* = K^{z\sigma}$, $T_2^* = K^{\sigma}$, $T_3^* = \prod_{j'=1}^{d'} g^{H(w_{j'})z}$, $T_4^* = g_1$;

3) Adversary $\mathcal{A}$ uses $W_0 = \{w_{j_0}\}_{j_0 \in [1,d]}$ and $W_1 = \{w_{j_1}\}_{j_1 \in [1,d]}$ as the challenge keyword sets, then $\mathcal{A}$ gives a challenge access policy. Finally, $\mathcal{C}$ selects 0 or 1 to generate the challenge index of W_b. It chooses $r', r'' \in \mathbb{Z}_p^*$ randomly, and sets $I_j^* = g^{zr'H(w_{j_b})}$, $I_1^* = g^{r''}$, $I_2 = g^{zr''}$, $I_3^* = g_1^{r'}$. The challenge index is $Index^* = \left\{\{I_j^*\}_{j \in [1,d]}, I_1^*, I_2^*, I_3^*\right\}$, and sends $Index^*$ to $\mathcal{A}$;

4) $\mathcal{A}$ repeats the inquiry in (2) until W_0 and W_1 cannot be queried;

5) Finally, $\mathcal{A}$ outputs guess c' of c, where $c' \in \{0,1\}$.

In the above security game, the adversary asks at most q times. In addition, the adversary needs to distinguish between $I_j^* = g^{zr'H(w_{j_0})}$ and $I_j^* = g^{zr'H(w_{j_1})}$. We consider simulating a game using $I_j^* = g^{\theta}$ instead of $I_j^* = g^{zr'}$ in real challenges. The adversary needs to distinguish g^{θ} from $g^{zr'}$, where θ is the randomly selected in $\mathbb{Z}_p^*$.

Next, we will conduct a detailed analysis of the $\mathcal{C}$ simulation. In the generic group model, the simulation of $\mathcal{C}$ is perfect as long as there is no unexpected collision. In other words, we think that the oracle query is a rational function $\delta = \eta/\xi$ in a variable $\theta, \alpha, \lambda_1, z, \sigma, r', r''$. When two queries correspond to different rational functions, since the values of these variables are randomly chosen, the rational function will have an unexpected collision, that is, when $\eta \neq \eta'\xi \neq \xi'$, then $\delta = \eta/\xi = \eta'/\xi' = \delta'$.

Our current condition is that this unexpected collision will not occur in G or G_T. For any pair of queries in a group with different rational functions η/ξ and η'/ξ' , If and only if the non-zero polynomial $\eta\xi' - \xi\eta' = 0$ holds, the collision will occur. Under the constraints, the probability of any such collision is $O(q^2/p)$ at most. Therefore, we assume that no such collision occurs and its probability is defined as $1 - O(q^2/p)$. Because θ only exists in $I_j^* = g^{\theta}$, If a collision occurs, it exists $\gamma \neq 0$, we have $\delta - \delta' = \gamma z r' - \gamma\theta$. After analysis, it is almost impossible for the adversary $\mathcal{A}$ to construct an inquiry about $\gamma z r'$: we have $\gamma z r' = \gamma\theta + \delta' - \delta$, for the query δ in $g^{\delta} = \varphi(\delta)$ and δ' in $g^{\delta'} = \varphi(\delta')$, The probability that the formula is true is $O(1/p)$. Therefore, the possibility that collision occurs is negligible. So, $\mathcal{A}$ is almost impossible to construct an inquiry on $\gamma z r'$.

Remark. Since the strategy is partially hidden, in order to ensure the validity of the obtained data, we can add the message verification as the proposal in [4]. In fact, in the algorithm of Encryption, we can add a term ζ and encrypt $\mathcal{M} \parallel \zeta$ instead of the message $\mathcal{M}$. The process is similar to [4] and is omitted here.

6 Performance Comparisons and Analyses

In this section, we will analyze the performance between the proposed scheme and several related literatures. And we simulate and compare the algorithms of the relevant literature and compare the time it takes.

First, we consider the performance of the structure from some aspects, including the scheme function and the computation overheads. We list the performance comparison (Table 1, Table 2,) between the proposed scheme and the related literatures [5, 11, 16, 24, 26]. The specific comparison items and results are as shown in the following tables:

Table 1: Scheme function comparison

Scheme	Keyword search	CP-ABE/KP-ABE	Access Policy	Large Universe
[11]	Single	CP-ABE	LSSS	No
[5]	Multiple	KP-ABE	LSSS	Yes
[16]	No	CP-ABE	Access Tree	No
[26]	Single	CP-ABE	LSSS	Yes
[24]	No	CP-ABE	And-Gate	no
ours	Multiple	CP-ABE	LSSS	Yes

It can be seen from Table 1 that our scheme is based on the prime order group on the large attribute universe to implement a ciphertext policy attributes-based searchable encryption scheme. Moreover, the proposed scheme supports multi-keyword search, however [11,26] only support single keyword search, and keyword searching is not supported in [16,24]. The structure of scheme [5] is based on KP-ABE. However, KP-ABE design is relatively close to a static scenario and cannot guarantee fine-grained access. CP-ABE, because the policy is embedded in the ciphertext, makes a granularity of the data that can be refined to the attribute level of encrypted access control. Since electronic medical records are generally stored in the cloud, the use of CP-ABE structure for encrypted storage of electronic medical cases is more suitable for medical data storage and fine-grained sharing. Thus the proposed solution uses CP-ABE more efficiently. The access policy of the proposed scheme is based on LSSS, which can achieve better privacy protection and secret sharing. The prime order group structure is more efficient than the combined order group. In "large universe" construction, the size of keyword space can be exponentially large, so it is much more desirable in practical applications. Overall, our solution enables efficient searchable encryption.

Table 2: Comparison of the computation overheads

Scheme	KeyGen	Enc	Trapdoor	Search	Dec
[11]	$(4k+1)E_1+(k+1)M_1$	$(12l+2)E_1$	-	-	$(6\|\mathcal{I}\|+1)P+4M_2+2M_2$
[5]	E_1	$(6l+2)E_1+(l+1)M_1+E_2$	$(11m+2)E_1+6mM_1+E_2$	$6\|\mathcal{I}\|P+4\|\mathcal{I}\|M_2+\|\mathcal{I}\|E_2$	-
[16]	$13kE_1+(8k+1)M_1$	$(12l+3)E_1+(8l+1)M_1$	-	-	$(4\|\mathcal{I}\|+2)P+(4\|\mathcal{I}\|+2)M_2$
[26]	$(6k+26)E_1+(3k+6)M_1$	$(8l+6)E_1+(3l+1)M_1+2P+(m+1)E_1$	$(5l+2)E_1$	$(2m'+2)P+2M_2$	$(12\|\mathcal{I}\|+2)P+2\|\mathcal{I}\|E_2+8\|\mathcal{I}\|M_2$
[24]	$(12k+10)E_1+(12k)M_1$	$(2l+7)E_1+(2l)M_1+4E_2+2M_2$	-	-	$12P+(8l+2)E_1+(8l)M_1+4E_2+10M_2$
ours	$(4k+4)E_1+(k+1)M_1$	$(7l+1)E_1+2lM_1+E_2+(m+3)E_1$	$(m'+2)E_1$	$(2m'+2)P+2M_2$	$(5\|\mathcal{I}\|+1)P+\|\mathcal{I}\|E_2+5\|\mathcal{I}\|M_2$

In the light of Pairing Based Cryptography (PBC) library [10], we mainly consider three time complexity algorithms including multiplication, exponential and pairing operation in group G and G_1. To be more specific, E_1 and M_1 are exponential and multiplication operation in group G, E_2 and M_2 are exponential and multiplication operation in group G_1, P is a pairing operation. In addition, k represents the number of attributes, l denotes the number of rows in a policy matrix, m is the number of keyword set and m' is the number of keyword in search. The authorized set $\mathcal{I}$ is used in access phase and its size $|\mathcal{I}|$ is determined by the complexity of the access policy. The environments of server is Windows 7 desktop PC system with 4th generation Intel Core i7-4790 @ 3.60GHz, and RAM is 8 GB. From table 2, we further compare the computational overheads incurred from KeyGen, Enc, Trapdoor, Search, and Dec, respectively, which denote the key generation, encryption, trapdoor generation, search and decryption. The length of an element in each group G and G_1 is set to 512 bits.

According to Table 2, it can be easily found that our solution has a better computing performance in the key generation phase. Since scheme [5] does not encrypt any valid information during the key generation phase, its computational cost is very low. In the encryption phase, the computational cost of the algorithm we proposed is lower than [11,16,24,26]. Since the schemes [11,16,24] do not support keyword search, we do not compare the computational costs of their trapdoor generation and search phases. Comparing with other solutions, our solution has lower trapdoor calculation cost and better search efficiency. The proposed solution supports multi-keyword search, which is a function that other programs cannot achieve. Since the search and decryption of the literature [5] is completed at the same stage, we do not compare the computational performance of the decryption phase. Compared with other literatures, the proposed scheme has lower decryption computational energy consumption.

In order to compare performance more clearly, we have drawn a performance comparison Figures 2-6 for the four phase of the key generation phase, the encryption phase, the trapdoor generation phase, the search phase and the decryption phase. In general, the proposed algorithm is more efficient than other literatures, and as the number of attributes increases, our algorithm consumes slightly less time than other literature algorithms. Moreover, the proposed solution can achieve lower computational cost and better multi-keyword search, and can be applied to practical electronic medical systems.

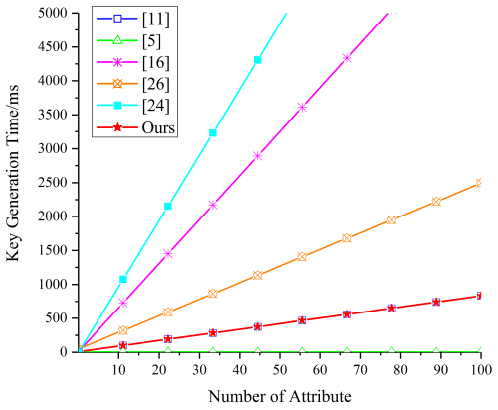


Figure 2: Key generation time

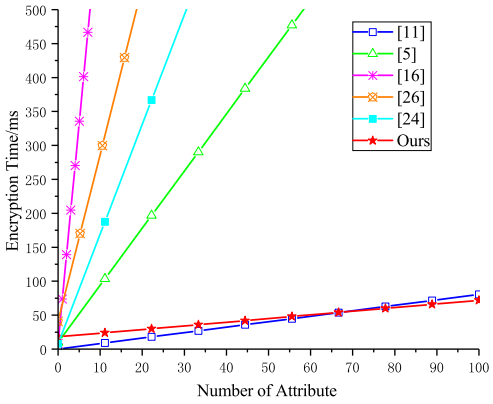


Figure 3: Encryption time

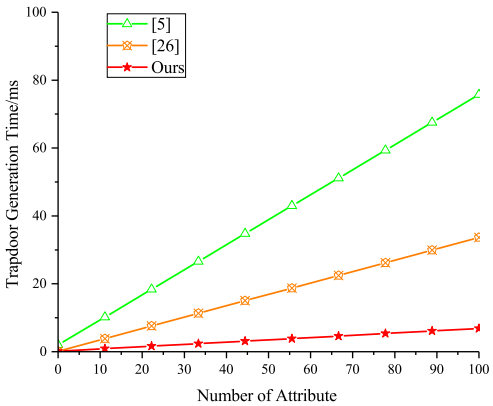


Figure 4: Trapdoor generation time

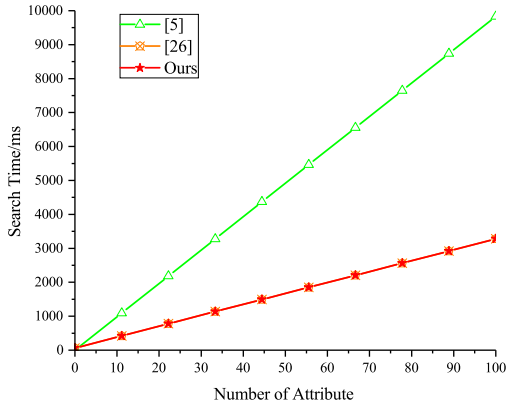


Figure 5: Search time

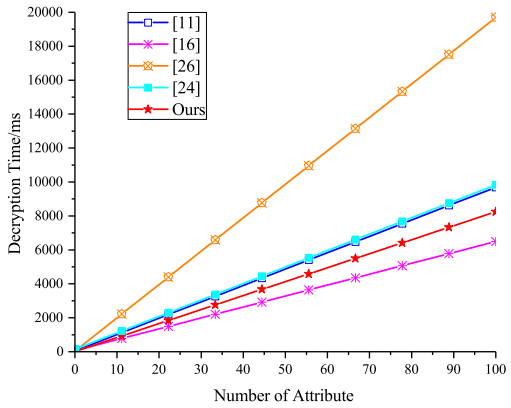


Figure 6: Decryption time

7 Conclusion and Future Work

In this paper, we propose a searchable encryption structure of ciphertext policy for large attribute universe. Our scheme is an expressive implementation of searchable encryption scheme established on prime order groups, which can effectively implement encrypted electronic medical records for multiple keywords search in cloud. This scheme greatly improves the efficiency for medical staff and protects data privacy of patients. Besides, the proposed algorithm supports efficient multi-keyword search for electronic medical records. Theoretical analysis proves that our scheme is selective indistinguishability security against chosen keyword-set attack in the standard model.

Acknowledgments

The authors thank the anonymous reviewers for their constructive comments and suggestions. This work was supported in part by the National Natural Science Foundation of China under Grants (51875457), the National Cryptography Development Fund under grant No. MMJJ20180209, and the International S&T Cooperation Program of Shanxi Province (2019KW-056).

References

[1] J. Baek, R. N. Safavi, and W. Susilo, "Public key encryption with keyword search revisited," in *Computational Science and Its Applications (ICCSA'08)*, pp. 1249–1259, 2008.

[2] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Advances in Cryptology*, pp. 506–522, 2004.

[3] J. W. Byun, H. S. Rhe, H.A. Park, and D. H. Lee, "Off-line keyword guessing attacks on recent keyword search schemes over encrypted data," in *Secure Data Management*, pp. 75–83, 2006.

[4] H. Cui, R. H. Deng, J. Lai, X. Yi, and S. Nepal, "An efficient and expressive ciphertext-policy attribute-based encryption scheme with partially hidden access structures, revisited," *Computer Networks*, vol. 133, pp. 157–165, 2018.

[5] H. Cui, Z. Wan, R. H. Deng, G. Wang, and Y. Li, "Efficient and expressive keyword search over encrypted data in cloud," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, pp. 409–422, May 2018.

[6] T. Feng, X. Yin, Y. Lu J. Fang, and F. Li, "A searchable cp-abe privacy preserving scheme," *International Journal of Network Security*, vol. 21, pp. 680–689, July 2019.

[7] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in *Applied Cryptography and Network Security*, pp. 31–45, 2004.

[8] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS'06)*, pp. 89-98, Nov. 2006.

[9] Q. Huang and H. Li, "An efficient public-key searchable encryption scheme secure against inside keyword guessing attacks," *Information Sciences*, vol. 403-404, pp. 1–14, Sep. 2017.

[10] F. Li and W. Wu, *Pairing-Based Cryptography*, Science Press, 2014.

[11] J. Li, Y. Shi, and Y. Zhang, "Searchable ciphertext-policy attribute-based encryption with revocation in cloud storage," *International Journal of Communication Systems*, vol. 30, no. 1, pp. e2942, 2017.

[12] X. Liu, T. Lu, X. He, X. Yang, and S. Niu, "Verifiable attribute-based keyword search over encrypted cloud data supporting data deduplication," *IEEE Access*, vol. 8, pp. 52062–52074, 2020.

[13] Z. Liu and Y. Fan, "Provably secure searchable attribute-based authenticated encryption scheme," *International Journal of Network Security*, vol. 21, pp. 177–190, Mar. 2019.

[14] Y. Miao, J. Ma, X. Liu, J. Weng, H. Li, and H. Li, "Lightweight fine-grained search over encrypted data in fog computing," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 772–785, 2019.

[15] S. Niu, L. Chen, J. Wang, and F. Yu, "Electronic health record sharing scheme with searchable attribute-based encryption on blockchain," *IEEE Access*, vol. 8, pp. 7195–7204, 2020.

[16] T. V. X. Phuong, G. Yang, and W. Susilo, "Hidden ciphertext policy attribute-based encryption under standard assumptions," *IEEE Transactions on Information Forensics and Security*, vol. 11, pp. 35–45, Jan. 2016.

[17] Y. Rouselakis and B. Waters, "New constructions and proof methods for large universe attribute-based encryption," in *Acm Sigsac Conference on Computer & Communications Security*, 2013. DOI: 10.1145/2508859.2516672.

[18] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in *International Conference on Theory and Applications of Cryptographic Techniques*, pp. 457–473, 2005.

[19] R. Sakai and J. Furukawa, "Identity-based broadcast encryption," *IACR Cryptology ePrint Archive*, vol. 2007, p. 217, 01 2007.

[20] C. Shen, Y. Lu, and J. Li, "Expressive public-key encryption with keyword search: Generic construction from kp-abe and an efficient scheme over prime-order groups," *IEEE Access*, vol. 8, pp. 93–103, 2020.

[21] M. Shen, B. Ma, L. Zhu, X. Du, and K. Xu, "Secure phrase search for intelligent processing of encrypted data in cloud-based iot," *IEEE Internet of Things Journal*, vol. 6, pp. 1998–2008, 2019.

[22] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceedings of IEEE Symposium on Security and Privacy*, pp. 44–55, May 2000.

[23] J. Wang, X. Yu, , and M. Zhao, "Fault-tolerant verifiable keyword symmetric searchable encryption in hybrid cloud," *International Journal of Network Security*, vol. 17, no. 4, pp. 471–483, 2015.

[24] Q. Wang, L. Peng, H. Xiong, J. Sun, and Z. Qin, "Ciphertext-policy attribute-based encryption with delegated equality test in cloud computing," *IEEE Access*, vol. 6, pp. 760–771, 2018.

[25] S. Wang, S. Jia, and Y. Zhang, "Verifiable and multi-keyword searchable attribute-based encryption scheme for cloud storage," *IEEE Access*, vol. 7, pp. 50136–50147, 2019.

[26] S. Wang, L. Yao, J. Chen, and Y. Zhang, "Ksabeswet: A keyword searchable attribute-based encryption scheme with equality test in the internet of things," *IEEE Access*, vol. 7, pp. 80675–80696, 2019.

[27] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization," in *Public Key Cryptography*, pp. 53–70, 2011.

[28] P. Xu, H. Jin, Q. Wu, and W. Wang, "Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack," *IEEE Transactions on Computers*, vol. 62, pp. 2266–2277, Nov. 2013.

[29] J. Yang, Y. Xu, Q. Wang, H. Pan, and J. Guan, "Data privacy protection technology for medical information," *Chinese Digital Medicine*, vol. 05, no. 8, pp. 50–54, 2010.

[30] W. Yau, S. Heng, and B. Goi, "Off-line keyword guessing attacks on recent public key encryption with keyword search schemes," in *Autonomic and Trusted Computing*, pp. 100–105, 2008.

[31] Z. Zhao, L. Sun, Z. Li, and Y. Liu, "Searchable ciphertext-policy attribute-based encryption with multi-keywords for secure cloud storage," in *Proceedings of the International Conference on Computing and Pattern Recognition (ICCPR'18)*, pp. 35–41, 2018.

[32] Q. Zhu, Y. Luo, and J. Le, "A survey of database encryption and ciphertext data query technology," *Journal of Donghua University*, vol. 33, no. 4, pp. 543–548, 2007.

[33] L. Zou, X. Wang, and S. Yin, "A data sorting and searching scheme based on distributed asymmetric searchable encryption," *International Journal of Network Security*, vol. 20, no. 3, pp. 502–508, 2018.

Biography

Qing Wu was born in Shandong Province, China in 1975. She received the M.E. degree in applied mathematics in 2005, and Ph.D. degrees in applied mathematics in 2009 from Xidian University, Xi'an, China. She is currently a professor with the School of Automation at Xi'an University of Posts and Telecommunications, China. Her research interests include network security, big data protection, computing security, and machine learning.

Xujin Ma is a master degree student in the school of automation, Xi'an University of Posts and Telecommunications. His research interests focus on big data privacy protection and network security.

Leyou Zhang is a professor in the school of mathematics and statistics at Xidian University, Xi'an China. He received his Ph.D. degree from Xidian University in 2009. From Dec. 2013 to Dec. 2014, he is a research fellow in the school of computer science and software engineering at the University of Wollongong. His current research interests include network security, computer security, and cryptography.

Yanru Chen is an associate professor, from the school of humanity and foreign languages of Xi'an University of Posts and Telecommunications, China. She got her B.A. degree from Xi'an Foreign Languages University in 2000, and M.A. degree in Linguistics from Xi'an International Studies University in 2007. Her current study fields include applied linguistics and big data management.

Integration of Quantization Watermarking and Amplitude-Thresholding Compression for Digital Audio Signal in the Wavelet Domain

Ming Zhao[1], Xindi Tong[2], and Jie Li[1]
(Corresponding author: Jie Li)
School of computer science, Yangtze University[1]
Jingzhou, Hubei province, China
Department of Mathematics and Information Engineering, The Chinese University of Hong Kong[2]
Email: 1104112257@qq.com
(Received Sept. 2, 2020; Revised and Accepted Jan. 10, 2021; First Online Apr. 24, 2021)

Abstract

Due to the advancement of technology and the rapid development of the Internet, digital information transmission has skyrocketed, and its acquisition and dissemination have become easier. Without the legal owner's permission, digital information is often stolen or turned into profit by illegal persons. This study proposes a combination of quantization watermarking and amplitude-thresholding compression technology for digital music (or audio) based on discrete wavelet transform (DWT). This technology is expected to reduce the carrying capacity of network transmission while protecting personal copyrights. Moreover, it is resistant to various malicious attacks.

Keywords: Digital Audio Watermarking; Compression; Discrete Wavelet Transform (DWT)

1 Introduction

An audio watermarking technology usually consists of the embedding and extraction techniques and satisfies three minimum requirements of audio watermarking standards set by the International Federation of Phonographic Industry (IFPI) requirements. According to the requirements of the IFPI, an audio watermarking technology should have three specifications [4, 5, 7, 8, 10, 11, 15, 16]:

1) Audio watermark should be imperceptible of original signal;

2) Signal-to-noise ratio (SNR) needs to be higher than 20 dB and the embedding capacity should be more than 20 bits-per-second (bps);

3) Watermark should be capable of resisting common attacks.

Internet development not only brings a lot of convenience but also relative risk. How to reduce the carrying amount of nature data and the hidden information in network transmission at the same time is an important issue. Audio compression technology is to sample or quantize digital audio to reduce the amount of audio data in order to save the time required for file storage and the communication bandwidth required for data transmission. The compressed audio quality must also be at a certain level. There are many types of audio compression technologies, including: MP3, WMA, WAV, turnpoint, threshold compression and so on. Some of these compression technologies can also be implemented in combination with transform domain based on their characteristics. In [18], authors proposed a new idea of integrating electrocardiogram watermarking and compression approach, which has never been researched before. ECG watermarking can ensure the confidentiality and reliability of a user's data while reducing the amount of data. In the evaluation, they apply the embedding capacity, bit error rate (BER), signal-to-noise ratio (SNR), compression ratio (CR), and compressed-signal to noise ratio (CNR) methods to assess the proposed algorithm. After comprehensive evaluation, the final results show that their algorithm is robust and feasible.

In this study, we integrate quantization watermarking technology with amplitude-thresholding compression for digital music (or audio) based on discrete wavelet transform (DWT). First of all, we perform DWT on each audio signal to embed private information into DWT lowest coefficients. Then, we obtain watermarked audios by inverse discrete wavelet transform (IDWT). At the same time, we adopt the amplitude-threshold compression to reduce the data amount of the embedded audio signal. In addition, the hidden information can be extracted without the original audio signal and the recovery of the compressed audio signal adopts cubic spline. In experiments, we evaluate

the appropriate threshold ε, embedding strength Q, and the robustness against various malicious attacks.

The rest of this paper is organized as follows. Section 2 reviews some preliminaries for later use. Section 3 presents the proposed integration of quantization watermarking and amplitude-thresholding compression for digital audio signal in the wavelet domain. Section 4 shows experimental results. Conclusions are drawn in Section 5.

2 Preliminaries

In this section, we review some preliminaries including Discrete Wavelet Transform, Cubic spline, and performance measurement for audio watermarking and signal compression.

2.1 Discrete Wavelet Transform

The wavelet transform maps a function in $L^2(R)$ onto a scale-space plane. Wavelets are obtained by a single prototype function (mother wavelet) $\psi(x)$ which is regulated with a scaling parameter and a shift parameter [3,12,13]. The discrete normalized scaling and wavelet basis function are defined as

$$\begin{aligned}\varphi_{i,n}(t) &= 2^{i/2} h_i \varphi(2^i t - n).\\ \psi_{i,n}(t) &= 2^{i/2} g_i \psi(2^i t - n).\end{aligned}$$

where i and n are the dilation and translation parameters; h_i and g_i are the low-pass and high-pass filters. Orthogonal wavelet basis functions not only provide simple calculation in coefficients expansion but also span $L^2(R)$ in signal processing. As a result, any audio signal $S(t) \in L^2(R)$ can be expressed as a series expansion of orthogonal scaling functions and wavelets. More specifically,

$$S(t) = \sum_{\ell} c_{j_0}(\ell)\varphi_{j_0,\ell}(t) + \sum_{k}\sum_{j=j_0}^{\infty} d_j(k)\psi_{j,k}(t),$$

where

$$c_j(\ell) = \int_R S(t)\varphi_{j,\ell}(t)dt$$

and

$$d_j(k) = \int_R S(t)\psi_{j,k}(t)dt$$

denote the sequences of low-pass and high-pass coefficients, respectively; j_0 be the integer to define an interval on which $S(t)$ is piecewise constant. Throughout this paper, the host digital audio signal $S(n), n \in N$, denoting samples of the original audio signal $S(t)$ at the n-th sample time, is cut into segments where DWT will be preformed. This can be done by exploiting orthogonal basis to implement DWT through filter bank. Figure 1 demonstrates how the input digital audio signal $S(n)$ is segmented into eight non-overlapping multi-resolution subbands by the seven-level DWT decomposition.

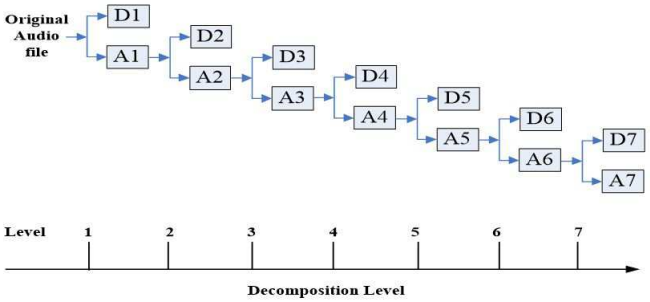


Figure 1: Seven-level discrete wavelet transformation

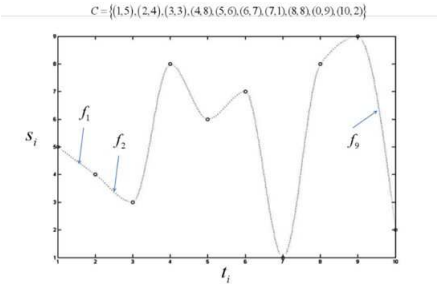


Figure 2: Cubic spline interpolation

2.2 Cubic Spline Interpolation

For a given dataset $C = \{(t_0, s_0), (t_1, s_1), ..., (t_N, s_N)\}$, the cubic function is formulated as [1,9,14]

$$f_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$$

Found that the N cloud gauge line collection of functions $\{f_i(t) | i = 1, ..., N\}$ as shown in Figure 2 to describe the entire set of data, where $f_i(t)$ must satisfy

$$\begin{aligned}f_i(t_i) &= s_i = f_{i-1}(t_i),\\ f'_i(t_i) &= f'_{i-1}(t_i),\\ f''_i(t_i) &= f''_{i-1}(t_i),\\ f''_1(t) &= f''_N(t) = 0\end{aligned}$$

2.3 Performance Measurement of Audio Watermarking and Signal Compression

In general, transparency is the key performance of audio watermarking. It is measured by signal-to-noise ratio (SNR) which are defined as follows [4,5,7,8,10,11,15,16]:

$$SNR = -10\log\left(\frac{\sum_{i=1}^{N}(\bar{s}_i - s_i)^2}{\sum_{i=1}^{N} s_i{}^2}\right)$$

In addition, we also apply relative root mean square error (rRMSE) and root mean square error (RMSE) to

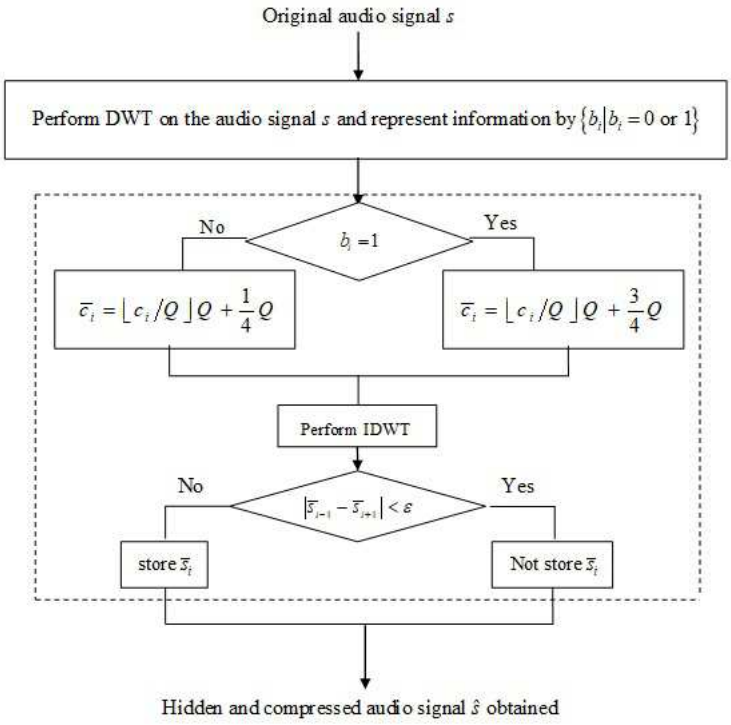


Figure 3: Flow chart of integration of watermarking and compression

judge the transparency, which is defined as follows:

$$\text{rRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\bar{s}_i - s_i}{s_i}\right)^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\bar{s}_i - s_i)^2}$$

where $\{s_i\}$ represents the original ECG signal for ECG, $\{\bar{s}_i\}$ represents the hidden (or modified) ECG signal.

In order to evaluate the quality of signal compression, compression ratio (CR) and percentage ratio difference (PRD) are utilized. Assume s is the original signal and $\breve{s}$ is the reconstruction from the compressed signal, then CR and PRD are defined as [2,6,17]

$$\text{CR} = \frac{\text{Datasizebeforecompression}}{\text{Datasizeaftercompression}}$$

$$\text{PRD} = \sqrt{\frac{\sum_{i=1}^{N}\left(\breve{s}_i - s_i\right)^2}{\sum_{i=1}^{N} {s_i}^2}} \cdot 100$$

where N is the number of testing samples in the signal s.

3 Proposed Method

In order to hide information into an audio and reduce the carrying amount on it when transmitting in the Internet, this section presents the proposed method integrating quantization watermarking and amplitude-thresholding compression for digital audio signal in the wavelet domain. Figures 3 and 4 show the flowchart of the proposed integration. The detail is introduced in the following.

3.1 Watermarking and Compression

In order to embed private information into audio signals conveniently, we utilize binary bits $\{b_i | b_i = 0 or b_i = 1\}$ to represent information that will be hidden, and then embed these binary bits to DWT coefficients of audio signals by quantization embedding technique which is proposed as follows. We use 7-level Haar DWT to decompose an ECG signal into eight non-overlapping sub-bands. Figure 1 shows the structure of the 7-level DWT decomposition. Taking into account the robust performance of the low-pass filtering, we embedded the watermark (binary bits $\{b_i | b_i = 0 or b_i = 1\}$) into the sub-band coefficients in the 7th level, which are the lowest-frequency coefficients. The embedding rule is based on the quantization technique [14,17].

$$\bar{c}_i = \begin{cases} \left\lfloor {c_i}/{Q} \right\rfloor Q + {3Q}/{4}, \text{if} b_i = 1 \\ \left\lfloor {c_i}/{Q} \right\rfloor Q + {Q}/{4}, \text{if} b_i = 0 \end{cases}$$

where c_i and $\{\bar{c}_i\}$ are the 7th low-frequency DWT coefficients before and after embedding, respectively; Q is the embedding strength; By applying the IDWT, the watermarked audio signal $\bar{S}$ is obtained.

To reduce the amount of data when transmitting in the Internet, we compress the embedded audio signal $\bar{s}$ by a

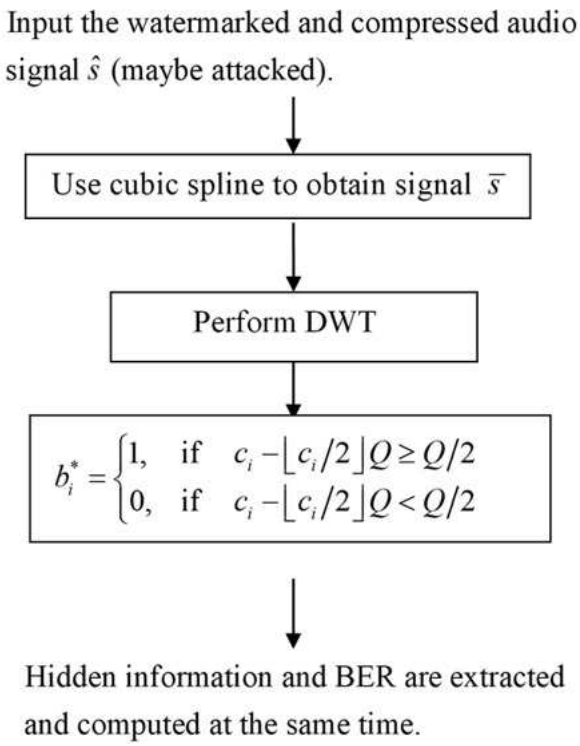


Figure 4: Flow chart of integration of watermarking and compression

threshold compression method which is as follows:

$$\hat{s}_0 = \bar{s}_0, \hat{s}_N = \bar{s}_N$$

$$\hat{s}_i = \begin{cases} \phi & \text{if } |\bar{s}_{i-1} - \bar{s}_{i+1}| < \varepsilon \\ \bar{s}_i & \text{otherwise} \end{cases}, i = \{1, \cdots N-1\}$$

where ε represents the threshold.

For the extraction of the hidden confidential information, we first recover the signal $\{\bar{s}_i\}_{i=0}^N$ from the compressed signal $\{\hat{s}_i\}_{i=0}^N$ by using the cubic function which is formulated as

$$f_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$$

Found that the N cloud gauge line collection of the functions $\{f_i(t) | i = 1, ..., N\}$ to describe the entire set of data, where $f_i(t)$ must satisfy

$$f_i(t_i) = \hat{s}_i = f_{i-1}(t_i), f'_i(t_i) = f'_{i-1}(t_i),$$
$$f''_i(t_i) = f''_{i-1}(t_i), f''_1(t) = f''_N(t) = 0$$

Next, we extract the hidden information from the DWT coefficients $\{\bar{c}_i\}_{i=0}^N$ of the recovered audio signal $\{\bar{s}_i\}_{i=0}^N$ according to the following rules:

$$b_i = \begin{cases} 1, \text{if} \bar{c}_i - \left\lfloor \bar{c}_i / Q \right\rfloor Q \geq Q/2 \\ 0, \text{if} \bar{c}_i - \left\lfloor \bar{c}_i / Q \right\rfloor Q < Q/2 \end{cases}$$

where Q is the same as the embedding strength (or secret key) in embedding; $b_i = 1$ or $b_i = 0$ is extracted binary bits or the embedded information equivalently.

4 Experimental Results

The evaluation of the proposed method is discussed in this section. Two types of audio signals, love song, folklore, and dance, are to be tested. These audio signals are 16-bit mono-type of length 11.6 seconds and sampling rate 44.1kHz.

From the results in Table 1, two observations are discussed. First, for the same threshold value, strong embedding strength has better compression due to the fact that the variation of the overall audio is small when the embedding strength is strong. As a result, the SNR, rRMSE and RMSE are worse. Second, for different thresholds, we found that CR value is increased when the threshold value is greater than the embedding strength. Restate, compression is better when the threshold value is greater than the embedding strength.

Common attacks are carried out after the embedding process with $Q = 6500$ and compression with $\varepsilon = 100, 500, 1000$. Based on the robustness is evaluated with the BER, three forms of attacks that apply to the audio signals will be explained in detail below.

1) Re-sampling: The sample rate, the number of samples of audio carried per second. The procedure converts an audio signal from a given sample rate to a different sample rate. In the proposed algorithm, watermarked audio signals are first decimated from 44.1kHz to 22.05kHz, and then interpolated to the original 44.1kHz. This step repeated two more times from 44.1kHz to 11.025kHz and 8kHz, and then back up to 44.1kHz. The BER under the re-sampling at-

Table 1: Performance test by different thresholds and quantization sizes

Audio	ε	Q	CR	PDR	SNR	BER	rRMSE	RMSE
love song	100	100	1.3889	1.2777	37.8712	1.0986	0.3735	70.3929
		500	2.0398	3.7400	28.5427	3.4668	0.5296	206.1828
		1000	3.2663	6.8321	23.3089	3.7119	0.6330	377.9388
		3000	5.6264	12.6983	17.9251	4.0049	0.7711	711.0507
		6500	12.1732	24.5075	10.5972	4.5974	0.9972	1.3591e+003
	500	100	3.7372	4.2549	27.4223	3.6387	1.0696	234.4087
		500	2.0398	3.7400	28.5427	3.4668	0.5296	206.1828
		1000	3.2663	6.8321	23.3089	3.7119	0.6330	377.9388
		3000	5.6264	12.6983	17.9251	4.0049	0.7711	711.0507
		6500	9.1636	23.5035	11.5974	4.5176	0.9972	1.3591e+003
	1000	100	15.8760	19.6754	14.1215	4.7373	5.4751	1.0840e+003
		500	8.0630	8.5770	21.333	4.4932	1.2893	472.8477
		1000	3.2663	6.8321	23.3089	4.7119	0.6330	377.9388
		3000	5.6264	12.6983	17.9251	5.0049	0.7711	711.0507
		6500	9.4136	23.5035	11.5174	5.5176	0.9972	1.3591e+003
dance	100	100	1.3676	0.9204	40.7209	1.0254	0.3558	48.5834
		500	2.2934	3.3630	29.4656	2.3701	0.4141	177.6400
		1000	3.6868	6.3754	23.9098	2.5176	0.5336	337.8759
		3000	6.7927	12.8518	17.8207	3.1279	0.7803	686.8794
		6500	9.8462	23.7338	11.4627	3.3965	0.9554	1.3129e+003
	500	100	5.0135	6.3998	23.8767	3.2734	2.0738	337.8303
		500	2.2934	3.3630	29.4656	4.3701	0.4141	177.6400
		1000	3.6868	6.3754	23.9098	5.5176	0.5336	337.8759
		3000	6.7927	12.8518	17.8207	5.1279	0.7803	686.8794
		6500	9.2482	23.7338	11.4327	5.3965	0.9554	1.3129e+003
	1000	100	13.8847	43.4587	7.2385	5.3965	12.3850	2.2941e+003
		500	9.7062	16.2784	15.7678	5.8105	2.2083	859.8685
		1000	3.6868	6.3754	23.9098	5.5176	0.5336	337.8759
		2048	6.7927	12.8518	17.8207	6.1279	0.7803	686.8794
		6500	9.2462	23.7338	11.4925	6.3965	0.9554	1.3129e+003

Table 2: BER (%) IN THE RE-SAMPLING ATTACK

Audio Type		Love Song			Dance		
Re-sampling Rate (kHz)		22.05	11.03	8	22.05	11.03	8
BER(%)	$\varepsilon = 100$	7.75	19.04	18.28	14.25	28.35	27.55
	$\varepsilon = 500$	9.75	20.04	20.28	16.25	29.85	29.76
	$\varepsilon = 1000$	10.84	23.12	23.18	16.97	30.47	30.65

Table 3: BER (%) IN THE LOW-PASS FILTERING ATTACK

Audio Type		Love Song		Dance	
Cut-off frequency (kHz)		3	6	3	6
BER(%)	$\varepsilon = 100$	40.64	25.13	36.24	21.22
	$\varepsilon = 500$	41.25	27.83	37.52	24.67
	$\varepsilon = 1000$	41.13	30.62	37.54	29.73

Table 4: BER (%) IN NOISE ATTACK

Audio Type		Love Song				Dance			
Noise in dB		-40	-30	-20	-15	-40	-30	-20	-15
BER(%)	$\varepsilon = 100$	6.84	9.14	14.26	17.78	5.91	9.21	14.46	17.52
	$\varepsilon = 500$	9.54	13.58	17.39	19.13	8.46	12.72	17.42	18.92
	$\varepsilon = 1000$	12.84	15.62	20.15	22.43	11.79	16.38	19.56	21.83

tacks are shown in Table 2. The data confirms that the proposed design results in a lower BER.

2) Low-pass filtering: A low-pass filter is a circuit that provides easy passage to low-frequency signals and difficult passage to high-frequency signals. Table 3 shows the BER information with a low-pass filter and cutoff frequencies at 3kHz and 6kHz.

