OIC-CERT
Computer Emergency Response Team

Enhancing the knowledge on cyber security among the OIC member countries

OIC-CERT JOURNAL OF CYBER SECURITY

Volume 2, Issue 1 February 2020

The Organisation of the Islamic Cooperation - Computer Emergency Response Team
www.oic-cert.org

ISSN 2636-9680
eISSN 2682-9266

Published by CyberSecurity Malaysia as the
OIC-CERT Permanent Secretariat

ISSN 2636-9680
eISSN 2682-9266

Editorial Panel

OIC-CERT Journal of Cyber Security

Volume 2, Issue 1
February 2020

Content

OIC-CERT Computer Emergency Response Team

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
1– 17

A Systematic Literature Review on the Security and Privacy of the Blockchain and Cryptocurrency

Aslinda Hassan[1], Mohd Zaki Mas'ud[2], Wahidah Md. Shah[3], Shekh Faisal Abdul-Latip[4], Rabiah Ahmad[5], Aswami Ariffin[6], and Zahri Yunos[7]
[1,2,3,4,5]Center for Advanced Computing Technology, Faculty of Information and Communications Technology, Universiti Teknikal Malaysia Melaka, Malaysia
[6,7]CyberSecurity Malaysia, Malaysia
[1]**aslindahassan@utem.edu.my,** [2]**zaki.masud@utem.edu.my,** [3]**wahidah@utem.edu.my,** [4]**shekhfaisal@utem.edu.my,** [5]**rabiah@utem.edu.my,** [6]**aswami@cybersecurity.my,** [7]**zahri@cybersecurity.my**

ARTICLE INFO

Article History
Received 31 May 2019
Received in revised form 15 Aug 2019
Accepted 25 Sep 2019

Keywords:
blockchain, cryptocurrency, SLR, security, vulnerabilities, threats

ABSTRACT

A blockchain can be summarized as a decentralized ledger of all transactions across a peer-to-peer network. It is the main technology behind the large number of diverse cryptocurrencies that are currently available in circulation. Since its introduction, the blockchain technology has shown promising application prospects and attracted lot of attention from both academia and industry. It also has become an obvious target to adversaries. In this paper, we conduct a systematic literature review on the security vulnerabilities and cyber-attacks to blockchain and cryptocurrency by searching and analyzing previous research papers indexed in reputable journal databases. Based on our findings, we then summarize the most common and critical security threats and attacks and the current countermeasures.

I. INTRODUCTION

The Blockchain technology has begun in 2008 when Satoshi Nakamoto proposed Bitcoin as a new and revolutionize conception of money. It is a purely peer-to-peer electronic cash that makes it possible to send payments directly to the intended recipients without relying to any third party [1]. According to Kobler et al. (2017), a blockchain can be described as a distributed ledger technology protocol that enables data to be exchanged directly between different parties without the need for a middle-man [2]. The participants anonymously interact with encrypted identities, and each transaction is subsequently added to a permanent transaction chain and distributed to all related nodes on the network. This allows for the potential of providing a trustworthy and secure platform to facilitate business activities. In other terms, blockchain is a chain of blocks that contain information [3]. Once recorded, the information inside of this chain becomes challenging to change, thus preventing tampering. The protocol is intended to make it easier for people to shift from centralized financial systems to a decentralized distributed network.

ISSN 2636-9680
eISSN 2682-9266

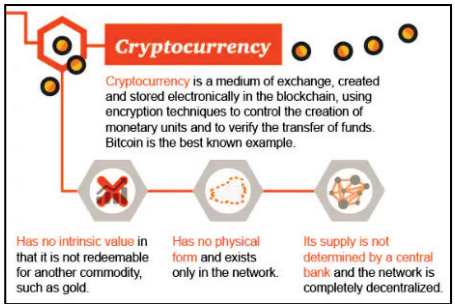

Fig. 1.: Infographic on cryptocurrency
Source: [4]

A blockchain is the foundation of all cryptocurrencies that are currently available in circulation [5]. A cryptocurrency, such as Bitcoin, is a medium of exchange, which is similar to the US dollar. Unlike the US dollar, however, a cryptocurrency is digital and uses encryption techniques to control monetary unit creation and verify the transfer of funds [4].

One of the best-known cryptocurrencies is Bitcoin, which is a decentralized virtual monetary unit that is based on peer-to-peer (P2P) network and not issued by a government or any organization [6]-[8]. After its introduction in year 2009, Bitcoin is the most successful cryptocurrency thus far. Given the Bitcoin's current value, it is obvious that Bitcoin has become a target for adversaries.

Currently, few existing surveys that have been done on a blockchain and cryptocurrencies. In particular, the survey in [8], [9] provides an extensive introduction of the blockchain and cryptocurrencies. The survey presented by [10] concentrates on security and privacy issues in the blockchain in general whereas the surveys in [11], [12] focus the review specifically on Bitcoin. However, these survey papers are done using the traditional narrative review method [13], [14]. In this paper, we present a survey based on the Cochrane Systematic Review [15], [16] specifically targeting the security and privacy aspects of the blockchain technology and cryptocurrency. In particular, we concentrate on the security challenges and their countermeasures regarding the key components of the blockchain technology and cryptocurrency.

II. RESEARCH METHODOLOGY

This section provides the methodology for the systematic review of the security and privacy in blockchain and cryptocurrency. According to Cochrane Collaboration [15], [16], a systematic review attempts to collect all documentation that suits pre-specified eligibility requirements to answer a particular research question. It uses definitive, systematic methods to reduce bias, thereby providing reliable discoveries from which conclusions can be drawn and decisions taken. In [17], a systematic literature review (SLR) is a process of identifying, evaluating and interpreting all available studies that are pertinent to a particular research question. Our review methodology is based on the guidelines proposed by Kitchenham and Charters, (2007). From [17], a systematic literature review consists of three primary phases: *planning*, *conducting* and *reporting*. The planning phase of the systematic reviews starts with the definition of a protocol that will guide the progress of the review. Our review protocol is based on the five steps in conducting a system review by Khan et al. (2003) in [18], as shown in **Fig. 1**.

The following subsections present a detailed description of the review protocol.

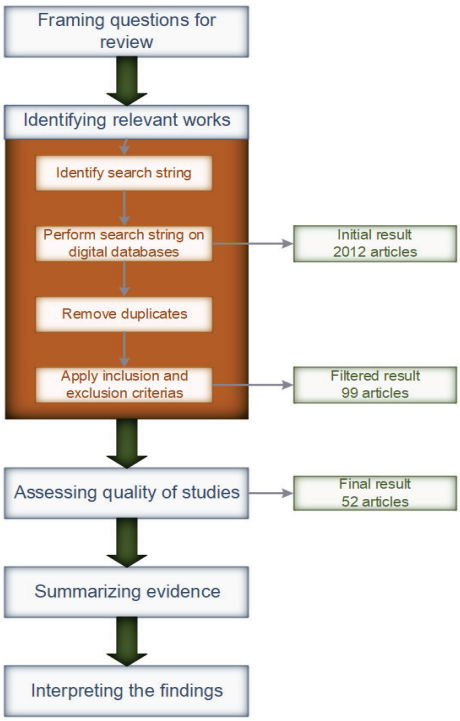

Fig. 2: SLR Methodology

A. Framing research questions for a review

In general, the main objective of this systematic literature review is to gain knowledge on the state of the art of the security in blockchain, specifically in cryptocurrencies. The systematic review also aims to look at the security threats in blockchain and any countermeasures proposed. Therefore, in order to have this knowledge in our investigation, we have defined the following research questions (RQ):

RQ1: What is the blockchain and its application in virtual currency and distributed ledger?

RQ2: What are threats/security vulnerabilities and countermeasures in the blockchain and cryptocurrency?

B. Identifying relevant work

After the research questions have been established, the next phase is to define the search strategy and search string. The primary goal of the search process is to identify journal articles on digital forensics in the blockchain with focusing on cryptocurrencies. The searching method included an automatic search provided by the digital libraries using a search string that is recurrently used by the researchers in this field.

Search Strategy

The searching process was started initially on August 2018 and with defining search strings. The search strings were composed of the following search terms:

RQ1	Fundamentals, blockchain, virtual currency, cryptocurrency, Distributed ledger.
RQ2	Blockchain, cryptocurrency, transaction, mining, threat, vulnerabilities, technologies, countermeasure.

Using the above search terms, we define the search strings and use them on online literature databased to find and collect relevant papers. We have considered four widely used online repositories for work: ACM Digital Library, IEEE Xplore Digital Library, SpringerLink, and ScienceDirect. Boolean logic (AND, OR) was added in the form of search operators (quotations, parentheses) to make the search results more relevant.

The definition of inclusion and exclusion criteria

As shown in **Fig. 1**, the original search produced 2012 papers because many of the papers were either duplicated, inadequate in quality or not affiliated to research questions. Due to the above reasons, we conducted additional filtration using the following inclusion and exclusion, as shown in **Fig. 1**.

ISSN 2636-9680
eISSN 2682-9266

Inclusion criteria	1. Articles from year 2013 – 2018 2. Articles related to security in cryptocurrencies and blockchain. 3. Articles must be published in a journal or a conference proceeding.
Exclusion criteria	1. Articles related to blockchain applications other than cryptocurrency such as healthcare, e-voting, etc. 2. Survey, news and commentary, patents, citation, book chapters, theses.

C. Summarizing the evidence

Fig. 3 until **Fig. 6** show the statistics of the selected publications after assessing the quality of the selected articles. As shown in **Fig. 2**, the final number of the selected publication is 52 publications. From 52 articles, 48% of the publications came from SpringerLink database and 35% came from IEEE database, as shown in **Fig. 3** and the rest came from ACM Digital Library and ScienceDirect. The highest number of articles for RQ 1 came from IEEE whereas for RQ2, the highest number of publications is from Springer.

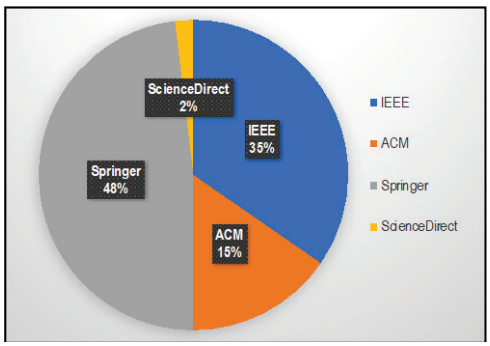

Fig. 3: Percentage of selected publications based on online database

Fig. 4 and **Fig. 5** displays the number of articles on blockchain and cryptocurrency published between the years 2014 and 2015. Although the selected period in our inclusion criteria is between year 2013 until 2018, as can be seen from both figures, researches on the blockchain and cryptocurrency began to emerge after 2013. Furthermore, the trend from both statistics shows that publication in blockchain and cryptocurrency have steadily risen over the years. In the beginning, between 2014 and 2015, there were average of four publications each year. However, the year 2017 and 2018 stand out because there were 13 and 23 publications, respectively. From **Fig. 5**, the articles on RQ 2 have the highest number in year 2018, which was 13 articles, compared to RQ 1.

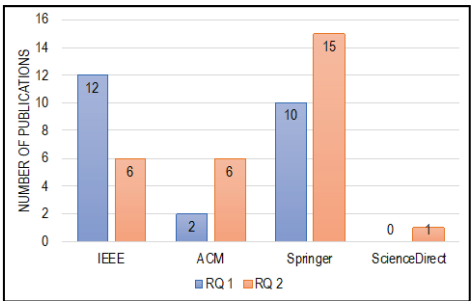

Fig. 4: Number of selected publications for each RQ by online database

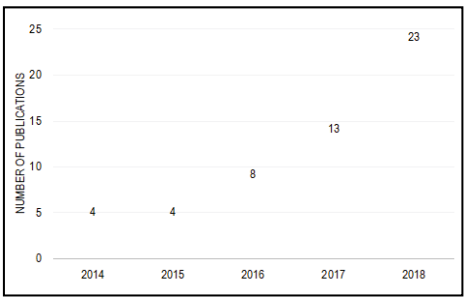

Fig. 5: Number of selected publications based on year of publication

As stated in previous section, RQ 2 focused on the threats and vulnerabilities of the blockchain and cryptocurrency. This is understandable since the first cryptocurrency, which was Bitcoin was rated as the best performing commodity in 2016 [19], with the market value of USD 1023 in January 2017. At the same time, the blockchain technology was introduced to many areas such as medicine, e-voting, the Internet of Things, etc. Since the

blockchain technology has been applied in many fields, users started to have concerns on its security since a number of security vulnerabilities and attacks have been recently reported. The most common example in Bitcoin security vulnerabilities is the Mt. Gox attack, where in March 2014, the criminals exploited transaction mutability in Bitcoin to attack Mt. Gox, the largest Bitcoin trading platform [20]. The attack caused the Mt. Gox to collapse with a value of 450 million dollars Bitcoin being stolen.