3) MP3 compression: MP3 compression is generally used to reduce file sizes. The bit rate of an MP3 file is a measure of the audio signal in a given period of time. Usually, the larger the bit rate, the better the sound quality. Table 4 contain experimental data while applying MP3 compression at different bit rates to the watermarked audio.

5 Conclusions

In this study, we propose the integration technology of the audio-signal quantization watermarking and amplitude-thresholding compression. The integration technology not only protect the security of private information but also reduce the amount of audio data transmission. We evaluate the appropriate the relationship between threshold and embedding strength Q. Furthermore, we test the robustness against common attacks. The future work is to find the optimal and Q between the CR and SNR.

References

[1] D. Boor and Carl, "A practical guide to splines," *Applied Mathematical Sciences New York Springer*, vol. 27, no. 149, pp. 157–157(1), 1978.

[2] M. Brito, J. Henriques, P. Gil, and M. Antunes, "A predictive adaptive approach to generic ecg data compression," in *IEEE International Workshop on Intelligent Signal Processing*, pp. 32–37, 2005.

[3] C. S. Burrus, R. A. Gopinath, and H. Gao, *Introduction to Wavelet Theory and Its Application*, New Jersey: Prentice-Hall, 1998.

[4] S. T. Chen, H. N. Huang, C. J. Chen, and G. D. Wu, "Energy-proportion based scheme for audio watermarking," *IET Signal Processing*, vol. 4, no. 5, pp. 576–587, 2010.

[5] S. T. Chen, G. D. Wu, and H. N. Huang, "Wavelet-domain audio watermarking scheme using optimization-based quantization," *IET Signal Processing*, vol. 4, no. 6, pp. 720–727, 2010.

[6] H. H. Chou, Y. J. Chen, Y. C. Shiau, and T. S. Kuo, "A high performance compression algorithm for ecg with irregular periods," in *IEEE International Workshop on Biomedical Circuits and Systems*, pp. S2/4–9–12, 2004.

[7] F. Djebbar and B. Ayad, "Energy and entropy based features for WAV audio steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 1, pp. 168–181, 2017.

[8] S. Gupta and N. Dhanda, "Audio steganography using discrete wavelet transformation (DWT) and discrete cosine transformation (DCT)," *IOSR Journal of Computer Engineering*, vol. 17, no. 2, pp. 32–44, 2015.

[9] C. A. Hall and W. W. Meyer, "Optimal error bounds for cubic spline interpolation," *Journal of Approximation Theory*, vol. 16, no. 2, pp. 105–122, 1976.

[10] M. C. Lee and C. Y. Lau, "Three orders mixture algorithm of audio steganography combining cryptography," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 4, pp. 959–969, 2018.

[11] H. Liu, J. Liu, X. Yan, P. Xue, S. Wan, and L. Li, "Centroid-based audio steganography scheme in wavelet domain," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 5, pp. 1222–1232, 2018.

[12] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 4, 1989.

[13] MathWords, *Dwt: Single-level 1-D Discrete Wavelet Transform*, Apr. 18, 2021. (`https://www.mathworks.com/help/wavelet/ref/dwt.html`)

[14] MathWords, *Spline: Cubic Spline Data Interpolation*, Apr. 18, 2021. (`https://www.mathworks.com/help/matlab/ref/spline.html`)

[15] R. M. Noriega, M. Nakano, B. Kurkoski, and K. Yamaguchi, "High payload audio watermarking: Toward channel characterization of MP3 compression," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 91–107, 2011.

[16] P. Parnami, K. Niwariya, and M. Jain, "Performance evaluation of DWT and LSB based audio steganography," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3, no. 12, pp. 167–170, 2016.

[17] S. G. Miaou, S. N. Chao, "Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework.," *IEEE Transactions on Bio Medical Engineering*, vol. 52, pp. 539–543, 2005.

[18] K. K. Tseng, X. He, W. M. Kung, S. T. Chen, M. Liao, and H. N. Huang, "Wavelet-based watermarking and compression for ecg signals with verification evaluation," *Sensors*, vol. 14, no. 2, pp. 3721–3736, 2014.

Biography

Ming Zhao received the Master degrees in computer aided Instruction from Huazhong Normal University, China in 2006, and he got Ph.D. in computer science and technology from Harbin Institute of Technology, China in 2015. He is currently an associate professor in Yangtze University, China. His research interests include computational intelligence, image and signal processing,

pattern recognition etc. He is an IEEE Senior Member.

Xindi Tong started her college life in 2019. She is currently a year-2 student in The Chinese University of Hong Kong. Her research interests includes Natural language processing, signal processing, artificial intelligence etc.

Jie Li received the Bachelor Degree in School of Computer Science from Yangtze University, China in 2020, and He is currently a graduate student at Yangtze University. His research interests include computational intelligence, few-shot learning, pattern recognition etc.

Digital Certificate of Public Key for User Authentication and Session Key Establishment for Secure Network Communications

Javad Saadatmandan and Amirhossein Rahimi

(Corresponding author: Amirhossein Rahimi)

Department of Mathematics, Qom Branch, Islamic Azad University, Iran
No. 223, ZIP area 11, Azarshahr Street, North Iranshahr Street, Karimkhan-e-Zand Avenue, Tehran, Iran
Email: Amir.Rahimi361@gmail.com

(Received Mar. 2, 2020; Revised and Accepted Nov. 10, 2020; First Online Apr. 23, 2021)

Abstract

In an insecure network environment, public key authentication in sending information is an important security challenge for user authentication. Many schemes based on cryptography have been proposed to solve the problem. However, previous schemes are vulnerable to various attacks and are neither efficient. Therefore, whereby needs more research; Therefore, Standard Digital Certificate Mechanism (SDCM) is an important primitive idea for authenticated key agreement. The (SDCM) is one of the simplest and the most convenient authentication mechanisms, which is more frequently required over insecure areas, such as computer and wireless networks, remote login, operation systems, and database management systems. The (SDCM) contains some user's public information, such as the information of user authentication's digital, the information of a digital birth certificate, general credentials (certificates), a digital signature of public information signed by a (CA), etc. Therefore must make design decisions that appear well-motivated but have unintended consequences, based on this concept, we propose user secure mutual authentication and session key establishment protocol based on combining the discrete logarithm (DL) problem, the Diffie-Helman assumption (DHA), and the one-way hash function to withstand the well-known attacks and to keep the same merits.

Keywords: Discrete Logarithm; Digital Signature; Public-key Digital Certificate; Secret Session Key Management; User Authentication

1 Introduction

Authenticity is verified by the theory of public key cryptography. So, public key authentication in sending information is a great research challenge. Therefore, the digital certificate technology is extremely important for computer and network system security, which it has enjoyed strong growth in recent years, but security threats and facing attacks in authentication have grown equally fast. Today, there are many potential attacks that are targeted at authentication including insider attack, masquerade attack, server spoofing attack, parallel session attack and many more. Digital certificate is one of the simplest and the most convenient authentication mechanisms over insecure networks. Therefore, many Internet applications are based on digital certificate mechanism, for example, government organizations, private corporations, database management systems, and school systems. But the public key digital certificate mechanism itself cannot be used to authenticate a user since a public-key digital certificate inclusives only public information and can be easily recorded. Many different data authentication and entity authentication as well as public key digital certificate are studied separately in literature. From well known digital certificates can be called the X.509 public key digital certificate, which in many cases is used in the public key structure (PKI) to ensure the authenticity of the public key and key agreement. Although public key digital authentication only includes general information, it can also be easily registered and reinstated once more. However, it is not advised to authenticate a user. To maintain the certificate framework, the public key infrastructure incurs a nontrivial level of system complexity and implementation costs. A key agreement scheme is said to provide the explicit key confirmation if one entity is assured that the second entity has actually computed the session key [10].

Key establishment is a process whereby two (or more) identities can establish a shared secret key (session key) after message interactions. There are two different approaches to key establishment between two identities. In one process, one identity generates a session key and securely transmits it to the other identity, this is known as encircling or key transmit. In order to obtain authentication for the key agreement protocol, public key certificate is often used in the traditional PKI setting. This

require the parties to provide and verify certificates whenever they want to use a specific public key and the management of public key certificates remains a technically challenging problem. Adi Shamir introduced the identity-based cryptography in 1984 [20]. His idea was to allow parties to use their identities as public keys. With the help of Private Key Generator (PKG), the users attain their private keys and perform cryptographic tasks subsequently, also users identify themself before joining the network. The user's identity (e.g. user's name or email address) will be the corresponding public key; a user only needs to know the "identity" of his communication partner and the public key of the PKG, which is extremely useful in cases like wireless communication.

In this way, in order to authenticate a digital signature of a message, the sender sends an encrypted message to a recipient. Where pre-distribution of authenticated public keys is infeasible. Of course authentication without the help of public key certificate is the major advantage of identity-based cryptography. Therefore, identity-based key agreement protocols without pairing may be more appealing in practice. Two-party authenticated key agreement (AK) protocol not only allows parties to compute a session key known only to them but also ensures the authenticity of the parties. This secret session key can be used to provide privacy and data integrity during subsequent sessions. A key agreement protocol is said to provide implicit key authentication (of Bob to Alice) if Alice is assured that no other entity besides Bob can possibly ascertain the value of the secret key [8, 11, 14]. A key agreement protocol that provides mutual implicit key authentication is called an authenticated key agreement protocol (or AK protocol).

Our proposed scheme is closely related to the ID-based cryptographic algorithms based on proposed Shamir [5, 20]. In this paper, we will introduce a new approach that is able to authenticate a user and also to establish a shared secret session key with the user 's communication partner using any general form of digital certificates, such as a digital authenticator, a digital birth certificate or a digital ID, etc. This kind of digital authentication is called to as Standard Digital Certificate Mechanism (SDCM). The (SDCM) contains some public information related to the user and a digital signature of this public information signed by a CA, which is used as a safe agent for authenticity of the user. Therefore, the digital signature of the SDCM is used as a secret token for each user. In fact, the SDCM does not include general information about any public key of the user, since the user does not have any private and public key pair. Therefore, key management this type of digital certificate is much easier to manage than the X.509 public key digital certificates and thus inherently viable for public-key environment. Naturally in PKI applications, the X.509 public-key digital certificate inclusives a status, also containing the user's public key, and a digital signature of the status. The difference between SDCM and public key digital certificate is that in a SDCM, the public information does not include any public key of the user. The owner of a SDCM can not reveal the original text of the signature of the SDCM to any authenticator. Instead the owner of a SDCM can estimate the response to the authenticator challenge because he has access to the digital signature authentication by responding to the verifier's challenge. Whereas, a secret session key is established between the authenticator and certificate owner during this interaction.

While in our proposed SDCM scheme, the user does not need to know any information of his/her communication participant. The public information of a SDCM, such as user's identity, can be transmitted and verified by each communication identity. Furthermore, this information is used to authenticate each other. In other words, the our proposed scheme supports general PKI applications, such as Internet e-commence, in which communication identities do not need to know each other 's previous communications. There are three basic identities in a digital certificate, which combining these methods can enhance the security level of a system. They are the following:

1) **Certificate Authority (CA):** CA is a trusted person or organization that digitally signs a status with its private key.

2) **The owner of a SDCM:** The owner of the SDCM is the person who receives the SDCM from a trusted certification authority via a secure channel. The owner needs to compute a valid "answer" in response to the verifier's challenged "question" in order to be authenticated and establish a secret session key.

3) **Designated Verifier:** The person verifier is the person who challenges the owner of a SDCM and validates the response by utilizing the owner's public information and CA's public key.

In this study, an effective and secure authenticated key agreement protocol is proposed based on a secure one-way hash function, discrete logarithm problem. By comparing the proposed algorithm with other similar algorithms, we found out that the proposed algorithm had a shorter run time, a lower computation and communication cost, and a more effective storage method. We also investigated the fundamental characteristics of hash functions by arguing that, as these functions cannot be executed computationally via inverse operators, their application in the proposed algorithm would provide further protection against known cyber attacks.

In most of the researches registered in the area of user authentication applications, a trusted authority is responsible for issuing identification card with user information, such as user name and a personal photo on the card, to each user. Each user can be successfully identified if the user owns a legitimate "paper certificate" and matches the photo on the card. The built identification cards of technology very difficult to be forged. A standarded digital certificate mechanism (SDCM) contains public information of the user and a digital signature of the public information signed by a trusted certificate authority.

Table 1: Notations

Symbol	Definition	Symbol	Definition
u_i	User	v_i	Verifier
p	Large Prime	g	Generator
x	Secret Session Key	K_{pub}	Public Key
d_{ID}	Secret Session Key	K_{u_i,v_i}	Secret Key
Not +	Operator XOR	r_{ID}	Random
(SDCM)	Standard Digital Certificate Mechanism	C_A	Certificate Authority
G	Cyclic Additive Group	F_P	Prime Finite Field
G_1, G_2	Multiplicative Cyclic Group	H(0)	Secure Scrambling Function
PKG	Private Key Generator	Z_P^*	Multiplication Group p
AK	Authenticated Key	$\|$	Concatenation Operation
(r_{u_i}, s_{u_i})	Signature Valid Components Pair	$X(\mathrm{mod}P)$	Remainder of X: p
ID	User ID (User Identiiy)	GCDH	Generalized Diffie-Hellman Assumption
DLP	Discrete Logarithm Problem	CDH	Computional Dffe-Hellman Assumption

The digital signature will never be leaked to the verifier. Therefore, the digital signature of a SDCM becomes a security basis that can be used for user authentication.

The remainder of the article is organized as follows. In Section 2, we present a brief review of the related work. In Section3, we describe the new proposed scheme and also in Session 4, we discuss the security analysis, compare the performance and efficiency of the propose scheme with other related schemes in Session 5, and finally concludes the paper in Section 6.

2 Related Works

From past until nowadays, the authenticity of the user and the establishment of key two basic services in secure communications have been, and always digital certificates has been a public key for authenticity of the user and the establishment of keys; in the past, a traditional digital signature to authenticate a certain message to receive the signer should be signed. However, in this method, sometimes, the private key of the signer was impaired. Because a malicious recipient can easily disclose the sender 's digital signature to any third party without the consent of the sender, and subsequently, anyone can access the signer's public key and valid digital signature. The current Internet environment is vulnerable to various attacks such as replay attack, guessing attack, modification attack, and stolen verifier attack. Therefore, extensive research at field digital certificates has been conducted in user authentication and key establishment areas, which both they are substantial services in secure communications. Therefore, we propose scheme which rely on the public-key digital certificates in providing user authentication and key establishment.

In 1989, Chaum and Antwerpen [3] introduced the irrefutable signature theory, which enables the signer to have a full control over his/her signature, in addition to requiring an undeniable signature to sign the message. However, this process could have prevented undesirable verifiers from validating signatures, The real issue of the irrefutable signature was that the signer needs to authenticate the verifier before helping the verifier to authenticate the irrefutable signature. Many different designated verifier signatures (DVS) are studied separately in literature. DVS was first introduced by M. Jakobsson, K. Sako, and R. Impagliazzo [6] and later introduced by Chaum [2] independently, both were introduced in 1996. A DVS provides authentication of a given message to a specified authenticator. One unique feature of a DVS is that a valid DVS can be generated by the "real" signer or designed by an authenticator, with this exclusive feature of a DVS from a traditional digital signature in two distinct ways:

1) Since the designated verifier to be convinced that the DVS is created by the actual signer, not by the verifier itself. Although traditional digital signatures can be verified by any authenticator for the DVS, without the help of a third part that can determine the actual signer even with private key identification.

2) A DVS provides the authenticity of a given message without the denial of a traditional digital signature. Of course, a DVS can replace the traditional digital signatures in most basic applications and provide services with deniability. In the scheme of Jakobsson, Sako, and Impagliazzo [6], a DVS scheme was proposed based on a non-interactive undeniable signature scheme with a limited requirement, but this scheme was not computationally inefficient and also scheme was found to be vulnerable to various forgery attacks.

A DVS can be established by setting the number of signers in a ring signature to two, as proposed in [15, 17]. However, a DVS based on ring signatures does not provide strong designated verifier properties.

In 2003, Saeednia, Kremer, and Markowitch [18] proposed a DVS flexible scheme based on the discrete log-

arithm of the Schnorr signature and Zheng 's signature. Recently, DVS schemes based on any bilinear map was proposed. The structure of a UDVS scheme (DVSBM) is based on a two-way design. Three new structures of UDVS are based on the signature of Schnorr [19] and RSA in Scheme Steinfeld, Wang, and Pieprzyk [21]. The UDVS was also proposed by Laguillaumie and Vergnaud [7], based on the signature of elgamal.

The concept of comprehensive DVS (UDVS) was proposed in [21]. A UDVS is a conventional digital signature with the additional functionality that allows the owner of a digital signature to convert the signature into a DVS of any designated authenticator at his choice. The structure of a UDVS scheme (DVSBM) was based on a bilinear map. Three new UDVS constructions based on the signature of Schnorr [19] and RSA [16] in Scheme Steinfeld, Wang, and Pieprzyk [21] were proposed. Also, The UDVS was proposed by Laguillaumie and Vergnaud [7], based on the signature of elgamal. Some other related research on the DVS and UDVS can be found in [12, 13]. There are three independent entities in each UDVS application:

1) Certification Authority (CA);
2) The owner of a digital signature;
3) Designated Verifier.

In a UDVS, the owner needs to convert the digital signature into an incomplete DVS (without interactions) in order to validate a message. The owner of a digital certificate requires a digital certificate, also interacts with a verifier in order to prove the knowledge of the digital certificate and to be authenticated by the verifier. Our proposed solution is based on the combination of a conventional digital certificate scheme DVS and the well-known (generalized) diffie- hellman assumption [1, 4] and the asymmetric cryptographic system.

3 The Proposed Scheme

Since authentication of user and key establishment of the two basic services are secure in communications, therefore we propose a Standard Digital Certificate Mechanism (SDCM) of public key for user authentication based on both the discrete logarithm (DL) problem on the finite field with Diffie-Hellman Assumption (DHA) [4] and session key establishment, that achieves almost all of the known security attributes with nice computational efficiency than reported other schemes. The primary merit of our scheme is in its simplicity and practicality for implementation under insecure communication links, and also performs the following features:

F_1. Non-repudiation;

F_2. Non-transferrability: For authentication based on a challenge / response, a response to a verifier's challenge cannot be transmitted into a response to another verifier's challenge, Because otherwise it would create impersonation of the user.

F_3. Free from maintaining verification table: The signature verification table is not stored in the server and also is secure against attacks of replaying previously intercepted requests. Instead it maintains only registration time of every the sign-in request message. This will reduce the server over head of maintaining large user data for authentication.

F_4. The computation cost is low since only one-way hash operations are required. "Server overhead decreases for authenticity". Because the server only holds the secret key.

F_5. Resists user impersonation attack (Unforgeability) and modification attack: Data is not transmitted in the main text form on the network "user anonymity and resistance to the modification attack ", and also in order to keep the secret information (secret key), a random number r_{ID} or a timestamp $T' - T \leq \triangle T$ are required. It is better to be resistant to such attacks, if the random number size is greater than the secret key. A valid response can only be generated by the certificate ownership who knows the digital signature of the SDCM.

F_6. Resistance to the collision and withstanding to compromise of data stored in smart card by a intruder "resistant to side- channel attacks". Note that providing security for token information is essential since this card is always vulnerable to attacks such as side-channel attacks, port attacks, and physical attacks "installing a chip at the port/gate". Implementation attacks can use the side channel information, such as timing measurement, power consumption, and faulty hardware, to extract secret keys from the token.

F_7. Provides mutual authentication and resists server spoofing, replay, reflection, parallel session, known session key attacks: In terms of effectiveness, our solution not only preserves secure key establishment, anonymous authentication but also can support mutual authentication (user-server) with agreement of verificated session key. A key agreement can prevent the insider, replay, parallel session, reflection, server spoofing, known session key and man in the middle attacks. The combination of a random number with a timestamp and hash value protects the authentication message against the parallel session attack. Therefore, the proposed protocol is safe against a parallel session attack.

From other the mutual authentication features, maintaining the user's anonymity of attacker based on the modular exponentiation (Identity Support). even an attacker can intercept a number of messages during a certain period, he cannot trace a user's physical position because our anonymity mechanism is a dynamic identification process, and generation of the session key is based on a nonce. Anonymity in the discussion of electronic payment systems, especially electronic money, has a special place.

F_8. Provides session key agreement: Users and the system can use the agreed session key to encrypt/decrypt their communicated messages using the asymmetric cryptosystem. Moreover, the session key is generated by the nonce and a hash function. Therefore, the session key confirms the forward secrecy. If the secret key of the server is disclosed, then the session keys are at the risk of being decoded. Therfore, achieve the goal of user authentication and key agreement with great assurance can prevent to compromise session key and consequently the well-known attacks, such as: the replay, side-channel, impersonation, parallel session, reflection, interleaving, insider, message modification, denial of service, server spoofing, registry center spoofing, test and error, plain text detection, dictionary, XOR inverse and man-in-the-middle attacks. Moreover, the performance analysis and the efficiency comparisons among our proposed scheme and other previously proposed schemes are shown in Table 2. It shows resistance to aforementioned attacks in various processes. Our scheme is an ideal scheme to achieve all the eight of aforementioned requirements in previous section, thus it is immune to various attacks. In contras, related schemes are fail to resist some attacks. This is mostly requiring for security enhancements in our proposed scheme, that is proved to be able to withstand the various possible attacks in Session 5.

3.1 User Authentication and Key Establishment Protocol

Our proposed scheme is mainly divided into two parts, namely, Registration at CA and Protocol.

3.1.1 Registration at CA

In the registration phase, user u_i wants to be the certificate owner and v_i be the verifier. It needs to register at a CA to obtain a SDCM. The CA generates an elgamal signature (r_{u_i}, s_{u_i}) for user's statement m'_{u_i} according to Equation (1), where m_{u_i} is the message digest of the statement m'_{u_i}.

Since the signature component r_{u_i} is a random integer and does not depend on m_{u_i}, it does not need to be kept secret. However, the signature component s_{u_i} is a function of the status. Each owner needs to keep it secret from the verifier in the authentication phase and also the CA knows the one-time secret session key shared between the users.

3.1.2 Protocol

The authentication and key establishment protocol contains the following four steps:

1) With assumption a large prime number p and a generator g in the order of $p-1$ of the Galois field GF(p). (g is a primitive element of the multiplicative group modulo p), the user u_i chooses the public key $k_{pub} \in [1, p-2]$, a random number $r_{ID} \in [1, p-1]$ and computes secret private key

$$x = k_{pub}^{r_{ID}} \mod p \tag{1}$$

as components

$$\begin{aligned} s_{u_i} &= g^{r_{ID}} + h(k_{pub} \| x) \bmod p, \\ r_{u_i} &= g^x \bmod p. \end{aligned}$$

Then the user u_i passes his user information digest

$$m'_{u_i} = r_{ID} s_{u_i} + r_{u_i} x \bmod p - 1. \tag{2}$$

2) After receiving m'_{u_i} and parameters (r_{u_i}, s_{u_i}) to the verifier v_i.

$$g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$$

where $x = k_{pub}^{r_{ID}} \mod p$ is the private key of the CA. If this equality holds true, the verifier v_i

$$c_{v_i} = r_{u_i}^{h_{v_i} \| x} \mod p$$

and send c_{v_i} to the user u_i. Otherwise, the user authentication fails and the protocol is stopped.

3) The user u_i first uses his secret s_{u_i} to compute the secret key

$$K_{u_i,v_i} = c_{v_i}^{(r_{ID}+k_{pub})x + h\left(r_{u_i} \| s_{u_i}\right)},$$

$K'_{u_i,v_i} = D\left(K_{u_i,v_i}\right)$ where $D\left(K_{u_i,v_i}\right)$ represents a key derivation procedure with K_{u_i,v_i} as an input. Then user u_i randomly selects an integer

$$c_{u_i} = r_{u_i}^{h_{u_i} \| x} \mod p$$

and the response

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right),$$

which secret session key

$$d_{ID} =< h_{ID}, s_{ID} >$$

is shared between u_i and v_i.

$$\begin{aligned} h_{ID} &= h(c_{v_i} \| c_{u_i} + x), \\ s_{ID} &= r_{u_i}^{h_{u_i} s_{u_i} + x}, \end{aligned}$$

so $H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right)$ represents a one-way keyedhash function under the key K'_{u_i,v_i} The user u_i sends SK and c_{u_i} back to v_i.

4) After receiving the SK and c_{u_i} from the user u_i, the verifier v_i uses his secret h_{v_i} to compute the shared secret key

$$K_{u_i,v_i} = c_{u_i}^{(r_{ID}+k_{pub})x + h\left(r_{u_i} \| s_{u_i}\right)} \mod p$$

Table 2: Comparison among various digital certificate schemes with our proposed scheme security features

Security Features	Scheme [2]	Scheme [18]	Scheme [21]	Scheme [7]	Proposed Scheme
F_1	Yes	Yes	Yes	No	No
F_2	No	Yes	Yes	Yes	Yes
F_3	No	Yes	No	Yes	Yes
F_4	Yes	No	Yes	Yes	Yes
F_5	No	No	No	Yes	Yes
F_6	No	Yes	Yes	Yes	Yes
F_7	Yes	No	Yes	Yes	Yes
F_8	Yes	Yes	No	Yes	Yes

$K'_{u_i,v_i} = D\left(K_{u_i,v_i}\right)$, and checks whether

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right)$$

is true. If this verification is successful, the certificate owner u_i is authenticated by the verifier v_i and a onetime secret session key $d_{ID} =< h_{ID}, s_{ID} >$ is shared between u_i and v_i. This shared key can provide perfect forward security, so proposed protocol is more computational efficient than the other comparable protocols.

In future section, we analyze the security implications of our proposed scheme against all possible attacks. In order to be authenticated successfully by the verifier v_i, in our protocol, the certificate owner needs to compute and send a valid pair (r_{u_i}, s_{u_i}) and SK to the verifier v_i in Steps (1) and (3). The parameters

$$g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p.$$

This pair of integers can be easily solved by anyone. However, we want to show that only the certificate owner u_i who knows the secret exponent of s_{u_i} can compute a valid SK. This is because the verifier v_i can compute the one-time secret key K_{u_i,v_i} used in generating the SK as $K_{u_i,v_i} = c_{u_i}^{(r_{ID}+k_{pub})x+h\left(r_{u_i}\|s_{u_i}\right)} \mod p$.

The certificate owner u_i who knows the secret exponent of s_{u_i} can also compute K_{u_i,v_i} as $K_{u_i,v_i} = c_{v_i}^{(r_{ID}+k_{pub})x+h\left(r_{u_i}\|s_{u_i}\right)} \mod p = K_{u_i,v_i} \mod p$. Thus, the certificate owner can interact with the verifier v_i and be authenticated successfully. Also each owner needs to keep the secret signature s_{u_i} from the verifier v_i in the authentication protocol. While proving knowledge of the secret component to the verifier v_i, the owner hides the secret component to the verifier during the authentication phase.

4 Proposed Scheme Security Analysis

In this section, to overcome the aforementioned weaknesses of previously schemes and some security loopholes, we going to analyze that the security of our proposed scheme is based on having both properties of the discrete logarithm problem over GF(p) and secure one-way hash function. Then, we will see that the proposed protocol certainly can restrict of well-known attacks with reasonable computational cost.

4.1 Attacks on Digital Signatures

Rivest [16] has divided the possible attacks on digital signature based on the information that the attacker has.

Key-Only Attack: Attacks in which the attacker only knows the public key of the signature and in fact can only control the authenticity of a digital signature.

Message Attack: Attacks in which the attacker, in addition to the signature owner of public key holder, also has samples of normal and equivalent signatures. Each of these attacks may result in the failure of the electronic sign-on system; the failure of a digital signature system has different interpretations.

1) Total Break: A complete failure means that the private key of the signatory is fully disclosed.
2) Universal Forgery: General forgery is realized in a way in which the attacker is not aware of the private key but can sign any desired message by the original owner.
3) Selective Forgery: Selective forgery means that an attacker can only sign a limited set of predefined messages.
4) Existential Forgery: Existential forgery means that an attacker can sign up to at least one message that is not predefined at a successful and credible time from the original owner. Since the attacker does not have much control over the signed messages, the probability that a signed message generating a message with a meaningful and meaningful meaning is very small, so this type of failures is not really difficult to load. They are not so important, since the defense of the proposed scheme from the various attacks by which previous techniques are suffered.

4.2 Security Analysis

We now analyze the points of vulnerability in the proposed protocol and provide desirable security attributes, such as known-key secrecy, SDCM forward secrecy, key-compromise impersonation resilience, and no key control to reduce the systems vulnerabilities.

1) **Key agreement secrecy:** The overture of one secret session key should not compromise other secret components such as h_{ID}, s_{ID} and private key x. Therefore key agreement can prevent to compromise session key and consequently the insider, replay, parallel session, reflection, server spoofing, and man in the middle attacks.

2) **Perfect forward secrecy:** x or k_{pub} is compromised, the attacker cannot get the secret session key $d_{ID} =< h_{ID}, s_{ID} >$.

3) **Known session-specific temporary information secrecy:** Some random private information such as private key x is used as an input of the session key generation function. The revelation of this private temporary information should not compromise the secrecy of (other) generated session key. Generally, this important security attribute requires that if the ephemeral secrets of a session are accidentally leaked to the adversary, the secrecy of the specific session key should not be affected. This revelation is reasonably not partial as it may happen in some practical scenarios. However, we find that their protocols do not offer an important security feature, namely known session specific temporary information secrecy, which considers the impact of ephemeral secrets exposure in affecting the secrecy of the session key. Therefore a user can not compute the secret key of the system from known information.

4) **No key control:** Neither entity should be able to force the session key to be a preselected value. Key escrow is desirable under certain circumstances especially in certain closed groups applications. For example, escrow is essential in situations where confidentiality as well as survey trail are legal requirements, such as secure communications in the health care profession.

5) **Resistance to the forgery attack (Unforgeability):** In order to perform a forgery attack, the attacker needs to present a valid pair (r_{u_i}, s_{u_i}) (r_{u_i}, s_{u_i}) alone in Step 1 cannot be used to authenticate the certificate owner since this pair of parameters can be solved easily by the attacker from Equation (2). However, it is computationally infeasible for the attacker to find the discrete logarithm of s_{u_i} because the security of the elgamal signature scheme. Therefore, it is computationally infeasible for the attacker to get a pair (r_{u_i}, s_{u_i}) to satisfy $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$, as well as without knowing the secret exponent of x and a random number r_{ID}; it would be infeasible for the attacker to compute K_{u_i,v_i} and forge a valid SK in Step 3. A valid response can only be generated by the certificate ownership who knows the digital signature of the SDCM. In summary, the security of the unforgeability of our proposed protocol is provided through combination of the security of the elgamal signature scheme (the discrete logarithm) and one-way hash function. Therefore, the proposed user authentication and key establishment protocol is secure against forgery attacks.

6) **Resistance to the user impersonation attack (Nontransferability):** For an entity called Alice, the compromise of an entity Alice's long-term private key will allow an adversary to impersonate Alice, but it should not enable the adversary to impersonate other entities to Alice, Because adversary couldn't use K_{u_i,v_i} to compute the identical session key as same as that of Alice. As due to the SDCM, a valid response SK can only be generated by a certificate owner who knows the secret digital signature component x such that of a random challenge selected by the verifier, or by who knows generator g to acquire (r_{u_i}, s_{u_i}). As the verifier selects a random challenge each time, the response is only valid for a one-time authentication. Since the digital signature of a SDCM is never passed to the verifier, the verifier cannot pass the complete SDCM to any third party. There is no privacy access problem in our protocol. Therefore, a valid response SK cannot be transferred into a response of another verifier's challenge. Also, our protocol enables a certificate owner to be authenticated and one-time shared secret keys K_{u_i,v_i} and $c_{v_i} = r_{u_i}^{h_{v_i}\|x \mod p}$ be established between u_i, who knows s_{u_i} such that $s_{u_i} = g^{r_{ID}+h(k_{pub}\|x)} \mod p$, and the verifier v_i through the authentication protocol. The former is used to generate the SK, and the latter is established shared secret key $d_{ID} =< h_{ID}, s_{ID} >$ between u_i and v_i, where be secret. Now if the adversary can catch the authentication information, and then the spoofing attack can be done, However adversary is unable to extract any of the nonce values from the eavesdropped login request. Since, the computation of discrete logarithm problem and so inverse of hash function is infeasible. Therefore, the proposed protocol is secure against user impersonation attack.

7) **Resistance to the modification attack:** In the proposed protocol, each authentication message is supported via a new secret randomized number $r_{ID} \in [1, p-1]$ and the public key $k_{pub} \in [1, p-2]$ and the secret private key $x = k_{pub}^{r_{ID}} \mod p$ which accompanied by a discrete logarithm problem and one-way hash function. Without this secret components, the attacker is unable to calculate the correct hash function value for certificating the valid pair (r_{u_i}, s_{u_i})

and parameters SK, $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$. Hence, the proposed protocol is resistance to the modification attack.

8) **Resistance to disclosure secret private key:** In Step 1, while the certificate owner presents s_{u_i} to the verifier v_i; the computation of secret x from s_{u_i} is infeasible since computation of x from the s_{u_i} is a discrete logarithm problem and so inverse of hash function. In Step 2, even if the secret private key x is disclosed, then the attacker would not be able to retrieve the valid pair (r_{u_i}, s_{u_i}), K_{u_i,v_i} and $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$ from x. Since, the computation of discrete logarithm problem and so inverse of hash function is infeasible. Also, the verifier cannot obtain the secret component x.

9) **Resistance to the server spoofing attack:** In this type of attack, an adversary cannot masquerade as a legal server since he cannot calculate the valid pair (r_{u_i}, s_{u_i}), K_{u_i,v_i}, SK and $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$ without first identifying the secret private key x, random number r_{ID}, and c_B. Therefore, the server would not be able to compute valid pair (r_{u_i}, s_{u_i}) without identifying the secret private key x, random number r_{ID}, and c_{v_i}. In addition, the shared session key between u_i and v_i same $d_{ID} =< h_{ID}, s_{ID} >$ is different for the same user at different sign-in sessions. Even if the secret private key x is prevail from Server S. The attacker cannot retrieve the valid pair (r_{u_i}, s_{u_i}), K_{u_i}, v_i and $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$ from x. Since, the computation of discrete logarithm problem and so inverse of hash function is infeasible. Hence the proposed scheme is secure against the server spoofing attack and it is breakable only by a legitimate user, Since the mutual authentication process prevents the spoofing attack completely.

10) **Resistance to the parallel session attack:** If the attacker can masquerade as legitimate user u_i by simply replaying a previously intercepted request massage

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{v_i}\right)$$

with in the valid time frame window. But attacker cannot compute the knowledge massage, whereas attacker cannot directly corrupt shared secret session key d_{ID} between u_i and v_i in the next step where

$$h_{ID} = H\left(c_{v_i} \| c_{u_i} + x\right), S_{ID} = r_{u_i}^{h_{u_i} s_{u_i} + x} \mod p$$

since knowledge massage does not contains any information to construct next process. So, the security of the proposed scheme certification message against the parallel attack would depend on the complexity of the logarithmic calculations over one-way hash function, discrete logarithm problem and the Diffie–Hellman key agreement protocol.

11) **Resistance to the insider attack:** If an attacker can masquerade as an immune insider server; and also, he obtains the confidential information $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$, then he would not be able to extract similar sensitive information $K_{u_i,v_i} = c_{v_i}^{(r_{ID}+k_{pub})x + h\left(r_{u_i} \| s_{u_i}\right)} \mod p$ and

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right).$$

Because hash function inverse and solving a discrete logarithm problem are computationally infeasible without knowing secret private key x and session key $d_{ID} =< h_{ID}, s_{ID} >$. In addition, the secret session key agreement acts against the insider attack procedures.

12) **Resistance to the replay attack:** (Re-execution Attack): Suppose the attacker has managed to impersonate the sign-in request message to replay the same sign-in message

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right)$$

to the server or by simply replaying the previously intercepted message confidential information $g^{m'_{u_i}} = x^{r_{u_i}} s_{u_i} \mod p$. However, he cannot derive the valid parameters (r_{u_i}, s_{u_i}). Because, it would not be easy for the server to discover the replay attack through examining the protocol combines with the random numbers and timestamp. The standard ways of preventing such online attack in practice are to either limit the number of failed runs, if the attacker re-executes an old message on the part of the server, then the server can easily discover the re-execution attack by including random number and timestamp in the login message. Hence the proposed scheme is protected from the replay attack.

13) **Resistance to the online dictionary attack** The online dictionary attack is very powerful since it can be performed online; therefore, the attacker does need to interact with the legitimate entities and can use a lot of computing power. The standard ways of preventing such online attacks in practice are to either limit the number of failed runs. Therefore, the online dictionary attack is easier to detect and limit, thus we claim that our improved scheme is secure in the general case.

14) **Resistance to the denial-of-service attack:** Suppose the attacker wants to send login request massage

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right)$$

continuously to keep server busy or the attacker resends the valid message that is sent previously in order to disturb or redirect the traffic flow, this attack is known as denial-of-service attack. The same process is repeated continuously by many other adversaries

to overload the server. Hence, this process holds the server accessibility for the valid users. But attacker cannot send login request massage; because knowledge massage is compute in the valid time frame window.

15) **Resistance to the man-in-the-middle attack:** An attacker interposes itself between two entities. Thus, the attacker can intercept, modify, inspect, or drop the login request massage

$$S_K = H\left(r_{ID}, x, k_{pub}, K'_{u_i,v_i}, d_{ID}, c_{v_i} \| c_{u_i}\right).$$

However, he would not be able to extract similar sensitive information. Because, hash function inverse and solving a discrete logarithm problem are computationally infeasible without knowing secret private key x and session key $d_{ID} =< h_{ID}, s_{ID} >$ In addition, the secret session key agreement acts against the man-in-the-middle attack procedures. In next section, we evaluate the performance and cost of proposed protocol under various scenarios.

5 Performance and Cost Analysis

and outputs of the hash function are all 160 bits long. A paper acknowledgment can be used as an user's authentication operator, but a public-key digital certificate cannot be used as an authentication operator in network applications. This is because a paper certificate cannot be easily forged and copied or counterparted, but a public-key digital certificate can be easily registered and restarted.

In proposal scheme, the owner of a SDCM never needs to disclose the digital signature of theSDCM in plaintext to the authenticator. Instead, the SDCM owner needs that proves knowledge of the digital signature by responding to the verifier's challenge. Therefore, the SDCM is an efficient knowledge of the digital signature that performs the authentication procedure of user locally and has low computational cost.

The major drawbacks of corresponsive schemes are demonstrate which they have higher computation and communication costs; due to the usage of rabin's public-key cryptosystem. Some of compared schemes cannot provide a function for session key agreement and ability of anonymity. Consequently, this schemes cannot prevent the insider attack. In contras, by performance and security analysis, the proposed scheme is shown to be very efficient based on having both properties of storage capacity and computation cost; and also it is superior to other schemes according to less computational complexity. Since, proposed protocol does not use complex public key cryptosystem. Furthermore, the resulting scheme enjoys many interesting properties and functionalities, such as anonymity (identity protection), as well as extremely low computation and communication cost, no verification table, mutual authentication, session key agreement. So, we demonstrate that our scheme can resist various attacks and provides better performance than existing schemes.

6 Conclusion

In this paper, we have proposed an effective and secure protocol (SDCM) based on the combination of the Discrete Logarithm (DL) problem, the Diffie-Helman Assumption (DHA) and a secure one-way hash function for key establishment and user authentication; consequently for verifying digital signatures. We also show that (SDCM) satisfies almost all of the essential security requirements as it is secure in the escrow mode against most of the well-known attacks with reasonable computational cost. Furthermore, by comparing the proposed scheme with other similar schemes, we find out that the efficiency of the proposed protocol was very high toward the previous proposed protocols that are vulnerable to different attacks. Whereas, the proposed protocol had a more effective storage method, a shorter run time, a lower computation and communication cost, and also it is more suitable for real-life applications. Also, it is not involved in any time consuming modular exponential computing and so reduces the server over head of maintaining large user data for authentication.

References

[1] E. Biham, D. Boneh, and O. Reingold, "Breaking generalized Diffie-Hellman modulo a composite is no easier than factoring," *Information Process Letters*, vol. 70, pp. 83-87, 1999.

[2] D. Chaum, *Private Signature and Proof Systems*, Patents, US5493614A, United States, 1996.

[3] D. Chaum and H. van Antwerpen, "Undeniable signatures," in *Advances in Cryptology (Crypto'89)*, Lecture Notes in Computer Science, vol. 435, pp. 212-217, Springer, 1989.

[4] W. Diffle and M. E. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, pp. 644-654, 1976.

[5] M. S. Hwang, J. W. Lo, S. C. Lin, "An efficient user identification scheme based on ID-based cryptosystem", *Computer Standards & Interfaces*, vol. 26, no. 6, pp. 565–569, 2004.

[6] M. Jakobsson, K. Sako, and R. Impagliazzo, "Designated verifier proofs and their applications," in *Advances in Cryptology (EUROCRYPT'96)*, Lecture Notes in Computer Science, vol. 1070, pp. 143-154, Springer, 1996.

[7] F. Laguillaumie and D. Vergnaud, "Designated verifier signatures: Anonymity and efficient construction from any bilinear map," in *International Conference on Security in Communication Networks*, Lecture Notes in Computer Science, vol. 3352, pp. 105-119, Springer, 2004.

[8] C. C. Lee, M. S. Hwang, L. H. Li, "A new key authentication scheme based on discrete logarithms", *Applied Mathematics and Computation*, vol. 139, no. 2, pp. 343-349, July 2003.