Therefore, with the increasing interest on security in the blockchain technology and cryptocurrency, researches on the threats and vulnerabilities of the blockchain and their countermeasure have become an emerging topic in year 2018.

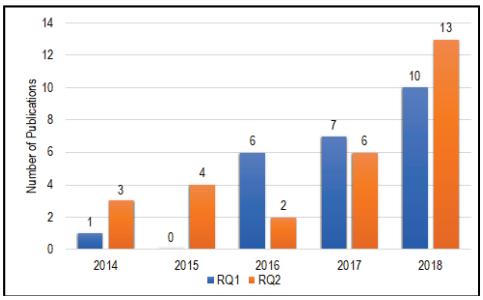

Fig. 6: Number of selected publications for each RQ by year of publication.

III. DISCUSSION AND ANALYSIS

A. Definition of blockchain and cryptocurrency

Various definitions have been used to conceptualize and define a blockchain and its application in cryptocurrencies. Kobler et al. (2016) outlined the definition of a blockchain as a distributed ledger technology protocol that enables data to be exchanged directly between different parties without the need for a middle-man [2]. From the online dictionary of Merriam-Webster [21], a blockchain is defined as "a digital database containing information (such as records of financial transactions) that can be simultaneously used and shared within a large decentralized, publicly accessible network." Merriam-Webster also quoted a definition from Iansiti and Lakhani (2017) in the blockchain definition. According to Iansiti and Lakhani (2017), "The technology at the heart of Bitcoin and other virtual currencies, blockchain is an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way. The ledger itself can also be programmed to trigger transactions automatically" [22].

From the above definitions, we can conclude that the definition of a blockchain must consists at least the following keywords:

1) distributed ledger
2) open and shared (publicly accessible)
3) verifiable
4) transaction
5) decentralized

A cryptocurrency is an application that utilizes the blockchain technology. To define cryptocurrency, we should look at the original definition of cryptocurrency or Bitcoin from Nakamoto (2008). In his whitepaper, cryptocurrency or Bitcoin is defined as "A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution" [1]. Nakamoto (2008) further define Bitcoin as the following:

> "We define an electronic coin as a chain of digital signatures. Each owner transfers the coin to the next by digitally signing a hash of the previous transaction and the public key of the next owner and adding these to the end of the coin. A payee can verify the signatures to verify the chain of ownership" [1].

In addition to Nakamoto definition, Merriam-Webster defines cryptocurrency as "any form of currency that only exists digitally, that usually has no central issuing or regulating authority but instead uses a decentralized system to record transactions and manage the issuance of new units, and that relies on cryptography to prevent

ISSN 2636-9680
eISSN 2682-9266

counterfeiting and fraudulent transactions" [21].

Therefore, to answer RQ 1, we look at the blockchain definitions in the selected publications for RQ 1 and see whether the definitions include the above keywords. However, the articles selected for RQ 1 must define the blockchain and cryptocurrency technologies based from the authors' own understanding of the technology. Survey articles are not used to answer RQ 1 since the definitions of the technologies are based from other researchers. **Fig. 7** shows the article categorization according to the authors' definition of the blockchain technology and cryptocurrency whereas **Fig. 8** present the number of articles based on the blockchain keywords in the abovementioned paragraph.

From **TABLE 1**, there is a significant overlap among the above-mentioned keywords in the blockchain definitions from the selected publications. As shown in **TABLE 1**, only authors from [31] and [40] use all five keywords for the blockchain definition whereas three out of the 24 articles use two of the keywords in their blockchain definition. Two of the 24 selected articles did not give any blockchain definition. The two papers focus only on their research in cryptocurrency.

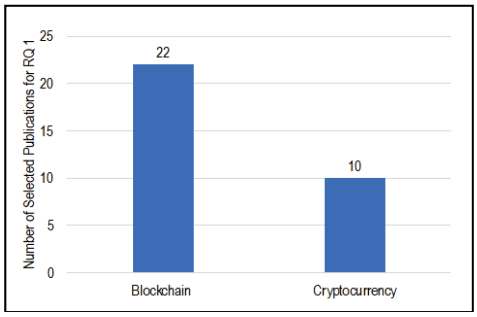

Fig. 7: Article categorization for RQ 1

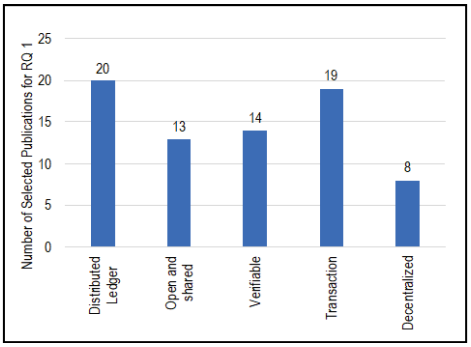

Fig. 8: Number of publications according to the blockchain keywords

TABLE 1: Definitions of the blockchain based on the selected keywords

References	Distributed ledger	Open and shared	Verifiable	Transaction	Decentralized	Technical Discussion/ Proposal related to cryptocurrency
[23]	√	√	√	√		√
[24]	√	√		√		√
[25]	√		√	√	√	√
[26]	√	√	√	√		√
[27]	√		√	√	√	√
[28]	√	√	√	√		√
[29]	√			√		
[30]	√	√	√	√		√
[31]	√	√	√	√	√	
[32]	√		√	√	√	
[33]	√			√		
[34]	√				√	
[35]	√	√		√		
[36]	√	√	√			
[37]			√	√		
[38]	√	√		√	√	
[39]	√	√	√			
[40]	√	√	√	√	√	
[41]	√		√	√		
[42]				√	√	√
[43]	√	√	√	√		
[44]						√
[45]						√
[46]	√	√		√		

B. RQ 2 - Threats, vulnerabilities and countermeasures for blockchain and cryptocurrencies

Demand on the application of Blockchain technology in securing online transaction and critical business increased dramatically. Blockchain has become most secured application for critical business infrastructure such as finance, transportation industries and medical. As the technology increased, blockchain also exposes to various possible security threats and vulnerabilities. Security threats can be defined in two categories i.e., deliberate and accidental. The threats which planned by a dedicated team with specific objective and target victim can be classified as deliberate threats. The unplanned or commonly known as accidental threats can be caused by natural disasters or any action which may create damage to any system. Deliberate threats also known as attack. Various type of threats possibly occurs in Blockchain technology including its application.

It is well accepted by expert that Blockchain possess with vulnerabilities due to drawbacks which possibly occur in software design, hardware requirements and protocol. **TABLE 2** below provide a summary of threats and vulnerabilities in blockchain from articles collected during the SLR search process to respond to RQ 2. For RQ 2, the articles collected must not only discuss the threat and vulnerabilities of cryptocurrencies, but the countermeasures as well. All threats and vulnerabilities as well as the countermeasures are categorized based on the Blockchain components as stated by Puthal et al. (2018) [47] (Refer to **Fig. 9**). It is important to note here that for each component posses with at least one possible threat.

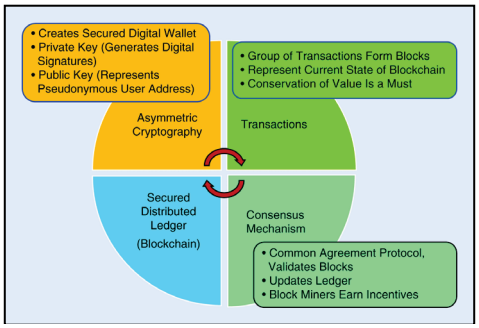

Fig. 9: The core component of a blockchain by Puthal et al. (2018)
Source: [47]

TABLE 2: Findings on threats and vulnerabilities of Blockchain and their countermeasures

Blockchain Component	Threats and Vulnerabilities	Countermeasures
Asymmetric Cryptography	Elliptic curve digital signature algorithm (ECDSA) for transaction authentication – unable to cope with ***quantum attack***. ECDSA is common signature algorithm used in Bitcoin – one of technology in blockchain. Blockchain operates as decentralized network which are much more temper resistant than centralized network. Researchers from National University Singapore (NUS) found out that Quantum Cryptography provide minimal number of entropies into system thus reduce noise. However, application of quantum crypto creates flaws due to asymmetric cryptography used for digital signature.	A new signature authentication scheme for blockchain by using the lattice based bonsai tree signature [48].

ISSN 2636-9680
eISSN 2682-9266

Blockchain Component	Threats and Vulnerabilities	Countermeasures
	The loss of private key during cybersecurity breach.	A private key safety model for safely keeping the sub elements of the private key under different span of operation profiles and adding a number of character salts as a common subsequence in each span. In addition, the authors use syntactic, semantic and cognitive safety control to enforce dependency among the spans [49].
	Weakened cryptographic primitives owing to either the advancement of cryptanalysis or the advancement of the attackers ' computing power [50]. Cryptographic primitives can be defined as well-established, low-level cryptographic algorithms that are considered the building blocks of a blockchain.	The authors in [50] recommended the following to avoid some types of primitive breakage: • Users should not reuse Bitcoin addresses • Use the least number of transactions per block • Migrate to new address types with string hashing and signature scheme. • Instead of using nested hashes for Address Hash and Main Hash, users should combine both hashes in a way that increases defense-in-depth. • Consider using a hardfork for a weakened Main Hash with re-designed headers and transactions, and without using any of the old primitives.
	A Bitcoin hierarchical deterministic (HD) wallet is a digital wallet that allows the creation of child keys from the master private/public key in a hierarchical form. However, ***HD wallet can be easily exploited*** where an attacker can easily retrieve the master private key using the master public key and any child private key [51]. ***Hardware-based HD wallet*** [52] ***is also vulnerable to a number of attacks*** since this type of wallet does not use a secure communication channel between the API and the hardware such as smart card and microcontroller.	A new HD wallet has been proposed by [51] that can remove the vulnerability and retain the master key property. For any chose parameter *m,* the proposed HD wallet is able to endure the vulnerability of the HD wallet up to m private keys with a master public key size of $O(m)$. In [52], the authors provided a solution that consists of three components: 1. The secure pre-setup phase. 2. The authentication and session key establishment protocol. 3. Encryption of sensitive parts.

Blockchain Component	Threats and Vulnerabilities	Countermeasures
Transactions	***Double spending*** In general, double spending defined as spending money twice due to transaction being copied at time (t). The non-equivocation contract proposed in [53] can suffer collusion attack where the sender conspires with the deposit beneficiary to transfer the deposit back to the sender if he decides to equivocate and double spend.	• Recipient-oriented transaction [54] • The authors in [53] introduce a low level cryptographic algorithm called *accountable assertion* to create a non-equivocation contract in case double spending. In this contract, any sender that attempts to equivocate and double spend will be penalized using the time-locked Bitcoin deposit, which is created by the sender. • In [55], the authors modify the non-equivocation contract proposed in [53] a signature generated from the payee's secret key to the time-locked deposit. Thus, if the sender decides to double spend, he will be penalized by the losing his deposit and the payee receives a compensation from the sender's deposit. • A mechanism is proposed in [56] to discourage double spending attempts in Bitcoin zero-confirmation transactions. The proposed mechanism generates a special type of outputs that enforces the disclosure of the private key in case of a double spending attempt [56].
	Malleability attack - the unique ID of a Bitcoin transaction is changed before it is confirmed on the Bitcoin network. In principal, malleable occurs if its output C can be transformed ("mauled") to some "related" value C by someone who does not know the cryptographic secrets that were used to produce C.	Create a malleability-resilient "refund" transaction based on the Bitcoin-based timed commitment scheme protocol [57]. Adding the hash of the intermediate transactions to the current transaction id [58].

ISSN 2636-9680
eISSN 2682-9266

Blockchain Component	Threats and Vulnerabilities	Countermeasures
	There are a number of Bitcoin transactions' properties that can be used to examine the characteristics of the Bitcoin transactions and how the transactions are performed since Bitcoin utilizes an open database which can be viewed and checked by anyone [59]. In addition, there are certain methods that can be used to determine the behaviors of the Bitcoin owners and in certain cases, the Bitcoin addresses can be linked to the real identity of the users. Furthermore, Bitcoin transactions are vulnerable to both active and passive attacks since the transaction is publicly exposed to the Internet.	The authors in [59] proposed a protocol of anonymizing Bitcoin transactions that is compatible with the current Bitcoin main network system. The proposed protocol has the following characteristics [59]: • It protects the Bitcoin address of the payer from the payee. • It does not allow any participant to lean the whole information of the chained transactions by dividing the information into several parts. • It can be cancelled at any state without any participant losing money in an honest majority condition. In [60], the authors propose a new method for increasing the Bitcoin anonymity by using a new primitive known as composite signature. The proposed method removes any cryptographic evidence of transfer of funds and obscures the connection between inputs and outputs. In [61], the authors proposed a framework that incorporates homomorphic Paillier encryption system to cover the plaintext amounts in the transactions, while the encrypted amounts will be checked by the Commitment Proof.
Consensus Mechanism	***Pitchfork attack*** – the use of merged mining attack against the other branch of a fork in a permissionless PoW cryptocurrency [62].	Provide countermeasures – the targeted miners can fork away empty blocks or use their mining power to launch a counter attack on the attacker [62].
	51% attack - a single miner's or a group of miners' hashing power accounts for more than 50% of the total hashing power of the entire blockchain.	• A random mining group selection technique - gives mining opportunity to a randomly selected group [63]. • Giving incentives based on psychological factors using gamification for the approved mining work [64]. • Increase the Bitcoin confirmation depth [65].