[9] Y. Li, W. Susilo, Y. Mu, and D. Pei, "Designated verifier signature:Definition, framework and new constructions," *Ubiquitous Intelligence and Computing*, vol. 4611/2007, Springer, 2007.

[10] I. C. Lin, M. S. Hwang, C. C. Chang, "A new key assignment scheme for enforcing complicated access control policies in hierarchy", *Future Generation Computer Systems*, vol. 19, no. 4, pp. 457–462, May 2003.

[11] I. C. Lin, C. C. Chang, M. S. Hwang, "Security enhancement for the simple authentication key agreement algorithm", in *Proceedings 24th Annual International Computer Software and Applications Conference (COMPSAC'00)*, 2000.

[12] H. Lipmaa, G. Wang, and F. Bao, "Designated verifier signature schemes: Attacks, new security notions and a new construction," *32nd International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science, vol. 3580, Springer, 2005.

[13] A. Mihara and K. Tanaka, "Universal designated-verifier signature with aggregation," in *Third International Conference on Information Technology and Applications (ICITA'05)*, 2005.

[14] H. H. Ou, M. S. Hwang and J. K. Jan, "A cocktail protocol with the authentication and key agreement on the UMTS", *Journal of Systems and Software*, vol. 83, no. 2, pp. 316-325, Feb. 2010.

[15] J. Ren and L. Harn, "Generalized ring signatures," *IEEE Transactions on Dependable and Secure Computing*, vol. 5, no. 3, pp. 155-163, 2008.

[16] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communnitions of ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978.

[17] R. Rivest, A. Shamir, and Y. Tauman, "How to leak a secret," in *Advances in Cryptology (ASIACRYPT'01)*, Lecture Notes in Computer Science, vol. 2248, Springer, 2001.

[18] S. Saeednia, S. Kremer, and O. Markowitch, "An efficient strong designated verifier signature scheme," in *International Conference on Information Security and Cryptology (ICISC'03)*, Lecture Notes in Computer Science, vol. 2836, pp. 40-54, 2003.

[19] C. Schnorr, "Efficient signature generation by smart cards," *Journal of Cryptology*, vol. 4, no. 3, pp. 161-174, 1991.

[20] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Advances in Cryptology (Crypto'84)*, Lecture Notes in Computer Science, vol. 196, pp. 47-53, Springer, 1984.

[21] R. Steinfeld, L. Bull, H. Wang, and J. Pieprzyk, "Universal designated verifier signatures," in *Advances in Cryptology (ASIACRYPT'03)*, Lecture Notes in Computer Science, vol. 2894, pp. 523-542, Springer, 2003.

Biography

Javad Saadatmandan received his Ph.D from the Department of Mathematics and Computer Sciences at Qom University in Qom, Iran. He is presently an assistant professor of mathematics at IAU, Qom, Iran. His research interest include cryptographic protocols and wavelet transforms.

Amir Hossein Rahimi received the B.Sc. degree in 2009 in applied mathematics from the University of Arak, Iran, and the M.Sc. degrees in 2014 in cryptography engineering from Malek-Ashtar University of Technology Isfehan. His current research interests include areas of communication theory, information security, cryptography, smart grid, steganography, digital signature and authentication protocols. He has published more than 10 papers in the fields mentioned. Also, he have been teaching mathematical sciences in universities of Qom province, Iran form 2015 until now. Email:Amir.Rahimi361@Gmail.Com.

Research on Network Security Intrusion Detection System Based on Machine Learning

Yin Luo

(Corresponding author: Yin Luo)

Sichuan TOP IT Vocational Institute, China

No. 2000, Xiqu Avenue, High-tech District, Chengdu, Sichuan 611743, China

Email: yinjielan@21cn.com

(Received May 17, 2019; Revised and Accepted June 20, 2020; First Online Apr. 25, 2021)

Abstract

This paper mainly analyzed the application of the machine learning method in the intrusion detection system (IDS). The support vector machine (SVM) algorithm parameters were improved by the adaptive particle swarm optimization (APSO) algorithm and the APSO-SVM algorithm, which obtains for intrusion detection. In feature selection, we will compare the proposed method with Relief and InfoGain methods. Experiments were carried out on the KDD CUP 99. The results showed that the proposed method greatly reduced the running time of the algorithm and improved the performance to a certain extent after the dimensionality reduction of features selected by Relief and InfoGain. Comparatively speaking, the feature extracted by Relief performed better in the algorithm. The comparison between SVM, particle swarm optimization (PSO)-SVM, and APSO-SVM algorithms demonstrated that the APSO-SVM algorithm had higher accuracy and lowered false alarm rate and missing alarm rate, i.e.,,, it had better performance in intrusion detection. The results show that the machine learning method is effective on IDS, which contributes to the further realization of network security.

Keywords: Intrusion Detection System; Machine Learning; Network Security; Particle Swarm Optimization; Support Vector Machine

1 Introduction

With the popularity of the network [10], it not only facilitates people's study, work and life but also brings a lot of security problems. The emergence of various viruses, vulnerabilities, and attacks poses a great threat to the security of individuals, enterprises, and even the country. Network security generally needs to ensure the integrity, availability, confidentiality, and controllability of information and prevent information from being leaked, tampered, or destroyed [17].

The current technologies used include access control [5], firewall [20], identity authentication [7], data encryption [24], *etc.*, but they can only carry out passive defense, not real-time monitoring.; therefore, intrusion detection system (IDS) [23] appears. IDS can detect potential threats in time by analyzing network information [13], which has been widely concerned by researchers. Kang *et al.* [12] designed an IDS using a deep neural network (DNN) and used a deep belief network (DBN) to pre-train the initial parameters of DNN [2, 6]. Through experiments, they found that the method had a high detection rate and could make a real-time response to attacks.

Pham *et al.* [18] designed a lightweight IDS, which converted the original network traffic into image data and then used a convolutional neural network (CNN) for detection. The experiment showed that the improved method could achieve 95% accuracy. Muhammad *et al.* [16] designed an IDS based on DNN, reduced the feature width using the stacked automatic encoder (AE), carried out experiments on KDD CUP 99, NSL-KDD, and AWID datasets, and found that the accuracy of the improved method reached 94.2%, 99.7%, and 99.9%, respectively.

Lee *et al.* [14] designed a hybrid IDS combining the C4.5 decision tree with weighted K-means, verified the method, and found that it had a detection accuracy of 98.68%. In this study, the support vector machine (SVM) algorithm in machine learning was studied and improved by combining the particle swarm optimization (PSO) method. The method of feature dimension reduction was also analyzed, and the proposed method was tested on the KDD CUP 99. The present study is conducive to the further development of IDS and better realization of network security.

2 Intrusion Detection System

According to different data sources, IDS can be divided into (1) the host-based IDS, which finds out the intrusion behavior and respond through the analysis of the system and application log, but it can not detect other hosts, not

suitable for the current complex network environment; (2) the application-based IDS, which is a refinement of the host-based IDS, mainly for an application; (3) the network-based IDS, which determines whether there are threats through the analysis of network packets, and it is most widely used because of its strong real-time performance and fast detection speed.

According to different detection principles, IDS can be divided into three categories. When the detection principle is anomaly detection [3], the IDS analyzes the characteristics of normal behavior, establishes a model, and determines an intrusion if there is a big difference between the behavior and normal behavior. The common methods include multivariate analysis and neural networks [1]. When the detection principle is misuse detection [8], the IDS analyzes the characteristics of abnormal behavior, establishes a feature library, and determines there is an intrusion if the feature that conforms to the feature library is detected. The common methods include pattern matching, expert system, *etc.*

The essence of IDS is a process of classification, i.e., distinguishing normal behaviors from intrusion behaviors. Therefore, the machine learning method has high availability in IDS [19]. This study mainly analyzes the application of the SVM method.

3 Feature Dimension Reduction Method

In intrusion detection, to reduce the dimension of data and improve the speed of detection, it is necessary to select features. A subset containing M features is selected from a set containing N features ($M < N$) to make the classification performance the best. Two common methods are introduced here.

Relief [22]: The method considers that good features can make the samples of the same class closer to each other and make the samples that do not belong to the same class farther away. The correlation between a feature and a class is represented by weight, and the weight lower than a threshold is removed. It is assumed that sample T is randomly selected from sample set S, and then samples X and Y are also selected and made closest to T; moreover, X and Y belong to the same class, and Y and T belong to different classes. For feature F, the distance between X and T and between Y and T on the feature is calculated. If the former is smaller than the latter, it indicates that the degree of distinction of the feature is good and the weight can be improved; otherwise, the weight is reduced. The updating formula of the weight is written as:

$$W(F) = W(F) + \frac{D(F,T,Y)}{n} - \frac{D(F,T,X)}{n}$$

where $W(F)$ refers to the weight of feature F. The distance between two samples and feature F can be written as:

$$D(F,I_1,I_2) = \frac{|value(F,I_1) - value(F,I_2)|}{\max(F) - \min(F)}$$

where $value(F,I_i)$ refers to the value of sample I_i on F. After n cycles, the feature with a larger weight has better classification performance, which can be used for intrusion detection.

InfoGain [4]: This method selects samples based on information entropy. It is assumed that there are s samples in sample set S, which can be divided into m classes, and class C_i contains S_i samples. The information entropy can be written as:

$$E(C) = -\sum_{i=1}^{m} \rho(C_i) \log_2 \rho(C_i)$$

where $\rho(C_i)$ refers to the probability that any sample belongs to class C_i, $\rho(C_i) = \frac{s_i}{S}$, and $E(C)$ represents the degree of uncertainty of classifying samples in C into m classes.

For feature F, when it is used for classifying S, the degree of uncertainty can be written as: $E(C|F)$. Suppose $F = \{F_1, F_2, \cdots, F_v\}$, then S is divided into: $S = (S_1, S_2, \cdots, S_v)$, and the conditional entropy can be obtained:

$$E(C|F) = \sum_{j=1}^{v} \rho(F_j) E(C|F = F_j),$$

where $\rho(F_j)$ refers to the occurrence probability of feature F_j. When the value of F is F_j, the conditional entropy can be written as:

$$E(C|F = F_j) = -\sum_{i=1}^{m} \rho_{i_j} \log_2 \rho_{i_j}$$

where $\rho_{i_j} = \frac{s_{i_j}}{s_j}$. After substitution, there is: .

$$E(C|F) = \sum_{j=1}^{v} \frac{S_{1_j} + S_{2_j} + \cdots + S_{m_j}}{S} (-\sum_{i=1}^{m} \frac{S_{i_j}}{S_j} \log_2 \frac{S_{i_j}}{S_j}).$$

The information gain of F is defined as $G(F)$, $G(F) = E(C) - E(C|F)$. The larger the $G(F)$ is, the larger the degree of distinction of F is, and the larger the contribution to the sample division is. In feature selection, the feature with larger $G(F)$ is selected for intrusion detection.

4 SVM Based Intrusion Detection Algorithm

4.1 Principle of SVM Algorithm

The SVM algorithm is a typical machine learning method, which can divide the data into two classes. In intrusion detection, the SVM algorithm can distinguish the

normal behavior of the network and intrusion behavior, which has good usability. If there is a data set $S = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ and its classification hyperplane is $wx + b = 0$, the maximum class interval is calculated: $\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i$, such that $y_i[(wx_i + b)] - 1 + \xi_i \geq 0$, where C is the penalty factors and ξ_i is the slack variable ($\xi_i \geq 0$). The Lagrange factor is introduced to solve the above equation; then, $\min_a \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j y_i y_j K(x_i, y_j) - \sum_{i=1}^{N} a_i$, such that $\sum_{i=1}^{N} a_i y_i = 0$, where $K(x_i, x_j)$ is the kernel function. Finally, the classification function can be written as:

$$f(x) = sgn(\sum_{i=1}^{N} a_i y_i K(x_i \cdot x) + b).$$

In selecting kernel function, the radial basis function (RBF) with good nonlinear mapping ability is selected:

$$K(x_i, x_j) = \exp(-\frac{|x_i - x_j|^2}{\rho^2}$$

In the SVM algorithm, the performance of the algorithm is mainly related to two parameters: penalty factor C and kernel parameter ρ. It is an important problem for the SVM algorithm to find the best parameter value and make the performance of the algorithm the best.

The SVM algorithm is mainly used for binary classification. There are many kinds of intrusion behaviors. To solve the problem of multi-classification, it can be divided into multiple binary classification problems. In this study, the one vs. Rest (OvR) method is used. In each training, it is assumed that there are N classes, samples from one class were positive, and the other samples were negative. In the test, if only one classifier predicts positive, it can be used as the classification result; if multiple classifiers predict positive, the one with the highest confidence is selected. This method only needs to train N classifiers, which needs less time and space.

4.2 Parameter Optimization of the SVM Algorithm

For the parameter optimization of the SVM algorithm, an adaptive particle swarm optimization (APSO) algorithm was designed to select parameters. For the traditional PSO, the value of the inertia weight w has a great impact on the performance of the algorithm. In this study, the value of w is combined with the fitness value of particles. It is assumed that the relative variation rate of the fitness value of particles is: $k = \frac{f_i(t) - f_i(t-1)}{f_i(t-1)}$, where $f_i(t)$ refers to the fitness value of particle i at the t^{th} iteration. The adjustment formula of w is:

$$w_i(t) = (1 + e^{-k})^{-1}$$

The value of w is controlled in $(0, 1)$. When $k = 0$, $w_i(t) = 0.5$. With the increase of $f_i(t)$ value, the value of $w_i(t)$ also increases. Such a method can make the algorithm converge better.

5 Experimental Analysis

5.1 Experimental Data Set

Experiments were carried out on the KDD CUP 99, and 10% of data sets were selected, including the following four types of intrusion.

1) DOS, which makes the network unable to provide normal services, such as land, smurf, *etc.*

2) Probe, which monitors or scans ports to obtain open services, such as saint, ipweep, *etc.*

3) R2L, which is illegal access to remote machines, such as imap, multihop, *etc.*

4) U2R, which can make unauthorized users become privileged users, such as loadmodule, rootkit, *etc.*

In the selected data sets, the number of intrusion behaviors is shown in Table 1.

Table 1: Experimental data sets

Intrusion Behavior	Training Set	Testing set
Normal	97278	60593
Probe	4017	4166
DOS	391458	229853
U2R	52	228
R2L	1126	16189

In KDD CUP99, each record contained 41-dimensional features and intrusion categories, for example, 0, tcp, http, SF, 177, 1985, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 28, 119, 1.00, 0.00, 0.04, 0.04, 0.00, 0.00, 0.00, 0.00, normal. As the second, third, and fourth features were characters, they needed to be transformed into numbers. The second feature was represented by numbers 0-2. The third feature was represented by numbers 0-60. The fourth feature was represented by numbers 0-10. Then, all the values were normalized and transformed to numbers in the range of 0-2. The formula is:

$$x' = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min}$$

where $y_{\max}$ and $y_{\min}$ are the maximum and minimum values of normalization and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of feature attributes.

5.2 Evaluation Index

According to the confusion matrix, the performance of the APSO-SVM-based IDS was evaluated, as shown in Table 2.

The evaluation indexes include:

Table 2: Confusion matrix

		Classification Results	
		Normal Sample	Abnormal Sample
The Real Situation	Normal Sample	TP	FN
	Abnormal Sample	FP	TN

1) Accuracy (ACC):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

2) False positive rate (FPR):

$$FPR = \frac{FP}{FP + TN} \times 100\%$$

3) False alarm rate (FAR):

$$FAR = \frac{FN}{TN + FP} \times 100\%$$

5.3 Experimental Results

Firstly, two feature selection methods, Relief and InfoGain, were compared. For KDD CUP 99, the top ten features were selected as the input of IDS, and the SVM algorithm was taken as an example to operate ten times. The operation time of the algorithm is shown in Table 3.

It was seen from Table 3 that the operation time of the algorithm became significantly shorter. When the 41-dimensional feature was used as input, the operation time of the algorithm was more than 30 s. After feature selection by Relief and InfoGain, the input was a ten-dimensional feature, and the operation time of the algorithm was less than 20 s. When Relief, InfoGain, and 41-dimensional feature were used as inputs, the average operation time of the algorithm was 14.39 s, 18.78 s, and 37.24 s, respectively. The ten-dimensional feature selected by Relief reduced the operation time of the algorithm by 61.36%, and the ten-dimensional feature selected by InfoGain reduced the operation time by 49.57%. In the aspect of the operation time, the feature selection result of Relief was better.

The ten-dimensional features selected by Relief and InfoGain were used as input, respectively. The operation repeated ten times, and the average value was taken. The performance of SVM, PSO-SVM, and APSO-SVM algorithms in detecting intrusions was compared, and the results are shown in Figures 1 and 2.

It was seen from Figure 1 that the ACC of the three algorithms was 87.42%, 92.34%, and 97.68%, respectively, i.e., the ACC of the APSO-SVM algorithm was the highest, which was 11.74% higher than the SVM algorithm and 5.78% higher than the PSO-SVM algorithm. The FRP of the three algorithms was 2.33%, 1.34%, and 0.17%, respectively, and the FAR was 21.22%, 12.36%, and 7.68%, respectively. It was found that the FPR and FAR of the SVM algorithm significantly decreased after optimization by the PSO algorithm and further decreased after further improvement by the PSO algorithm.

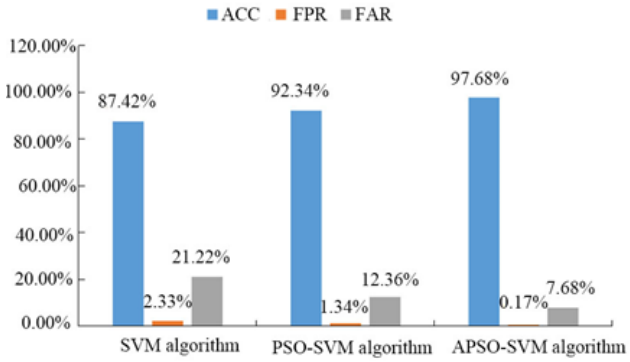


Figure 1: Comparison between algorithms when the feature selected by Relief is used

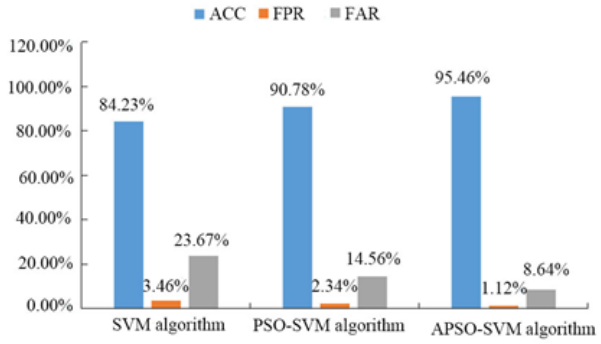


Figure 2: Comparison between algorithms when the feature selected by InfoGain is used

It was seen from Figure 2 that the ACC of the three algorithms was 84.23%, 90.78%, and 95.46%, respectively, the FPR was 3.46%, 2.34%, and 1.12%, respectively, and the FAR was 23.67%, 14.56%, and 8.64%, respectively. Compared with Figure 1, the ACC of the algorithm decreased, and the FPR and FAR increased, when the feature selected by InfoGain was used. It was concluded that the performance of the algorithm was better when the feature selected by Relief was used as the input.

Table 3: Comparison of the operation time of the algorithm

	1	2	3	4	5	6	7	8	9	10
Relief	13.24	14.68	13.68	15.12	14.11	15.18	14.26	15.41	13.96	14.21
InfoGain	19.33	19.12	18.78	18.59	19.03	18.72	19.33	17.64	19.21	18.07
41-dimensional Features	36.78	37.12	36.77	36.81	37.23	38.07	36.95	37.08	38.11	37.45

6 Discussion

The machine learning method is a simulation of human learning by computers [9], which is related to knowledge such as artificial intelligence, biology, and statistics. Its goal is to establish a learning machine from the existing data and classify or predict the unknown data. Up to now, it has been well applied in many fields, such as image processing [21], data classification [15], prediction [11], *etc.* This paper mainly analyzed the SVM algorithm in machine learning and its application in IDS.

Aiming at the problem of parameter optimization of the SVM algorithm, this paper selected the PSO algorithm and improved the SVM algorithm to obtain the APSO-SVM algorithm. Then, in feature selection, to reduce the feature dimension, the performance of Relief and InfoGain algorithms was compared, and the experiment was carried out on the KDD CUP 99 data set.

First of all, the features selected by Relief and InfoGain both significantly reduced the operation time of the algorithm, but the performance of Relief was better as it reduced the operation time of the SVM algorithm by 61.36%, greatly improving the efficiency of the algorithm. Then, in the aspect of the specific performance of the algorithm, when the feature selected by Relief was used, the accuracy of the algorithm became higher, and the false positive rate and false alarm rate became lower, which verified that Relief had a better performance in feature selection. Then, in the aspect of the optimization of the SVM algorithm, the performance of the algorithm significantly improved after optimization by the PSO algorithm and further improved after further optimization by the PSO algorithm. It was seen from Figure 1 that the ACC of the APSO-SVM algorithm was 5.78% higher, the FPR was 87.31% lower, and the FAR was 37.86% lower compared with the PSO-SVM algorithm. It was concluded that the APSO-SVM algorithm designed in this study presented an excellent performance in detecting intrusions.

Though this study has obtained some achievements from the research of the machine learning based-IDS, there are still some shortcomings. In future research, works, including studying more machine learning methods, verifying IDS in the real network environment, and further optimizing the performance of the SVM algorithm, need to be completed.

7 Conclusion

IDS was studied using the SVM algorithm in this paper, an APSO-SVM algorithm was designed for intrusion detection, and experiments were carried out on the KDD CUP 99. The results are as follows.

1) The features selected by Relief and InfoGain both reduced the operation time of the algorithm, and the performance of Relief was better.

2) In terms of accuracy, the performance of the feature selected by Relief was better than that by InfoGain, and the accuracy of the APSO-SVM algorithm was the highest, reaching 97.687%.

3) In terms of false positive rate and false alarm rate, the feature selected by Relief was better, and the false positive rate and false alarm rate of the APSO-SVM algorithm were lower.

It is concluded that the IDS that selects features with Relief and detects intrusions with the APSO-SVM algorithm has better performance, which can be further promoted and applied in practice.

References

[1] N. Abd, K. M. A. Alheeti, S. S. Al-Rawi, "Intelligent intrusion detection system in internal communication systems for driverless cars," *Webology*, vol. 17, no. 2, pp. 376, 2020.

[2] D. S. Abdul Minaam and E. Amer, "Survey on machine learning techniques: Concepts and algorithms," *International Journal of Electronics and Information Engineering*, vol. 10, no. 1, pp. 34–44, 2019.

[3] T. B. Adhi, R. Kyung-Hyune, "HFSTE: Hybrid feature selections and tree-based classifiers ensemble for intrusion detection system," *IEICE Transactions on Information & Systems*, vol. 100, no. 8, pp. 1729-1737, 2017.

[4] M. Alkasassbeh, "An empirical evaluation for the intrusion detection features based on machine learning and feature selection methods," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 22, pp. 5962-5976, 2017.

[5] M. Cheminod, L. Durante, L. Seno, F. Valenza, A. Valenzano, "A comprehensive approach to the automatic refinement and verification of access control

policies," *Computers & Security*, vol. 80, pp. 186-199, 2018.

[6] A. Dewanje and K. A. Kumar, "A new malware detection model using emerging machine learning algorithms," *International Journal of Electronics and Information Engineering*, vol. 13, no. 1, pp. 24–32, 2021.

[7] R. Divya, V. Vijayalakshmi, "Analysis of multimodal biometric fusion based authentication techniques for network security," *International Journal of Security & Its Applications*, vol. 9, no. 4, pp. 239-246, 2015.

[8] S. Elhag, A. Fernández, A. H. Altalhi, S. Alshomrani, "A multi-objective evolutionary fuzzy system to obtain a broad and accurate set of solutions in intrusion detection systems," *Soft Computing*, vol. 23, no. 4, pp. 1321-1336, 2019.

[9] M. I. Jordan, T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-260, 2015.

[10] A. Kak, "Computer and network security," *Friend of Science Amateurs*, vol. 31, no. 9, pp. 785-786, 2017.

[11] U. Kanewala, J. M. Bieman, A. Ben-Hur, "Predicting metamorphic relations for testing scientific software: A machine learning approach using graph kernels," *Software Testing Verification & Reliability*, vol. 26, no. 3, pp. 245-269, 2016.

[12] M. J. Kang, J. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," *Plos One*, vol. 11, no. 6, pp. e0155781, 2016.

[13] N. Khamphakdee, N. Benjamas, S. Saiyod, "Improving intrusion detection system based on snort rules for network probe attacks detection with association rules technique of data mining," *Journal of ICT Research & Applications*, vol. 8, no. 3, pp. 234-250, 2015.

[14] W. Lee, S. Oh, "Efficient feature selection based near real-time hybrid intrusion detection system," *KIPS Transactions on Computer and Communication Systems*, vol. 5, no. 12, pp. 471-480, 2016.

[15] N. Milosevic, A. Dehghantanha, K. K. R. Choo, "Machine learning aided Android malware classification," *Computers & Electrical Engineering*, vol. 61, pp. 266-274, 2017.

[16] G. Muhammad, M. S. Hossain, S. Garg, "Stacked autoencoder-based intrusion detection system to combat financial fraudulent," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1-1, 2020.

[17] E. U. Opara, O. A. Soluade, "Straddling the next cyber frontier: The empirical analysis on network security, exploits, and vulnerabilities," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 10–18, 2015.

[18] V. Pham, E. Seo, T. M. Chung, "Lightweight convolutional neural network based intrusion detection system," *Journal of Communications*, vol. 15, no. 11, pp. 808-817, 2020.

[19] N. A. H. Qaiwmchi, H. Amintoosi, A. Mohajerzadeh, "Intrusion detection system based on gradient corrected online sequential extreme learning machine," *IEEE Access*, vol. PP, no. 99, pp. 1-1, 2020.

[20] X. Song, "Firewall technology in computer network security in 5G environment," *Journal of Physics Conference Series*, vol. 1544, pp. 012090, 2020.

[21] S. A. Tsaftaris, M. Minervini, H. Scharr, "Machine learning for plant phenotyping needs image processing," *Trends in Plant Science*, vol. 21, no. 12, pp. 989-991, 2016.

[22] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, pp. S1532046418301400, 2017.

[23] G. B. White, E. A. Fisch, U. W. Pooch, "Cooperating security managers: a peer-based intrusion detection system," *IEEE Network*, vol. 10, no. 1, pp. 20-23, 2015.

[24] R. Yadav, V. Kapoor, "A hybrid cryptography technique for improving network security," *International Journal of Computer Applications*, vol. 141, no. 11, pp. 25-30, 2016.

Biography

Luo Yin, born on July 20, 1982, holds a master's degree and is an associate professor of Sichuan top information technology vocational college. He is interested in big data and artificial intelligence.

Artificial Neural Network Model for Decreased Rank Attack Detection in RPL Based on IoT Networks

Musa Osman[1], Jingsha He[1], Fawaz Mahiuob Mohammed Mokbal[1,2], and Nafei Zhu[1]
(Corresponding author: Nafei Zhu)
Faculty of Information Technology, Beijing University of Technology[1]
Beijing 100124, China
Faculty of Engineering and Information Technology, Taiz University[2]
Taiz, Republic of Yemen
Email:znf@bjut.edu.cn
(Received March 26, 2020; Revised and Accepted Nov. 10, 2020; First Online Apr. 17, 2021)

Abstract

Internet of Things (IoT) cyber-attacks are growing day by day because of the constrained nature of the IoT devices and the lack of effective security countermeasures. These attacks have small variants in their behavior and properties, implying that the traditional solutions cannot detect the small mutant variations. Therefore, a robust detection method becomes necessary. One of the common attacks is routing protocol for low power and lossy network attacks, which has not been well investigated in the literature. In this paper, we propose an artificial neural network (ANN) model for detecting decreased rank attacks, which includes three phases: Data pre-processing, Feature extraction using random forest classifier, and an artificial neural network model for the detection. The proposed model has been tested in multi and binary detection scenarios using the IRAD dataset. The results obtained are promising with accuracy, precision, false-positive rate, and AUC-ROC scores of 97.14%, 97.03%, 0.36%, and 98%, respectively. The proposed approach is efficient and outperforms previous methods of precision, recall, and F-score metrics.

Keywords: 6LoWPAN; Attacks; Detection Technique; IoT; RPL; Security

1 Introduction

Internet of Things (IoT) is a system of interconnected devices, machines and related software services. The core elements of IoT are the sensors and actuators, which are used to collect and actuate data. These devices use many communication techniques such as Bluetooth, WiFi, LoRa, IEEE802.15.4, *etc.* Many technologies are classified under IoT such as smart homes, smart cities, smart healthcare, *etc.* Moreover, IoT is expected to be the next generation of worldwide network, where a large number of things are expected to be part of the Internet [9, 22, 23].

To make the things a part of the Internet, a routing protocol for low power and lossy network (RPL) has been developed by IETF [5] to perform routing over IPv6 over Low-power wireless personal area network (6LoWPAN). Furthermore, RPL forms the topology in a mathematical graph model which is known as a directed acyclic graph (DAG) without directed cycles. In a DAG, all nodes are connected in a way that the traffic is routed through the nodes via one or more routes and there is no cyclic round within the DAG. In the DAG, there are one or more destination oriented directed acyclic graph (DODAG) in which there will be one node named the sink node or the border router (6BR) [6]. Moreover, within the DAG, several instances may also exist and each instance may have one or more DODAG. Figure 1 shows the RPL network with one instance and two DODAG in each instance. Besides, RPL DODAG is constructed by four control messages, DODAG Information Solicitation (DIS), DODAG Information objects (DIO), Destination Advertisement Objects (DAO), and DODAG acknowledges (DAO ACK).

The DIO message is advertised by a root node or a node in a DODAG which contains such information as RPL instance ID, DODAGs ID, DODAGs version number, RPL mode of operation, the rank of sending node, and the objective function used, and other control information. If the sender is a root, then the DIO contains information to create the DODAG. If the sender is not a root node, it means that this node wants to join the DODAG [17]. When a node wishes to join a DODAG and, for a while, doesn't receive any DIO message, it starts to broadcast DIS messages looking for an existing DODAG. While DIO

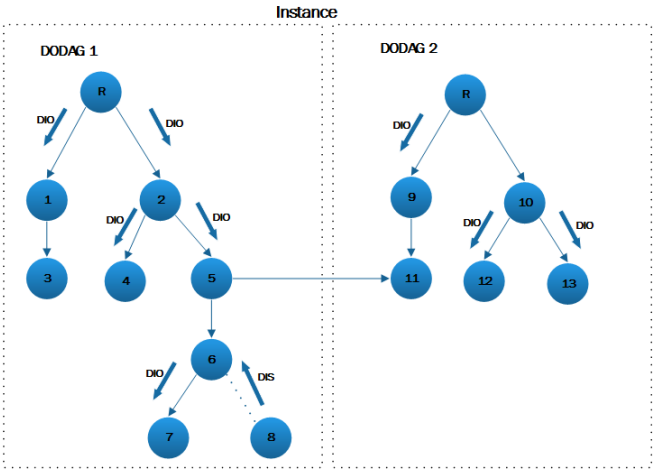


Figure 1: RPL network (one instance two DODAG)

and DIS are used to maintain the upward path towards the root node, DAO is used to construct the downward path from the root to the leaf nodes or children [6, 21].

The RPL protocol is vulnerable to different types of attacks, some of which are originated from wireless sensor networks (WSN) and traditional networks [10, 11]. Moreover, RPL has its specific vulnerabilities [6]. Consequently, these attacks require robust and versatile techniques to protect valuable resources and data. Rank attack is one type of RPL attacks in which the malicious node advertises a false rank which will be the best parent for the benign nodes. Literatures [7, 17] investigated the rank attack as three types of attacks: Decreased rank attack (DR) in which the malicious node advertises a rank that is lower than the other nodes, which will make it the best parent node [12], resulting in attracting a large part of the traffic, increased rank attack (IR) in which the node that in reality is close to the root node advertises a higher rank, forcing nodes to choose other parents [7], and worst parent attack in which the malicious node advertises its correct rank but selects the worst parent for itself. Attacks against routing protocol in IoT need more attention from research to better protect IoT networks and devices from such attacks. Some research showed that lightweight solutions are the best solutions. From our point of view, machine learning techniques provide a viable approach for detecting these attacks because IoT devices generate a tremendous amount of data, rendering a robust detection mechanism a necessity. In this paper, a robust artificial neural network-based multilayer perceptron (MLP) model is proposed to detect RPL attacks such as decreased rank attack, along with feature selection using the random forest (RF) classifier [2]. IRAD dataset is used as a benchmark [23] for training and validation of the proposed model.

The proposed model has successfully surpassed several tests on the held-out testing dataset and achieved promising results with accuracy, precision, detection probabilities, false-positive rate, false-negative rate, and area under the ROC curve (AUC) scores of 97.01%, 97.03%, 97.01%, 4.6%, 1.6% and 98%, respectively.

The rest of this paper is organized as follows. Section 2 reviews some related work. Section 3 describes the proposed method. Section 4 presents the experimental results and Section 5 shows the conclusion and future work.

2 Related Work

Attacks against IoT devices have increased significantly, affecting the availability of both traditional networks and IoT devices. Recent research on RPL attacks has been focused on the detection and mitigation of different types of attacks. Furkan *et al.* [23] prepared a real IoT dataset using the COOJA simulator called the IoT Routing Attack Dataset (IRAD) which contains three types of attacks: Version number attacks (VN), decreased rank attacks (DR) and hello flood attacks (HF). They employed artificial neural networks model for classification to obtain good accuracies like 94.9% in the DR model, 99.5% in the HF model and 95.2% in the VN model. Another dataset was generated by Verma *et al.* [20] for IoT and was named RPL-NIDDS17 which is specially developed for IoT routing attacks. It contains seven types of routing attacks such as clone ID, hello flooding, local repair, selective forwarding, sinkhole, blackhole and sybil in the IoT field comprised of 20 features and 2 labeling attributes. The authors used five deep learning techniques to evaluate the complexity of this dataset, such as naive Bayes (NB), decision tree (DT), logistic regression (LR), expectation-maximization (EM) clustering and artificial neural networks (ANN), and achieved accuracies of 80.71%, 94.07%, 79.79%, 77.17% and 93.99%, respectively. Ahmet *et al.* [1] proposed a lightweight technique to mitigate the effect of version number attacks in RPL by using two techniques. One is the elimination of any version number updates (VN) coming from leaf nodes and the other, called a shield, makes the node change the VN depending on its neighbors with a better rank. It was claimed that the delay caused by the attacker can be shortened up to 87% and the average power consumption can be reduced up to 63%. In addition, the control message overhead can be lowered up to 71% and the data packets delivery ratio can be increased up to 86%. Mayzaud *et al.* [13] investigated the effect of VN attacks in a network with 20 nodes. In the work, the authors claimed that the control overhead can be increased by up to 18 times.

The authors also reported that the delivery ratio of packets was reduced by 30% and the location of the attacker could affect the consistency of the network. If the attacker is close the root, the effect of the attack is less than if it is far away from the root. Nikravan *et al.* [15] first analyzed the RPL routing protocol and proposed a lightweight technique to mitigate VN attacks. The technique relies on using the identity based offline/online signature (IBOOS) scheme which is divided into two phases,

Table 1: Sample of the dataset features and instances

No	Time	Source	Destination	Length	Info	Trans Rate(per 1000 ms)	Reception Rate(per 1000 ms)	TR/ RR	Sources Count Per Sec	Destination Count Per Sec
1	00.00	1	9999	64	2	0.039	0.195	0.2	39	195
2	0.003289	1	9999	64	2	0.039	0.195	0.2	39	195
3	0.006555	1	9999	64	2	0.039	0.195	0.2	39	195
4	0.009851	1	9999	64	2	0.039	0.195	0.2	39	195
5	0.013153	1	9999	64	2	0.039	0.195	0.2	39	195
6	0.016411	1	9999	64	2	0.039	0.195	0.2	39	195

i.e., the online phase where most of the heavy computational operations are performed and the online phase where it performs a lightweight scheme. Snehal *et al.* [3] designed an IDS for detecting wormhole attack using received signal strength indicator (RSSI) which is converted to distance and by using Euclidean distance method so as to compare the distance between a node and its neighbors. If the distance is more than the transmission range of the node, it is identified as an attacker node. The proposed IDS has good results in a small number of nodes. The detection of rank attack has also been investigated by Usman *et al.* [18] through using a root-based statistical intrusion detection system to detect rank attacks by applying statistical algorithms to comparing the rank of the nodes.

Under normal conditions, the number of nodes is small and there is no mobility and the model can achieve high accuracy. However, when the number of nodes increases, the accuracy decreases. Kfoury *et al.* [8] proposed an IDS using a self-organizing map to detect three types of RPL attacks: Hello flood, sinkhole and version number attacks. However, there is no clear implementation of this IDS method and the power consumption is not clear from the study. Dvir *et al.* [4] proposed an IDS based on cryptographic techniques to avoid false rank and claimed that these techniques had high computational overhead which would affect the IoT device's power consumption. Besides, it is also vulnerable to other attacks such as those discussed in paper [16].

3 Proposed Methodology

The proposed model named multi-layer RPL attack detection (MLRPL) is composed of three modules that work together to perform the detection of RPL attacks. The first module is data pre-processing, the second is feature selection and the third is the artificial neural network for attack detection.

3.1 Dataset

The dataset used to evaluate the MLRPL model is the IRAD dataset [23] which consists of three types of attacks: VN attack, DR attack and HF attack. Each attack appears in a separate CSV file consisting of 18 features with 1048575 samples in total (579944 malicious and 468630 benign). The label feature is binary (0 is benign and 1 is decreased rank attack). Table 1 shows the sample records of the IRAD DR attack dataset. Then the dataset of the decreased rank attack and the version number attack is combined in one dataset named RPL attack dataset for categorical classification. The new dataset consists of 2997150 records and 18 features. The label feature is categorical (0 means benign, 1 means DR attack and 2 means VN attack) encoded using a one-hot encoder. Table 2 shows a subdivision of the RPL dataset.

Table 2: Subdivision of the RPL attacks dataset

Category	Malicious	Benign	Total
DR attack	468631	579944	1048575
VN attack	503326	545249	1048575
Total	971957	1125193	2097150

3.2 Features Selection

To improve the performance of the model, feature selection is used to extract the most important features from the dataset. The feature selection process contributes most in the prediction of the model, moreover, it reduces the training and validation time and increases the performance of the model. In general, pre-processing is performed applying the dataset before running the artificial neural networks model. To identify features of high importance, information gain is used to evaluate the gain for each variable and a random forest (RF) classifier is trained on the entire dataset. By using entropy shown in Equation (1) as a measure of information gain while splitting samples at each node of a tree, we assumed that features with low entropy were strong signals for identifying the most relevant features, which is summarized in Table 3.

$$\begin{cases} E_s = & \frac{n_l}{n} E_l + \frac{n_r}{n} E_r \\ E_l = & -\sum_{i \in c} p_{il} \log p_{il} \\ E_r = & -\sum_{i \in c} p_{ir} \log p_{ir} \end{cases} \quad (1)$$

where p_{ir} is the proportion of samples of the left split, p_{ir} is the proportion of samples of the right split, n_i is the number of samples in the left split, and N represents the total number of samples.

Table 3: The important features selected from the dataset

No.	Selected Feature	Score
8	DIO	0.04499746
7	DAO	0.05063975
6	Transmission Rate (per 1000 ms)	0.07437906
5	TR / RR	0.07781687
4	Trans Total Duration Per Sec	0.09315402
3	Trans Average Per Sec	0.09966616
2	Rcv Total Duration Per Sec	0.17506193
1	Rcv Average Per Sec	0.188617
0	Rcv Total Duration Per Sec	0.17506193

Before fitting the random forest, the dataset was preprocessed. Firstly, we manually dropped features that did not affect an artificial neural network model such as Source, Destination, *etc.* Then, we checked out the missing values and split the dataset into training and test set (70% for training and 30% for the test). Subsequently, feature scaling was performed so that they could be compared based on common grounds. Thereafter, the preprocessed data is fitted into the Random Forest (RF) Classifier for selecting the most important features. As the result, the best 10 features were selected based on the importance score as shown in Table 3. Figure 2 depicts the selected features and their scores. Furthermore, RPL protocol depends on three types of control message which are DIO, DAO and DIS. So, DIS was included in the selected feature set despite its low score rate of 0.00648158.