Blockchain Component	Threats and Vulnerabilities	Countermeasures
	Crypto jacking or drive-by mining – a new web-based attack that uses people's devices (computer, smartphones, tablets and servers) to secretly mine cryptocurrencies without their consent or knowledge [66].	In [67], the authors proposed a detection approach called MineSweeper based on the cryptographic functions of the cryptojacking codes through static analysis and monitoring of CPU cache during run time.
	Selfish mining – it is an attack on the integrity of the Bitcoin network where a group of miners do not publish and distribute a valid solution to the rest of Bitcoin network to invalidate the honest miners work. The main idea behind the selfish mining strategy is to force the honest miners into performing wasted computations on blocks that are destined to not be part of the blockhain.	In [68], the authors propose a modification to the Bitcoin protocol that prohibits selfish mining by ensuring that mining pools smaller than ¼ of the total mining power cannot profitably engage selfish mining.
Cryptocurrency's Networking (for distributing the distributed ledger)	**P2P-layer anonymity** vulnerabilities that allow transactions to be linked to users' IP addresses with accuracies over 30% [69].	Dandelion++ is lightweight, scalable, that uses 4-regular anonymity graph that offers anonymity gains [69].
	Routing attacks – partitioning the Bitcoin network, slowing down the Bitcoin network [70].	Provide short term and long term countermeasures. Examples of short term measures include increase the diversity of the node connections and measure round trip time. Examples of long term measures include encrypt Bitcoin communication and use UDP connections [70].
	Bitcoin nodes with anomalous behavior patterns for illegal interests.	• A behavior pattern clustering algorithm to address the problem of clustering node behaviors in blockchain networks [71]. • In [72], the authors use specific transaction patterns to cluster nodes that are owned by the same entity. The proposed method converts the network properties into tables with attributes for more efficient data extraction from large Bitcoin network.
	DDoS attack is a common type of attack that occurs in many cyber platforms.	A decentralized protocol for anonymously finding partners and provides evidence of the agreement that can be leveraged if a party abort [73].

ISSN 2636-9680
eISSN 2682-9266

Blockchain Component	Threats and Vulnerabilities	Countermeasures
	Deanonymization attacks are the attacks that focused on unreachable Bitcoin nodes – Bitcoin nodes that are nodes that do not accept incoming connections and hidden behind NAT. The attacks depend on the nodes consecutive block-requests.	If the victim nodes request blocks in a non-consecutive manner, then it will not be possible for an adversary to estimate their Blockchain height and link sessions [74].
	To use the Bitcoin network to enable command and control communications for botnets.	• The use of Software Defined Networking (SDN) to assist in detecting malware-related anomalies at the network level [75] • Researchers and law enforcement should cultivate working relationships with registrars and ISPs to enable rapid response time to malware threats [75].

IV. CONCLUSION

This systematic review is intended to explore a fundamental view of cryptocurrency under the blockchain technology by addressing two main research questions. In this review, we examined 64 articles between the years 2014 and 2018 and categorized these publications based on the defined research questions. Furthermore, based on this systematic review, we identified research challenges. The main findings of this review are as follows:

RQ 1: The review shows that the blockchain is an emerging topic with common understanding of the blockchain definition. We also found that more than 50% out of 25 articles are more focus on the blockchain technology itself rather than relating the blockchain technology with cryptocurrency.

RQ2: We identified 17 security threats and vulnerabilities in the blockchain technology in cryptocurrency and categorized them based on the main components of the blockchain technology, which are asymmetric cryptography, transactions, proof-of-work, mining, and cryptocurrency's network. Out of 28 articles, only one publication provides a countermeasure on pitchfork attack on the blockchain proof-of-work. Furthermore, from our review, there are a number of attacks targeted on the cryptocurrency's networking such as routing attack, DDoS, and deanonymization attack. However, we found that there is only one publication addressing each of the attack.

Based on our review, we came to a conclusion that the blockchain technology is under imminent threat, especially in cryptocurrency. Despite its trustworthy architecture and the use of the cryptography, adversaries are still able to find vulnerabilities in this technology. From the findings in the systematic literature review, we also found that many researchers are experimenting with the cryptocurrency's vulnerabilities and threats but not many researchers provide countermeasures for the vulnerabilities and threats. To ensure that the blockchain technology is able to perform according to its proposed implementation, more countermeasures are needed to address the vulnerabilities and threats.

V. ACKNOWLEDGEMENT

This research was supported by CyberSecurity Malaysia. We thank our colleagues from CyberSecurity Malaysia who provided insight and expertise that greatly assisted the research. A high appreciation to Digital Forensics and Computer Networking (INSFORNET) research group under Center for Advanced Computing Technology (C-ACT); and Faculty of Information and Communication Technology (FTMK) the use of the existing facilities to complete this research.

VI. REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Bitcoin*, 2008.

[2] D. Kobler, M. Koch, and J. Seffinga, "The Blockchain (R)evolution - The Swiss Perspective," 2017.

[3] A. Kharpal, "Blockchain: What is it and how does it work?," *Trade.io*, 2018. [Online]. Available: https://www.cnbc.com/2018/06/18/blockchain-what-is-it-and-how-does-it-work.html. [Accessed: 29-May-2019].

[4] PwC, "Making sense of Bitcoin and blockchain: PwC," *February*, 2016. [Online]. Available: https://www.pwc.com/us/en/industries/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html. [Accessed: 29-May-2019].

[5] R. Houben and A. Snyers, "Cryptocurrencies and blockchain - Legal context and implications for financial crime, money laundering and tax evasion," no. July, p. 103, 2018.

[6] C. Kaminski, "Online peer-to-peer payment: PayPal primes the pump, Will Banks Fol," *N.C. Bank. Inst.*, vol. 1, no. 1, pp. 375–404, Apr. 2003.