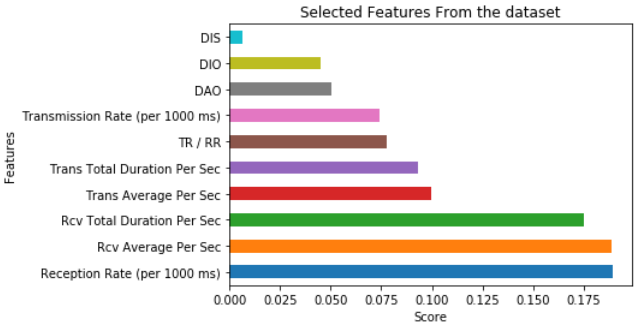


Figure 2: The selected feature and score

3.3 The ANN Model

Artificial neural network (ANN) attempts to mimic the human brains. ANN consists of a set of units named as neurons that are interconnected together to form layers [14]. ANN consists of an input layer, one or more hidden layers and an output layer. Each layer consists of several neurons. The model has one input layer which receives the data as a vector of features (xi) and produces the result as a vector of (yi) which is $y_i \in \{0, 1\}$ where 0 means benign and 1 means malicious. The output of each previous layer and the bias b value is computed by a nonlinear activation function f, which takes a weighting w_n from the previous layer as an input for the next layer and the calculation follows Equation (2).

$$a_n = f\left(\sum_{i=0}^{n} w_i x_i + b\right) \tag{2}$$

The neurons at the hidden layer(s) has activation functions, i.e., RelU and Tanh. The output layer neurons have the f_z activation function, i.e., sigmoid. The output of the sigmoid function is a binary output which is 0 or 1 calculated using Equation (3).

$$sigmoid = \frac{1}{(1+e^x)} \tag{3}$$

The main model which is used to detect the anomaly in the IoT network is based on artificial neural network (ANN) named as a multilayer perception technique for detecting RPL attacks (MLRPL) with input layer consisting of 20 neurons and three hidden layers. The first hidden layer has 50 neurons, the second hidden layer has 150 neurons and the third layer has 20 neurons. All these layers use rectified linear function (ReLU) as the activation function, and the output layer uses logistic function (Sigmoid) as the activation function in binary classification case and the Softmax function for categorical classification. For the loss function, mean square error (MSE) is used which is the sum of squared distances between the target variable and predicted values. Moreover, stochastic gradient descent (SGD) optimizer is used for optimizing the loss function with suitable properties. Table 4 shows the performance of the MLRPL model in the case of the three optimizers (SGD, Adam, and Adadelta optimizer). From the results, it can be concluded that SGD is the best optimizer in the RPL attack dataset as it can be inferred from the values of the metrics. To train the MLRPL model, grid search is used for tuning the best parameters, and 64 is the batch size whereas 700 is the best number of epochs.

4 Results and Discussions

4.1 Performance Evaluation Metrics

The performance of the proposed model was evaluated using various measurement metrics such as accuracy, detection rate (DR), precision and F1 score. Accuracy is a ratio of a number of correct predictions to the total number of samples and it is counted for both training and validation datasets. DR is the ratio of intrusions detected by the model. Another estimator is the precision which is the ratio between the correct positive results (TP) to all positive results predicated by the model. An additional estimator is the recall which is correct positive results to all samples that are supposed to be identified as positive.

Another measure of the quality of the model is the F1-score which is a consistent mean between precision and recall [14, 19]. The formulas are defined as follows:

$$
\begin{aligned}
Accuracy &= \frac{(TP+TN)}{(TP+TN+FP+FN)} \\
DR &= \frac{TP}{(TP+FN)} \\
Precision &= \frac{TP}{(TN+FP)} \\
F1-score &= \frac{2TP}{2(TP+FP+FN)} \\
AreaUndertheCurve &= \frac{1}{2}\left(\frac{TP}{(TP+FN)}+\frac{TN}{(TN+FP)}\right)
\end{aligned}
$$

where TN is true negative which denotes that a benign case was correctly labeled as benign, FP is false positive which points that a benign case was incorrectly labeled as an attack. As for the performance metrics, FN is false negative which indicates that an attack is incorrectly identified as benign, TP is true positive that indicates that an attack is correctly identified as an attack.

4.2 Experiment Results

The artificial neural network model was trained and tested using the method proposed in both cases for the multi and binary detection problems, respectively. The results are shown in the following.

4.2.1 Binary Classification Results

As a result of our proposed model, in the case of decreased rank attack dataset, for the binary classification case, the training and testing accuracy obtained is 97.14% and 97.01%, respectively, as shown in Figure 3. Furthermore, Figure 4 shows the loss function performance over time. Figure 5 shows the results of the confusion matrix while the receiver operating characteristic (ROC) is shown in Figure 6. The results are summarized in Table 4, showing that excellent precision can be obtained.

4.2.2 Multi-classification Results

In the case of the multi-classification problem, we used the same model (MLRPL) with the same parameters with Softmax as the activation function and the categorical cross-entropy as the loss function. Table 6 shows the classification results obtained, where 0 is benign, 1 is DR attack and 2 is VN attack. Furthermore, Figure 7 shows the training and testing accuracy for the model in which the accuracy obtained is 96.59% for the training phase and 96.39% for the testing phase. In Figure 8 the loss function in training and testing is shown. Based on these results it can be concluded that MLRPL can achieve high accuracy in both training and testing for DR attacks and VN attacks. The precision of MLRPL in detecting DR attacks is also high in the case of multi-classification as

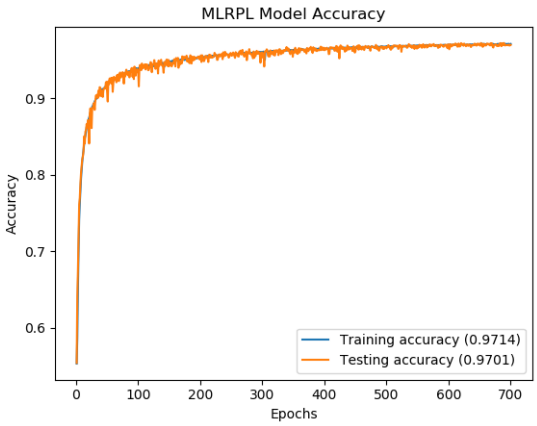


Figure 3: Model accuracy (training and testing)

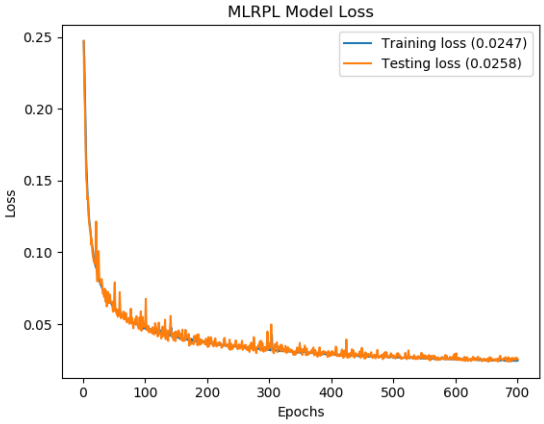


Figure 4: Model log loss values over time

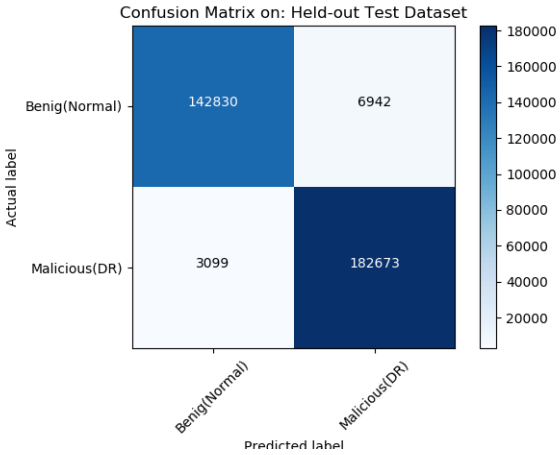


Figure 5: Confusion matrix on test dataset

Table 4: Classification report

	Precision	Recall	F1-score	Support
0	0.9788	0.9536	0.9660	149772
1	0.9634	0.9833	0.9733	185772
Avg./Total	0.9703	0.9701	0.9700	335544

Table 5: Result comparison MLRPL with other classifiers

Optimizer	**Benign**			**Malicious**		
	Precision	Recall	f1-score	Precision	Recall	F1-score
MLRPL	0.9788	0.9536	0.9660	0.9672	0.9882	0.9776
KNN	0.91	0.90	0.91	0.92	0.93	0.92
SVM	0.95	0.92	0.93	0.93	0.96	0.94
Random forest	0.96	0.95	0.95	0.96	0.97	0.96

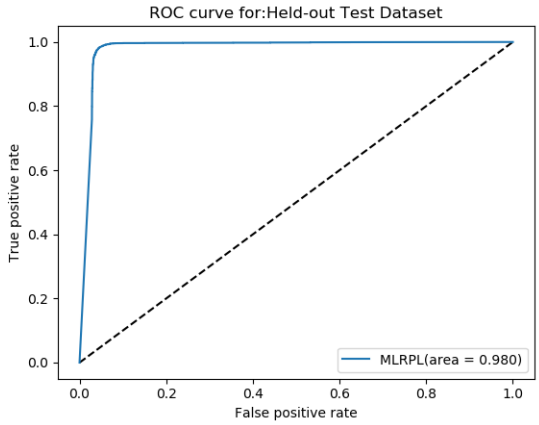


Figure 6: Receiver operating characteristic

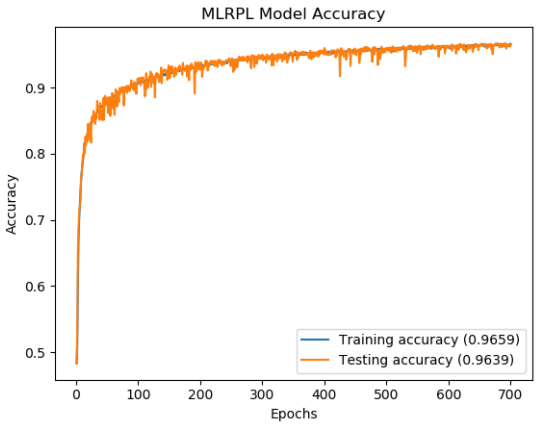


Figure 7: Model accuracy (training and testing)

in the binary classification. Table 6 shows the comprehensive multi-classification results generated based on the confusion matrix. Also, we performed a new experiment using the same dataset with different machine learning algorithms such as KNN, SVM and RF Classifier. Table 5 shows the classification results.

Table 6: Multi-classification report

	Precision	Recall	F1-score	Support
0	0.95608	0.9702	0.9630	334163
1	0.9651	0.9629	0.9640	191685
2	0.9792	0.9525	0.9657	166212
Avg./Total	0.9641	0.9639	0.9639	692060

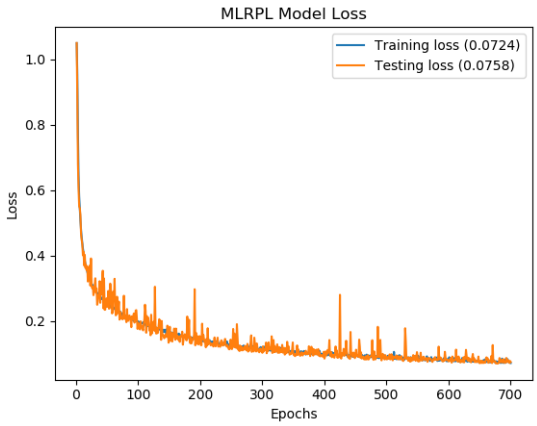


Figure 8: Log loss values over time

4.3 Comparison with Related Methods

To evaluate the proposed scheme, the MLRPL model is compared to some related methods. The comparison is

Table 7: Comparison of related work

Models	Precision	Recall	F1-score	Accuracy
MLRPL (binary)	97.14%	97.88%	95.36%	97.01%
Furkan *et al.* [23]	94.9%	95%	96%	94%
Abhishek *et al.* [20]	93.99	-	-	-

applied to proposed work by Furkan *et al.* [23] and by A. Verma *et al.* [20] which were discussed in the related work. Table 7 are the comparison results which show that the MLRPL model is better in the case of accuracy, precision, and F1-score. Moreover, the MLRPL model is also more efficient in the form of training time (number of epochs) and the complexity of the model (number of neurons and the number of layers).

5 Conclusion

In this paper, we proposed a machine learning model for detecting decreased rank attacks. The proposed model consists of three steps, namely data collection, feature extraction using random forest classifier and classification. Experiment results revealed that the proposed approach can achieve better results than other related methods. The results obtained from the MLRPL model indicate the fact that accuracy can be further improved. We believe that artificial neural network techniques provide the best direction for detecting and preventing routing attacks for both traditional networks and IoT networks. However, it is worth mentioning that better accuracy can be further achieved by conducting more experiments. It is clear that securing IoT is still in its infancy and, therefore, more solutions and additional research can be pursued to develop more effective solutions to secure IoT data and the networks.

Acknowledgment

The work in this paper has been supported by National Key Research and Development Program of China under grant 2019QY(Y)0601. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] A. Aris, S. B. Ö. Yalçın, and S. F. Oktuğ, "New lightweight mitigation techniques for RPL version number attacks," *Ad Hoc Networks*, vol. 85, pp. 81–91, 2019.

[2] T. T. Gao, H. Li, and S. L. Yin, "Adaptive convolutional neural network-based information fusion for facial expression recognition," *International Journal of Electronics and Information Engineering*, vol. 13, no. 1, pp. 17-23, 2021.

[3] S. Deshmukh-Bhosale and S. S. Sonavane, "A real-time intrusion detection system for wormhole attack in the RPL based internet of things," *Procedia Manufacturing*, vol. 32, pp. 840–847, 2019.

[4] A. Dvir, T. Holczer, and L. Buttyan, "Vera-version number and rank authentication in RPL," in *IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems*, pp. 709–714, Oct. 2011.

[5] O. Gaddour and A. Koubaa, "RPL in a nutshell: A survey," *Computer Networks*, vol. 56, no. 14, pp. 3163–3178, 2012.

[6] J. Granjal, E. Monteiro, and J. S. Silva, "Security for the internet of things: A survey of existing protocols and open research issues," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1294–1312, 2015.

[7] A. Kamble, V. S. Malemath, and D. Patil, "Security attacks and secure routing protocols in RPL-based internet of things: Survey," in *International Conference on Emerging Trends & Innovation in ICT (ICEI'17)*, pp. 33–39, Feb. 2017.

[8] E. Kfoury, J. Saab, P. Younes, and R. Achkar, "A self organizing map intrusion detection system for RPL protocol attacks," *International Journal of Interdisciplinary Telecommunications and Networking (IJITN'19)*, vol. 11, no. 1, pp. 30–43, 2019.

[9] A. Kumari, V. Kumar, M. YahyaAbbasi, and M. Alam, "The cryptanalysis of a secure authentication scheme based on elliptic curve cryptography for IoT and cloud servers," in *International conference on advances in computing, Communication Control and Networking (ICACCCN'18)*, pp. 321–325, Oct. 2018.

[10] C. T. Li, M. S. Hwang, "A lightweight anonymous routing protocol without public key en/decryptions for wireless ad hoc networks",*Information Sciences*, vol. 181, no. 23, pp. 5333–5347, Dec. 2011.

[11] C. T. Li, M. S. Hwang and Y. P. Chu, "An efficient sensor-to-sensor authenticated path-key establishment scheme for secure communications in wireless sensor networks", *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 8, pp. 2107-2124, Aug. 2009.

[12] A. Mayzaud, R. Badonnel, and I. Chrisment, "A taxonomy of attacks in RPL-based internet of things," *International Journal of Network Security*, vol. 18, no. 3, pp. 459-473, 2016.

[13] A. Mayzaud, A. Sehgal, R. Badonnel, I. Chrisment, and J. Schonwalder, "A study of RPL dodag version attacks," in *IFIP International Conference on Au-*

tonomous Infrastructure, Management and Security, pp. 92–104, 2014.

[14] F. M. M. Mokbal, W. Dan, A. Imran, L. Jiuchuan, F. Akhtar, and W. Xiaoxi, "MLPXSS: An integrated xss-based attack detection scheme in web applications using multilayer perceptron technique," *IEEE Access*, vol. 7, pp. 100567–100580, 2019.

[15] M. Nikravan, A. Movaghar, and M. Hosseinzadeh, "A lightweight defense approach to mitigate version number and rank attacks in low-power and lossy networks," *Wireless Personal Communications*, vol. 99, no. 2, pp. 1035–1059, 2018.

[16] H. Perrey, M. Landsmann, O. Ugus, T. Schmidt, and M. Wahlisch, "Trail: Topology authentication in RPL," in *Proceedings of International Conference on Embedded Wireless Systems and Networks*, pp. 59-64, 2013.

[17] A. Raoof, A. Matrawy, and C. H. Lung, "Routing attacks and mitigation methods for RPL-based internet of things," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1582–1606, 2018.

[18] U. Shafique, A. Khan, A. Rehman, F. Bashir, and M. Alam, "Detection of rank attack in routing protocol for low power and lossy networks," *Annals of Telecommunications*, vol. 73, no. 7-8, pp. 429–438, 2018.

[19] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in *International Conference on Wireless Networks and Mobile Communications (WINCOM'16)*, pp. 258–263, Oct. 2016.

[20] A. Verma and V. Ranga, "Evaluation of network intrusion detection systems for RPL based 6lowpan networks in IoT," *Wireless Personal Communications*, vol. 108, no. 3, pp. 1571–1594, 2019.

[21] A. Verma and V. Ranga, "Mitigation of dis flooding attacks in RPL-based 6lowpan networks," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 2, pp. e3802, 2020.

[22] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in internet-of-things," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, 2017.

[23] F. Y. Yavuz, D. Unal, and E. Gul, "Deep learning for detection of routing attacks in the internet of things," *International Journal of Computational Intelligence Systems*, vol. 12, no. 1, pp. 39–58, 2018.

Biography

Musa Osman is a PhD student at Beijing University of Technology (BJUT), China. He received his BSc in computer science at University of Gazira, Sudan, and MSc in Information System at Osmania University,India. His main research interests are security issues in the Internet Of Things mostly based on RPL protocol, Machine Learning, and Artificial Neural Network.

Jingsha He received his bachelor's degree in computer science from Xi'an Jiaotong University in China and his Master's and Ph.D. degrees in computer engineering from the University of Maryland at College Park in US. He is currently a professor in the Faculty of Information Technology at Beijing University of Technology (BJUT) in Beijing, China. Prior to joining BJUT in August 2003, Prof. He worked for several multi-national companies in the US, including IBM Corp., MCI Communications Corp. and Fujitsu Laboratories, during which he published more than 10 papers and received 12 US patents. Since joining BJUT in August 2003, Prof. He has published over 270 papers in scholarly journals and international conferences, received 66 patents and 33 software copyrights in China and finished 9 books. He has been the principal investigators of more than 30 research projects. Prof. He's research interests include information security, wireless networks and digital forensics.

Fawaz Mokbal received his BS degree in Computer Science from Thamar University, Yemen, and MS degree in Information Technology from the University of Agriculture, Pakistan. He is currently pursuing Ph.D. studies in Computer Science and Technology with Beijing University of Technology, China. He has won numerous international and university awards and he is the author and reviewer with various SCI, EI, and Scopus indexed journals. His area of interest includes Machine Learning, Artificial Neural Networks, Web Application Security, and Security issues in IoT.

Nafei Zhu received her B.S. and M.S. degrees from Central South University, China in 2003 and 2006, respectively, and her Ph.D. degree in computer science and technology from Beijing University of Technology, China in 2012. From 2015 to 2017, she was a postdoc as well as an assistant researcher in the Trusted Computing and Information Assurance Laboratory, State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences in China. She is now on the faculty of Information Technology in Beijing University of Technology. Dr. Zhu has published over 20 research papers in scholarly journals and international conferences (16 of which have been indexed by SCI/EI/ISTP). Her research interests include information security and privacy, wireless communications and network measurement.

RingCoin: An Accountable Mix for Achieving Bitcoin Anonymity

Albert Kofi Kwansah Ansah[1,2], Fengli Zhang[1], and Daniel Adu-Gyamfi[3]
(Corresponding author: Albert Kofi Kwansah Ansah)
School of Information and Software Engineering, University of Electronic Science and Technology of China[1]
610054, Chengdu, Sichuan, P.R. China
Department of Computer Science and Engineering, University of Mines and Technology[2]
P. O. Box TK 237, Tarkwa, Ghana
Department of Computer Science and Informatics, University of Energy and Natural Resources[3]
P. O. Box 214, Sunyani, Ghana
Email: afkansah@umat.edu.gh
(Received Feb. 11, 2020; Revised and Accepted Oct. 5, 2020; First Online Apr. 24, 2021)

Abstract

True anonymity may not be fully satisfied in Bitcoin. Linking transactions to input and output addresses to reveal user identity is possible. The author proposes a mixing scheme that modifies the mix coin protocol to hide the mapping of input transactions to addresses from the mixing server. The author used a ring signature scheme and an append-only log to achieve anonymity and accountability. The scheme ensures that transactions do not reveal any linkage between input transactions and retrieving addresses, and also, the mix does not store input transactions-addresses mapping. RingCoin achieves a higher degree of user anonymity against malicious mix and dishonest nodes.

Keywords: Accountable Mix; Anonymity; Bitcoin; Blockchain; Ring Signature

1 Introduction

Privacy is reasonably provided by generic banking which clients enjoy by default. Bitcoin, a decentralised financial electronic system is touted to provide fair privacy through pseudonymous addresses since its inception in 2008 [14, 15]. Users could use new address for each transaction to enhance privacy. However, address reuse, tracking payments, IP address monitoring and linkage to nodes, transaction graph analysis etc put pseudonymity friable and easily compromised [12, 19, 23, 24, 27]. [17] showed that multiple pseudonymous addresses can be linked to a single user. Thus, anonymity is a momentous requirement for cryptocurrencies.

Mixing is one of the techniques to upsurge privacy in Bitcoin from its weak pseudonymity. Mixes offer users the opportunity to mix their transactions with other users' transactions to achieve anonymity. Mixes, on the other hand, may learn the input transactions to addresses mapping of users. To curb this, mixes can be cascaded but this may increase the transaction cost. Anonymity against mixes should be of prior consideration in mixing services. RingCoin considered this property and thus employs a ring signature scheme and public append-only log to achieve anonymity and accountability. RingCoin does not learn nor store input transactions to addresses mapping of users. Combining ring signature and append-only public log with warranty scheme to implement RingCoin conform to fundamental design principles of the current Bitcoin system. Users will use the warranty scheme to incriminate RingCoin if it erred. Implementation of public log in RingCoin allows users to check the number of users using the mix simultaneously 'a posteriori'. Our scheme meets the mix properties (Section 1.2) and inherits the main advantages of centralised mixing scheme hence, it is easy to deploy and scalable to accommodate larger users. Mixing is deemed to complete in a few hours with mixing fee expected to be less than 0.5%.

1.1 Motivation and Contribution

Bitcoin users are challenged with threats of leakage of real identities and security issues through transaction aggregation and analysis [7, 10]. If dishonest node links transactions to real identities, all financial behaviour of same could be tracked. Approach to link transactions to real users using suitable heuristics has been presented in [1]. Bitcoin is increasingly hitting large userspace, therefore, solutions to deal with privacy threats cannot be overemphasised.

Some privacy-enhancing techniques and anonymity schemes are available in literature [3,9–11,16,26,29]. However, there are two major drawbacks on most of the ap-

proaches.

1) Users are to trust a third-party to mix bitcoins without stealing;

2) Mix may store input transactions to addresses mapping information of users. Mixes may reveal mapping information for gain or under compulsion.

[2] proposed Mixcoin protocol as a currency mix adding an accountable mechanism to facilitate anonymous payment in Bitcoin and to expose theft but the anonymity of users is still not protected because it exposes mapping information to the mix which can be leaked in the future as mix acts maliciously. In [25], Valenta et al. proposed Blindcoin accountable mix based on blind signatures attempting to solve input transactions-addresses mapping problem in [2]. [9] has not been largely explored with digital coins. In [16], traceable pseudonymise was used to develop a privacy-preservation validation approach to achieve anonymity but it is computationally expensive.

Our main contributions:

1) We point out that mixcoin suffers from possible leakage of inputs transactions to addresses mapping of users.

2) We utilise features of ring signature [18] to solve user information linkage problem in mixcoin and used an append-only public log to offer accountability with warranty scheme.

3) We proposed RingCoin as a centralised mixing scheme to unlink inputs transaction(s) to address(es) mapping based on ring signature.

4) We show that using a ring signature will provide anonymity and the public log will lead to theft detection.

RingCoin's contribution thus eliminates the privacy threats from the input transactions-addresses mapping exposed in [2] that could potentially be used to deanonymised its users. RingCoin exhibits all the properties of an ideal mix.

1.2 Mixing Services and Related Works

[5,19,23,24] provided an extensive survey on security and attacks in cryptocurrencies and security in blockchain. [4] introduced anonymous communication mixes. Mixing services have been used to solve deanonymisation problem in Bitcoin [6]. An ideal mix possesses certain properties, albeit most current mixes are deficient with some of the properties. These properties are seen in the ensuing paragraph.

Accountability: Mix should be accountable to its users by issuing a warranty in the form of a contract. This gives users the confidence to participate in the mixing process. Users can use the warranty to dent the image of the mix if it fails to send the funds on scheduled time or steal the bitcoins.

Anonymity: It must be certified that only the user should know the input transaction(s) and address(es) mapping. This must be guaranteed in a mix scheme.

Mixing fees: An implementation to incentivise mixes. Fees must be reasonably fixed to promote expediency for many users to use mixing services. Randomise, all or nothing fees could be applied to retain the entire value of the transaction.

Scalability: Mixes should scale to accept a larger number of user transactions and large anonymity sets. The larger the anonymity set, the higher the obscurity.

DoS Attacks: Mixes should be resilient to resist denial-of-service (DoS) attacks from malicious nodes.

Compatibility: Ideal mix should be backwards compatible with the current Bitcoin system.

Mix indistinguishability: Passive adversaries should not link mix to the user to determine which mix a user is interacting with.

Mixing approach to enhance anonymity has been proposed in [11]. Many mixes have not proven to offer full anonymity or accountability. [2] has all the mix properties but failed to provide anonymity against a malicious mix. Mixcoin has access to input transaction to addresses mapping and may be revealed to deanonymise users. CoinJoin [12] and coinswap [13] provided backward compatibility with Bitcoin but offered smaller anonymity sets. [12] generates a joint transaction with users and ensures anonymity. [13] is a fair exchange mixing service but it is unable to ensure that input and output addresses cannot be linked by the intermediary. [25] used blind signatures and partial warranty to provide unlinkability. CoinParty [28] provides unlinkability to Bitcoin by imploring threshold ECDSA protocol and mixing peers.

CoinShuffle [21] uses a multiparty sorting protocol to achieve anonymous mixing built on [12]. It implements a blaming process which seeks to decline misbehaving users from future mixing. There is no mixing fee hence large-scale Sybil attack would be easy to delay rounds of mixing. Xim espoused a P2P network mixing service utilizing a blockchain broadcast to amass interested users to partake in the mixing process. It has to wait for users to accept a request to mix hence it takes a couple of hours to finish mixing. CoinShuffle++ [22] is built on CoinShuffle. ValueShuffle [20] extends [22] and utilises DiceMix. Funds are concealed through credential transaction and use stealth address to protect address information.

1.3 Organisation

Section two introduces the ring signature scheme. Section three presents a formal model of RingCoin scheme. We present our RingCoin protocol in Section four and section five sees the analysis of the scheme. We conclude the paper in section six.

2 Preliminaries

2.1 Ring Signatures

Ring signature has the property of signer ambiguity. Signer assembles set of public keys $U_R=\{pk_i\}_1^n$, $n\geq 2$ to create a ring signature [18]. The private key sk is used to generate the ring signature. The verifier is unable to determine who signed but it is convinced that the private key that produced the signature has its public key in U_R. Ring signature scheme is defined by three procedures:

1) KenGen(1^λ): Where λ is a security parameter. It outputs a secret key sk and public key pk, $\{(pk_i, sk_i)\}_i^n$.

2) RingSign$_{sk}$ (m, U$_R$, sk$_s$): On an input of message $m\in(0, 1)^*$, set of public keys $U_R=\{pk_i\}_1^n$, private key sk of i^{th} ring member who is the actual signer. It outputs a ring signature σ.

3) RingVer (m, pk$_i$, σ): On an input of message m, and signature σ which includes $U_R=\{pk_i\}_1^n$. All possible signers verify signature σ and output true or error symbol $\perp$.

Verification satisfies soundness and completeness condition for any message m. Completeness condition holds for any k. RingVer (m, RingSign$_{sk}$ (m, U$_R$, sk$_s$)) = 1, else output error symbol $\perp$.

3 The Formal Definition of RingCoin Scheme

We include the formal definition of RingCoin and its security notions. All the symbols used are defined in Table 1 for simplicity. The design of RingCoin scheme is guided the following properties of ring signature *i.e.* Unforgeability, Linkability, and Correctness. Unconditional anonymity with n number of ring members: For an adversary to recognise the actual signer from ring members, it requires a probability of no more than 1/n ($Pr_{unring}\leq 1/n$). n defines the anonymity set and is arbitrarily defined by the signer. As n becomes increasingly large, chances for an adversary to decipher the signer become negligible.

3.1 Design Goals

Under the system model and security requirements, our design goal is to propose a secure anonymous mix (M_x), a scheme that ensures user input transactions and addresses mapping anonymity. The scheme ensures the privacy of the users and their transactions by making it as difficult as possible to link input transactions and addresses to reveal users real identities. RingCoin mixing scheme is required to assure anonymity and confidentiality against adversary as well as providing accountability. Users are protected by the warranty as a security measure against the mix. Figure 1 presents the high-level model of RingCoin scheme. Figure 2 shows RingCoin design framework with the Bitcoin blockchain. As seen in the Figure 2, RingCoin (1) and users (2) communicate bi-directionally to get user transactions anonymised by following the protocol in Figure 3, thus Algorithm 1 and then broadcast the anonymised transaction into the blockchain network (3) to be mined (4) and then added to a block (5).

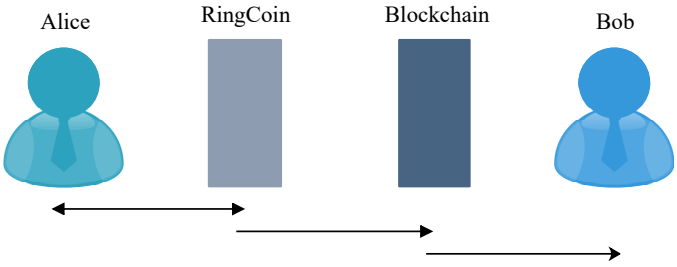


Figure 1: RingCoin model overview - High level

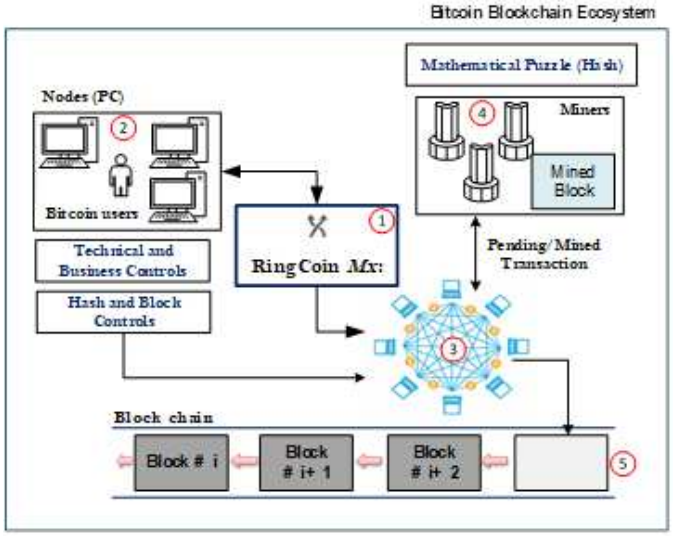


Figure 2: RingCoin design framework

3.2 RingCoin Protocol Settings

RingCoin provides a mixing protocol with accountability. Users are given a signed warranty as a guarantee to ensure that funds would be transferred as scheduled. Users can use it to incriminate the mix. Users negotiate with the mix to choose mix parameters. The mix accepts transactions from a user into an escrow account with each user specifying retrieving address(es). Users interact with the mix over an anonymous channel with proposed mix parameters (v, ρ, ω, n, T1$_{app}$, T2$_{app}$, T3$_{app}$, T4$_{app}$, k$_{out}$). RingCoin implements approximate times (Ti $\pm$ 0.1%) μsec, $1\leq i\leq 4$ to allow little flexibility for both users (U_A) and RingCoin (M_x). This is not to give the mix an urge to steal bitcoins or take undue advantage to deliberately delay the mixing process but it is to sanitise the system

Table 1: Basic notation

Symbol	Description
M_x	Mixing server (RingCoin)
U_A	A user wishing to transfer coin (BTC)
$U_A^{'}$	Users anonymous identity for publishing to a public log
U_{A_k}	Secret commitment/encryption function of U_A
$invU_{A_k}$	Inverse of U_{A_k}
k_{in}	Input address of U_A to move funds from
k_{out}	Output address to transfer funds to
k_{esc}	M_x's unique escrow address for each U_A to accept coins from k_{in}
$k_{esc}^{'}$	M_x use to pay to k_{out}
n	The nonce for determining payment of radomised mixing fee
v	Chunk to mix
ω	Number of blocks required to verify user payment by M_x
ρ	Mixing fees rate payable by U_A
$T1_{app}$	Time for U_A to move chunk v to k_{esc}
$T2_{app}$	Time for M_x to publish ring signature nominal R_{mt} to public log
$T3_{app}$	Time for $U_A^{'}$ to reveal k_{out} to public log
$T4_{app}$	Time for M_x to move coins from k_{esc} to k_{out}
M_{param}	M_x parameters which is a tuple $\{v, \rho, \omega, T1_{app}, T2_{app}, T3_{app}, T4_{app}\}$
BCN	Random beacon function which is publicly verifiable
sk_{mx}	M_x private key
pk_{mx}	M_x public key
R_{mt}	Ring signature nominal
$\rightarrow$	Input direction
$\leftarrow$	Output direction

to curb haste of user publishing incriminating evidence against the mix for not transferring funds at $T4_{app}$. RingCoin requires a safety margin of several blocks to guarantee that transactions can be included by $T4_{app}$. RingCoin may decide to accept or otherwise any negotiation from a user to mix bitcoins. We note that this is all or nothing negotiation with the user.

RingCoin generate escrow address k_{out} plus users free parameters and encapsulate it as a warranty signed with its private key and send it back to the user. Users are not obliged to transfer bitcoins when they receive RingCoin's signed warranty. RingCoin then deletes the users' records in the event of any rejection. Once the user transfers coins to M_x by $T1_{app}$, the onus lies on RingCoin to transfer funds to k_{out} specified by the user at $T4_{app}$ unless the funds are booked as mixing fee. This may happen if the funds to be transferred was not enough to cover the mixing and mining fee. To ensure forward anonymity against future data, both RingCoin and users destroy the annals. Users can only publish incriminating evidence against the mix if funds are not transferred to k_{out} by $T4_{app}$.

3.3 Generating RingCoin Ring Signature

RingCoin's ring signature is based on RSA Ring Signature Scheme. Given transaction message $(mt)$to be signed, set of hopeful ring members public keys are assembled as $P = \{pk_i\}_{\forall 2 \leq i \leq n} = \{pk_1, \ldots, pk_n\}$ and private key sk_s. A trapdoor one-way permutation g_i is specified by each ring member's public key pk_i and sk_s also specifies trapdoor information for computing g_s^{-1}.

1) Signer computes symmetric secret key k by hashing mt: $k = H(mt)$.

2) Signer uniformly and randomly picks an initialization worth z from $\{0, 1\}^b$.

3) Signer randomly and independently picks $x_i \ \forall$ ring members $1 \leq i \leq n$, $i \neq s$ from $\{0, 1\}^b$ to compute $y_i = g_i(x_i, pk_i)$.

4) Signer solves ring equation for y_s which satisfies the ring equation below and efficiently computable given arbitrary values:

$$z = C_{k,z}(y_1, ..., y_n = R(k, z, y_1, ..., y_n) \quad (1)$$

5) Signer inverts trapdoor permutation using his knowledge of his trapdoor in g_s on y_s to obtain x_s: $x_s = g_s^{-1}(y_s, sk_s)$

6) Ring signature on the transaction message mt is output by the $(2n+1)$-tuple as $R_{mt} = (pk_1, \ldots, pk_n, z, x_1, \ldots, x_n)$

Ring Signature verification:
Given R_{mt}=(pk_1, ….. , pk_n, z, x_1, …., x_n) and mt, trapdoor permutations are applied. $\forall_{1\leq i\leq n}$ verifier computes: y_i=g_i (x_i, pk_i).
Obtaining k to reveal output address, the verifier (M_x) hashes mt: k= H (mt).
Verifier (M_x) checks y_i's. If equation (1) is satisfied, RingCoin accepts the signature as valid. Otherwise returns error symbol $\perp$.

3.4 Security of the Ring Signature

The ring signature scheme espoused in RingCoin is secured against adaptive chosen message attacks in the ideal cypher model (assuming each public key specifies a trapdoor one-way permutation). There will be a noticeable difference between the distribution of randomly chosen and computed x_i when g_i is not permutable and can lead to the identification of the real signer among possible signers. RingCoin randomly chooses x_i hence, the identification of the real signer is hard to achieve.

4 The RingCoin Scheme

The primary procedure for mixing funds is that M_x first publicises mix parameters. User (U_A) shows interest and M_x responds with a limited warranty. U_A moves bitcoins to the escrow address (k_esc) specified by M_x on the warranty. M_x concludes warranty and publishes it to the public log. U_A reveals output address (k_out) in the ring signature and M_x moves coins to k_out. We remark the following notation; $(m)_c$ is a message committed with a function c and m_c is a signature on a message with signing key. The signature is sent with the message to aid in message recovery.

4.1 Protocol Description – How it Works

The protocol runs through nine stages, thus stage 0 to 8 as follows:

- STG 0: Setup

M_x publishes mix parameter (M_{param}) publicly *i.e.* (v), (ω), (ρ), ($T1_{app}$, $T2_{app}$, $T3_{app}$, $T4_{app}$). $T3_{app}$, and $T4_{app}$ were precluded from mixcoin.

- STG 1: U_A makes an offer to M_x

U_A shows interest and sends offer with M_{param} and ring signature tag R_{mt}=(k_{out}, n) to M_x. Nonce n is randomly selected primarily for receiving fees. R_{mt} is encrypted with commitment function U_{A_k} known to U_A. This is to keep parameters of R_{mt} hidden. k is a ring signature symmetric encryption key (section 3.3). The inverse $invU_{A_k}$ is known to U_A. The designation of the offer is M_{param}, $[R_{mt}]_{UAk}$. Mixcoin did not anonymise the retrieving addresses as our scheme has implemented here using ring signature (R_{mt}).

- STG 2: M_x: Response to U_A's offer

M_x decides either to accept or reject the offer. An acceptance means M_x sends limited warranty to U_A. Warranty comprises R_{mt}, k_{esc} (to receive coins from U_A) and M_{param} which is signed by sk_{mx} (M_x private key) *i.e.* $\{M_{param}, k_{esc}, [R_{mt}]_{UAk}\}_{SKmx}$. U_A cannot convalesce signed warranty directly due to the inclusion of other parameters in the warranty. Note that M_x provides unique k_{esc} to each user. This is salient to M_x for payment verification. Unlike Mixcoin where mix only sends signed warranty, RingCoin sends back warranty plus R_{mt}. In the event that M_x rejects U_A's offer, U_A will destroy the retrieving address(es).

- STG 3: Payment to M_x

U_A makes an honest transfer of v bitcoins to M_x's k_{esc} from her k_{in} address at $T1_{app}$. In the event that U_A is unable to move v coins by $T1_{app}$, the protocol is terminated by both parties.

- STG 4: M_x public acknowledgement

M_x signs $\{[R_{mt}]_{Uak}\}_{SKmx}$ and publishes ring signature tag to public log by $T2_{app}$ when U_A transfers funds. This process completes the warranty. Note that $T2_{app}$ is long enough time between $T1_{app}$ and $T2_{app}$. This allows transactions to have at least ω blocks for M_x to con?rm U_A's reimbursement. Here, M_x publicly concedes that U_A transferred v at the slated time to k_{esc}. Third-parties can authenticate from the public log that M_x successfully completed warranty at $T2_{app}$. This is not implemented in Mixcoin and it is RingCoin's contribution. In the event that M_x is unable to complete warranty, U_A may publicise incriminating evidence by publishing the warranty *i.e.* $\{M_{param}, k_{esc}, [R_{mt}]_{UAk}\}_{SKmx}$. U_A can incriminate M_x only when M_x is unable to publicly log R_{mt} by $T2_{app}$. U_A can incriminate M_x using the limited warranty, transfer (v, k_{in}, k_{esc}) in the block by $T4_{app}$ and failure to publish signed R_{mt} to public log by $T2_{app}$ as shreds of evidence. Third-parties could see that M_x signed warranty, k_{esc} received funds, and R_{mt} was not published to public log by $T2_{app}$, hence M_x is not trustworthy. If U_A breaches protocol, he cannot publish incriminating evidence against M_x. This is exclusive to RingCoin and was not implemented in Mixcoin.

- STG 5: U_A anonymously reveals output address

U_A applies reverse $invU_{A_k}$ to recover signed tag $\{R_{mt}\}_{sk_{mx}}$=$\{k_{out}, n\}$. U_A connects as $U_A^{'}$ anonymously and reveals k_{out}. $U_A^{'}$ then posts signed tag to public log. M_x verifies the authenticity of the output address(es). Signed R_{mt} is seen as a proof of knowledge that shows that M_x signed the commitment. $U_A^{'}$ must publish R_{mt} to the public log by $T3_{app}$, otherwise M_x may choose to refund coins to U_A or retain it. Upon $U_A^{'}$ successfully revealing k_{out} by $T3_{app}$, M_x computes BCN ($T3_{app}$, ω, n)

for each pair (k_{out}, n). This determines the address(es) to take the mixing fees from. The public can verify the beacon function using entropy collected from blockchain to produce uniform number range [0, 1]. If input value $\geq \rho$ per mix parameters, the protocol continues execution to the next stage else chunk for retrieving address(es) is retained by M_x as a mixing fee.

- STG 6: Funds paid to retrieve addresses by M_x

In time $T4_{app}$, an honest M_x transfers v BTC to all retrieving addresses once they have passed beacon function. M_x is unable to determine input transactions to retrieving addresses from the same user. M_x only knows one user signed the ring but does not know which one actually signed. Hence no mapping of input transactions and addresses. In the event when M_x steals coins, then M_x is dishonest at the time $T4_{app}$.