[7] G. F. Hurlburt and I. Bojanova, "Bitcoin: Benefit or curse?," *IT Prof.*, vol. 16, no. 3, pp. 10–15, May 2014.

[8] A. Manimuthu, V. Raja Sreedharan, G. Rejikumar, and D. Marwaha, "A literature review on Bitcoin: Transformation of crypto currency into a global phenomenon," *IEEE Engineering Management Review*. 2019.

[9] Y. Yuan and F. Y. Wang, "Blockchain and cryptocurrencies: Model, techniques, and applications," *IEEE Trans. Syst. Man, Cybern. Syst.*, 2018.

[10] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, 2017.

[11] M. C. Kus Khalilov and A. Levi, "A survey on anonymity and privacy in Bitcoin-like digital cash systems," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 3, pp. 2543–2585, 2018.

[12] M. Conti, E. Sandeep Kumar, C. Lal, and S. Ruj, "A survey on security and privacy issues of Bitcoin," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 4, pp. 3416–3452, 2018.

[13] B. McRae, "Library guides: Systematic literature reviews for education: Different types of literature review," 2018.

[14] G. Natal, "LibGuides: Literature review: lit review types," 2016.

[15] S. Chapman, "What are cochrane reviews? - Evidently Cochrane," 2014. [Online]. Available: https://www.evidentlycochrane.net/what-are-cochrane-reviews/. [Accessed: 30-May-2019].

[16] J. P. T. Higgins, S. Green, and (editors), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. 2011.

[17] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering version 2.3," *Engineering*, vol. 45, no. 4ve, p. 1051, 2007.

[18] K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, "Five steps to conducting a systematic review," *Journal of the Royal Society of Medicine*. 2003.

[19] J. Adinolfi, "And 2016's best-performing commodity is ... Bitcoin? - MarketWatch," 2016. [Online]. Available: https://www.marketwatch.com/story/and-2016s-best-performing-commodity-is-bitcoin-2016-12-22. [Accessed: 02-Mar-2019].

[20] J. Adelstein and N.-K. Stucky, "Behind the biggest Bitcoin heist in history: inside

ISSN 2636-9680
eISSN 2682-9266

the implosion of Mt. Gox," *Dly. Beast*, pp. 1–5, 2016.

[21] Merriam Webster, "Merriam Webster," *Online Dictionary*. 2016.

[22] M. Iansiti and R. K. Lakhani, "The truth about blockchain," *Harvard Business Review*, 2017. [Online]. Available: https://hbr.org/2017/01/the-truth-about-blockchain. [Accessed: 01-Mar-2019].

[23] D. Patel, J. Bothra, and V. Patel, "Blockchain exhumed," in *ISEA Asia Security and Privacy Conference 2017, ISEASP 2017*, 2017.

[24] H. Sun, H. Mao, X. Bai, Z. Chen, K. Hu, and W. Yu, "Multi-blockchain model for central bank digital currency," in *Parallel and Distributed Computing, Applications and Technologies, PDCAT Proceedings*, 2018.

[25] R. Bhatia, P. Kumar, S. Bansal, and S. Rawat, "Blockchain -the technology of crypto currencies," in *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2018, pp. 372–377.

[26] S. Singh and N. Singh, "Blockchain: Future of financial and cyber security," in *Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016*, 2016.

[27] P. W. Chen, B. S. Jiang, and C. H. Wang, "Blockchain-based payment collection supervision system using pervasive Bitcoin digital wallet," in *International Conference on Wireless and Mobile Computing, Networking and Communications*, 2017.

[28] P. Urien, "Towards secure Bitcoin fast trading: Designing secure elements for digital currency," in *Proceedings of the 2017 3rd Conference on Mobile and Secure Services, MOBISECSERV 2017*, 2017.

[29] N. Chalaemwongwan and W. Kurutach, "State of the art and challenges facing consensus protocols on blockchain," in *International Conference on Information Networking*, 2018.

[30] I. Alqassem and D. Svetinovic, "Towards reference architecture for cryptocurrencies: Bitcoin architectural analysis," in *Proceedings - 2014 IEEE International Conference on Internet of Things, iThings 2014, 2014 IEEE International Conference on Green Computing and Communications, GreenCom 2014 and 2014 IEEE International Conference on Cyber-Physical-Social Computing, CPS 20*, 2014.

[31] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, 2017.

[32] Y. Xinyi, Z. Yi, and Y. He, "Technical characteristics and model of blockchain," in *2018 10th International Conference on Communication Software and Networks, ICCSN 2018*, 2018, pp. 562–566.

[33] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, and J. Wang, "Untangling blockchain: A data processing view of blockchain Systems," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 7, pp. 1366–1385, Jul. 2018.

[34] M. Milutinovic, W. He, H. Wu, and M. Kanwal, "Proof of luck: An efficient blockchain consensus protocol," *Proc. 1st Work. Syst. Softw. Trust. Exec. - SysTEX '16*, pp. 1–6, 2017.

[35] K. Brünnler, D. Flumini, and T. Studer, "A logic of blockchain updates," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[36] H. F. Ouattara, D. Ahmat, F. T. Ouédraogo, T. F. Bissyandé, and O. Sié, "Blockchain consensus protocols: Towards a review of practical constraints for implementation in developing countries," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 2018.

[37] M. R. Biktimirov, A. V. Domashev, P. A. Cherkashin, and A. Y. Shcherbakov, "Blockchain technology: Universal structure and requirements," *Autom. Doc. Math. Linguist.*, 2018.

[38] M. Swan, "Blockchain temporality: Smart contract time specifiability with blocktime," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence*

and Lecture Notes in Bioinformatics), 2016.

[39] S. Bhardwaj and M. Kaushik, "Blockchain—technology to drive the future," in *Smart Innovation, Systems and Technologies*, 2018, vol. 78, pp. 263–271.

[40] W. Yang, S. Garg, A. Raza, D. Herbert, and B. Kang, "Blockchain: Trends and future," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11016 LNAI, pp. 201–210.

[41] Q. Zhang, P. Novotny, S. Baset, D. Dillenberger, A. Barger, and Y. Manevich, "LedgerGuard: Improving blockchain ledger dependability," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10974 LNCS, pp. 251–258.

[42] Y. Kawase and S. Kasahara, "Transaction-confirmation time for Bitcoin: A queueing analytical approach to blockchain mechanism," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.

[43] G. Pîrlea and I. Sergey, "Mechanising blockchain consensus," 2017.