- STG 7: Thievery is detected and U_A publishes incriminating evidence

If M_x fails to transfer funds to U_A's retrieving address(es) at $T4_{app}$, the user assumes that theft has occurred and he is free to publish implicating information about M_x *i.e.* commitment function U_{A_k}, $invU_{A_k}$, warranty $\{M_{param}, k_{esc}, [R_{mt}]_{Uak}\}_{SKmx}$, transaction (v, k_{in}, k_{out}) in the blockchain by $T4_{app}$, signed $\{k_{out}, n\}_{sk_{mx}}$, and transaction $\left(v, k'_{esc}, k_{out}\right)$ is not present in the blockchain by $T4_{app}$. Third parties can verify that the '*nominal*' was signed with sk_{mx} and recover same with $invU_{A_k}$. Public log and the blockchain could be checked to see if the protocol was duly followed.

- STG 8: Protocol terminates. User and RingCoin destroy all annals and formally end protocol.

We remark that STG 2, STG 3, and STG 5 are consistent with Mixcoin. Figure 3 presents the flowchart of RingCoin Protocol. RingCoin's execution algorithm is presented in Algorithm 1.

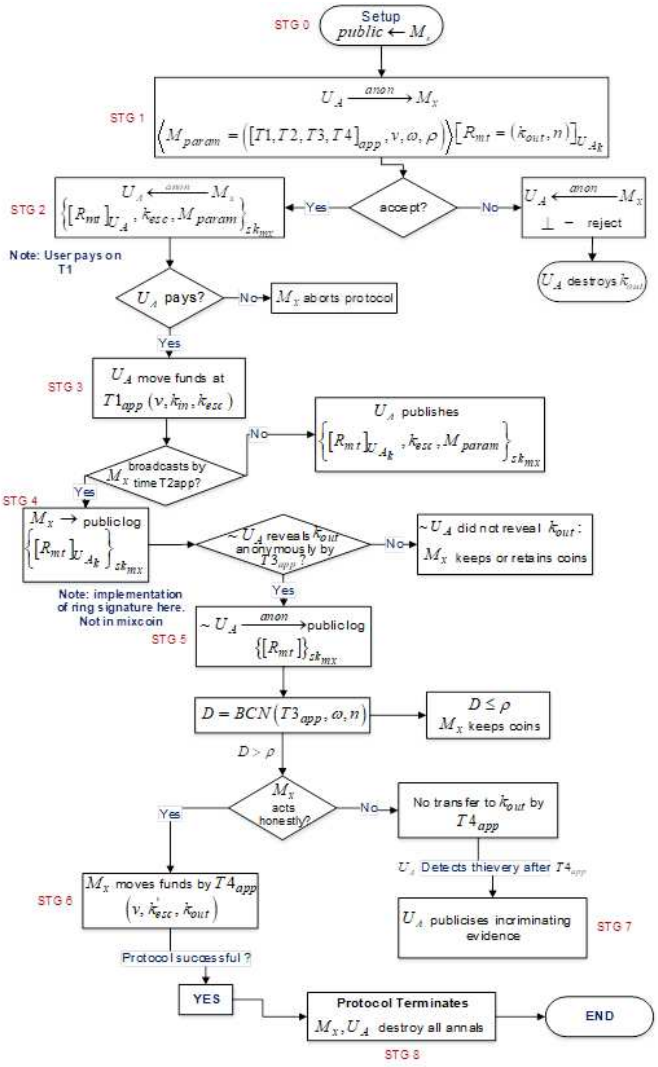


Figure 3: The RingCoin protocol

5 Discussion

The section discusses the properties of the RingCoin scheme. These properties are accountability, anonymity, RingCoin as an attacker and denial-of-service (DoS) attack. It also discusses mixing fees, scalability, compatibility, and advantages of mixes. Overheads of RingCoin is also discussed here.

5.1 Accountability

RingCoin is accountable to its users by issuing a warranty in the form of a contract which gives users the confidence to participate in the mixing. This property exposes M_x in case it cheats. Users can use it to dent the image of M_x if it fails to send funds. M_x could cheat when determining retrieving addresses. Users use this warranty to protect themselves and therefore, an important property in RingCoin. RingCoin sends limited warranty to users after agreeing the offer spewing that 'if the user fulfils his payment, RingCoin publishes signed R_{mt} to the public log by agreed times'. To make sure valid paid R_{mt} are only duly logged, RingCoin will wait for users to make appropriate payment before publishing R_{mt} to the public log. Third-parties could use this log to verify that RingCoin acknowledged users payment and also signed R_{mt}. This is a strong view to incriminate RingCoin if it fails to send funds retrieving addresses that have gone through Beacon successfully by the agreed time. RingCoin chooses a more flexible time due to the possibility of network jitter and may not necessarily be misconduct or deliberate to transfer coins by agreed times. Beacon (STG5) which is publicly verifiable determines the collection of the arbitrary fee and third-party can verify that Beacon correctly computed the function.

Algorithm 1 RingCoin execution

Begin
Function: *RingCoin*
Setup() //STG 0
public mix parameters $\leftarrow M_x$
$M_x \overset{anon}{\longleftarrow} U_A$ //STG 1
if ($M_x =$ "*accept*") //STG 2 **then**
  $U_A \overset{anon}{\longleftarrow} M_x$
else
  U_A destroys k_{out}
end if
if ($U_A =$ pays) **then**
  U_A sets $T1_{app}$ //STG 3
  U_A moves funds
else
  M_x aborts protocol
end if
if (M_x broadcast by $T2_{app}$ **then**
  public log $\leftarrow M_x$ //STG 4
else
  U_A publishes $\{ \mathrm{k_{esc}}, [\mathrm{R_{mt}}]_{\mathrm{UAk}}, \mathrm{M_{param}} \}_{\mathrm{SKmx}}$
end if
if ($\sim U_A$ reveals k_{out} by $T3_{app}$) **then**
  public log $\overset{anon}{\longleftarrow} \sim U_A$ //STG 5
else
  $\sim U_A$ did not reveal k_{out}
  $D = BCN(T3_{app}, \omega, n) \leftarrow M_x$ //computes mixing fee for each pair (k_{out}, n)
end if
if ($D \leq \rho$) **then**
  M_x keeps coins
else
  if ($D > \rho$) **then**
    and
    if (M_x acts honestly) **then**
      M_x moves funds by $T4_{app}$ // STG 6
    else
      No transfer to k_{out} by $T4_{app}$
    end if
  end if
end if
if (U_A detects thievery after $T4_{app}$ //STG 7 **then**
  U_A publishes incriminating evidence
end if
if (protocol is successful) **then**
  Protocol Terminates //STG 8
end if
M_x and U_A destroy all annals
End

5.2 Anonymity

Mixes can act as adversaries if they are able to store annals of users which can potentially be used to deanonymise same. RingCoin provides anonymity guarantee that mixing server learns nothing about input transactions and addresses pairings due to the use of ring signatures scheme ($\mathrm{R_{mt}} = (k_{out}, n)$), therefore, input and output addresses cannot be linked with its transactions.RingCoin guarantees anonymity against mix adversarial. We analysed the anonymity and describe a threat model, active and passive adversarial threats.

5.2.1 Threats Model

We considered an attacker trying to have ample information about anonymity set of pre-mixing input addresses which feeds the retrieving addresses. In RingCoin, the use case is a single linkable input address transferring a chunk of funds to unique retrieving addresses. In blockchain, every dishonest node could be a possible global adversary. A dishonest node can delay, block, or flood the network with dummy messages. We assume that a dishonest node has the ability to flood the network causing denial-of-service (DoS) attack to detain valid transactions from going through or refuse to send funds at stipulated times. This may cause a failure of the operation. Causing DoS attack in our protocol may be difficult because RingCoin disfavours any attempt to replay messages. RingCoin users interact only with the mix, hence if a user refuses to comply with the protocol would not affect other users. If an attacker could control a significant amount of mining pool, then it stands a higher probability to block transactions else it would be practically intolerable to block transactions for a greater time period. A malicious node may deny transactions from entering the blockchain at stipulated times or messages from being posted on the public log by carrying out DoS attack. The implementation of this attack in practice would be rather thorough since it may require the attacker to control a substantial portion of the Bitcoin mining pool.

5.2.2 Global Passive Attacker

This is perhaps the weakest adversarial model. The public Bitcoin blockchain makes Bitcoin transactions accessible to all and the public log of RingCoin also allows anyone to verify entries. This makes every node a passive adversary. It is assumed that passive adversary can resolve with sufficient probability of which transactions are of $\mathrm{M_x}$ characteristics by naively observing v and escrow address k_{esc}. Mix indistinguishability [2] makes passive adversary unable to link escrow addresses to mixes. Passive adversaries should not link mix to the user to determine which mix a user is interacting with but only to observe a stream of similar escrow addresses making it look like a one mix serving large users with v. Mix indistinguishability delivers strong anonymity and extends users anonymity set to all users in different mixes using the same mix parameters. This is considered in a future paper.

Side channel [25] properties such as timing, network layer information, interaction with public log etc could cause leakage of user information to a passive adversary. Avoidance of timing analysis is paramount against attacks. Interaction with RingCoin is devoid of predictabil-

ity. If a signed warranty is published by M_x and retrieving addresses are learned thereafter, a malicious mix could link the warranty to the retrieving addresses. User input transactions and addresses could not be linked if they negotiate their own set of parameters with M_x. If the deadline for revealing output address(es) is precise, a passive adversary may trivially map retrieving addresses to its input transactions. We assume that users connection is through secure anonymity network. To keep U_A and U'_A from being linked may be difficult pragmatically against an adverse adversary in this current work and is considered in future work.

5.2.3 Active Attacker

An active attacker could compromise sections of input transactions and addresses for pragmatic operation. Active attackers are strongest which has the potential to link escrow addresses to its mix. This makes an active adversary potential to distinguish escrow addresses that matches M_x. In RingCoin, anonymity set is still non-compromised addresses pairs. It is quite expensive to curb this attack in terms of mixing fees.

When v is sent from K_{in} to K_{out}, receiver learns that K_{in} interacted with M_x. In the same vein if a sender sent v to K'_{esc} and consequently to K_{out}, an attacker may also learn that the sender interacted with M_x. An active attacker may use flooding attack to learn other addresses interacting with the same mix for every v sent via M_x. Mixes could be unduly forced to pay transaction fees. The effectiveness of this attack depends on the ratio of transaction fee per chunk τ to average mixing fee per chunk $v\rho$, *i.e.* $\left(\frac{\tau}{v\rho}\right)$. Thus $\left(\frac{\tau}{v\rho}\right)$ is spent by M_x to service transaction fees. For each transaction v, M_x pays $\left(\frac{1}{\rho} * \frac{\tau}{v\rho}\right)$. The probability of retaining each v is $\rho * \frac{\tau}{v\rho}$ maximised to 1 if the mixing fees are sufficient to cater for transaction fees. An active adversary shall have to perform with M_x to link up to $(1-\rho) * 2 + \frac{\tau}{v\rho}$, thus $\left[Adv^{mix}_{active} \leq (1-\rho) * 2 + \frac{\tau}{v\rho}\right]$. An attacker would have to generate an enormous portion of M_x traffic to achieve this attack setting.

5.2.4 RingCoin as an Attacker

Here we consider RingCoin to be an adversary itself. This is possible if the Mix is unduly compelled or compromised to reveal input transactions to addresses. This is impossible in RingCoin due to ring signature implementation to anonymise the pair. [2] reveals all pairings because the mix has access to them. In RingCoin, even if M_x is compromised it would still not weaken anonymity guarantees since ring signature tag $R_{mt} = (k_{out}, n)$ cannot be deanonymised by M_x. This is our main contribution and thus an advantage over Mixcoin.

5.2.5 Denial-of-Service (DoS) Attack

Mixes should be resilient to resist DoS attacks by a handful of malicious nodes. This property was duly expressed in RingCoin. In RingCoin, users interact with mixing server only, therefore, not complying with the protocol cannot affect other users and the mixing process would not be slowed either. DoS can be unleashed by an attacker to deny users from getting into the blockchain timely or deny messages from being posted to the public log. The adversary should control the sizable amount of Bitcoin block mining pool to launch this attack on RingCoin.

5.3 Mixing Fee

Fees are reasonably and randomly fixed to determine the remuneration users pay to promote expediency for many users to mix freely. Randomise, all or nothing fees is applied to retain entire value from a small percentage of the transaction. Users pay the incentives to M_x based on a certain percentage fixed by M_x. Delays may sometimes be inevitably associated and this is why RingCoin did not put stringent times for the mix operation.

5.4 Scalability and Compatibility

Mixes must scale to accommodate larger number of users. RingCoin is able to add more users to perform the mix. For effective scalability, load balancing is considered using different servers operating with the same cryptographic settings. An ideal mix should be backwards compatible with the current Bitcoin system. We reiterate that RingCoin deployment requires no changes to the existing Bitcoin protocol. Thus, RingCoin is backwards compatible with existing Bitcoin.

5.5 Advantage of Mixes

Funds theft is averted by group transactions because customers can check and verify for their address included in the outputs. The mapping of input transactions to addresses and transaction tracking are vetoed by ring signature in our mixing service. In a centralised mixing scheme, expense in deployment and upgrade is on the downside. RingCoin requires non computationally complex cryptography. RingCoin scheme ensures that transactions do not reveal any linkage between input transactions and retrieving addresses and the mix does not store input transactions and addresses mapping. RingCoin achieves a higher degree of user anonymity against malicious mix and dishonest nodes. The scheme's security and privacy are ensured through the standard ring signature setting. To highlight the advantages of our proposed scheme, we compare RingCoin to other proposed schemes on the backdrop of the properties of an ideal mix as presented in Table 2.

Table 2: The comparison results of RingCoin with other proposed schemes

Schemes	Accountability	Scalability	Architecture	Bitcoin Compatibility	Resilience to DoS	Anti-theft	Anonymity against Mix Server
RingCoin	√	√	service	√	√	√	√
Mixcoin [2]	√	√	service	√	√	x	x
Blindcoin [25]	√	√	service	√	√	x	√
Coinjoin [12]	-	weak	-	√	x	√	-
Coinswap [13]	-	x	p2p	√	√	partially	-

5.6 Overheads

RingCoin mixing requires two messages between $U_A \leftrightarrow M_x$, two other messages posted to the public log and two Bitcoin messages. An overhead to this protocol is the messages published to the public log. These overheads are time bounded to be posted. To be consistent with double spending property in the Bitcoin setting, the time between deadlines are carefully chosen to allow the previous message to be at least ω (normally $\omega = 6$) blocks deep into the blockchain. This seeks to prevent double spending. RingCoin implements T_{app} to alleviate this problem. As in Bitcoin, the time between blocks is about 10mins, therefore, an average time for a message to be six (6) blocks deep into the blockchain is expected to be about an hour. Variation in the time to each block cannot be overemphasised, therefore, deadlines are set further apart. This makes T_{app} espoused in RingCoin novel. RingCoin protocol needs four timelines and the allowable time between them is a dominating factor. This factor determines the length of time protocol takes to run. Table 3 compares the overhead cost of our proposed scheme to other schemes.

If Bitcoin implements public log message posting, it will attract some additional transaction fees plus the standard transaction fees. Typically 0.0001 BTC per 1000 bytes [8]. Exact message size will depend on the implementation but could judiciously be 4000-5000 bytes for 0.0004-0.0005 BTC as the total cost per message. The total cost to users would be mixing fee plus a transaction fee. For a chunk v of 0.1 BTC, 0.01 fee rate shall apply plus taking transaction fee of 0.0004 BTC per message. It will effectively cost the user a total of 0.0014 BTC or 0.14% to mix. In RingCoin, multiple mixing rounds are not needed to achieve anonymity guarantee. The downside is the additional cost that comes with posting messages to public log but we believe the benefit outweighs the value of paying for anonymity loss.

6 Conclusions

RingCoin presented a mixing scheme that meets anonymity properties and compatible with Bitcoin. Our proposed protocol modifies mixcoin mixing protocol to further enhanced anonymity using ring signatures schemes incorporated with public append-only record and commitment to guarantee accountability. RingCoin scheme ensures that there is no linkage between input transactions and address(es) pairs and offers properties of an ideal mix which include accountability, anonymity, resilience, scalability, compatibility, and incentives. RingCoin scheme does not store annals of input transaction and addresses mapping. The introduction of ring signature hides inputs transactions and addresses mapping from the mix and append-only log keeps mix accountable. RingCoin requires non computationally complex cryptography.

Acknowledgments

The author gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in bitcoin," in *International Conference on Financial Cryptography and Data Security*, pp. 34–51, 2013.

[2] J. Bonneau, A. Narayanan, A. Miller, J. Clark, J. A. Kroll, and E. W. Felten, "Mixcoin: Anonymity for bitcoin with accountable mixes," in *International Conference on Financial Cryptography and Data Security*, pp. 486–504, 2014.

[3] S. C. Chang and J. L. Wu, "A privacy-preserving cloud-based data management system with efficient revocation scheme," *International Journal of Computational Science and Engineering*, vol. 20, no. 2, pp. 190–199, 2019.

[4] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp. 84–90, 1981.

[5] M. Conti, E. S. Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of bitcoin," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3416–3452, 2018.

[6] W. Fang, X. Z. Wen, Y. Zheng, and M. Zhou, "A survey of big data security and privacy preserving," *IETE Technical Review*, vol. 34, no. 5, pp. 544–560, 2017.

Table 3: The comparison results of RingCoin with other proposed schemes

Scheme	Traffic	Fee	Rounds	Time (min)
RingCoin	6 transactions	$v\rho$	one round	10 x ω x 6
Mixcoin [2]	2 transactions	$v\rho*$ round	multiple rounds	10 x ω x 2 x round
Blindcoin [25]	4 transactions	$v\rho$	one round	10 x ω x 4
Coinjoin [12]	1 transaction	-	one round	Negotiation + 10 x ω
Coinswap [13]	4 transactions	-	one round	10 x ω x 4

[7] A. K. M. B. Haque and M. Rahman, "Blockchain technology: Methodology, application and security issues," *International Journal of Computer Science and Network Security*, vol. 20, no. 2, 2020.

[8] S. Kasahara and J. Kawahara, "Effect of bitcoin fee on transaction-confirmation process," *Journal of Industrial and Management Optimization*, vol. 13, no. 5, pp. 1-22, 2017.

[9] A. Kumar, "Design of secure image fusion technique using cloud for privacy-preserving and copyright protection," *International Journal of Cloud Applications and Computing (IJCAC'19)*, vol. 9, no. 3, pp. 22–36, 2019.

[10] X. Li, Y. Mei, J. Gong, F. Xiang, and Z. Sun, "A blockchain privacy protection scheme based on ring signature," *IEEE Access*, vol. 8, pp. 76765–76772, 2020.

[11] Y. Liu, X. Liu, C. Tang, J. Wang, and L. Zhang, "Unlinkable coin mixing scheme for transaction privacy enhancement of bitcoin," *IEEE Access*, vol. 6, pp. 23261–23270, 2018.

[12] G. Maxwell, "Coinjoin: Bitcoin privacy for the real world," in *Development & Technical Discussion*, 2013. (`https://bitcointalk.org/index.php?topic=279249.0`)

[13] G. Maxwell, "Coinswap: Transaction graph disjoint trustless trading," *Development & Technical Discussion*, 2013. (`https://bitcointalk.org/index.php?topic=321228.0`)

[14] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2019. (`https://bitcoin.org/bitcoin.pdf`)

[15] D. A. Nair, "The bitcoin innovation, crypto currencies and the leviathan," *Innovation and Development*, vol. 9, no. 1, pp. 85–103, 2019.

[16] O. O. Olakanmi and A. Dada, "An efficient privacy-preserving approach for secure verifiable outsourced computing on untrusted platforms," *International Journal of Cloud Applications and Computing (IJCAC'19)*, vol. 9, no. 2, pp. 79–98, 2019.

[17] F. Reid and M. Harrigan, "An analysis of anonymity in the bitcoin system," in *Security and Privacy in Social Networks*, pp. 197–223, 2013.

[18] R. L. Rivest, A. Shamir, and Y. Tauman, "How to leak a secret," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 552–565, 2001.

[19] D. Ron and A. Shamir, "Quantitative analysis of the full bitcoin transaction graph," in *International Conference on Financial Cryptography and Data Security*, pp. 6–24, 2013.

[20] T. Ruffing and P. Moreno-Sanchez, "Valueshuffle: Mixing confidential transactions for comprehensive transaction privacy in bitcoin," in *International Conference on Financial Cryptography and Data Security*, pp. 133–154, 2017.

[21] T. Ruffing, P. Moreno-Sanchez, and A. Kate, "Coinshuffle: Practical decentralized coin mixing for bitcoin," in *European Symposium on Research in Computer Security*, pp. 345–364, 2014.

[22] T. Ruffing, P. Moreno-Sanchez, and A. Kate, "P2p mixing and unlinkable bitcoin transactions.," in *NDSS*, pp. 1–15, 2017.

[23] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka, "Security services using blockchains: A state of the art survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 858–880, 2018.

[24] S. Shalini and H. Santhi, "A survey on various attacks in bitcoin and cryptocurrency," in *International Conference on Communication and Signal Processing (ICCSP'19)*, pp. 0220–0224, 2019.

[25] L. Valenta and B. Rowan, "Blindcoin: Blinded, accountable mixes for bitcoin," in *International Conference on Financial Cryptography and Data Security*, pp. 112–126, 2015.

[26] Q. Wang, X. Li, and Y. Yu, "Anonymity for bitcoin from secure escrow address," *IEEE Access*, vol. 6, pp. 12336–12341, 2017.

[27] H. H. S. Yin, K. Langenheldt, M. Harlev, R. R. Mukkamala, and R. Vatrapu, "Regulating cryptocurrencies: A supervised machine learning approach to de-anonymizing the bitcoin blockchain," *Journal of Management Information Systems*, vol. 36, no. 1, pp. 37–73, 2019.

[28] J. H. Ziegeldorf, F. Grossmann, M. Henze, N. Inden, and K. Wehrle, "Coinparty: Secure multi-party mixing of bitcoins," in *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, pp. 75–86, 2015.

[29] J. H. Ziegeldorf, R. Matzutt, M. Henze, F. Grossmann, and K. Wehrle, "Secure and anonymous decentralized bitcoin mixing," *Future Generation Computer Systems*, vol. 80, pp. 448–466, 2018.

Biography

Albert Kofi Kwansah Ansah is pursuing a PhD degree in Computer Science and Technology at the School of Computer Science and Engineering in University of Electronic Science and Technology of China, Chengdu, PR China. He holds an M.S. degree in Mobile Computing and Communication from the University of Greenwich, London, UK in 2008 and PG Dip in Networks and Communication from the Westminster College, London, UK in 2006. He is a Lecturer in the Computer Science and Engineering Department at the University of Mines and Technology in Ghana. His research interests cover blockchain technology, Cryptocurrencies, privacy-Preservation and security, RFID systems, computer systems and Networks. He is a member of IEEE, International Association of Engineers (IAENG) and the Internet Society (ISOC) UK and Ghana chapters.

Fengli Zhang is a Professor in the School of Information and Software Engineering, University of Electronic Science and Technology of China. She received her Ph.D. degree from the University of Electronic Science and Technology of China (UESTC) in 2007, and M.S. degree in 1986. She joined the DataBase System and Mobile Computing (DBMC) laboratory of the University of Illinois at Chicago, USA from 2000-2002 as a visiting scholar. She has published more than sixty papers and translated and/or edited five books as a partaker.

Daniel Adu-Gyamfi obtained his PhD in Software Engineering and ME in Computer Science and Technology from the University of Electronic Science and Technology of China. He also obtained his BSc (hons.) in Computer Science from the University for Development Studies in Ghana. He has been a faculty member in Department of Computer Science and Informatics at the University of Energy and Natural Resources in Ghana since 2014. He has published over 20 articles in prestige refereed international journals and conferences. He has also reviewed for several refereed international conferences and journals including those from BMC, Elsevier and Springer. His current research areas include mobile data management and network security technologies. He is a member of ACM SIGCHI, ACM SIGSPATIAL and among others.

Research on Network Security Risk Assessment Method Based on Improved Analytic Hierarchy Process

Gang Wang
(Corresponding author: Gang Wang)
Shaanxi Police College, China
No. 1, Qiyuan 2nd Road, Weiyang District, Xi'an, Shaanxi 710021, China
Email: gwiays@126.com
(Received Mar. 29, 2019; Revised and Accepted July 13, 2020; First Online Apr. 25, 2021)

Abstract

Risk assessment can help understand network security. This paper mainly analyzed the analytic hierarchy process (AHP) method, improved the AHP method with the fuzzy operator, applied the improved AHP method to risk assessment, and took a local network as an example to evaluate its network security risk. The results showed that the probability of low risk in the network was the highest, 28%. Among the indicators established, the more important ones were transmission relay failure, software, and hardware failure, no data backup, *etc.*. The above aspects should be strengthened in the network security construction. The experimental analysis results verify the effectiveness of the improved AHP method in risk assessment, which provides some reliable bases for the formulation of defense strategy.

Keywords: Analytic Hierarchy Process; Fuzzy Operator; Network Security; Risk Evaluation

1 Introduction

With the popularity of the Internet [8], it plays a more and more important role in people's life and work and provides great convenience to many fields, such as politics, economy, entertainment, *etc.* [2]; however, the issue of network security is also becoming more prominent: the number of spam mails increases rapidly, harmful information spreads faster, and the attack means of hackers also becomes more complex and diversified [10]. The increasing network security incidents have brought a great threat to society and the economy, and the network security issues have been paid more attention to by researchers [5, 7, 15, 20].

Risk assessment can help managers understand the current and future risks of the network [1] and provide some reliable bases for establishing defense strategies, which are more conducive to the safe operation of the network. The method of risk assessment has also been widely concerned by researchers [14]. Xu *et al.* [22] designed a method based on non-cooperative differential game theory, which regarded the process of risk assessment as a differential game of optimal resource control. The experiment found that the method was feasible. Deng *et al.* [4] proposed a method based on the rough set and gene expression program to mine security risks and predicted and analyzed risk levels. The experiment showed that the method had a high mining efficiency and strong practicability.

Wang *et al.* [21] improved the factor analysis of information risk (FAIR) with the Bayesian network and obtained a FAIR-BN model. The experiment showed that the model was more accurate, flexible and extensible, and had the potential to provide solutions for decision-making. Based on the network penetration test, Sun *et al.* [19] generated an attack graph, calculated the attack probability of the atomic node, used the Markov chain to calculate the attack transition probability, and selected the best attack path to realize the evaluation of network security risks. The simulation results showed that the method could make an objective response to the actual situation of the network. Based on the analytic hierarchy process (AHP) and fuzzy operator, this paper designed an improved AHP method, applied it to risk assessment, and made an experimental analysis to understand the reliability of the method. This paper makes some contributions to the better realization of network security.

2 Risk Assessment Method

At present, the commonly used risk assessment methods can be divided into four types.

1) Qualitative analysis method: Based on the work experience and theoretical knowledge of the assessors, the risk level is divided according to some evaluation standards and similar cases in the past. However, this method generally has strong personal subjectiv-

ity. The specific methods include the historical comparison method, expert evaluation method [25], *etc.*

2) Quantitative analysis method: Risk factors were represented by specific values. With strong objectivity, this method can help people to observe and analyze the evaluation results clearly. The specific methods include fuzzy comprehensive evaluation method [24], back-propagation (BP) neural network [23], grey model [3], *etc.*

3) Qualitative and quantitative analysis combined method: The quantitative method is applied for quantifying the risk factors that can be quantified, and the qualitative method is used for analyzing the factors that cannot be quantified. The combination can more comprehensively describe the whole evaluation process.

4) Model evaluation method: The method evaluates the whole system with the model analysis tool. This method can find out the unknown vulnerable points and the security risks in the system. The specific methods are information flow model, fault tree model [17], graph model [13], *etc.*

The risk of network security involves software, hardware, environment, *etc.* [6], which is generally analyzed from three aspects. The first aspect is assets. Assets refer to valuable information, data, and resources, and risk assessment is related to the importance of assets. The second aspect is threats. Threats refer to the possibility of causing negative impacts on assets based on weakness, for example, threats from viruses and hackers. Risk assessment is related to the possibility of threats. The third aspect is vulnerabilities. Vulnerabilities refer to the weak links of assets that may be threatened, such as the deficiency of network software, hardware and defense measures. Once the deficiencies are used, assets will be damaged. Risk assessment is related to the severity of vulnerabilities.

3 Improved AHP Method

3.1 Basic Principle of AHP

AHP is a multi-criteria decision-making method [16]. In the risk assessment of network security, many factors are difficult to quantify; therefore, it is necessary to adopt appropriate methods to evaluate the importance of these factors to realize the assessment of network risk. There are four steps in AHP.

1) Building a hierarchical structure model:
Building a model aims to analyze the risk problem in detail. Generally speaking, the model includes three layers, as shown in Figure 1. The target layer has only one element, which is the predetermined goal of the problem. The criterion layer refers to a series of intermediate links involved in achieving the goal. The scheme layer is the optional scheme and measures needed to achieve the goal.

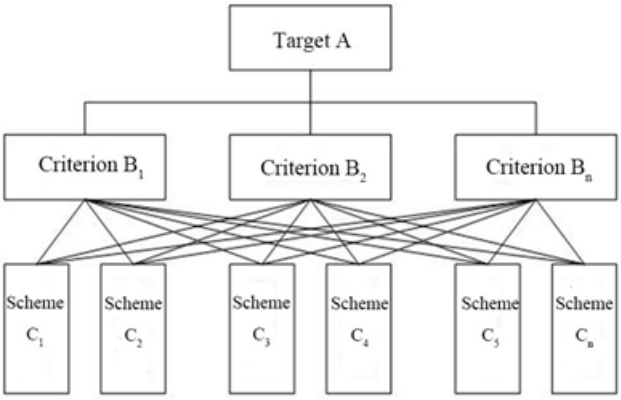


Figure 1: A hierarchical structure model

2) Establishing a judgment matrix:
AHP requires to calculate the relative importance of different factors layer by layer and quantify it into a judgment matrix. For example, scheme B is associated with criterion A of the last layer. The judgment matrix can be written as:

$$A\_B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

where b_{ij} is the importance of b_i and b_j relative to A. AHP adopts the Saaty1-9 scale to quantitatively describe the value, as shown in Table 1. The value can be obtained through the Delphi method.

3) Calculating single hierarchical arrangement:
Firstly, all the elements are normalized; then, they were added up according to the row, and $\bar{w_i} = \sum_{i=1}^{n} \bar{b_{ij}}$ is obtained, where $\bar{w_i}$ refers to the feature vector of the matrix and $\bar{b_{ij}}$ is the element obtained after normalization. $\bar{w_i}$ is normalized again, and weight $w_i = \frac{\bar{w_i}}{\sum_{j=1}^{n} \bar{w_i}}$ is obtained. The weight of every layer is calculated; then, the result of a single hierarchical arrangement is obtained. However, in the actual calculation process, there may be some inconsistency in the matrix; thus, it is necessary to check the consistency of the matrix after the calculation.

Firstly, a consistency test index (CI) is established, $CI = \frac{\lambda_{\max} - n}{n-1}$, where $\lambda_{\max}$ refers to the maximum feature value of the matrix. $CR = \frac{CI}{RI}$ is calculated, where RI refers to the average random consistency index. The values given by Satty are shown in Table 2. If the calculated result is $CR \leq 0.1$, the matrix has consistency; otherwise, the matrix needs correction.

Table 1: The Saaty1-9 scaling method

Importance scale	Explanation
1	b_i is no less important than b_j
3	b_i is a little important than b_j
5	b_i is significantly more important than b_j
7	b_i is strongly more important than b_j
9	b_i is extremely more important than b_j
2, 4, 6, 8	The median values of the above judgment
Reciprocal	The ratio of the importance of b_j to the importance of b_i is $b_{ji} = \frac{1}{b_{ij}}$

Table 2: Comparison table of consistency check

Matrix Order	RI
1	0
2	0
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45

4) Calculating total hierarchical arrangement:
After calculating the single hierarchical arrangement and performing a consistency test on the matrix of each layer, all the factors of every layer are calculated. Based on the combined weight of the target layer, the total hierarchical arrangement is performed.

3.2 Improved AHP

AHP can not directly estimate the risk level. This paper improves AHP with the fuzzy operator. In the risk assessment, the risk is divided into five levels, namely very low (VL), low (L), medium (M), high (H), and very high (VH). The membership of different indicators is represented by the Gaussian membership function, and the formula is as follows:

$$\begin{aligned} f_{VL}(x) &= e^{-\frac{x^2}{2\times 0.1^2}} \\ f_L(x) &= e^{-\frac{(x-0.25)^2}{2\times 0.1^2}} \\ f_M(x) &= e^{-\frac{(x-0.5)^2}{2\times 0.1^2}} \\ f_H(x) &= e^{-\frac{(x-0.75)^2}{2\times 0.1^2}} \\ f_{VH}(x) &= e^{-\frac{(x-1)^2}{2\times 0.1^2}} \end{aligned}$$

Then, the membership matrix of the criterion layer can be written as:

$$R = \begin{bmatrix} f_{VL}(x_1) & f_L(x_1) & f_M(x_1) & f_H(x_1) & f_{VH}(x_1) \\ f_{VL}(x_n) & f_L(x_n) & f_M(x_n) & f_H(x_n) & f_{VH}(x_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{VL}(x_m) & f_L(x_m) & f_M(x_m) & f_H(x_m) & f_{VH}(x_m) \end{bmatrix}$$

According to the criterion layer weight W_i and matrix R_i, a fuzzy evaluation matrix is obtained:

$$D_i = W_i \times R_i$$

According to the level one fuzzy comprehensive evaluation matrix and based on the target layer weight W, a level two membership matrix is established, written as:

$$\begin{aligned} S &= \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} \\ &= \begin{bmatrix} W_1 \times D_1 \\ W_2 \times D_2 \\ \vdots \\ W_n \times D_n \end{bmatrix} \end{aligned}$$

The level two fuzzy comprehensive evaluation matrix of the target layer can be written as:

$$A = W \times S,$$

i.e., the final risk level.

4 Experimental Analysis

Taking a local network as an example, the improved AHP was applied to the risk assessment of network security. The risk level has five levels, and the division is shown in Table 3.

First of all, the corresponding hierarchical structure model needed to be established. Considering the assets, threats, and vulnerabilities of the network, the corresponding scheme layer indicators were determined. A total of 15 indicators were determined. The established model is shown in Table 4.

Table 3: Risk classification

Value-at-risk	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1
Level	VL	L	M	H	VH

Five network security experts were invited to determine the weight. First, the judgment matrix of the criterion layer relative to the target layer was established by the 1-9 scaling method, as shown in Table 5.

According to Table 5,

$$A\_B = \begin{bmatrix} 1 & 1/3 & 5 \\ 3 & 1 & 7 \\ 1/5 & 1/7 & 1 \end{bmatrix}$$

Through calculation, the maximum feature value $\lambda_{\max}$ of the matrix is 3.06, and the RI value is 0.52; then, its consistency indicator is:

$$CR = \frac{CI}{RI} = \frac{3.06 - 3}{3 - 1} = 0.03 < 0.1,$$

which shows that the obtained matrix satisfies the consistency. After normalization, the weight of each layer is obtained:

$$\begin{aligned} C_1 &= 0.28 \\ C_2 &= 0.65 \\ C_3 &= 0.07 \end{aligned}$$

, , . The weight of each layer is calculated one by one using the same method. After the consistency test, the final results are shown in Table 6.

It was seen from Table 6 that the weight of threat was the largest in the criterion layer, followed by asset and vulnerability, which indicated threat brought the greatest risk in the network and was the most important in risk assessment. Among the indicators of the scheme layer, in addition to the three indicators related to assets, the indicators with higher weight were C6 (transmission relay failure), C11 (no data backup), and C12 (no relay link protection), which indicated that these indicators had important impacts on the risk in the risk assessment.

Then, the experts evaluated the indicators and took the average values. The results are shown in Table 7.

According to Table 7, based on the level one fuzzy comprehensive evaluation, the evaluation results of different risk factors are obtained:

$$\begin{aligned} B_1 &= (0.15\ 0.22\ 0.15\ 0.31\ 0.07) \\ B_2 &= (0.17\ 0.32\ 0.21\ 0.08\ 0.01) \\ B_3 &= (0.17\ 0.21\ 0.22\ 0.09\ 0.04) \end{aligned}$$

On this basis, the total fuzzy evaluation results are calculated. Finally, the risk evaluation result of the network is:

$$A = [0.17\ 0.28\ 0.15\ 0.08\ 0.02].$$

The probability of risks in this network is shown in Figure 2.

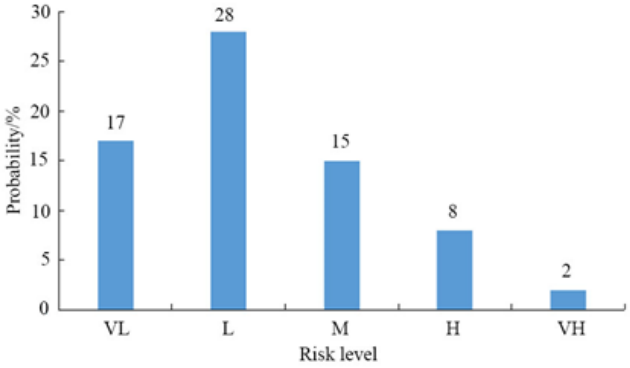


Figure 2: Network security risk level

As shown in Figure 2, the probability of the very low risk in the local network was 17%, the probability of the low risk was 28%, the probability of the medium risk was 15%, the probability of the high risk was 8%, and the probability of the very high risk was 2%. Overall, the probability of the low risk in the network was the highest, but there was also the possibility of the high risk. Therefore, it is necessary to strengthen the protection of the network and realize the safe operation of the network through technologies such as intrusion detection.

5 Discussion

Network security refers to protecting network hardware, software, and information data from being damaged or leaked due to unexpected and malicious factors to ensure uninterrupted network service. For network security, many technologies and methods have been applied, such as firewall [18], intrusion detection [9], situation awareness [11], data encryption [12], *etc.* However, before arranging these methods, the security risk of the network needs to be evaluated first to understand the security situation of the network better and make a reasonable arrangement.

In this study, the AHP method was improved to make it have a good application in the network security risk assessment. An experiment analysis was carried out by taking a local network as an example. The evaluation results of the improved AHP method showed that the network assets had an important impact on the network security, and the probability of the high risk in assets was highest, 31%, and the weight was 0.28. The above results showed that the protection of assets was a very important part of network security. From the perspective of threat, it was found that the most important threat was transmission relay failure. In the local network, its transmission depended on the transmission network. However, due to the mismatch of optical power, the transmission process

Table 4: The hierarchical structure model

Target layer	Criterion layer	Scheme layer
Network security risk assessment index system A	Asset B1	Deliberate theft C1
		Deliberately tamper with C2
		File missing C3
	Threat B2	Hardware and software failure C4
		Operation failure C5
		Transmission relay fault C6
		Malicious code C7
		Hacker attacks C8
	Vulnerability B3	Equipment aging C9
		Unprotected core disk C10
		No dada backup C11
		No relay link protection C12
		Insufficient anti-virus measures C13
		Insufficient anti-attack capability C14
		Multiple links are connected by C15

Table 5: The judgment matrix of level one indicators

	B1	B2	B3
B1	1	1/3	5
B2	3	1	7
B3	1/5	1/7	1

Table 6: Determination of indicator weight

Criterion layer	Weight	Scheme layer	Weight
B1	0.28	C1	0.33
		C2	0.33
		C3	0.33
B2	0.65	C4	0.22
		C5	0.16
		C6	0.32
		C7	0.15
		C8	0.15
B3	0.07	C9	0.11
		C10	0.09
		C11	0.27
		C12	0.25
		C13	0.05
		C14	0.05
		C15	0.18

Table 7: Expert risk assessment results

Scheme Layers	Assessment Results				
	VL	L	M	H	VH
C1	0.12	0.24	0.57	0.05	0.02
C2	0.33	0.25	0.31	0.07	0.04
C3	0.18	0.21	0.41	0.1	0.1
C4	0.22	0.44	0.21	0.08	0.05
C5	0.23	0.31	0.23	0.15	0.08
C6	0.05	0.2	0.2	0.55	0
C7	0.26	0.25	0.27	0.15	0.07
C8	0.31	0.27	0.25	0.09	0.08
C9	0.22	0.26	0.24	0.18	0.1
C10	0.07	0.45	0.21	0.21	0.06
C11	0.05	0.12	0.11	0.67	0.05
C12	0.08	0.11	0.05	0.48	0.28
C13	0.21	0.23	0.22	0.21	0.13
C14	0.22	0.21	0.25	0.18	0.14
C15	0.23	0.21	0.24	0.15	0.17

might be interrupted frequently. Also, with the development of urban construction, the growth of construction engineering has aggregated artificial cutting, which led to frequent transmission failures. Another threat with high weight was hardware and software failure; therefore, in the construction of network security, it is necessary to replace the damaged equipment in time and increase the maintenance and management of network equipment to reduce such risks.

Overall, for network threats, the probability of the low risk was the highest, 32%. Finally, from the perspective of vulnerability, it mainly showed the possibility of the medium risk, 22%. The relatively important vulnerabilities are "no data backup" and "no relay link protection". Therefore, it is necessary to strengthen the management of these two items, i.e., setting up a good protection link for the relay link in the network and strengthen the data backup, to reduce the security risk of the network. On the whole, the overall risk of the local network studied was low. According to the results of risk evaluation, the proposed network security measures include:

1) Reducing the human-made cable damage to avoid transmission relay failure;

2) Strengthening the backup of important data;

3) Improving the ability to prevent viruses and attacks and deploying the corresponding firewall and intrusion detection system.

In this study, although some achievements have been made in the research of network security risk assessment, there are still some shortcomings, which need to be solved in future workd. For example, the division of risk hierarchy should be divided in more detail to more comprehensively describe the security risk of the network; the AHP method should be further optimized to reduce the subjectivity of expert evaluation; the correctness of the method should be verified in more data sets.

6 Conclusion

Aiming at the risk assessment of network security, this study designed an improved AHP method and applied it to a local network. Through an evaluation by the improved AHP method, it was found that the probability of the low risk in the network was high, 28%, followed by the very low risk, 17%. Then, according to the risk of the network, the network security measures were discussed. The results show that the improved AHP method has a good performance in risk assessment and can be further promoted and applied in practice.

References

[1] O. O. Abimbola, B. Akinyemi, T. Aladesanmi, G. A. Aderounmu, K. B. Hamidja, "An improved stochastic model for cybersecurity risk assessment," *Computer and Information Science*, vol. 12, no. 4, pp. 96, 2019.

[2] N. Athavale, S. Deshpande, V. Chaudhary, J. Chavan, S. S. Barde, "Framework for threat analysis and attack modelling of network security protocols," *International Journal of Synthetic Emotions*, vol. 8, no. 2, pp. 62-75, 2017.

[3] J. Chen, Z. Zhou, Y. Tang, Y. He, S. Zhao, "Research on network security risk assessment model based on grey language variables," in *IOP Conference Series: Materials Science and Engineering*, vol. 677, no. 4, pp. 042074, 2019.

[4] S. Deng, D. Yue, X. Fu, A. Zhou, "Security risk assessment of cyber physical power system based on rough set and gene expression programming," *IEEE/CAA Journal of Automatica Sinica*, vol. 2, no. 4, pp. 431-439, 2015.

[5] A. Dewanje and K. A. Kumar, "A new malware detection model using emerging machine learning algorithms," *International Journal of Electronics and Information Engineering*, vol. 13, no. 1, pp. 24–32, 2021.

[6] D. Henshel, M. G. Cains, B. Hoffman, T. Kelley, "Trust as a human factor in holistic cyber security risk assessment," *Procedia Manufacturing*, vol. 3, pp. 1117-1124, 2015.

[7] S. Islam, H. Ali, A. Habib, N. Nobi, M. Alam, and D. Hossain, "Threat minimization by design and deployment of secured networking model," *International Journal of Electronics and Information Engineering*, vol. 8, no. 2, pp. 135–144, 2018.

[8] A. Kak, "Computer and network security," *Friend of Science Amateurs*, vol. 31, no. 9, pp. 785-786, 2017.

[9] M. J. Kang, J. W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," *Plos One*, vol. 11, no. 6, pp. e0155781, 2016.

[10] Z. M. King, D. S. Henshel, F. Liberty, L. Flora, M. G. Cains, B. Hoffman, C. Sample, "Characterizing and measuring maliciousness for cybersecurity risk assessment," *Frontiers in Psychology*, vol. 9, 2018.

[11] Y. B. Leau, A. A. Khudher, S. Manickam, S. Al-Salem, "An adaptive assessment and prediction mechanism in network security situation awareness," *Journal of Computer Science*, vol. 13, no. 5, pp. 114-129, 2017.

[12] J. Li, "A symmetric cryptography algorithm in wireless sensor network security," *International Journal of Online Engineering*, vol. 13, no. 11, pp. 102, 2017.

[13] S. Liu, Y. Liu, "Network security risk assessment method based on HMM and attack graph model," in *17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD'16)*, Shanghai, pp. 517-522, 2016.

[14] T. Ncubukezi, "A Proposed: Integration of the monte carlo model and the bayes network to propose cyber security risk assessment tool for small

and medium enterprises in South Africa," *International Journal of Computer Science and Information Security*, vol. 3, no. 18, pp. 152-155, 2020.

[15] E. U. Opara, O. A. Soluade, "Straddling the next cyber frontier: The empirical analysis on network security, exploits, and vulnerabilities," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 10–18, 2015.

[16] A. A. Salo, R. P. Hämäläinen, "On the measurement of preferences in the analytic hierarchy process," *Journal of Multi-Criteria Decision Analysis*, vol. 6, no. 6, pp. 309-319, 2015.

[17] M. Sarbayev, M. Yang, H. Wang, "Risk assessment of process systems by mapping fault tree into artificial neural network," *Journal of Loss Prevention in the Process Industries*, vol. 60, pp. 203-212, 2019.

[18] X. Song, "Firewall technology in computer network security in 5G environment," *Journal of Physics Conference Series*, vol. 1544, pp. 012090, 2020.

[19] F. Sun, J. Pi, J. Lv, T. Cao, "Network Security risk assessment system based on attack graph and Markov chain," *Journal of Physics Conference Series*, vol. 910, pp. 012005, 2017.

[20] A. Tayal, N. Mishra and S. Sharma, "Active monitoring & postmortem forensic analysis of network threats: A survey," *International Journal of Electronics and Information Engineering*, vol. 6, no. 1, pp. 49–59, 2017.

[21] J. Wang, M. Neil, N. Fenton, "A bayesian network approach for cybersecurity risk assessment implementing and extending the FAIR model," *Computers & Security*, vol. 89, pp. 101659, 2019.

[22] H. T. Xu, R. J. Lin, "Resource allocation for network security risk assessment: A non-cooperative differential game based approach," *China Communications*, vol. 04, no. v.13, pp. 136-140, 2016.

[23] Z. Xu, J. Li, S. Xiao, Y. Yuan, "Study on security risk assessment of power system based on BP neural network," *Journal of Computational and Theoretical Nanoscience*, vol. 13, no. 8, pp. 5277-5280, 2016.

[24] B. Yi, Y. P. Cao, Y. Song, "Network security risk assessment model based on fuzzy theory," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 4, pp. 3921-3928, 2020.

[25] A. Youssef, "A delphi-based security risk assessment model for cloud computing in enterprises," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 1, pp. 151-162, 2020.

Biography

Wang Gang, born on Oct 8,1976, holds a master's degree and is an associate professor of shaanxi police college. He is interested in network security and big data.

S-PPOC: Multi-scheme Privacy-Preserving Outsourced Classification

Kwabena Owusu-Agyemeng, Zhen Qin,Hu Xiong,Tianming Zhuang, Liu Yao, and Zhiguang Qin
(Corresponding author: Kwabena Owusu-Agyemeng)

School of Information and Software Engineering, University of Electronic Science and Technology of China
No. 2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, Sichuan 611731, China
Email: cooljacko@gmail.address

(Received Mar. 4, 2020; Revised and Accepted Oct. 12, 2020; First Online Apr. 20, 2021)

Abstract

Human activity recognition HAR has gained tremendous research attention aiming at understanding the human activities in the field of ubiquitous healthcare. Deep learning has progressively been applied to extract, constructed features, and train models accurately for high-level data representations required for medical diagnosis. However, these deep learning models are affected by efficiency, accuracy and are also confronted with data privacy dilemmas. This paper proposes an architecture dubbed S-PPOC, a combination of batch normalization and polynomial approximation of Rectified Linear Unit (ReLU) activation function for privacy-preserving classification. This is to enforce both the confidentiality of the manipulated datasets and the model's efficiency to address these problems successfully. With the application of S-PPOC, the computational evaluator is capable of securely training a classification model over data encrypted with different fully homomorphic encryption schemes contributed by multiple Data Providers. Our scheme is proven to be secured in the honest-but-curious threat model compared to other existing cutting-edge privacy-preserving deep learning models.

Keywords: Batch Normalization; Classification; Human Activity Recognition; Privacy-Preserving; ReLU Activation Function

1 Introduction

Human activity recognition [12, 24] based on Wireless Body Area Network [40,41] (WBAN) has gained tremendous research attention in the field of ubiquitous healthcare [17, 32]. WBAN [33] as an extension of our traditional wireless sensor networks (WSN) through Human activity recognition detects, interprets and monitors human behavior, activity types and patterns, either occasional or habitual with the aid of wearable devices attached to the human body with the aim of transforming lives [20, 23, 28]. Numerous real-world applications such as medical diagnosis [13,24], athletic competition, nursing and smart homes has exploited and successfully benefited from the potential of the huge amount of data generated from these sensor-based human activity recognition.

To provide useful information to data owners and patients, storage and processing requirements of the massive amount of data generated by sensor-based HAR systems still remain an imperative task in the ubiquitous healthcare space. In the scenario of personal healthcare, fitness records monitored by WBAN [34], specific deep leaning models are trained on aggregated datasets from different parties for better representations and accurate diagnosis, which may be expensive or complicated for a single data owner to achieve. Fortunately, cloud computing comes in handy as a novel computing paradigm to eliminate the need to maintain expensive computing resources, dedicate storage and software at a nominal management cost. Recently, patients and hospitals have exploited the promising potential of cloud computing [10, 39, 45] to remotely store and train their deep learning models.

Arguable, deep learning models has been applied on the massive data generated by WSNs [49] through their low-capacity sensor to efficiently solve the evolving challenges during big data processing and complex representations in Human Activity Recognition. However, there are still prevalent deficiencies confronting HAR models:

1) Better representations of the models requires data aggregated from different parties which may be expensive or completed for a single party to achieve.
2) Security [4] of the aggregated datasets.
3) Privacy and security of the model. In addition, sensor nodes of WBAN [48] have challenges with energy consumption of these wearable devices which is outside the scope of this paper.

Furthermore, training of deep learning models requires access to the unprotected datasets which is customarily privacy sensitive. Unfortunately, cloud-based classification would not be preferred by the data owners without guarantee for confidentiality, privacy and security [6, 7, 16, 51]

of the outsourced datasets. Since the threats in the cloud jeopardize some of the basic security requirement. Therefore, it is essential to protect these confidential data before outsourcing any computation to the untrusted third party [29, 46]. One of the promising solutions is for users to encrypt their private data before outsourcing computations to the Data Service Provider (DSP). However, due to organizational policies mutually untrusted parties would like to encrypt their data with different fully homomorphic encryption schemes before the untrusted third party can aggregate encrypted datasets and perform any computations. The Multi-key FHE [30] offers a suitable scheme for aggregating the datasets under different keys in such domains. This powerful scheme has been integrated with deep learning in the pioneering architecture of [25], known as Multi-key privacy-preserving deep learning MK-PPDL which built a convolutional neural network on Multikey fully homomorphic encryption to process inference queries.

Existing endeavors on collaborative privacy preservation in outsourcing computations are mostly based on encrypting the datasets with same public keys [21, 22]. Considering a more practical scenario where outsourced datasets are from diverse locations and parties. For example, the network infrastructure of a communication network may belong to multiple providers; Base stations of a sensor network may be owned by multiple medical institutions. To be more precise, we are considering the following scenario in this work.

- For n mutually non-colluding data providers $P_1, P_2, \cdots, P_n$, every data owner $P_i (i \in [1, n])$ encrypt private data with their respective scheme before outsourcing the ciphertext to a Data Service Provider before allowing the Computational Evaluator to concatenate the datasets before any computations are performed.
- Computational Evaluator performs part of the computational classification over the concatenated data. All contributed data and intermediate results still remains confidential.
- There is no interaction between Data Owners, Data Services Provider and Data Providers.
- Immediately the classifier is trained, the computational evaluator can respond to the client query without learning any of the intermediate results.

In this domain, we focuses on multi-scheme privacy-preserving outsourced deep learning classification over data encrypted with *different encryption schemes or even different public keys*.

The multi-scheme privacy-preserving outsourced classification models are confronted with three fundamental issues:

P1. The MK-FHE scheme in [25] was built upon provable secure NTRU encryption and had a limitation that the maximal number of participants in a computation had to be known at the time a key was generate. Thus making it computationally expensive and undesirable for large-scale *deeper* neural networks.

P2. Classification over sensitive data is extremely complicated, since it involves additions, multiplications and the nonlinear (ReLU, Sigmoid) polynomial approximations, which leads to low efficiency and degraded accuracy in practice.

P3. Over the past decade, numerous privacy-preserving collaborative deep learning frameworks with the polynomial approximation of activation functions have been developed [43]. Training is unstable due to approximated polynomial activation functions but only support inference. Privacy-preserved training is considered in [11] which also utilizes polynomial approximation (e.g. Standard Chebyshev) to outwit the difficulty of activations. Thus, it suffers from accuracy loss and training instability.

In this paper, we propose a novel architecture dubbed **M**ulti-**S**cheme **P**rivacy-**P**reserving **O**utsourced deep learning **C**lassification in Cloud Computing(**S-PPOC**). This is to resolve the three outlined concerns in the existing framework [25]. The inherent strategy is to demonstrate that, Computational Evaluator is capable of utilizing the outsourced ciphertext for the evaluation of privacy-preserving classification model for deep neural network with depth greater than two and respond to Data Requesters predicted query. To end up with a construction being FHE friendly while keeping accuracy at a higher level, the multiplicative depth of the deep neural network is to be maintained reasonable low.Hence, our technique is much more efficient than the general secure multi-party evaluation [9, 25] which is impractical when subjected to larger number of trails. In addition, we present a security analysis of our proposed S-PPOC.

We advance the state-of-the-art technique of improving [9, 25] with the following main contributions:

- We propose a secure S-PPOC, which allows the data analysts to make privacy queries with higher accuracy predictions for patient's diagnosis without leaking sensitive information. In S-PPOC, the massive amount of HAR datasets from different data providers are securely stored and aggregated in the cloud server and can be used to build classifiers with the aid of the deep neural network. In this setting, the cloud server is not required to choose a type of fully homomorphic encryption scheme for data providers, rather DP can encrypt their data with their preferred choice of FHE encryption scheme. With the aid of the classifiers on the cloud server, data analyst can perform privacy-preserving disease diagnosis.
- To improve efficiency and circumvent the classification errors that dramatically degrade the accuracy

of the models, S-PPOC combine the polynomial approximation of ReLU [14] activation function with batch normalization without information leakage to facilitate secure outsourced multi-party computation of ciphertexts.

- To validate the efficiency and accuracy of S-PPOC, we demonstrate the theoretical analysis and experimental simulations that indicate the feasibility and practically secure model required to achieve the privacy-preserving classification in the semi-but-honest threat model.

The rest of the paper is organized as follows. In Section 2, we provide an overview of the related work. Section 3 gives some notations and presents the definition of homomorphic encryption and classification related to this paper. The system architecture is given in Section. 4. The collaborative privacy-preserving classification system based on the Multi-scheme fully homomorphic encryption scheme is illustrated in Section 5, while analyzing security in Section 6. Finally, the whole paper is concluded in Section 7.

2 Related Works

Most of the existing privacy-preserving methods in machine learning are based on secure multiparty (SMC) computations, fully homomorphic encryption, garbled circuits or combination of both. Concretely, the structure of privacy-preserving machine learning can basically be divided into two main categories:

1) Training phase, which requires an efficient algorithm to solve the optimization problem by using secure samples;

2) The classification phase, which predicts the classification of new samples. The efficiency of this proposals strongly depends on the complexity of the deep neural network *i.e.* the approximation of the activation function, number of layers, and number of neurons per each of the layers for the classification function.

Previous endeavors focused on the first or the second categories. Our proposed framework is directed towards the protection of the privacy of the training datasets, parameter of the classifiers and the intermediate results.

2.1 Privacy-Preserving Training

Several related works on the problem of privacy-preserving deep neural network learning [18] have recently been presented. For example, Graepel *et al.* established that some of the basic machine learning algorithms, such as classification can be computed efficiently over a small scale encrypted datasets. However, as the input size grows the efficiency will also be degraded rapidly. Theoretically, most of the issues confronting multiparty privacy preserving deep learning can be resolved with the aid of secure multi-party computation (SMC) protocols, but the extremely computational complexities and high communication usually makes it impractical, whiles the two-party case is not an exception. In [47], with the adoption of double homomorphic encryption scheme(BGN), Yuan *et al.* proposed a system in which the training phase was securely delegated to the data service provider for efficient computations in the multi-party scenario. In multiparty scenario, back-propagation network learning has been one of the widely used learning methods. In [3, 50] applied BGN in other to support secure computation operations and also realized a high-order back-propagation algorithm computation model training in data service provider. Chen *et al.* [2] also propose a privacy preserving back-propagation network learning algorithm for two-party scenarios with the use of secure scalar and secrete sharing to protect data and the intermediate results against leakages during the training of the deep neural network model. Though this scheme is applied in the two-party settings, however extending it to the multi-party models might not be trivial. In [43], Takabi *et al.* proposed a privacy-preserving back-propagation algorithm for the multi-party deep learning over arbitrarily partitioned datasets base on BGN homomorphic encryption. One of the bottlenecks of this scheme is where all the data owners must stay online and work interactively do control and noise in the ciphertext and also to decrypt the ciphertext of the intervening parameters in each iteration.

2.2 Privacy-Preserving Classification

Classification is one of the most fundamental task in deep learning and data mining, let consider prediction of medical diagnosis with generated data from wireless sernsor networks [5, 19] (WSN), models generated from existing data and if new patient record are similar to the existing data used for the model, then we can obtain a correct prediction for the new patients diagnosis. However, information from a predictors datasets are sensitive, these include, temperature, gender, age, medical history and symptoms. However, to preserve the privacy of the patients datasets, it is important to propose a privacy-preserving medical diagnosis model. Deep neural network model generalization performance is highly stimulated by the quality and volume of datasets used in the training process. As a result of this, privacy-preserving machine learning via secure multiparty computations (SMC) provides a promising solution by allowing multiple entities to train various models, researchers have recently proposed SecureML, CryptoDL, CryptoML, and schemes for secure delegation of iterative collaborative machine learning to the third party cloud servers. In MK-PP, which is the most popular scheme in the collaborative privacy-preservation deep learning settings is based on multi-key fully homomorphic encryption. In this scheme data owners encrypts their own datasets with different public keys and uploads them to the data service provider whiles the cloud server implement the

deep learning algorithm by the use of multi-key FHE. Based on the related works described above, it is obvious that most of the deep learning secure multiparty based approaches refer to semi-homomorphic encryption or multi-key fully homomorphic encryption.

The existing major interesting challenges confronting these scheme is whether all the FHE schemes from the data providers can be made multi-key to enable deep learning privacy preservation computations and is it also possible to homomorphically evaluate a classifier $\mathcal{C}$ on a ciphertext. the adaptation of a classification algorithm to render it compatible with the fully homomorphic encryption scheme whiles maintaining accuracy and efficiency. Lastly, most of the existing deep neural networks and its utility is very low with non-linear layers more than 2 (since it is predominant in current applications).

3 Preliminaries

This section defines some notations and review some cryptographic scheme used in our architecture. The notations are listed in Table 1 Fully homomorphic encryption (FHE) supports meaningful unlimited addition and multiplication computations over encrypted data with results of addition and multiplications which is equivalent to computations over plaintext by the application of the corresponding ciphertext directly without decryption.

Table 1: Notations of our protocol

Acronym	Description
CE	Computational Evaluator
CSP	Crypto service provider
(pk_i, sk_i)	Data provider public and secret key
(pk_0, sk_0)	CSP public and secret key
$\mathcal{C}$	Classifier
$\mathcal{C}(x, \theta)$	Prediction of $\mathcal{C}$ with input x and θ

3.1 Homomorphic Encryption

FHE scheme is $\varepsilon_i = (KeyGen_i, Enc_i, Dec_i, Eval_i)$ with a Key generation algorithm $KeyGen$ takes a parameter λ as input and outputs a public key pk_i and a private key pk_i *i.e.* $(pk_i,\ sk_i) \leftarrow KeyGen(\lambda)$. The encryption algorithm Enc_i takes as input public key pk_i and plaintext x and returns the corresponding ciphertext ψ *i.e.*, $\psi \leftarrow Enc(pk_i, x)$. Decryption algorithm Dec_i takes the private key sk_i as input and ciphertext ψ and returns x as plaintext, *i.e.* $\psi \leftarrow Dec(pk_i, \psi)$. Evaluation algorithm $Eval$ takes pk_i as input, a circuit $\mathcal{C} \in \mathcal{C}_{\prod}$, and a tuple of ciphertext $\{\psi_1, \psi_2, \cdots, \psi_n\}$ and returns a ciphertext ψ_i *i.e.* $c \leftarrow Eval(pk, \mathcal{C}, \psi_1, \psi_2, \cdots, \psi_n)$,such that $\mathcal{C}(x_1, x_2, \cdots, x_n) = Dec(sk, \psi)$, where $\mathcal{C}_{\prod}$ is a permitted circuit set and $\psi_i \leftarrow Enc(pk, x_i)$, $i \in [1, n]$. The circuits $\mathcal{C}(\in \mathcal{C})$ might be any circuit with different functions.For example $\mathcal{C}$ is decryption circuit $D_{\prod}(D_{\prod} \in C_{\prod})$. The decryption operation can be performed by the Eval.$\prod$ is therefore bootstrappable in this instance.

3.2 Classification in Machine Learning

Most of the existing machine learning algorithms are basically used for classification [15]. Classification models can be categorized into two major phases: inference stage and decision stage. During the inference stage, training sets are used to learn a model for the computation of the feature vectors whiles in the decision stage these feature vectors are applied for optimal class assignments. On the other hand, classification models can also be viewed as implementing a set of discriminant functions, $f_1(x), \cdots, f_n(x)$ mapping each input x directly inot one of the N class labels, denoted by B_n, where the input vector x with k attributes: $x = (x_1, x_2, \cdots, x_k) \in \mathbb{R}^k$. we therefore choose B_n if $f_n(x) = max f_i(x)$, not B_n.

Training in the classification algorithm involves two main groups: supervised and unsupervised. Supervised classifiers contains targeted features or labels, whiles the unsupervised classifiers experience a dataset containing many feature samples helping the algorithm to learning useful knowledge of the structure of these datasets.

4 Problem Scenario

Consider a scenario in a Ubiquitous healthcare system: A healthcare center has created a deep learning solution which is capable of assisting medical diagnosis, financial discovery and financial review of a patient. Interested hospitals might want to improve the model and also take advantage of the services offered by the model. This can be achieved by contributing their data for the collaborative training of the model. Privacy issues and fear of reverse-engineering of their solutions increases the unwillingness of the client to send their data to the other parties, therefore making this server-sided services impossible.

To preserve the privacy of Data Providers data, all parties would have to crowd-source their data in an encrypted form in such a way that only the Data Provider could decrypt. In this settings the client due to organizational policies is required to encrypt their data with their choice of encryption scheme or even keys different from the other Data Providers. Under a typical encryption scheme, the organization could not process this input in any meaningful way.In this domain, basic solution to this problem would require the Data Provider's input is not revealed to the Data Service Provider and the weights of the model are also not revealed to the client. Data Service Provider should be able to support arbitrary deep models with common operations which include CNN and ReLU.Our approach is efficient and capable of preserving the privacy of Data Provider's data without sacrificing security or accuracy with the aid of fully homomorphic encryption (FHE).

4.1 System Model for S-PPOC

In this section, we give some intuition behind our multi-scheme privacy-preserving classification in the cloud S-PPOC. A secured multi-party deep learning model follows the insight of prior works [25]. In order to offer, confidentiality, privacy and security of the data, parameters and models of the neural network whiles improving generalization of the model, we adopt *Multi-scheme fully homomorphic encryption,DBLP:journals/iacr/LiL13* (MS-FHE). We incorporate the idea of outsourcing technique with each of the Data Providers encrypting their datasets with their choice of fully homomorphic encryption. Our protocol consist of the following components and entities.

- *Data Providers*(DP): Each of the Data providers $P_i \in [1, n]$ outsourced their data to the Computational Evaluator CE for storage and also allow computational operations on these stored datasets. The datasets may contain sensitive information. The Data Providers $P_i \in [1, n]$ encrypts their sensitive data before outsourcing it to the CE to prevent privacy leakages.

- *Honest-but-curious Computational Evaluator*(CE): Computational Evaluator provides services such as storage and response to queries for Data Providers and Data Requestors. CE can basically represent a medical institution in this scenario. The CE stores the datasets encrypted with different fully homomorphic encryption or even different public keys chosen independently by the Data Providers. The CE therefore aggregates the ciphertext and uses the concatenated ciphertext to construct a classification model. The trained classifier is then used by the CE for data analytics or predications for the Data Requestors query.

- *Honest-but-curious crypto service provider* (CSP): The Crytpto Service Provider is capable of providing only crypto services to the Data Requestors. For example, the CSP is responsible for handling the ciphertext and also performs the decryption of the ciphertext sent by the CE.

- *Data Requestors* (DR): The DR is capable of gathering feature vectors of their records. In situations where DR wants to query the CE for prediction service securely, then DP encrypt their queries before outsourcing to the CE.

The interactive scenario between the entities and components is illustrated in Figure 1.

4.2 Threat Model

In this settings the entities involved in the privacy-preserving process *i.e.* Data Providers, Data Requestors, Computational Evaluator and Crypto Service Provider are expected not to collude with each other. S-PPOC is based on honest-but-curious model, in this settings malicious attackers may exist around Data Providers and steal information during outsourcing of training datasets. Then, we consider the Computational Evaluator to be curious-but-honest. During the training process, Computational Evaluator may be curious about participant's local datasets. CP may be strictly following the training protocol whiles trying to violate and disclose participant's sensitive information.

4.3 Design Goals

As a privacy-preserving collaborative deep learning model, S-PPOC allows the Computational Evaluator to process and construct a privacy-preserving classifier over the collaborative Data Providers and response to the Data Requestors prediction queries. S-PPOC should meet the following requirement to address the security and adversary models:

- *Classifier accuracy*: S-PPOC should be capable of classifying correctly for every query from the Data Requestors whiles making accurate predictions with high probabilities.

- *Privacy-Preservation*: Provision of privacy guarantee by the system should be assured without disclosing confidential information about Data Provider's and the classification model. To achieve this security goal, training and prediction stage should be performed in the encrypted settings. The Data Requestor is the only who will be capable of obtaining the decrypted intermediate results by the application of the private key after the Computational evaluator has responded to predictive query. The Crypto Service Provider is not known and thereby hidden from the Data Providers.

- *Flexibility*: In S-PPOC, crypto service provider is not a fixed service provider. In this domain, the CSP can be different parties or institutions publishing their corresponding different schemes or different public keys based on different functions or motivations.

5 S-PPOC Scheme

5.1 Main Idea

S-PPOC scheme focuses on the training of a classification model over aggregated datasets from multiple Data Providers with the aim of offering a confidential prediction services for Data Requestors with this classification model. In this instance, a set of n mutually non-colluding DPs $P_1, P_2, \cdots, P_n$ outsourced their encrypted data to the CE for storage whiles allowing it to perform some computational operations on these concatenated datasets. In other to support the computation over the encrypted

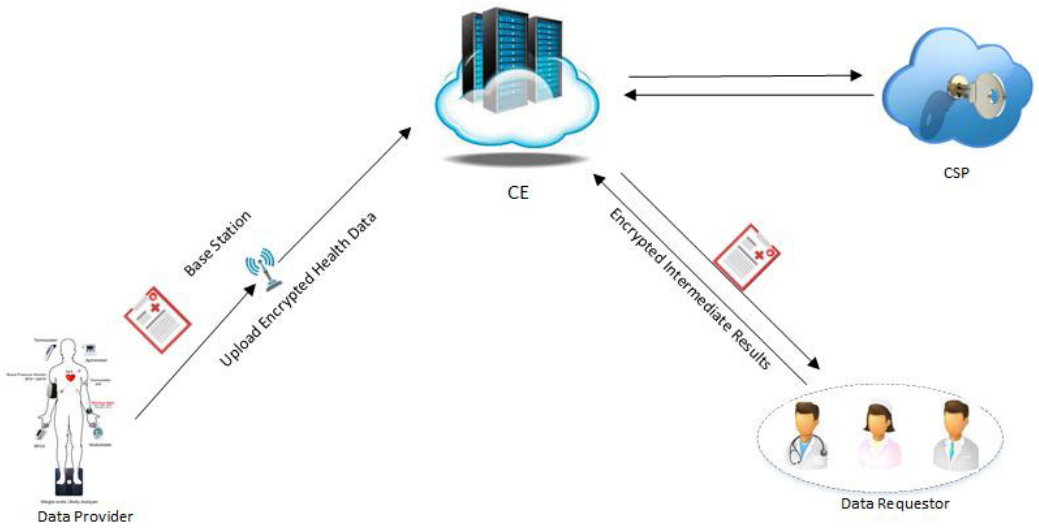


Figure 1: System model

data with possibly different fully homomorphic encryption schemes or with even different public keys, Multi-scheme Fully homomorphic encryption scheme will be used as our privacy-preserving technique to concatenate the encrypted dataset from DP before outsourcing to the Computational evaluator. Application of these outsourced concatenated ciphertext enables the CE to construct a classification model whiles storing and maintaining the model in an encrypted form. The classification model in the CE is therefore ready to respond to queries from DP.

5.2 Overview of S-PPOC Scheme

This section gives an overview of the S-PPOC scheme and demonstrate how to achieve the data outsourcing and secure classification.

- *Secure Data Outsourcing*: Encrypted data is been outsourced to the CE. DP are required to encrypt their data to preserve the confidentiality of the sensitive datasets based on their choice of Fully Homomorphic encryption. In this paper, the method for the aggregated confidential data processing is based on [27] ε^i which is capable of concatenating all the different public keys into ciphertext under same public key.

- *Secure training*: CE then used the outsourced concatenated datasets to build a classification model. The ciphertext encrypted with possibly different keys requires the CE and CSP to jointly train the classification model. The model is stored in an encrypted form in the CE after classification. CE then uses this model to provide secure prediction services to the DP query.

- *Secure Prediction*: Computational Evaluator CE run a classifier, denoted as $\mathcal{C}$ over encrypted database $\text{Enc}(\mathcal{X})$ and returns the result $\psi \leftarrow Evaluate(\mathcal{C}, \{\langle pk_i \psi_i \rangle\})$ which is finally send to the Data Requestor.

- *Secure Extraction*: One the Data Requestor obtains a prediction $\psi \leftarrow Evaluate(\mathcal{C}, \{\langle pk_i \psi_i \rangle\})$, the data is then decrypted with their private key sk_i to get the prediction.

5.3 Detail Design of S-PPOC

This section present a more detailed description of S-PPOC scheme which is basically divided into three phases: Privacy-preserving training classified, Generating Privacy-preserving prediction query from Data Requestors, preserving-preserving extracting results. The overall construction of our scheme is shown in Algorithm 3.

Stage 1: Privacy-preserving training of the classifier. In this section, we discuss the proposed S-PPOC model. The main work is directed towards the training of classifier over encrypted setting with data contributed from multiple Data Providers $\{P_1, P_2, \cdots, P_n\}$. For example, Alice encrypting Db_1 [3], Bob encrypted Db_2 with [42] whiles Eve also encrypt Db_3 with [8]. In this scenario, all parties are encrypting their data with different fully homomorphic encryption scheme. For a lucid discussion, initially a setup process for all the schemes is independently initialized and distribute the system security parameters. Meanwhile, in this stage, according to different target or motivation, determine a Crypto Service Provider entity with special function. Therefore, once the CSP is confirmed. It then distribute a public or private key pairs $\{pk_i, sk_i\}$.

In this process, each of the Data Providers independently generates a pair of public and pri-

vate keys $pk_i, sk_i \leftarrow KeyGen(1^{\lambda})$. Then they encrypts their secrete fields to generate cipher: $r_i = pub(r_i)||Enc(pk_i sec(r_i)$. All Data Providers outsource their encrypted data ψi, along with respect to encryption key pk_i to Computational Evaluator. CE construct an encrypted database $Enc(\mathcal{X}) \triangleq \{\langle pk_i, \psi_i \rangle\}$ from each Data Provider.

Polynomial approximation of ReLU. Our proposed model S-PPOC for a privacy-preserving classification on deep neural network has three major requirements: data privacy, efficiency with reasonably low multiplicative depth and accuracy which is close to the state-of-the-art convolutional neural network. The ReLU function and max pooling functions which have a high multiplicative depth in the CNN architecture are incompatible with the efficiency requirement of S-PPOC.

Table 2: Polynomial approximation of ReLU

Degree	Polynomials
2	$0.1992 + 0.5002X + 0.1997X^2$
3	$0.1995 + 0.5002X + 0.1994X^2 - 0.0164X^3$
4	$0.1500 + 0.5012X + 0.2981^2 - 0.0004X^3 - 0.0388X^4$
5	$0.1488 + 0.4993X + 0.3007X^2 + 0.0003^3 - 0.0168^4$
6	$0.1249 + 0.5000X + 0.3729X^2 - 0.0410X^4 + 0.0016X^6$

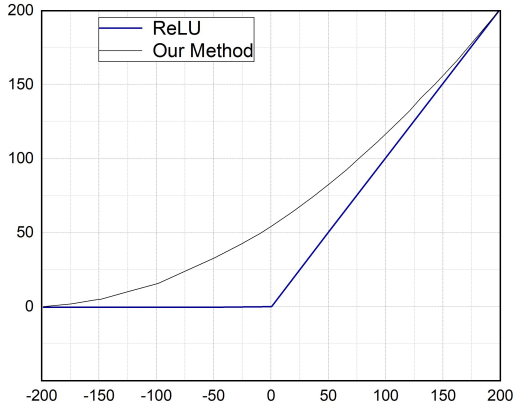


Figure 2: Approximaiton of ReLU in the 2 degree

In this stage, we modify the CNN model by replacing the high multiplicative depth layers *i.e.* ReLU and Max pooling by a low multiplicative depth polynomial layers into the CNN model, with reduction in degradation of the accuracy of the classification. Our aim is to approximate the ReLU function. We therefore focus on approximating the derivative of the ReLU function instead of approximating the ReLU function. The derivative of the activation function is like step function and is non-differentiable in point 0. In situations where the function is continuous or non-infinitely differentiable function we can approximate it more accurately. We therefore simulate the structure of the derivative of the ReLU function. Sigmoid activation function is bounded, continuous and infinitely differentiable function, it′s structure is like the derivative of the ReLU function in the large intervals. We therefore approximate the sigmoid function with the polynomial, find the integral of the polynomial and use it as the activation function. To achieve our goal, we combine the polynomial approximation of the ReLU with batch normalization [14] whiles replacing the max-pooling by the sum-pooling which has a null multiplicative depth. On the other hand, before each ReLU layer, we add a batch normalization layer in order to have a restricted stable distribution at the entry of the ReLU. The batch normalization layers are therefore added to the training and the classification stage to avoid high accuracy degradation between the training stage and the classification stage due to numerous modifications to the CNN model.

Algorithm 1: Batch Normalizing Transformation

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_1, \cdots, m\}$;Parameters to be learned:γ, β

Output: $\{y_i = BN\gamma, \beta(x_i)\}$

$\mu_\beta \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$ // mini-batch mean

$\sigma_\beta^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_\beta)^2$ // mini-batch variance

$\hat{x} \leftarrow \frac{(x_i - \mu_\beta)}{\sqrt{\sigma_\beta^2 + \varepsilon}}$ // normalize

$y_i \leftarrow \gamma\hat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i)$ // scale and shift

Our proposed S-PPOC proposed scheme is a composition of privacy-preserved feedforward propagation (Algorithm 2) and backpropagation (Algorithm 3) with the details in Algorithm 3.

Stage 2: Generating Privacy-preserving prediction query. On receiving the query $\mathcal{C}$ from the Data Provider, the Computational evaluator generates a response by evaluating the query on the encrypted ciphertext. Using Evaluate algorithm, where the circuit $\mathcal{C}$ being evaluated is the query. This yields an encryption of the results of the deep learning process:$\psi_i \leftarrow Evaluate(\mathcal{C}, Enc(\mathcal{X}))$ $= Evaluate(\mathcal{C}, \{\langle pk_i, \psi_i \rangle\}_{i=1,\cdots,n}.|$.

Stage 3: Privacy-preserving extracting result. Once the transformed query results have been computed, the Computational Evaluator sends it back to the DP. When the Data Provider receives the encrypted query. The ciphertext ψ_i decrypts to $\mathcal{C}(\mathcal{X})$. with $Decrypt(sk_1, \cdots, sk_n; \psi) = \mathcal{C}(X)$.

Algorithm 2: Privacy-Preserved Forward Propagation

Input: $\{\psi_1, \psi_2, \cdots, \psi_n\} = Enc(\mathcal{X})$ $\phi = \{W^1, W^2, b^1, b^2\}$;
Output: $z^2, z^3; a^2, a^3$

for $i = 1, 2, \cdots, i_{max}$ **do**
  for $i = 1, 2, \cdots, (n-1)$ **do**
    Compute;
    $z^{(2)}_{j_1 j_2 \cdots j_n} = W^{(1)}_{\alpha} \cdot X + b^{(1)}_{j_1 j_2 \cdots j_n}$;
    $a^{(2)}_{j_1 j_2 \cdots j_n} = f(z^{(2)}_{j_1 j_2 \cdots j_n})$;
    $z^{(3)}_{i_1 i_2 \cdots i_n} = W^{(2)}_{\beta} \cdot a^{(2)} + b^{(2)}_{i_1 i_2 \cdots i_n}$;
    $h_{(i_1 i_2 \cdots i_n) W,b}(X) = a^{(3)}_{i_1 i_2 \cdots i_n} = f(z^{(3)}_{i_1 i_2 \cdots i_n})$;
  call Algorithm 3 for backpropagation

Algorithm 3: Privacy-Preserved Backpropagation

Input : $(\mathcal{X})$ $\{W^1, W^2, b^1, b^2\}; z^2, z^3; a^2, a^3; \eta$
Output: $\phi = \{W^1, W^2, b^1, b^2\}$

for $i = 1, 2, \cdots, i_{max}$ **do**
  for *example* $= 1, 2, \cdots, N$ **do**
    for $i_n = 1, 2, \cdots, I_1 \times \cdots, I_N$ **do**
      Calculate;
      $\sigma^{(3)}_i = (a^{(3)}_i \cdot (1 - a^{(3)}_i)) \cdot \sum_{j=1}^{I_1 \times I_2 \times \cdots \times I_N} g_{ij}(a^{(3)}_j - y_j)\ \sigma^{(2)}_{j_1 j_2 \cdots j_n} = (\sum_{i_1=1}^{I_1} \cdots \sum_{i_n}^{I_n} w^{(2)}_{\lambda j_1 j_2 \cdots j_n} \cdot \sigma^{(3)}_{i_1 i_2 \cdots i_n}) f'(z^{(2)}_{j_1 j_2 \cdots j_n})$;
      $i_n = 1, 2, \cdots, I_N$
      $\Delta b^{(2)}_{i_1 i_2 \cdots i_n} = \Delta b^{(2)}_{i_1 i_2 \cdots i_n} + \sigma^{(3)}_{i_1 i_2 \cdots i_n}$
      $\Delta w^{(2)}_{i_1 i_2 \cdots i_n j_1 j_2 \cdots j_n} = \Delta w^{(2)}_{i_1 i_2 \cdots i_n j_1 j_2 \cdots j_n} + a^{(2)}_{j_1 j_2 \cdots j_n} \cdot \sigma^{(3)}_{i_1 i_2 \cdots i_n}$
      $b^{(1)}_{j_1 j_2 \cdots j_n} = \Delta b^{(1)}_{j_1 j_2 \cdots j_n} + \sigma^{(2)}_{j_1 j_2 \cdots j_n}$
      $\Delta w^{(1)}_{j_1 j_2 \cdots j_n i_1 i_2 \cdots i_n} = \Delta w^{(1)}_{j_1 j_2 \cdots j_n i_1 i_2 \cdots i_n} + x_{i_1 i_2 \cdots i_n} \cdot \sigma^{(2)}_{j_1 j_2 \cdots j_n}$
  $W = W - \eta \cdot (\frac{1}{N} \Delta w)$;
  $b = b - \eta \cdot (\frac{1}{N} \Delta b)$;

6 Security Complexity Analysis

The privacy characteristics of S-PPOC is analyzed against possible attacks by various entities involve in our privacy-preserving model. We define the semantic security as follows:

Definition 1. *(Semantic Security(SS), IND-CPA). A public-key encryption scheme ε = (KeyGen, Enc, Dec) is semantically secure, if for any stateful PPT adversary $A = (A_1, A_2)$, its advantage $Adv^{SS}_{\varepsilon.A}(K) := |Pr[Exp^{SS}_{\varepsilon.A}(K) = 1] - \frac{1}{2}]$ is negligible, where the experiment $Exp^{SS}_{\varepsilon.A}(K)$ is defined as follows:*

$Exp^{SS}_{\varepsilon.A}(K)$

$b \leftarrow 0, 1$

$(pk, sk) \leftarrow keyGen(1^k)$

$(m_0, m_1) \leftarrow A_1(pk)$

$c \leftarrow Enc_{pk}(m_b)$

$b' \leftarrow A_2(c)$

if $b' = b$. return 1:

else, return 0

The messages m_0, m_1 are called chosen plaintext, whiles c is called the challenge ciphertext. The length of two messages are required which must be equal, *i.e.* $|m_0| = |m_1|$. If the messages length is not equal. If the message length is not equal, A_2 can pad the shorter message into the equal length of the other. Thus, from the Definition 1, a property of semantic security can be obtained as follows

Property 6.1: Given the ciphertext, any computation be efficiently performed on the plaintext, can also be efficiently performed with the ciphertext.

Honest-but-curious crypto service provider. For the honest-but-curious crypto service provider CSP, it uses the private key sk_i to get blinded plaintext. Since the CE randomly chooses some randomness as blinded factor, and adds it to a fully homomorphic ciphertext before outsourcing to the CSP. Hence, CSP cannot gain any additional information on the blinded ciphertext.