[44] E. Heilman, F. Baldimtsi, and S. Goldberg, "Blindly signed contracts: Anonymous on-blockchain and off-blockchain Bitcoin transactions," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[45] C. Boyd and C. Carr, "Fair client puzzles from the Bitcoin blockchain," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[46] R. Dennis, G. Owenson, and B. Aziz, "A temporal blockchain: A formal analysis," in *2016 International Conference on Collaboration Technologies and Systems (CTS)*, 2016, pp. 430–437.

[47] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and G. Das, "Everything you wanted to know about the blockchain: Its promise, components, processes, and problems," *IEEE Consum. Electron. Mag.*, vol. 7, no. 4, pp. 6–14, Jul. 2018.

[48] W. Yin, Q. Wen, W. Li, H. Zhang, and Z. Jin, "An anti-quantum transaction authentication approach in blockchain," *IEEE Access*, 2017.

[49] H. ur Rehman, U. A. Khan, M. Nazir, and K. Mustafa, "Strengthening the Bitcoin safety: a graded span based key partitioning mechanism," *Int. J. Inf. Technol.*, pp. 1–7, Oct. 2018.

[50] I. Giechaskiel, C. Cremers, and K. B. Rasmussen, "On Bitcoin security in the presence of broken cryptographic primitives," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[51] G. Gutoski and D. Stebila, "Hierarchical deterministic Bitcoin wallets that tolerate key leakage," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.

[52] A. Gkaniatsou, M. Arapinis, and A. Kiayias, "Low-level attacks in Bitcoin wallets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.

[53] T. Ruffing, A. Kate, and D. Schröder, "Liar, liar, coins on fire!," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*, 2015, pp. 219–230.

[54] H. Lee, M. Shin, K. S. Kim, Y. Kang, and J. Kim, "Recipient-oriented transaction for preventing double spending attacks in private blockchain," in *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON 2018*, 2018.

[55] X. Yu, M. T. Shiwen, Y. Li, and R. Deng Huijie, "Fair deposits against double-spending for Bitcoin transactions," in *2017 IEEE Conference on Dependable and Secure Computing*, 2017.

[56] C. Perez-Sola, S. Delgado-Segura, G. Navarro-Arribas, and J. Herrera-

ISSN 2636-9680
eISSN 2682-9266

Joancomartí, "Double-spending prevention for Bitcoin zero-confirmation transactions," *Int. J. Inf. Secur.*, pp. 1–13, Nov. 2018.

[57] M. Andrychowicz, S. Dziembowski, D. Malinowski, and Ł. Mazurek, "On the malleability of Bitcoin transaction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.

[58] U. Rajput, F. Abbas, R. Hussain, H. Eun, and H. Oh, "A simple yet efficient approach to combat transaction malleability in Bitcoin," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.

[59] D. A. Wijaya, J. K. Liu, R. Steinfeld, S. F. Sun, and X. Huang, "Anonymizing Bitcoin transaction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[60] A. Saxena, J. Misra, and A. Dhar, "Increasing anonymity in Bitcoin," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[61] Q. Wang, B. Qin, J. Hu, and F. Xiao, "Preserving transaction privacy in Bitcoin," *Futur. Gener. Comput. Syst.*, 2017.

[62] A. Judmayer, N. Stifter, P. Schindler, and E. Weippl, "Pitchforks in cryptocurrencies:," in *International Workshop on Cryptocurrencies and Blockchain Technology - CBT'18*, Barcelona, Catalonia.: Springer, Cham, 2018, pp. 197–206.

[63] J. Bae and H. Lim, "Random Mining Group Selection to Prevent 51% Attacks on Bitcoin," in *Proceedings - 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, DSN-W 2018*, 2018.

[64] Y. Kano and T. Nakajima, "A new approach to mining work in blockchain technologies," in *Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia - MoMM2017*, 2017, pp. 107–114.

[65] A. Fehnker and K. Chaudhary, "Twenty percent and a few days – Optimising a Bitcoin majority attack," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10811 LNCS, pp. 157–163.

[66] O. N. Toronto and C. Canada, "MineSweeper: An in-depth look into drive-by cryptocurrency mining and its defense," in *CCS18*, 2018.

[67] R. K. Konoth *et al.*, "MineSweeper," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security - CCS '18*, 2018, pp. 1714–1730.

[68] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[69] G. Fanti *et al.*, "Dandelion++: Lightweight cryptocurrency networking with formal anonymity guarantees," in *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems - SIGMETRICS '18*, 2018, pp. 5–7.

[70] M. Apostolaki, A. Zohar, and L. Vanbever, "Hijacking Bitcoin: Routing attacks on cryptocurrencies," in *Proceedings - IEEE Symposium on Security and Privacy*, 2017, pp. 375–392.

[71] B. Huang, Z. Liu, J. Chen, A. Liu, Q. Liu, and Q. He, "Behavior pattern clustering in blockchain networks," *Multimed. Tools Appl.*, 2017.

[72] T. H. Chang and D. Svetinovic, "Improving Bitcoin ownership identification using transaction patterns analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.

[73] G. Bissias, A. P. Ozisik, B. N. Levine, and M. Liberatore, "Sybil-resistant mixing for Bitcoin," in *Proceedings of the 13th Workshop on Privacy in the Electronic Society - WPES '14*, 2014, pp. 149–158.

[74] I. Deep Mastan and S. Paul, "A new approach to deanonymization of

unreachable Bitcoin nodes," pp. 277–298, Nov. 2018.

[75] S. T. Ali, P. McCorry, P. H. J. Lee, and F. Hao, "ZombieCoin 2.0: managing next-generation botnets using Bitcoin," *Int. J. Inf. Secur.*, vol. 17, no. 4, pp. 411–422, Aug. 2018.