Honest-but-curious computational evaluator. We therefore give a security analysis for our proposed model S-PPOC, proving that it is semantically secure.We assume that an adversary A chooses two plaintexts $m_0 = (m_{01}, m_{02}, \cdots, m_{0k})$, $m_1 = (m_{11}, m_{12}, \cdots, m_{1k}) \in P = [0, 1]$, and some circuit $\mathcal{C} \in \mathcal{C}$ with size $k = play(k)$.

Then A sends two plaintext m_0, m_1 and circuit $\mathcal{C}$ to the honest-but-curious computational evaluator CE (the challenger). The DSP tosses a fair coin $b \in [0.1]$, and outputs $\psi^* \leftarrow Evaluate(\mathcal{C}, pk_i, \psi_i, y_{b1}, y_{b2}, \cdots, y_{bk})$ where $x_{bi} \leftarrow Encrypt(pk, m_{bi})$. However, A knew the encrption x_{bi} and circuit $\mathcal{C}$, it then compute the Evaluate. This shows that the role of A and CE are the same. The CE only guesses the encrypted plaintext correctly with negligible advantage, even though some of the encrypted datasets are stored with the same plaintext. Based on the Property 6.1 we obtain the following theorem.

Theorem 1. *The S-PPOC scheme represented in Section 5 is semantically secure, if only the uderlying public encryption scheme is semantically secured.*

Algorithm 4: Overall Scheme of S-PPOC

Input: $\{m_1, m_2, \cdots, m_n\}$, $\{W^1, W^2, b^1, b^2\}$; $iteration_{max}$ learning rate η
Output: $\phi = \{W^1, W^2, b^1, b^2\}$

Data Provider P_i $(i \in [1, k])$ does:;
Initialize the Models parameters randomly;
Sample a key tuple:$(pk_i, sk_i, ek_i) \leftarrow KeyGen(1^k)$;
Encrypt private data objects: m_i $W_i^{(j)}$ $\psi_i \leftarrow pub(m_i)||Enc(pk_i, sec(m_i))$ $d_i \leftarrow Enc(pk_i, W_i^{(j)})$, $g \leftarrow Enc(pk_i, t_i)$;
Crowdsource the private data to Data Service Provider:$(pk_i, ek_i, \psi_1, d_i, g_i)$;
Data Service Provider: $Enc(\mathcal{X}) \triangleq \{\langle pk_i, \psi_i \rangle\}$ an encrypted database which is the aggregation of public keys and ciphertext tuples. Execute *Algorithm 2* and *3* over the ciphertext domain;
Update the encrypted parameters: $\phi = \{W^1, W^2, b^1, b^2\}$;;
Send the learning results γ^* to the Data providers:;
The Data providers $P_i, P_2, \cdots, P_k$ do:;
The Data providers $P_i, P_2, \cdots, P_k$ jointly run a secure multiparty protocols to calculate ;
Evaluate$(\mathcal{C}), \{\langle pk_i, \psi_i \rangle\}$

In this paper, our proposed S-PPOC scheme will allow multiple data providers to outsource fully homomorphic ciphertext to a computational evaluator CE for storage and privacy-preserving data processing. Data Providers encrypt their datasets with different fully homomorphic encryption schemes to preserve the confidentiality of the private data. The application of the S-PPOC scheme involves the use of CSP and CE to collaboratively train a classification model over the datasets from these different entities. This classification model is stored in the encrypted form in CE. Which will be used to provide query response services for the Data Requestors.

7 Performance Evaluation

Our prototype is implemented in PySEAL [44]. PySEAL is python wrapper for the Microsoft Research's homomorphic encryption implementation, the Simple Encrypted Arithmetic Library (SEAL) homomorphic encryption library. The implementation of our privacy-preserving deep learning training model and classification were conducted on a computer NVIDIA GeForce GTX 780M 4096 MB @ 3.4 GHz, Intel Core i5 and 16 GB 1600 MHz DDR3 RAM and installed with Ubuntu Server 16.04 64-Bit Version.

7.1 Datasets

In this paper, the training and testing datasets are chosen from four benchmark datasets encrypted and outsourced to the cloud representing the typical problem of human activity recognition.

SBHARPT: SBHARPT [1, 37] is publicly available at [36]. The human activity recognition signals are generated with smartphones with embedded triaxial accelerometer and gyroscope. This devices are placed on the waist and with a constant frequency rate of 50Hz to collect 12 different kinds of activity signals such as walking, walking downstairs, walking upstairs, sitting, standing and laying. and 6 possible activity transitions: stand-to- sit, sit-to-lie, lie-to-sit, sit-to-stand, lie-to-stand and stand-to-lie. There are 815,614 records of sensor data.

OPPORTUNITY: Dataset from the opportunity [38] activity recognition is a composition of atomic activities generated with a sensor-based environment in excess of 27,000. Opportunity dataset comprises of recordings of 12 subjects with the application of 15 networked sensor systems with 72 sensors of 10 modalities, integrated into the WBAN attached to the human body. This experiment we consider the sensors on the body, which include initial measurement units and 3-axis accelerometer. Each of the sensor channels are therefore treated as an individual channels with a total of 113 channels. Opportunity dataset captures different postures and gestures whiles ignoring the null class. This is a composition of 18-class classification problem.

PAMAP2: PAMAP2 [35] physical activity monitoring dataset contains based on 18 different physical activities such as cycling, walking, Nordic walk, dancing, lie, sit, stand, run, iron, vacuum clean, rope jump, ascend and descend stairs, playing soccer were recorded from 9 participants (1 female and 8 males). In addition, a variety of leisure activities such as computer work, fold laundry, watch TV, drive car, clean house. The gyroscope, Accelerometer, magnetometer, temperature and heart rate data were recorded by 9 subjects wearing 3 inertial measurement unit with a heart rate monitor. The 3 calibre wireless inertial measurement units (IMU) with a sampling frequency of 100Hz: 1 IMU is placed over the wrist on the dominant arm, 1 IMU on the chest, and 1 IMU also on the dominant side of the ankle. The HR-monitor with a sampling frequency of 9Hz is therefore used to monitor the system over 10 hours.

7.2 Comparison Evaluation

In this section, with the consideration of privacy-preserving of human activity recognition applications, we apply our proposed S-PPOC model on the above mentioned public datasets. We compare our result with existing privacy-preserving deep neural network models and classifiers in terms of communication overhead and computational cost. The existing approaches are also based

on secure multi-party computations and fully homomorphic encryption.

Recently proposed schemes based on secure multiparty computations algorithm are [31] and [11]. Mohassel *et al.* presented SecureML which enables data providers distribute the private data among the two non-colluding servers in a distributed setting therefore falling under the two-server model where diverse neural network models are evaluated on the collaborative data using secure two-party computation. The authors use Yao's Garbled Circuit to securely perform deep learning. Two of the most important advantages of this setting are that,

1) Data owners can distribute their data inputs among the two non-colluding servers in the setup phase without engaging in any future computations.

2) It also benefit from a combination of efficient methods for Boolean computations such as the garbled circuits and oblivious Transfer extension and arithmetic computations.

They perform experiments on Arcene and MNIST datasets and report the results. As presented in Table 3 and Figure 3, S-PPOC significantly outperforms [31] and [11] in all aspects. It is important to note that, [26] decided to keep the computational cost and the communication cost of the data providers at a minimal level. The proposed algorithm classifies one instance at each prediction round. S-PPOC on the other hand is capable of classifying a bath of instances in each round with size 6152MB or larger.

To provide a fair comparison, we implemented the deep neural network with 2 hidden layers with 128 neurons in each of the layers without any convolutional layers using S-PPOC. In a 100 instances they reported 16 seconds as their running time whiles S-PPOC reported 10 seconds. By increasing the input batch size, the running time increases sub-linearly in [31] and [11] whereas in S-PPOC, the running time remains the same even when the input batch size becomes larger. Furthermore, S-PPOC does not require any communication between data providers and the cloud server for the provision of privacy-preserving predictions.

Table 3: Comparison with the state-of-the-art framework

Models	SBHARPT	PAMAP2	OPPORTUNITY
[31]	0.863	0.87	0.84
[11]	0.915	0.906	0.901
[26]	0.920	0.916	0.922
S-PPOC	0.954	0.935	0.955

Generally, existing secure multi-party deep learning privacy-preserving solutions have one of the biggest communication overheads since such schemes requires an interaction between the data providers and cloud server for the secure computations. [31] has a huge communication cost of 595.55MB for a relatively small convolutional neural network where as S-PPOC is only 257.4MB for the same work. Note that since the data providers participate in the training of the model and the computations, information about the deep neural network model may possibly be compromised. For example, the data providers can learn information such as the structure of the layers, activation functions and the number of the layers therefore demonstrating the potential for security breaches.

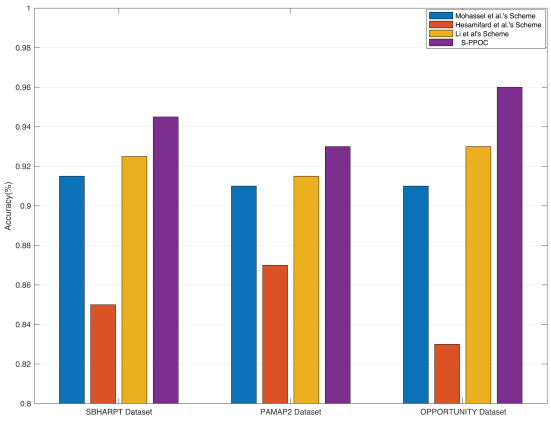


Figure 3: Performance comparison of S-PPOC with state-of-the-arts models on public datasets

8 Conclusion and Future Work

In this paper, we propose a S-PPOC scheme will allows multiple human activity recognition data providers to outsource fully homomorphic ciphertext to a computational evaluator CE for storage and privacy-preserving data processing. Data Providers encrypt their datasets with different fully homomorphic encryption schemes to preserve the confidentiality of the private data. The application of the S-PPOC scheme involves the use of CSP and CE to collaboratively train a classification model over the datasets from these different entities. This classification model is stored in the encrypted form in CE. Which will be used to provide query response services for the Data Requestors.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No.61672135), the Frontier Science and Technology Innovation Projects of National Key R&D Program (No.2019QY1405), the Sichuan Science and Technology Innovation Platform and Talent Plan (No.20JCQN0256), and the Fundamental Research Funds for the Central Universities (No.2672018ZYGX2018J057).

References

[1] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *The 21st European Symposium on Artificial Neural Networks*, 2013. (https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2013-84.pdf)

[2] A. Bansal, T. Chen, and S. Zhong, "Privacy preserving back-propagation neural network learning over arbitrarily partitioned data," *Neural Computing and Applications*, vol. 20, no. 1, pp. 143–150, 2011.

[3] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(leveled) fully homomorphic encryption without bootstrapping," *ACM Transactions on Computation Theory*, vol. 6, no. 3, pp. 13:1–13:36, 2014.

[4] D. Chen, N. Zhang, N. Cheng, K. Zhang, Z. Qin, and X. S. Shen, "Physical layer based message authentication with secure channel codes," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 5, pp. 1079–1093, 2018.

[5] D. Chen, N. Zhang, R. Lu, N. Cheng, K. Zhang, and Z. Qin, "Channel precoding based message authentication in wireless networks: Challenges and solutions," *IEEE Network*, vol. 33, no. 1, pp. 99–105, 2018.

[6] D. Chen, N. Zhang, R. Lu, X. Fang, K. Zhang, Z. Qin, and X. Shen, "An LDPC code based physical layer message authentication scheme with prefect security," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, pp. 748–761, 2018.

[7] D. Chen, N. Zhang, Z. Qin, X. F. Mao, Z. Qin, X. Shen, and X. Y. Li, "S2M: A lightweight acoustic fingerprints-based wireless device authentication protocol," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 88–100, 2017.

[8] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pp. 169–178, 2009.

[9] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. E. Lauter, M. Naehrig, and J. Wernsing, "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proceedings of the 33nd International Conference on Machine Learning*, vol. 48, pp. 201–210, 2016.

[10] S. Gong, B. B. Yin, Z. Zheng, and K. Y. Cai, "Adaptive multivariable control for multiple resource allocation of service-based systems in cloud computing," *IEEE Access*, vol. 7, pp. 13817–13831, 2019.

[11] E. Hesamifard, H. Takabi, and M. Ghasemi, "Cryptodl: Deep neural networks over encrypted data," *CoRR*, vol. abs/1711.05189, 2017.

[12] Y. L. Hsu, S. C. Yang, H. C. Chang, and H. C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[13] H. Hui, X. Li, and Y. Sun, "A triboelectric motion sensor in wearable body sensor network for human activity recognition," in *The 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4889–4892, 2016.

[14] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, pp. 448–456, 2015.

[15] A. N. Jaber, Z. M. Fadli, and H. Othman, "A conceptual model using the elliptic curve Diffie-Hellman with an artificial neural network over cloud computing," *The National Conference for Postgraduate Research*, 2016. (https://www.researchgate.net/publication/308548189_A_Conceptual_Model_Using_the_Elliptic_Curve_Diffie-Hellman_With_An_Artificial_Neural_Network_Over_Cloud_Computing)

[16] A. N. Jaber and M. F. B. Zolkipli, "Use of cryptography in cloud computing," in *IEEE International Conference on Control System, Computing and Engineering*, pp. 179–184, 2013.

[17] S. Kim, S. Yeom, O. J. Kwon, D. Shin, and D. Shin, "Ubiquitous healthcare system for analysis of chronic patients' biological and lifelog data," *IEEE Access*, vol. 6, pp. 8909–8915, 2018.

[18] O. A. Kwabena, Z. Qin, T. Zhuang, and Z. Qin, "Mscryptonet: Multi-scheme privacy-preserving deep learning in cloud computing," *IEEE Access*, vol. 7, pp. 29344–29354, 2019.

[19] C. Lai, H. Li, R. Lu, and X.(Sherman) Shen, "SE-AKA: A secure and efficient group authentication and key agreement protocol for LTE networks," *Computer Networks*, vol. 57, no. 17, pp. 3492–3510, 2013.

[20] C. T. Li, M. S. Hwang, "A lightweight anonymous routing protocol without public key en/decryptions for wireless ad hoc networks",*Information Sciences*, vol. 181, no. 23, pp. 5333–5347, Dec. 2011.

[21] C. T. Li, M. S. Hwang, Y. P. Chu, "Further improvement on a novel privacy preserving authentication and access control scheme for pervasive computing environments", *Computer Communications*, vol. 31, no. 18, pp. 4255–4258, Dec. 2008.

[22] C. T. Li, M. S. Hwang, Y. P. Chu, "A secure and efficient communication scheme with authenticated key establishment and privacy preserving for vehicular ad hoc networks", *Computer Communications*, vol. 31, no. 12, pp. 2803-2814, July 2008.

[23] C. T. Li, M. S. Hwang and Y. P. Chu, "An efficient sensor-to-sensor authenticated path-key establishment scheme for secure communications in wireless sensor networks", *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 8, pp. 2107-2124, Aug. 2009.

[24] J. Li, L. Tian, H. Wang, Y. An, K. Wang, and L. Yu, "Segmentation and recognition of basic and transi-

tional activities for continuous physical human activity," *IEEE Access*, vol. 7, pp. 42565–42576, 2019.

[25] P. Li, J. Li, Z. Huang, T. Li, C. Z. Gao, S. M. Yiu, and K. Chen, "Multi-key privacy-preserving deep learning in cloud computing," *Future Generation Computer Systems*, vol. 74, pp. 76–85, 2017.

[26] T. Li, Z. Huang, P. Li, Z. Liu, and C. Jia, "Outsourced privacy-preserving classification service over encrypted data," *Journal of Network and Computer Applications*, vol. 106, pp. 100–110, 2018.

[27] Z. Li and T. H. Lai, "On evaluating circuits with inputs encrypted by different fully homomorphic encryption schemes," *IACR Cryptology ePrint Archive*, vol. 2013, p. 198, 2013.

[28] C. H. Ling, C. C. Lee, C. C. Yang, and M. S. Hwang, "A secure and efficient one-time password authentication scheme for WSN", *International Journal of Network Security*, vol. 19, no. 2, pp. 177-181, Mar. 2017.

[29] L. H. Liu and Y. Liu, "A note on one outsourcing scheme for large-scale convex separable programming," *International Journal of Electronics and Information Engineering*, vol. 12, no. 4, pp. 155-161, 2020.

[30] A. López-Alt, E. Tromer, and V. Vaikuntanathan, "On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption," in *Proceedings of the 44th Symposium on Theory of Computing Conference*, pp. 1219–1234, 2012.

[31] P. Mohassel and Y. Zhang, "Secureml: A system for scalable privacy-preserving machine learning," in *IEEE Symposium on Security and Privacy*, 2017. (`https://eprint.iacr.org/2017/396.pdf`)

[32] T. Muhammed, R. Mehmood, A. Albeshri, and I. Katib, "Ubehealth: A personalized ubiquitous cloud and edge-enabled networked healthcare system for smart cities," *IEEE Access*, vol. 6, pp. 32258–32285, 2018.

[33] Z. Qin, L. Hu, N. Zhang, D. Chen, K. Zhang, Z. Qin, and K. K. R. Choo, "Learning-aided user identification using smartphone sensors for smart homes," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7760–7772, 2019.

[34] Z. Qin, Y. Wang, H. Cheng, Y. Zhou, Z. Sheng, and V. C. M. Leung, "Demographic information prediction: A portrait of smartphone application users," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 3, pp. 432–444, 2018.

[35] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *The 16th International Symposium on Wearable Computers*, pp. 108–109, 2012.

[36] J. L. Reyes-Ortiz, D. Anguita, L. Oneto and X. Parra, *Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set*, 2015. (`https://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions`)

[37] J. L. Reyes-Ortiz, L. Oneto, A. Ghio, A. Samà, D. Anguita, and X. Parra, "Human activity recognition on smartphones with awareness of basic activities and postural transitions," in *International Conference on Artificial Neural Networks (ICANN'14)*, vol. 8681, pp. 177–184, 2014.

[38] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. del R. Millán, "Collecting complex activity datasets in highly rich networked sensor environments," in *Seventh International Conference on Networked Sensing Systems*, pp. 233–240, 2010.

[39] H. Saleh, H. Nashaat, W. Saber, and H. M. Harb, "IPSO task scheduling algorithm for large scale data in cloud computing environment," *IEEE Access*, vol. 7, pp. 5412–5420, 2019.

[40] S. Shen, J. Qian, D. Cheng, K. Yang, and G. Zhang, "A sum-utility maximization approach for fairness resource allocation in wireless powered body area networks," *IEEE Access*, vol. 7, pp. 20014–20022, 2019.

[41] S. Sodagari, B. Bozorgchami, and H. Aghvami, "Technologies and challenges for cognitive radio enabled medical wireless body area networks," *IEEE Access*, vol. 6, pp. 29567–29586, 2018.

[42] D. Stehlé and R. Steinfeld, "Faster fully homomorphic encryption," in *International Conference on the Theory and Application of Cryptology and Information Security*, vol. 6477, pp. 377–394, 2010.

[43] H. Takabi, E. Hesamifard, and M. Ghasemi, "Privacy Preserving Multi-party Machine Learning with Homomorphic Encryption," *Proceedings of the Workshop on Private Multi-Party Machine Learning*, no. Nips, pp. 1–5, 2016.

[44] A. J. Titus, S. Kishore, T. Stavish, S. M. Rogers, and K. Ni, "Pyseal: A python wrapper implementation of the SEAL homomorphic encryption library," *CoRR*, vol. abs/1803.01891, 2018.

[45] H. Xiong, H. Zhang, and J. Sun, "Attribute-based privacy-preserving data sharing for dynamic groups in cloud computing," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2739–2750, 2019.

[46] C. Yang, Q. Chen, Y. Liu, "Fine-grained outsourced data deletion scheme in cloud computing," *International Journal of Electronics and Information Engineering*, vol. 11, no. 2, pp. 81–98, 2019.

[47] J. Yuan and S. Yu, "Privacy preserving back-propagation neural network learning made practical with cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 212–221, 2014.

[48] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, "Synergy of big data and 5g wireless networks: Opportunities, approaches, and challenges," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 12–18, 2018.

[49] N. Zhang, P. Yang, S. Zhang, D. Chen, W. Zhuang, B. Liang, and X. S. Shen, "Software defined networking enabled wireless network virtualization: Challenges and solutions," *IEEE Network*, vol. 31, no. 5, pp. 42–49, 2017.

[50] Q. Zhang, L. T. Yang, Z. Chen, P. Li, and M. J. Deen, "Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2896–2903, 2018.

[51] H. Zhu, R. Lu, X. Shen, and X. Lin, "Security in service-oriented vehicular networks," *IEEE Wireless Communications*, vol. 16, no. 4, pp. 16–22, 2009.

Biography

Owusu-Agyemang Kwabena is currently a Ph. D. candidate in the School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC). He received his MSc. degree from Coventry University in 2012. His research interests include machine learning, data mining, big data analysis, applied cryptography, blockchain technology and medical image processing.

Zhen Qin is currently an associate professor in the School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC). He received his Ph.D. degree from UESTC in 2012. He was a visiting scholar in the Department of Electrical Engineering and Computer Science at Northwestern University. His research interests include network measurement, mobile social networks, wireless sensor networks and medical image processing.

Hu Xiong received the Ph.D. degree from the School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC) in 2009. He is currently a Full Professor with the School of Information and Software Engineering, UESTC. His research interests include applied cryptography and cypberspace security. He is a member of IEEE.

Yao Liu is currently an associate professor in the School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC). She received her Ph.D. degree from UESTC in 2016. Her research interests include machine learning, social networks, network security and data mining.

Tianming Zhuang is currently an undergraduate in Glasgow College, University of Electronic Science and Technology of China (UESTC). His research interests include big data analysis and data mining.

Zhiguang Qin is the full professor of the School of Information and Software Engineering in University of Electronic Science and Technology of China (UESTC), where he is also Director of the Key Laboratory of New Computer Application Technology and Director of UESTC-IBM Technology Center. His research interests include medical image processing, computer networking, information security, cryptography, information management, intelligent traffic, electronic commerce, distribution, and middleware.

Blockchain Data Sharing Scheme Based on Searchable Agent Re-Encryption

Tao Feng[1], Hongmei Pei[1], Pengshoou Xie[1], and Xiaoqing Feng[2]
(Corresponding author: Hongmei Pei)
School of Computer and Communication, Lanzhou University of Technology[1]
36 Pengjiaping Road, Qilihe District, Lanzhou, Gansu, China
School of Cyber Engineering, Xidian University[2]
Email: fengt@lut.cn, 2319444123@qq.com
(Received Jan. 3, 2020; Revised and Accepted July 15, 2020; First Online Apr. 17, 2021)

Abstract

The record books on the traditional blockchain of global transaction information for any nodes joining the blockchain system are public. This paper introduces searchable encryption and proxy re-encryption to propose blockchain data sharing scheme based on searchable proxy re-encryption, which solves the problems of conventional blockchain data privacy disclosure and rapid query. Firstly, we define the level of data users on the blockchain transactions to achieve access control of data. Secondly, to lighten the burden of storage blockchain, we use proxy encryption to encrypt the data before storing it in the external database, which achieves data sharing. Furthermore, binary search is performed by computing the inner product of the search token vector and index vector. Finally, we demonstrate the security and effectiveness of our scheme under the model of oracle.

Keywords: Blockchain; Data Privacy; Inverted Index Structure; Proxy Re-encryption; Searchable Encryption

1 Introduction

Since the emergence of cloud computing, secure data sharing in a distributed setting has long been a hot and challengingtopic. In the context of cloud storage, Saeid Rezaei *et al.* [16] proposed an efficient data sharing scheme based on attribute encryption, which transferred most of the decryption computing load to the cloud service provider and realized the revocation of lightweight data. Considering the fact that users and cloud providers usually belong to different administrative or security domains,the difficulty of cloud based data sharing lies in how much trust users can place on cloud service providers. As we all know that blockchain [21] is a distributed database of accounts, which makes data stored on the blockchain not be modified. In bitcoin blockchain in literature [4], information representing the electronic cash is attached to a digital address. Digital signatures is a method for the Internet which is similar to traditional signatures. Hwang *et al.* [7] investigated some multiple digital signatures and summarized the advantages and disadvantages of these schemes, as well as the problems to be solved in these schemes. Users can digitally sign this information and transfer rights to another user so that bitcoin block chain publicly can record transfers rights to allow all participants in the network to independently verify the validity of transactions. However, the public-private key pair required for digital signature comes from the bitcoin wallet [18], so it is particularly important to ensure the security of the wallet. In the blockchain system, users can store transactions directly or change their own property. However, the traditional blockchain globally books for any node joining the block chain system is transparent, the users use the public key as an address [9] to protect their anonymity and identity, it is clearly far from adequate.

Data sharing plays an important role in our daily life, espically in medical Patients' information is quickly shared with multiple attending doctors without their information being leakage, it is very critical for nations and individuals. Chung *et al.* [3] studied data privacy protect and sharing scheme using the attribute-based proxy re-encryption, which can make the data owner entrust to the rights of the encryption cloud server to share data, the data owner need not always online, but this scheme introduced the third-party cloud server, data security and privacy need to be further improved. There is no doubt that blockchain is widely used in other industries and privacy protection because of its transparency, traceability, anti-tamper and decentralization.

The data privacy of blockchain means that blockchain can provide the confidential attribute for data stored in the blockchain or particularly sensitive data. However, due to various inference attacks, sensitive transaction data or aliases will be associated with the real identity of real users even if pseudonyms are used, so there is a risk of privacy disclosure [5, 13]. Such privacy disclosure

would damage the confidentiality of trading information. Therefore, confidentiality and privacy cause significant challenges to blockchain and its applications involving sensitive transactions and private data. While bitcoin's blockchain supports anonymity by providing pseudonyms to ensure pseudonyms, it does not provide unlinkability protection for users' transactions. Unlinkability resists anti-anonymous inference attack [15], which links all the information of users together to discover the real identity of users under the existing background knowledge.

Related work. Currently, there are some research results on the privacy protection of blockchain data. As the blockchain of bitcoin cannot guarantee the anonymity of users, Bonneau *et al.* [1] proposed a mixed coin program in 2014, but this scheme introduced centralized accountability detection service, which may expose users' privacy. JUZIX introduced group signature in the alliance chain [6] to protect the anonymity of users. However, this solution requires the introduction of a third party as the group administrator. In literature [23], a ring signature scheme is proposed to hide the address of the sender, but the inadequacy of this scheme is that the signer cannot be disclosed in case of disputes.

Then,some researchers used encryption methods to reslove the privacy leakage of blockchain. A homomorphic encryption scheme in the literature [17] was proposed to protect the data privacy of blockchain, but this scheme could only realize simple addition and multiplication operations, and the computational efficiency of complex functions was relatively low. Wang *et al.* [19] proposed a scheme of attribute encryption in the bitcoin system, which realized the supervision of transaction information through attribute encryption, but increased the storage burden of blockchain. After that, in literature [2], a trusted and private keyword searchable privacy protection scheme based on blockchain is proposed, which uses searchable encryption to realize the search of encrypted data. In literature [20], a data sharing and traceability scheme based on blockchain is proposed, which adopts a double-chain blockchain structure, one chain is used to store the original data, and the other chain is used to store the generated transaction data. Proxy re-encryption is used to realize data sharing, and encryption technology is used to protect data privacy. But it increased the cost of protecting the privacy of data.

Searchable encryption is very suitable for ciphertext search environment, providing two great convenience for data security sharing [10]. Then Tian *et al.* [11] proposed a blockchain privacy protection scheme based on searchable symmetric encryption, which protected the fairness of both parties. If the user was dishonest, he could not get data from the server, and if the server was dishonest, he would be punished. The processing of data sharing is relatively complicated. Zhang Peng *et al.* [22] proposed a decentralized sharing and privacy protection system based on blockchian, which mainly used in medical field. It realized data sharing and the privacy of data. But token is not very security. Medical personnel not only can access related data by tokens, others also can access if they gain token.

Our contribution. From the literature above, many researches has been proposaled, while data sharing and privacy protection are not completely resolved. This paper proposes a new privacy protection scheme for blockchain data based on searchable agent re-encryption. Our contributions are described as follows:

1) Through data transactions on the blockchain, the level of data users is defined to achieve access control of data. We store the index structure, the level of data and search token on the blockchain, which makes sure of the security and tamperability.
2) In order to lighten the burden of storage blockchain, we use proxy encryption to encrypt the data before storing in the external database, which allows users to convert encryption data into using their own public key encryption cipher so that users can decrypt with their own private key in the further. Even if the malicious nodes obtained ciphertext, they cannot decrypt data.
3) Computing the inner product of the search token vector and index vector to find matching index entries, mining nodes on the blockchain can perform binary search. The efficiency of search is improved rapidly.
4) Solving the problem of searching ciphertext through keywords in agent re-encryption realizes the privacy protection of mixed encryption data.

Organize. In the second part, we introduce the preliminary knowledge. We introduce the security model in the third part. In the fourth part, we introduce the realization process of the algorithm in detail. The proof and analysis of security are introduced in the fifth part. In the sixth part, we summarize the article.

2 Preliminary Knowledge

2.1 Bilinear Mapping

Definition 1. *Let G_1 and G_2 be multiplication cycle group of prime order, be a random generating element of the group G_2. There exists a bilinear pair mapping that satisfies the following properties:*

1) Bilinear: For $\forall(P,Q) \in G_1$ and $\forall(a,b) \in Z_q^*$, $e(P^a,Q^b) = e(P,Q)^{ab}$is true.
2) Non-degenerative: $e(P,Q) \neq 1$.
3) Computability: For $\forall(P,Q) \in G_1$ there is a polynomial time algorithm for calculatione(P,Q).
4) It follows that $g_T \overset{def}{=}, e(g,g)$ generates G_T.Let $x = (x_1,x_2,...,x_n)$ and $y = (y_1,y_2,...,y_n)$ be two n-dimensional vectors. The bilinear map is computed as follows:

$$\begin{aligned} e(g_1^x, g_2^y) &= e(g_1^{x_1}, g_2^{y_1}) \cdot ... \cdot e(g_1^{x_n}, g_2^{y_n}) \\ &= e(g_1,g_2)^{x_1y_1+...x_ny_n} \end{aligned}$$

2.2 Data Transaction Structure

The most popular application of blockchain is bitcoin blockchain [14], which is a public blockchain ledger to support financial transactions of bitcoin's digital currency. The miner node verifies the validity of the trading information and adds it to the blockchain. In this paper, we can effectively use the blockchain technology to share data with index and encryption key by using the improved consistency algorithm [12].

To ensure that the content of a data transaction is credible, antitamper, and traceable, this article treats data as a type of transaction. Firstly, the data consumer divides the data into n blocks. During the transaction, the data owner needs to sign the previous transaction with a random numeric hash. Then the level of authority, the signature, the agent of the next data user re-encrypts the key and the keyword index of the data will be appended to the hash value of the data block. After that, the data is submitted to the next data consumer. The permission level authorized by the data owner to the next data user should be indicated, which is usually divided into three levels: Level 1, which has the right to use data, the right to grant data use and the right to distribute data; Level 2, with the use of data and the right to use data; Level 3, with permission to use data. In Figure 1, we show the data transaction record structure.

2.3 Searchable Symmetric Encryption

$EKS = (KeyGen, Encrypt, TokenGen, Test)$ is described as follows:

1) $(MK, pp) = KeyGen(1^\lambda)$: Input security parameters, the algorithm according to the security parameters to form a key MK.
2) $I_w = Enc(MK, w, pp)$: Use the generated key MK and keyword to generate the key ciphertext index.
3) $TK = TokenGen(MK, w, pp)$: Using the generated key MK and keywords generates the search token.
4) $b = Test(MK, I_w, pp, T_w)$: The index and search token are matched and the search result 0,1 is successfully returned.

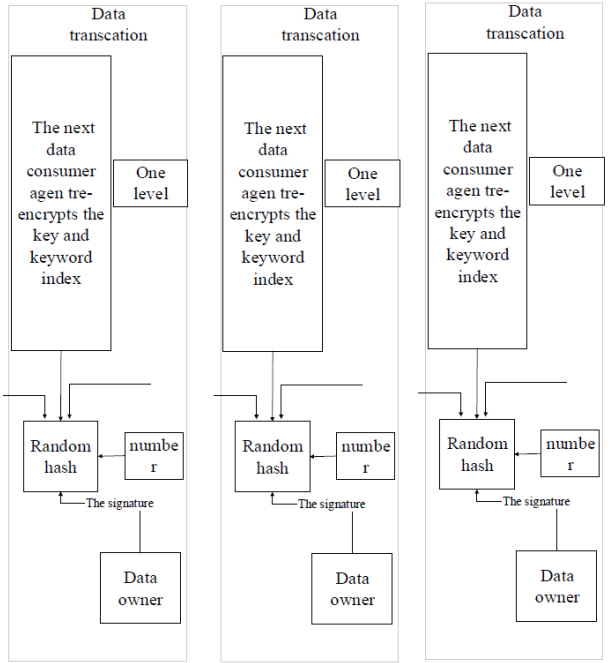


Figure 1: The data transaction record structure

2.4 Inverted Index Based on Data Structure

Inverted index structures are data structures that store content to maps, such as the keywords for a set of documents. The purpose of inverted index structure is to allow fast searching of full text. It is document retrieval, the most popular data structure used in large-scale search engines. The inverted index structure based on the data structure [24] is shown in Figure 2. Documents D=$\{d_1, d_2, ...d_n\}$ are encrypted and stored in an external database, and each document contains a set of keywords. Let $w = \{w_1, ..., w_i, *, ..., *\}$ is the keyword set in document D, where $|W| = n$ is the total number of keywords. $D_{i\in[n]} \subseteq D$ represents a set of documents containing keywords. As shown in Figure 2, keywords and document sets are encrypted using encryption algorithms Enc and E, which respectively are stored on external databases and blockchains. Note that Enc and E are different encryption schemes. Enc is the proposed scheme that supports linear searchable encryption, and E can be any secure encryption scheme. The keyword index of the encrypted document can be described as I=$\{I_1, I_2, ...I_n\}$=$Enc(w_1, ..., w_i, *, ..., *)$, which can be any keyword. The corresponding data consumer can use a search token $TK = TokenGen(0, ..., 0, w_i, 0, ..., 0,)$ that including a document keyword, where represents the generation of the search token. In this case we can use two vectors $(w_1, ..., w_i, *, ..., *)$ and $(0, ..., 0, w_i, 0, ..., 0,)$. Therefore, the miner node on the blockchain can perform binary search by calculating the inner product of the search token vector and index vector to locate the matched index item. If so, the server continues to ex-

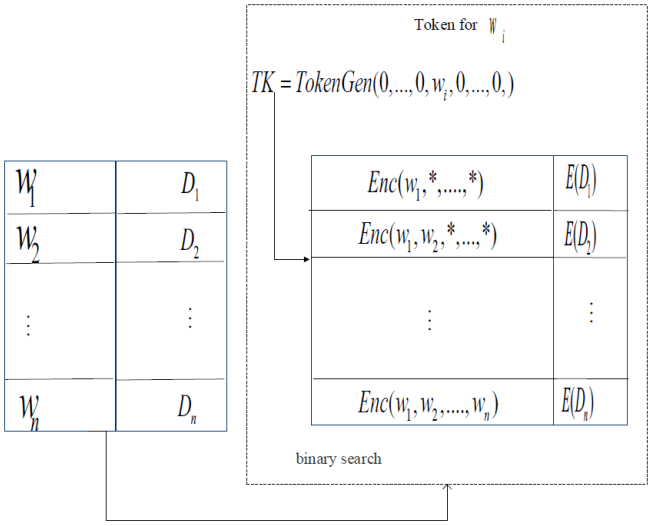


Figure 2: Keyword structure

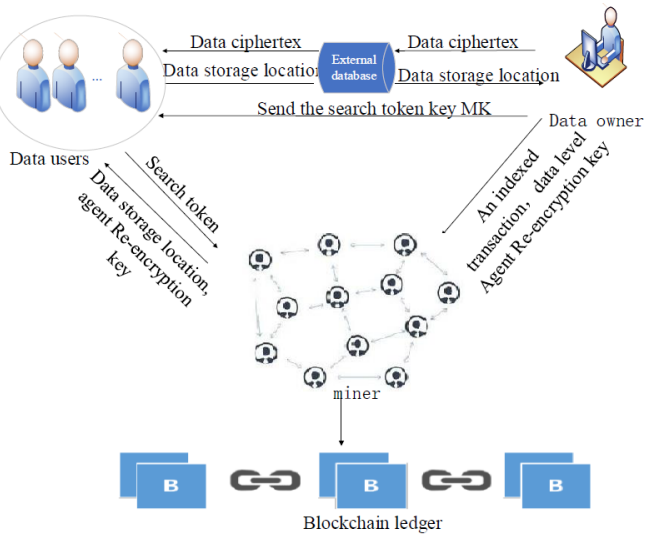


Figure 3: System model

amine the search token with the first half of the index; Otherwise, the server continues to check the search token with the second half of the index.

The pseudocode of the binary search algorithm is shown as follows. The search token TK for a given keyword $w_i$and the encrypted index I, the algorithm opens the corresponding index entryI_i and returns $E(D_i)$ or does not find. The algorithm $Test(MK, I_w, pp, TK)$ outputs a byte indicating whether the index I_i matches the query keyword corresponding to the given search token TK. The pp in Algorithm 1 represents the public parameter of our scheme.

Algorithm 1 Pseudocode of Binary search algorithm.

```
Begin
input C;TK;pp
output round = 0; head = 1; end = n; mid = 0; result = ?
repeat;
mid= [(head + end)/2];
if test(TK ;I_mid ; pp) = 1 then
    end = mid - 1;
else
    end = mid + 1;
    round++ ;
end if
until head > end;
if I_head.value = φ then
    result = I_head.value
end if
return result ;
```

3 System Model

The model of the blockchain data privacy protection system based on agent searchable encryption is shown in Figure 4, which includes four types of participating entities: data owner, data user, external database, and miner.

Data owner: Re-encrypt the data with the agent, extract the keyword of the data, and generate the keyword index. Finally, the data owner forms the transaction (identity and keyword index after hash) and broadcasts it to the blockchain network, then miners add the transaction to the block after verification.

Data consumer: With his own client public key, data consumer requests access to data. Firstly, he initiates a transaction (attach its own identity, search tokens) to the blockchain. After the miner verifies that the search Token and index match successfully, the location of the file and the proxy re-encryption key are sent to the data consumer. Then the data consumer requests the data ciphertext from the external database.

External database: Store encrypted raw data.

Miner: The miner validates the transaction and verifies whether the index matches the search Token, and then sends the storage location of the data and the proxy re-encryption key to the data consumer.

3.1 Threat Model

In the proposed scheme, the external data users may obtain ciphertext database storage, then guess the clear content. Miners are "honest but curious" node, when executing search, they may guess key words of the user's private data according to the data submitted by the user search token.

3.2 Security Model

The security model is defined by A game played between A probabilistic polynomial time adversary A and challenger C. If the probability of the polynomial time opponent guessing the keyword corresponding to the trapdoor in the game is negligible, then the scheme can resist the keyword guessing attack under the random predictor model. The formal description is shown in Figure 5.

Definition 2. *If for any adversary A in polynomial time, there is a negligible function* $\varepsilon(\lambda)$ *that makes* $Adv_{\Pi,A}^{PP1}(\lambda) \stackrel{def}{=} |\Pr[Expts_{PP1,\Pi,A}^{(0)} = 1] - \Pr[Expts_{PP1,\Pi,A}^{(1)} = 1] \leq \varepsilon(\lambda)$ *equal, where* $Expts_{PP1,\Pi,A}^{(b)}$ *defined as shown in Algorithm 2, so searchable encryption scheme* $\Pi = (\text{Setup}, \text{Enc}, \text{Token}, \text{Test}, \text{Query})$ *is against plaintext attack.*

The following is the formal definition of the model:

Initialization: Challenger C runs the initialization algorithm to generate the public parameters and the master key, and sends the public parameters to the adversary.

Stage 1: The adversary adaptively proposes the following polynomial query. $Enc(w_i, MK, pp)$ And $TokenGen(TK_1, TK_2, sk, pp)$ request: The opponent can ask the random oracle.

Ask. The adversary can request any keyword ciphertext.

Challenge. The adversary submits two keywords $W_0^*$and W_1^* to challenger C, challenger C sends $I^* \leftarrow Enc(W_b^*, MK, pp)$ on condition that is the adversary cannot ask $Query(TK_{w_{i,j}}, I_{w_0^*}, pp) \neq Query(TK_{w_{i,j}}, I_{w_1^*}, pp)$.

Phase 2. Phase 1 is repeated.

Guess. Adversary A outputs b's guess b'. If opponent's guess is equal to b, adversary A will win the game.

If for any polynomial time adversary A, there exists an negligible function equal, where defined as shown in Figure 2. Therefore, searchable encryption scheme (Setup, Enc, Token, Test, Query) is against a key word attack.

Described as $Expts_{pp1,\Pi,A}^{(b)}$ experiment:

Definition 3. *If for any adversary A in polynomial time, there is a negligible function* $Adv_{\Pi,\text{A}}^{PP2}(\lambda) \stackrel{def}{=} |\Pr[Expts_{PP2,\Pi,\text{A}}^{(0)} = 1] - \Pr[Expts_{PP2,\Pi,\text{A}}^{(1)} = 1] \leq \varepsilon(\lambda)$ *equal, where* $Expts_{PP2,\Pi,\text{A}}^{(b)}$ *defined as shown in Algorithm 3, so searchable encryption scheme* $\Pi = (\text{Setup}, \text{Enc}, \text{Token}, \text{Test}, \text{Query})$ *is against a key word attack.*

The following is the formal definition of the model:

Algorithm 2 $Expts_{PP1,\Pi,A}^{(b)}$ experiment

1. $Setup(1^\lambda)$
2. $(w_0^*, w_1^*, state) \leftarrow A(1^\lambda)$, where$w_0^*, w_1^* \subseteq W_\lambda$ and $|w_0^*|=|w_1^*|=n$;
3. $I \leftarrow Enc(MK, \text{pp}, w_i)$;
4. $b' \leftarrow A^{Enc(MK,\text{pp},w_i),TokenGen(MK,\text{pp},w_i),(I^*,state)}$;
5. For all tokens query $Query(TK_{w_{i,j}}, I_{w_0^*}, pp) \neq Query(TK_{w_{i,j}}, I_{w_1^*}, pp)$;

Then Output b', otherwise output$\perp$;

Initialization: Challenger B runs the initialization algorithm to generate the public parameters and the master key, and sends the public parameters to the adversary.

Stage 1: The adversary adaptively proposes the following polynomial. $Enc(w_i, MK, pp)$ and $TokenGen(TK_1, TK_2, MK, pp)$ request: The opponent can ask the random oracle.

Ask. An adversary can request any keyword.

Challenge. The opponent submits two keywords and sends them to challenger C, then challenger C sends $TK^* = TokenGen(W_b^*, j^*, MK, PP)$ to the opponent. The restriction is that the opponent cannot ask $Query(TK_{w_0^*}, I_i, pp) \neq Query(TK_{w_1^*}, I_i, pp)$.

Phase 2. Phase 1 is repeated.

Guess. Adversary A outputs b's guess. If opponent's guess is equal to b, Adversary A will win the game.

Described as $Expts_{pp2,\Pi,A}^{(b)}$ experiment in Algorithm 3.

Algorithm 3 $Expts_{pp2,\Pi,A}^{(b)}$ experiment

1. $Setup(1^\lambda)$
2. $(w_0^*, w_1^*, j^*, state) \leftarrow A(1^\lambda)$,where $w_0^*, w_1^* \subseteq W$;
3. $TK^* = TokenGen(MK, w_b^*, j^*, pp)$, where $j \in [n]$;
4. $b' \leftarrow A^{Enc(MK,\text{pp},w_i),TokenGen(TK_i,MK,\text{pp},w_i),(TK^*,state)}$, where $b^{'} \in \{0, 1\}$;
5. For all ciphertext query $(w_{i,j}, j)$, where$j \in [n]$ $Query(TK_{w_0^*}, I_i, pp) \neq Query(TK_{w_1^*}, I_i, pp)$

Then Output b', otherwise output $\perp$;

4 Algorithm Construction

Any effective operation in the block chain system will be recorded on the block chain in the form of transactions. In the privacy protection scheme of the block chain data constructed based on the searchable agent encryption, the specific scheme includes the following 9 steps:

1) Initialization $(G_1, G_2, e, g, H, H_B, \text{sk}) \leftarrow Setup\left(1^\lambda\right)$: Given security parameters, generate order $q$bilinear group G_1andG_2, randomly select group

G_1generator$g$, bilinear paire : $G_1 \times G_1 \to G_2$, hash function$H : \{0,1\}^* \to G_1$, public system parameter $\mathrm{PP}(G_1, G_2, e, g, g_1, H)$. In the scheme, the data publisher selects random number$\beta \in {Z_p}^*$ and initializes the calculation of public $PK_i = \mathrm{g}^{\beta}$, $MK = (M_1, M_2)$ and private keys $SK_i = \beta$, whereM_1, M_2 are the matrix of full rank.

2) Encryption $C_z \leftarrow E\left(D, g^{\beta}\right)$: Firstly, the data$p_i$ extracts the key of the data, and then uses the public key to obtain the ciphertext. The encryption process is as follows: Select a random number $r_i \in Z_p^*$, and re-encryption calculation is described as follows:

$$\begin{aligned} C_z &= (c_1, c, c_3) \\ c_1 &= g^{r_i} \\ c_2 &= D \cdot e(g_1, g^{\beta})^{r_i} \\ c_3 &= H(H(c_1)||H(c_1||c_2)). \end{aligned}$$

It is then stored to an external server, which sends the location D_{Loc} of the data store to the data owner. Data plaintext calculation is as follows:

$$D = c_2 / e(c_1, {g_1}^{\beta}) \tag{1}$$

3) Re-encryption key generation: $rk_{i \to j} \leftarrow (rk_1, rk_2, rk_3)$, when the data user wants to access the data, the data password is converted to. The data user can use his private key to decrypt the data originally encrypted by the public key. We then select the random number, and the calculation process of the re-encryption key is as follows:

$$\begin{aligned} rk_1 &= g^{r_j} \\ rk_2 &= {g_1}^{-r} pk_j^{r_j} \\ rk_3 &= H(H(rk_1)||H(rk_1||rk_2)). \end{aligned}$$

4) Index generation: Key words $I \leftarrow Enc\left(MK, \mathrm{pp}, w_i\right)$: $w = \{w_1, w_2, ...w_n\}$ are collected by the data owner. Input key MK and output encrypted index $I = (I_1, I_2, ..., I_n)$, each of which is generated as follows: Based on the security KNN technology modified by Cao *et al.*, this paper generates (n+1)-dimension vector $p_i = (p_{i,1}, p_{i,2}, ..., p_{i,n})$, which is divided into two vectors (p_i', p_i'') and extends to $p_i = (\beta, 1)$, where $\beta = H(w_i)$. Similarly generate (n+1)-dimension vector $q_i = (q_1, ..., q_n)$, which is divided into two vectors (q_i', q_i'') and extend to $q_i = (r, \alpha)$, where for keyword w_i encryption of these two vectors $I_i = ({g^{M_1}}^T{p_i}', {g^{M_2}}^T{p_i}'')$=$(I_{i,1}, I_{i,2})$. The data publisher creates a new transaction attaching an index,and broadcasts the transaction to the blockchain system. Transactions with indexes are added to the block after verification by the miner node, just like normal transactions in the blockchain.

5) The generation of the search token $TK \leftarrow TokenGen\left(MK, \mathrm{pp}, w_i\right)$: The data querier builds the search token. The specific construction process is as follows: The data querier will input the keyword w_i and I, MK, PP. The search token algorithm for the keyword is as follows:

$$TK = ({g^{{M_1}^{-1}}}^{q_i}{}', {g^{{M_2}^{-1}}}^{q_i}{}'', g^{r\beta+\alpha})$$

The data querier generates the transaction and attaches the search token $TK$to the transaction. Finally, send it to the surrounding miner node.

6) Test$\{0,1\} \leftarrow Test\left(I, TK, pp\right)$: After receiving the search token submitted by the data user, the miner node first performs index matching in its stored copy of the blockchain. The matching algorithm is as follows:

$$e(I_{i,1}, TK_1) \cdot e(I_{i,2}, TK_2) = e(g, TK_3) \tag{2}$$

7) Question $(TK, I, PP) \to E(D_{LOC})$ or terminate. With the keyword token and encrypted index, the miner node of the blockchain performs a binary search Algorithm 1. Return the storage location of the encrypted file.

8) Ciphertext conversion $W_j \leftarrow (c_1, c_2, c_3, rk_{i \to j})$:

$$W_1 = c_1, W_2 = c_2 \cdot e(c_1, rk_2), W_3 = rk_1 \tag{3}$$
$$W_4 = H(H(W_1)||H(W_1||W_2)||H(W_2||W_3)). \tag{4}$$

9) Decryption$D \leftarrow D\left(W_i, SK_i\right)$:

$$D = W_2 / e(W_1, W_3)^r \tag{5}$$

5 Security Certification and Analysis

5.1 Correctness Analysis

Theorem 1. *First prove the correctness of Equations (1) and (5).*

Proof. According to and bilinear mapping, it can be known that:

$$\begin{aligned} D &= \frac{c_2}{e(c_1, g_1^{\beta})} \\ &= \frac{De(g_1, g^{\beta})^{r_i}}{e(g^{r_i}, g_1^{\beta})} \\ &= \frac{De(g_1, g)^{\beta r_i}}{e(g, g_1)^{\beta r_i}} \\ &= D. \end{aligned}$$

Equation (1) is obtained. □

Since the data is converted into ciphertext by Equations (3) and (4), it can be obtained according to the decryption:

$$\begin{aligned} D &= \frac{W_2}{e(W_1,W_3)^r} \\ &= \frac{c_2 \cdot e(c_1, rk_2)}{e(c_1, rk_1)^r} \\ &= \frac{D \cdot e(g_1, g^\beta)^{r_i} \cdot e(g^{r_i}, g_1^{-\beta} pk_j^{r_j})}{e(g^{r_i}, g^{r_j})^\beta} \\ &= \frac{D \cdot e(g_1, g^\beta)^{r_i} \cdot e(g^{r_i}, pk_j^{r_j}) \cdot e(g^{r_i}, g_1^{-\beta})}{e(g^{r_i}, g^{r_j})^\beta} \\ &= \frac{D \cdot e(g^{r_i}, pk_j^{r_j})}{e(g^{r_i}, g^{r_j})^\beta} \\ &= \frac{D \cdot e(g^{r_i}, g^{sk_j r_j})}{e(g^{r_i}, g^{r_j})^\beta} \\ &= \frac{D \cdot e(g,g)^{r_i \beta r_j}}{e(g,g)^{\beta r_i r_j}} \\ &= D. \end{aligned}$$

Equation (5) is proved.

Theorem 2. *If the binary search algorithm is executed correctly, the searchable encryption scheme satisfies* ***Definition 2.***

Enter a keyword search token TK and a keyword index ciphertext entry, where the left side of Equation (2) can be calculated as follows:

$$\begin{aligned} & e(I_{i,1}, TK_1) \cdot e(I_{i,1}, TK_2) \\ = \; & e(g^{{M_1}^T p_i'}, g^{{M_1}^{-1} q_i'}) e(g^{{M_2}^T p_i'}, g^{{M_2}^{-1} q_i'}) \\ = \; & e(g,g)^{p_i' q_i' + p_i'' q_i''} \\ = \; & e(g,g)^{(r,\alpha)(\beta,1)} \\ = \; & e(g,g)^{r\beta+\alpha} \\ = \; & e(g, TK_3). \end{aligned}$$

5.2 Security Certificate

Lemma 1. *Searchable scheme statistics privacy when plaintext and keyword predicates.*

5.2.1 Proof of Plaintext Privacy

Lemma 2. *Searchable scheme statistics when plaintext privacy.*

Proof. Suppose A is a computationally infinite adversary that performs polynomial number of queries on encryption and search token generation under the oracle. We proved that adversary A in the experiment $(pp, TK) \leftarrow Setup(\lambda)$ the observation of is lose to the observation of.

We respectively use these two distribution $W_0^* = (w_{0,1}^*, ..., w_{0,n}^*)$ and $W_1^* = (w_{1,1}^*, ..., w_{1,n}^*)$ to represent the challenge of two key words issued by the adversary.Given the hash function and the key $MK = (M_1, M_2)$, we can assume $View_{PP1}^{(b)} = ((I_{1,1}^*, I_{1,2}^*), ..., (I_{n,1}^*, I_{n,2}^*)) = ((g^{{M_1}^{T p_i'}}, g^{{M_2}^{T p_i''}}), \cdots, (g^{{M_1}^{T p_n'}}, g^{{M_2}^{T p_n''}}))$ for $b \in \{0,1\}$, which for $b = 0$ there is$(w_1, ..., w_n) = (w_{0,1}^*, ..., w_{0,n}^*)$,for $b = 1$ there is $(w_1, ..., w_n) = (w_{1,1}^*, ..., w_{1,n}^*)$. Look at the M_1^TandM_2^T uniformly selected from $Z_q^{(n+1)\times(n+1)}$, therefore the distribution p'_i and the distribution p''_i are uniform. We further prove that the joint distribution $M_1^T p_1', ..., M_1^T p_n'$ and $M_2^T p_1'', ..., M_2^T p_n''$ are also uniform. The above two joint distribution can be expressed as$P' {M_1}^T$ and$P'' {M_2}^T$, where $P'$and P'' represent two existing vectors $(p'_1, p'_2, ..., p'_n)$ and $(p''_1, p''_2, ..., p''_n)$ respectively. Finally, we know that $P' {M_1}^T$ and $P'' {M_2}^T$ are uniformly distributed, as long as $(p'_1, p'_2, ..., p'_n)$ and $(p''_1, p''_2, ..., p''_n)$ are linearly dependent.

We first think of vector $(p'_1, p'_2, ..., p'_n)$ as linearly dependent. H is a function that resists hash collisions. The probability of $H(w_j)$ and $H(w_i)$ is same that can be negligible. If for any two vectors $(p'_1, p'_2, ..., p'_n)$ and $(p''_1, p''_2, ..., p''_n)$ is the same, then these two vectors can be also linearly dependent. And since each $x_{i,j}$ of p_i is uniformly chosen from Z_q, the probability that the vector $p_{n-1} = (H(w_1), ..., H(w_{n-1}), x_{n-1,n}, 1)$ and the vector $p_n = (H(w_1), ..., H(w_{n-1}), H(w_n), 1)$ are the same is $1/q$. So the probability of linear dependence of the vector $(p'_1, p'_2, ..., p'_n)$ is $1/q$ at most, which is negligible.It implies that the probability of linear dependence of the vector $(p''_1, p''_2, ..., p''_n)$ is $1/q$ at most. Therefore, the statistical difference λ of $View_{PP1}^{(0)}$ and $View_{PP1}^{(1)}$ is negligible. □

5.2.2 Predicate Privacy Proof

Lemma 3. *Searchable scheme statistics predicate privacy when $n \geq 2$ keyword privacy. Predicate private is also called the security of protecting search pattern which proposed by Kamara and Papamanthou [8].*

Proof. Suppose A is A computationally infinite adversary that performs polynomial number of queries on encryption and search token generation under the oracle.We proved that the observation of and of adversary A in the experiment are close to. We represent these two distributions by $View_{PP2}^{(0)}$ and $View_{PP2}^{(1)}$. (w_0^*, j) and (w_1^*, j) represent the opponent's challenge to search token oracle the. Given the hash function H and the key $MK = (M_1, M_2)$, we can assume $View_{PP2}^{(b)}$=(TK_1^*, TK_2^*, TK_3^*)=$(g^{{M_1}^T p_i'}, g^{{M_2}^T p_i''}, g^{r\beta+\alpha})$. $r, \alpha \xleftarrow{R} Z_q$ for $b \in \{0,1\}$, there is $w_b^* = w_0^*$, and for $b = 1$,there is $w_b^* = w_1^*$. The observations is that only TK_3^* is relevant to w_b^*,TK_1^* and TK_2^* is only related to the serial number of choice random. The serial number of select random is the same for $b \in \{0,1\}$. So for the $b \in \{0,1\}$, we can know that is statistically uniform. This implies that statistical difference λ of $View_{PP2}^{(0)}$ and

$View_{PP2}^{(1)}$ can be negligible. □

5.3 Privacy Analysis

Data privacy. The basic concept of data privacy requires the data to be outsourced should not be revealed to any unauthorized blockchain miner node including external database. In our scheme, data is encrypted by re-encryption. The user who has data level(one) and re-encryption can encrypt. So our scheme protect data privacy.

Plaintext privacy. The basic concept of plaintext privacy requires any association between frequent keywords and encrypted dataset from the index can be deduced, which may learn the mian content of a document. Our searchable index is encrypted by MK key, so our scheme against chosen keyword attack so that it can protect plaintext privacy.

Predicate privacy. The basic concept of predicate private requires any association between frequent keywords and the corresponding token. Our tokens are randomly generated in the scheme. Therefore, the design of our scheme protects predicate private (search pattern). In conclusion,the scheme can guarantee the privacy of search pattern on the blockchain.

5.4 Scheme Comparison and Analysis

This part first presents the comparison between our scheme and that of other literatures. Among them, public key searchable encryption is adopted in literature [2] to realize the privacy protection of data and make fast query for users. Literature [11] a searchable symmetric encryption scheme based on block chain. In literature [20], proxy re-encryption is adopted to realize data sharing and privacy protection, but the specified user cannot query the data quickly and effectively. In this paper, searchable encryption and agent re-encryption are combined to realize data sharing and data tracking, and users can quickly retrieve data information. Scheme comparison are shown in Table 1.

The complexity mainly includes data back and searchable encryption. The traceability complexity is $o(nM)$, where n is the data block, and M is the consistent node when requesting the data transaction again. The index complexity is calculated as $o(n^2)$, the search Token complexity is calculated as $3E + H$, and the test complexity is calculated as $2E + P + H$,where E is exponential function,H is hash function and P is bilinear mapping. The time complexity of binary search algorithm is $O(n \log n)$.

6 Conclusions

The global ledger of transactions stored in blockchain technology is public so that any node joining the blockchain network can get a complete copy. Therefore, the data privacy of blockchain needs to be further strengthened and improved. This paper introduces searchable encryption and proxy re-encryption technology to propose a privacy protection scheme for block chain data based on searchable proxy re-encryption. Use proxy re-encryption technology to encrypt data and store it in the external database, the storage load of blockchain is reduced well, meanwhile not only the data is shared in multi-user mode, but also the data privacy is protected. The miner node on the blockchain can perform binary search by calculating the inner product of the search token vector and index vector to locate the matched index item, which improves the query efficiency of data users. Finally, the effectiveness and security of the proposed scheme are proved by the model of the random oracle model.

In the future, we will improve our scheme to rank keywords, and realize fuzzy search. Although our scheme solve the problem of data sharing and fast search, the corresponding computational complexity will increase. Furthermore, we will improve our scheme in computational performance and search pattern.

Acknowledgments

This research was supported by The National Natural Science Foundation of China (No.61462060, No. 61762060) and The Network and Information Security Innovation Team of Gansu Provincial Department of Education Lanzhou University of Technology (No.2017C-05).

References

[1] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten, "SoK: Research perspectives and challenges for bitcoin and cryptocurrencies," in *IEEE Symposium on Security and Privacy*, pp. 104–121, 2015.

[2] C. Cai, X. Yuan, and C. Wang, "Towards trustworthy and private keyword search in encrypted decentralized storage," in *IEEE International Conference on Communications (ICC'17)*, pp. 1–7, 2017.

[3] P. S. Chung, C. W. Liu, and M. S. Hwang, "A study of attribute-based proxy re-encryption scheme in cloud environments," *International Journal Network Security*, vol. 16, no. 1, pp. 1–13, 2014.

[4] Y. Ding, D. Luo, H. Xiang, C. Tang, L. Liu, X. Zou, S. Li, and Y. Wang, "A blockchain-based digital advertising media promotion system," in *International Conference on Security and Privacy in New Computing Environments*, pp. 472–484, 2019.

[5] J. DuPont and A. C. Squicciarini, "Toward deanonymizing bitcoin by mapping users location," in *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, pp. 139–141, 2015.

Table 1: Scheme comparison

Schemes	Encryption Method	Token Security	Data Privacy	Data Share	Data Back	Predicate Security
Paper [12]	hash	×	✓	✓	×	×
Paper [14]	symmetry	×	✓	✓	×	×
Paper [13]	agent re-encryption	–	✓	✓	✓	–
Our scheme	agent re-encryption	✓	✓	✓	✓	✓

[6] M. Fröwis, A. Fuchs, and R. Böhme, "Detecting token systems on ethereum," in *International Conference on Financial Cryptography and Data Security*, pp. 93–112, 2019.

[7] M. S. Hawng and C. C. Le, "Research issues and challenges for multiple digital signature," *International Journal of Network Security*, vol. 1, no. 1, pp. 1–7, 2005.

[8] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 965–976, 2012.

[9] E. Karafiloski and A. Mishev, "Blockchain solutions for big data challenges: A literature review," in *IEEE EUROCON 17th International Conference on Smart Technologies*, pp. 763–768, 2017.

[10] C. C. Lee, S. T. Hsu, M. S. Hwang, *et al.*, "A study of conjunctive keyword searchable schemes," *International Journal Network Security*, vol. 15, no. 5, pp. 321–330, 2013.

[11] H. Li, H. Tian, F. Zhang, and J. He, "Blockchain-based searchable symmetric encryption scheme," *Computers, Electrical Engineering*, vol. 73, pp. 32–45, 2019.

[12] L. Luu, V. Narayanan, K. Baweja, C. Zheng, S. Gilbert, and P. Saxena, "SCP: A computationally-scalable byzantine consensus protocol for blockchains," *Cryptology ePrint Archive*, vol. 20, no. 20, pp. 2016, 2015.

[13] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," in *Proceedings of the Conference on Internet Measurement Conference*, pp. 127–140, 2013.

[14] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2009. (`https://bitcoin.org/bitcoin.pdf`)

[15] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder, *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*, 2016. ISBN-10: 0691171696.

[16] S. Rezaei, M. A. Doostari, and M. Bayat, "A lightweight and efficient data sharing scheme for cloud computing," *International Journal of Electronics and Information Engineering*, vol. 9, no. 2, pp. 115–131, 2018.

[17] W. She, Z. H. Gu, X. K. Lyu, Q. Liu, Z. Tian, and W. Liu, "Homomorphic consortium blockchain for smart home system sensitive data privacy preserving," *IEEE Access*, vol. 7, pp. 62058–62070, 2019.

[18] L. Wang, J. Gao, and X. Li, "Efficient bitcoin password-protected wallet scheme with key-dependent message security," *International Journal of Network Security*, vol. 21, no. 5, pp. 774–784, 2019.

[19] Y. Wang and J. Gao, "A regulation scheme based on the ciphertext-policy hierarchical attribute-based encryption in bitcoin system," *IEEE Access*, vol. 6, pp. 16267–16278, 2018.

[20] Z. Wang, Y. Tian, and J. Zhu, "Data sharing and tracing scheme based on blockchain," in *The 8th International Conference on Logistics, Informatics and Service Sciences (LISS'18)*, pp. 1–6, 2018.

[21] D. Yaga, P. Mell, N. Roby, and K. Scarfone, "Blockchain technology overview," *arXiv Preprint arXiv: 1906.11078*, 2019. (`https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8202.pdf`)

[22] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, "Fhirchain: Applying blockchain to securely and scalably share clinical data," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 267–278, 2018.

[23] R. Zhang, R. Xue, and L. Liu, "Security and privacy on blockchain," *arXiv Preprint arXiv: 1903.07602*, 2019. (`https://arxiv.org/pdf/1903.07602.pdf`)

[24] R. Zhang, R. Xue, T. Yu, and L. Liu, "Dynamic and efficient private keyword search over inverted index–based encrypted data," *ACM Transactions on Internet Technology (TOIT'16)*, vol. 16, no. 3, p. 21, 2016.

Biography

Tao Feng received the M.E. degree in control theory and control engineering from Lanzhou University of Technology in 1999, and Ph.D. degree in computer architecture from Xidian University in 2008. He is currently a Full Professor and a Ph.D. Supervisor with Lanzhou University of Technology. His main research interests include information security, provable theory of security protocols, wireless network security, and sensor network security. He is a member of the China Computer Federation and China Cryptography Federation.

Hongmei Pei born in Tianshui, Gansu province, is a

master's student of Lanzhou university of the technology.Her research interests is Network and information security.

Pengshou Xie born in Jan, he is a professor,a supervisor of master student at Lanzhou university of the technology, His research interests is Security on Internet of Things

Xiaoqing Feng received the B.S. degree in information and computing science from Xidian University, Xi?an, China, in 2016. She is currently studying for the Ph.D. degree in cyberspace security, Xidian University. Her research interests include blockchain, consensus mechanism, and blockchain applications.

Analysis of Shim's Attacks Against Some Certificateless Signature Schemes

Zhengjun Cao[1] and Olivier Markowitch[2]
(Corresponding author: Zhengjun Cao)
Department of Mathematics, Shanghai University, Shangda Road 99, Shanghai, 200444, China[1]
Computer Sciences Department, Université Libre de Bruxelles, 1050 Bruxelles, Belgium[2]
Email: caozhj@shu.edu.cn
(Received Dec. 3, 2019; Revised and Accepted May 15, 2020; First Online Apr. 12, 2021)

Abstract

We show that the Shim's attacks [Information Sciences, 296 (2015), 315-321] against some certificateless signature schemes are flawed because the adversary is just transforming a valid signature into a new valid signature, but both are bound to *the same identity* and *the same message*. This kind of attack against signature schemes has no practical significance because the relationship between the signer and the message (who undertakes the ultimate responsibility of signing a message) is not changed. We also find the proposed attack against Choi *et al.*'s scheme is false because the adversary has to use the signer's secret key, but it is unavailable to the attacker. We think what is a true forgery against a digital signature should be explicitly specified in future works.

Keywords: Certificateless Signature; Existential Forgery; Man-in-Middle Attack; Selective Forgery

1 Introduction

In the classical public key cryptography, one has to verify the facticity of involved public key in order to resist man-in-middle attack. The work of public key verification depends essentially on public key infrastructure (PKI). If the amount of users is very large, the public key manager could become the bottleneck of the whole system.

Identity-based encryption was introduced by Shamir [19], in which one can encrypt messages using an intended user's identity information. Of course, some system public parameters should be invoked. ID-based cryptography has been intensively studied since the Boneh-Franklin's breakthrough work [5] using pairings over elliptic curves. ID-based cryptography intends to remove the managing center by directly invoking an intended user's ID and viewing it as the replacement of the common public key. But it has to introduce a new participant, key generation center (KGC), who is in charge of generating the secret key for each user in the system [2–4, 6, 9]. As the KGC generates the secret key for all users, a complete trust must be placed on the KGC.

In 2003, Al-Riyami and Paterson [1] introduced the new paradigm—certificateless cryptography, in order to relieve the KGC dependency, which was just a variant of ID-based cryptography [7, 10]. Its key generation process is split between the KGC and the user. Notice that in certificateless cryptography the identity information no longer forms the entire public key [11, 12, 15, 17, 18, 21]. That is, the user's public key is not discoverable from only the user's identity string and the KGC's public key. Thus, the user's public key must be published by himself. It is not authentic. The identity string and the KGC's public key can be used to verify that the involved public key belongs to the true entity. Frankly, in certificateless cryptography it is still a challenge to efficiently create basic trust between participants. Therefore, the so-called key-replacement attack against certificateless cryptographic schemes is broadly adopted

Li *et al.* [8, 14, 16] had presented three certificateless signature schemes to demonstrate the powerful strength of certificateless cryptography. In 2015, Shim [20] argued that the three signature schemes can not resist public-key-replacement attack, and proposed the revisited security model.

In this note, we show that the attacks against the three schemes are flawed, because the relationship between the signer and the message (who undertakes the ultimate responsibility of signing a message) is not changed at all. We also find the attack against Choi *et al.*'s scheme is false because the adversary has to use the signer's secret key, but it is not available to him. Besides, we find the security model for certificateless signature described in the literatures is somewhat artificial, which results in the contentious attacks.

This paper is organized as follows. In Section 2, we present an explicit comparison of three kinds of cryptography, *i.e.*, the general public key cryptography, ID-based cryptography, and certificateless cryptography. In Section 3, we clarify what is a true forgery against signature schemes. In Section 4 and Section 5, we show that the

Table 1: Comparison of three kinds of cryptography

	public-key cryptography	ID-based cryptography	certificateless cryptography
ID number	——	invoked, certificate-checking for a fresh ID number	invoked, certificate-checking for a fresh ID number
user's public key/parameters	invoked, certificate-checking for a fresh *public key*	——	invoked, no certificate-checking for *public parameters*
secret key	set by the user, *exclusive*	totally assigned by some social institute, *nonexclusive*	partially assigned by some social institute, *partially exclusive*

Shim's attacks are not sound.

2 Comparison of Three Kinds of Cryptography

In a certificateless encryption or signature scheme, a user's public parameters, which are directly issued by its owner, are different from the general public key. As we know, the general public key must be notarized, signed and issued by managing center, while public parameters are but auxiliary numbers invoked by algorithms. Usually, one cannot confirm the true owner of the public parameters invoked in a certificateless encryption or signature scheme (see Table 1 for the comparison of three kinds of cryptography).

We here would like to stress that in certificateless cryptography [11, 13, 16], the invoked ID number must be legitimate. Actually, the credibility of the invoked ID number originates just from its associated certificate, which is issued by the relevant government department, such as passport issuing authority. This means the so-called certificateless cryptography is *not totally certificateless*. It has to make use of the certificate associated to a certain ID in order to build trust. Some researchers have misunderstood the essence and proposed several false attacks against certificateless cryptography by replacing users' IDs optionally.

3 What Is a True Forgery against Signature Schemes

The verification of a digital signature is achieved by means of signer's public key. The verifier must get the signer's public key via a secure and trustable channel. Otherwise, it seems impossible to create a signature algorithm if the invoked public key is doubtful. So, attacks against digital signature can be classified into:

Only public key known attack. The adversary only knows the signer's public key and the related verifying procedure.

Previous signatures known attack. The adversary knows the signer's public key and some message/signature pairs made by the signer previously. In practice, it is usual to assume that an adversary can access to these resources at least.

Chosen-message attack. The adversary is permitted to choose messages and ask the signer for their signatures. Making use of these valid message/signature pairs, the adversary tries to forge a new valid message/signature pair.

What's the true meaning of forging signatures? A signature always corresponds to a message. It could be possible to forge a signature for a special message, or signatures for some special messages. So, the forgeries can be classified into the following kinds.

Existential forgery. The adversary can forge signatures for a special message.

Selective forgery. The adversary can choose some messages and forge signatures for them. These messages may have no any regular type or format, and the amount of these messages could be impossible to estimate.

Arbitrary forgery. Although the adversary cannot retrieve the signer's secret key, he can generate signatures for any message.

Total break. The adversary can retrieve the signer's secret key.

Notice that we do not consider whether an adversary can transform a valid signature $\{ID; m; \sigma\}$ into another valid signature $\{ID; m; \sigma'\}$, where both signatures do correspond to the same message m and the same entity ID. That means the original signer undertakes the ultimate responsibility for the signed message. From the practical point of view, *the relationship between the entity and the message are not changed at all.* Such a forgery by altering a part of data in a valid signature is not a true one. We find, however, the work [20] have neglected the commonplace guideline and proposed some trivial attacks against the three signature schemes [8, 14, 16].

Table 2: Differences between the original signatures and the forged signatures

Liu-Au-Susilo signature	
CL-Sign (original)	Output the signature $\{ID, m, (g^x, g_1^x), (V, R_\pi, R_m)\}$.
CL-Sign (forged)	Pick y and output the signature $\{ID, m, (g^{xy}, g_1^{xy}), (V^y, R_\pi^y, R_m^y)\}$.
Li-Chen-Sun signature	
CL-Sign (original)	Output the signature $\{ID_A, m, (X_A, Y_A), (U, V)\}$.
CL-Sign (forged)	Pick t and output the signature $\{ID_A, m, (tX_A, tY_A), (U, tV)\}$.

4 A False Attack Against Choi *et al.*'s Signature

In 2007, Choi *et al.* [8] proposed a certificateless signature scheme, which can be described as follows.

4.1 Description of The Signature Scheme

Setup. Let $\mathbb{G}_1, \mathbb{G}_2$ be bilinear groups of prime order q, $P \in \mathbb{G}_1$ be a generator, $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ be the bilinear map. $\mathsf{H}_1 : \{0,1\}^* \to \mathbb{G}_1$, $\mathsf{H}_2 : \mathbb{G}_1 \to \mathbb{Z}_q^*$, $\mathsf{H}_3 : \{0,1\}^* \to \mathbb{Z}_q^*$ are three hash functions. Pick $s < q$ and compute $P_{pub} = sP$. The system parameters are set as $q, \mathbb{G}_1, \mathbb{G}_2, \hat{e}, P, P_{pub}, \mathsf{H}_1, \mathsf{H}_2, \mathsf{H}_3$. The master key is s.

Key-extraction. For Alice's identity ID_A, the system manager computes

$$D_A = s\,\mathsf{H}_1(ID_A)$$

and returns it to her.

Public parameters. Alice picks $x_A < q$, computes $PK_A = x_A P$, and publishes it as her public parameter.

Secret key deriving. Alice computes her secret key

$$SK_A = \frac{D_A}{x_A + \mathsf{H}_2(PK_A)}$$

Signing. For m, Alice picks $r < q$ and computes

$$U = r\mathsf{H}_1(ID_A), \quad V = (r + \mathsf{H}_3(m, U))SK_A.$$

The resulting signature is $\{ID_A; m; PK_A, U, V\}$.

Verifying. Check that

$$\begin{aligned} & \hat{e}(V, PK_A + \mathsf{H}_2(PK_A)P) \\ = \; & \hat{e}(U + \mathsf{H}_3(m,U)\mathsf{H}_1(ID_A), P_{pub}). \end{aligned}$$

4.2 The False Attack

In 2015, Shim [20] presented an attack against the signature scheme. Suppose that an adversary obtains Alice's valid signature $\{ID_A; m; PK_A, U, V\}$. He then picks $x'_A < q$ and computes

$$PK'_A = x'_A P, \quad V' = \frac{x_A + \mathsf{H}_2(PK_A)}{x'_A + \mathsf{H}_2(PK'_A)} V$$

The forged signature is $\{ID_A; m; PK'_A, U, V'\}$.

Note that the adversary has to invoke x_A. As for this invoking, it writes (see page 320, line 8, [20]): *"Then $\mathcal{A}^I$, who knows x_A, computes $\cdots$"* But x_A is the secret key chosen by the original signer. Clearly the adversary $\mathcal{A}^I$ cannot obtain x_A, or $x_A V$, given $\{U, V, ID_A, PK_A\}$, even along with the key D_A. Thus, the adversary cannot finish the above computation.

5 On Other Flawed Forgeries

In the same paper [20], Shim also presented some attacks against the two schemes [14, 16]. We now only describe the main differences between the original and the forged signatures (see Table 2).

Note that the relationship between the true signer with the identity ID and the true message m is not changed. The so-called attack is just transforming a legal signature into a new legal one, but the new is still bound to the original signer, who should undertake the ultimate responsibility of signing the true message, not a wrong message. From the practical point of view, a verifier does only care about who is the true signer of a signature on a message. He does not consider whether there are plenty of legal signatures on a same message bound to its original signer. So, the above attacks make no sense from a practical perspective.

6 Conclusion

We point out that the Shim's attacks against certificateless signature schemes are false. We want to stress that a forgery by altering a part of data in a valid signature while keeping the relationship between the true signer and the true message, is not a true one. But the commonplace guideline is often neglected in some literatures, and the inattention usually gives rise to misunderstanding. We think what is a true forgery against a signature scheme should be explicitly specified in future works.

Acknowledgment

We thank the National Natural Science Foundation of China (61411146001). We are grateful to the reviewers for their valuable suggestions.

References

[1] S. Al-Riyami and K. Paterson, "Certificateless public key cryptography," in *Proceedings of 9th International Conference on the Theory and Application of Cryptology and Information Security, Advances in Cryptology*, pp. 452–473, Dec. 2003.

[2] D. Boneh and X. Boyen, "Short signatures without random oracles," *Journal of Cryptology*, vol. 21, no. 2, pp. 149–177, 2008.

[3] D. Boneh and X. Boyen, "Efficient selective identity-based encryption without random oracles," *Journal of Cryptology*, vol. 24, no. 4, pp. 659–693, 2011.

[4] D. Boneh, R. Canetti, and S. Haleviand J. Katz, "Chosen-ciphertext security from identity-based encryption," *SIAM Journal on Computing*, vol. 36, no. 5, pp. 1301–1328, 2007.

[5] D. Boneh and M. Franklin, "Identity-based encryption from the Weil pairing," in *Proceedings of 21st Annual International Cryptology Conference, Advances in Cryptology*, pp. 213–229, Aug. 2001.

[6] D. Boneh, A. Raghunathan, and G. Segev, "Function-private identity-based encryption, hiding the function in functional encryption," in *Proceedings of 33rd Annual Cryptology Conference, Advances in Cryptology*, pp. 461–478, Aug. 2013.

[7] S. F. Chiou, H. T. Pan, E. F. Cahyadi, and M. S. Hwang, "Cryptanalysis of the mutual authentication and key agreement protocol with smart cards for wireless communications," *International Journal of Network Security*, vol. 21, no. 1, pp. 100–104, 2019.

[8] K. Y. Choi, J. H. Park, J. Y. Hwang, and *et al.*, "Efficient certificateless signature schemes," in *Proceedings of 5th International Conference on Applied Cryptography and Network Security*, pp. 443–458, June 2007.

[9] J. Coron, "A variant of boneh-franklin IBE with a tight reduction in the random oracle model," *Design, Codes and Cryptography*, vol. 50, no. 1, pp. 115–133, 2009.

[10] L. Deng, H. Huang, and Y. Qu, "Identity based proxy signature from RSA without pairings," *International Journal of Network Security*, vol. 19, no. 2, pp. 229–235, 2017.

[11] B. Hu, D. Wong, Z. Zhang, and X. Deng, "Key replacement attack against a generic construction of certificateless signature," in *The 11th Australasian Conference on Proceedings of Information Security and Privacy (ACISP'06)*, pp. 235–246, July 2006.

[12] L. C. Huang, T. Y. Chang, and M. S. Hwang, "A conference key scheme based on the Diffie-Hellman key exchange," *International Journal of Network Security*, vol. 20, no. 6, pp. 1221–1226, 2018.

[13] X. Huang, W. Susilo, Y. Mu, and F. Zhang, "On the security of certificateless signature schemes from asiacrypt 2003," in *Proceedings of 4th International Conference, Cryptology and Network Security*, pp. 13–25, Dec. 2005.

[14] X. Li, K. Chen, and L. Sun, "Certificateless signature and proxy signature schemes from bilinear pairings," *Lithuanian Mathematical Journal*, vol. 45, no. 1, pp. 76–83, 2005.

[15] T. C. Lin, T. Y. Yeh, and M. S. Hwang, "Cryptanalysis of an id-based deniable threshold ring authentication," *International Journal of Network Security*, vol. 21, no. 2, pp. 298–302, 2019.

[16] J. K. Liu, M. H. Au, and W. Susilo, "Self-generated-certificate public key cryptography and certificateless signature/encryption scheme in the standard model," in *Proceedings of ACM Symposium on Information, Computer and Communications Security*, pp. 273–283, Mar. 2007.

[17] L. Liu, W. Kong, Z. Cao, and J. Wang, "Analysis of one certificateless encryption for secure data sharing in public clouds," *International Journal of Electronics and Information Engineering*, vol. 6, no. 2, pp. 108–113, 2017.

[18] H. T. Pan, E. F. Cahyadi, S. F. Chiou, and M. S. Hwang, "Research on batch verification schemes for identifying illegal signatures," *International Journal of Network Security*, vol. 21, no. 6, pp. 1062–1070, 2019.

[19] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Proceedings of Advances in Cryptology (CRYPTO'84)*, pp. 47–53, Aug. 1984.

[20] K. A. Shim, "Security models for certificateless signature schemes revisited," *Information Sciences*, no. 296, pp. 315–321, 2015.

[21] Y. L. Wang, J. J. Shen, and M. S. Hwang, "A survey of reversible data hiding for VQ-compressed images," *International Journal of Network Security*, vol. 20, no. 1, pp. 1–8, 2018.

Zhengjun Cao, associate professor, with the Department of Mathematics, Shanghai University, received his PhD degree in applied mathematics from Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He had served as a post-doctor at Computer Sciences Department, Université Libre de Bruxelles. His research interests include cryptography, discrete logarithms and quantum computation.

Olivier Markowitch, professor, with the Computer Sciences Department, Université Libre de Bruxelles, is an information security advisor of his university. He is working on the design and analysis of secure multi-party computation protocols, as well as on the design and analysis of digital signature schemes.

Guide for Authors
International Journal of Network Security

IJNS will be committed to the timely publication of very high-quality, peer-reviewed, original papers that advance the state-of-the art and applications of network security. Topics will include, but not be limited to, the following: Biometric Security, Communications and Networks Security, Cryptography, Database Security, Electronic Commerce Security, Multimedia Security, System Security, etc.

1. Submission Procedure

Authors are strongly encouraged to submit their papers electronically by using online manuscript submission at http://ijns.jalaxy.com.tw/.

2. General

Articles must be written in good English. Submission of an article implies that the work described has not been published previously, that it is not under consideration for publication elsewhere. It will not be published elsewhere in the same form, in English or in any other language, without the written consent of the Publisher.

2.1 Length Limitation:

All papers should be concisely written and be no longer than 30 double-spaced pages (12-point font, approximately 26 lines/page) including figures.

2.2 Title page

The title page should contain the article title, author(s) names and affiliations, address, an abstract not exceeding 100 words, and a list of three to five keywords.

2.3 Corresponding author

Clearly indicate who is willing to handle correspondence at all stages of refereeing and publication. Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address.

2.4 References

References should be listed alphabetically, in the same way as follows:
For a paper in a journal: M. S. Hwang, C. C. Chang, and K. F. Hwang, ``An ElGamal-like cryptosystem for enciphering large messages," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 445--446, 2002.

For a book: Dorothy E. R. Denning, *Cryptography and Data Security*. Massachusetts: Addison-Wesley, 1982.

For a paper in a proceeding: M. S. Hwang, C. C. Lee, and Y. L. Tang, ``Two simple batch verifying multiple digital signatures," in *The Third International Conference on Information and Communication Security (ICICS2001)*, pp. 13--16, Xian, China, 2001.

In text, references should be indicated by [number].

Subscription Information

Individual subscriptions to IJNS are available at the annual rate of US$ 200.00 or NT 7,000 (Taiwan). The rate is US$1000.00 or NT 30,000 (Taiwan) for institutional subscriptions. Price includes surface postage, packing and handling charges worldwide. Please make your payment payable to "Jalaxy Technique Co., LTD." For detailed information, please refer to http://ijns.jalaxy.com.tw or Email to ijns.publishing@gmail.com.