ISSN 2636-9680
eISSN 2682-9266

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
19 – 29

Cloud Forensic Challenges and Recommendations: A Review

Warusia Yassin[1], Mohd Faizal Abdollah[2], Rabiah Ahmad[3], Zahri Yunos[4], and Aswami Ariffin[5]
[1,2,3]Centre for Advanced Computing Technology, Faculty of Information and Communications Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[4,5]CyberSecurity Malaysia, Cyberjaya, Malaysia
[1]**s.m.warusia@utem.edu.my**

ARTICLE INFO

Article History
Received 22 May 2019
Received in revised form 15 Aug 2019
Accepted 25 Sep 2019

Keywords:
cloud computing, forensic investigation, challenges, recommendation, forensic phases

ABSTRACT

Cloud computing becomes more popular since the emergence of the Fourth Industrial Revolution (IR 4.0) as almost all internet services are highly dependent on high-end networks of server computers. The large-scale used on the internet around the world may cause the cloud server to be highly exposed to cyber threats and it is very difficult to apply forensic method specifically in conducting cloud forensic investigation. Subsequently, the lack of digital investigation may increase the threats towards cloud environment. Consequently, the cloud forensic investigation needs to be recognized for any incident happened in cloud services. Thus, this paper will review the the challenges in conducting a forensic investigation on cloud computing and the challenges are described according to cloud forensic investigation phase, which are identification, collection, examination and analysis, and lastly reporting and presentation. Moreover, recommendation to overcome current cloud forensic challenges which were specified by previous researches also being provided. This review will be beneficial to the community in order to overcome the challenges of cloud forensic investigation in the future.

I. INTRODUCTION

Cloud is a technology that is no longer new, and the technology has already been used for various services. The continuous increase in the volume and detail of data captured by establishments such as Internet of Things (IoT), has produced an overwhelming flow of data whether the data are in a structured or unstructured format. However, many customers remain reluctant to move their business IT infrastructure completely to a cloud environment. This is because security is one of the main concerns of customers and unknown threat need to be considered. The issues in security are also related to the ability to perform digital investigations in cloud sector [1]. With the rising acceptance of cloud computing, the attacker is starting to target cloud services and the incident will probably increase in the future. Furthermore, an attacker might leak confidential information from a victim by abusing a cloud storage service that allows users to store documents and images and access them through endpoint devices such as smartphone [2].

Cloud computing technology provides demanding usage of computing resources with minimal effort of management and cloud service provider interaction [3]. The cloud service uses virtualized resources that can be accessed by common users without running out of resources [4].

ISSN 2636-9680
eISSN 2682-9266

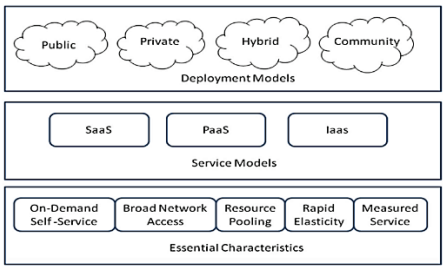

Fig.1: NIST Cloud Model [5]

The National Institute of Standards and Technology (NIST) defines cloud computing as a model with which to enable convenient, on-demand network access to a shared pool of configurable computing resources such as networks, servers, storage, applications, and services that can be rapidly provisioned and released with minimal management effort [4].

There are several types of cloud that are currently provided by the cloud service provider. A cloud infrastructure that is owned by a cloud service provider is called a public cloud. The service provider is responsible to manage the cloud while distributing and selling the cloud resources to other companies [6]. In a private cloud, the cloud infrastructure is for the exclusive use of one company only. Thus, the company owns the cloud and uses the resources. Thus, the company, or a contracted company, is responsible for maintaining the cloud [6]. A cloud infrastructure that is owned and used by several companies can be called a community cloud. This type of cloud service is managed by the organization or a third party [6]. Most hybrid clouds combine public cloud with private cloud. Although the hybrid cloud uses multiple types of clouds, each of the modules still functions separately [6].

TABLE 1: Types of Cloud

Author	Public	Private	Hybrid	Community
(Sharma, 2016) [7]	/	/	/	
(Park et al., 2018) [8]	/	/	/	/
(Ho et al., 2018) [30]	/	/		
(Delport, 2013) [6]	/	/	/	/
(Birk and Wegener, 2011) [9]	/	/	/	/
(Doran, 2014) [10]	/	/	/	
(Galvan, 2013) [4]	/	/	/	/
(Alex and Kishore, 2017) [5]	/	/	/	/

TABLE 1 shows that most authors mentioning and describe types of clouds that has been made by the Cloud Service Provider based on the customer needs. This shows that most common types of cloud will and probably become a target of the attacker with malicious intent to steal the data from Cloud Service Provider.

There are three types of cloud computing service models which are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [6].

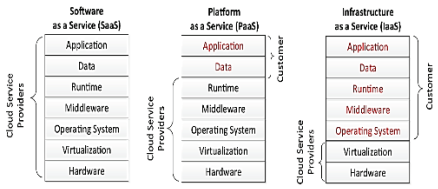

Fig. 2: Layers Architecture of Cloud Service [11]

In the Infrastructure as a Service (IaaS) model, the customer uses the virtual machine provided by the CSP for installing his own system on it. The system can be used like any other physical computer with a few limitations. However, the additive power over the system comes along with additional security obligations. Platform as a Service (PaaS) offerings provide the capability to deploy application packages created using the virtual development environment supported by the CSP. For the efficiency of Software Development Process this service model can be propellant. In the Software as a Service (SaaS) model, the customer makes use of a

service run by the CSP on a Cloud infrastructure. In most of the cases this service can be accessed through an API for a thin client interface such as a web browser.

II. CLOUD FORENSIC

Today, digital forensics has become more popular with law enforcement recognizes its function as to exploit criminal in cybercrime section. This also includes gathering the evidence, including digital devices such as smartphones, computer and smart sensors with can help police investigations. However, with the current tools that are sometimes not capable of analyzing the evidence because of compatibility issues, encryption or lack of training causing digital forensics to become inferior to be applied. Also, because of data management issues, most of the data evidence needs to be analyzed in a longer period of time that takes weeks to several months [12].

Although digital forensics has been established for several years, there is no specific or consistent methodology that can become a guide especially for cloud technology. With the increasing number of digital evidence that has been captured into laboratories, digital forensic methodology needs to be prioritized first in order to reduce the risk of evidence to be questioned during judicial proceedings [13,14].

Many authors have presented their understanding regarding cloud forensic by using the model, framework, layer or even process. However, all of these are included in the phases of cloud forensic investigation. Many authors have discussed about the phases and they are shown in **TABLE 2**.

TABLE 2: Number of Phases in Cloud Forensic

Author & Year	Research Title	Number of Phases
(Alex and Kishore, 2017) [5]	Forensics framework for cloud computing	4 Phases
(Martini and Choo, 2012) [14]	An integrated conceptual digital forensic framework for cloud computing, Digital Investigation	4 Phases
(Raju and Geethakumari, 2017) [15]	An advanced forensic readiness model for the cloud environment	4 phases
(Martini and Choo, 2013) [14]	Cloud storage forensics: OwnCloud as a case study	4 Phases
(Quick and Choo, 2013) [17]	Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata?	5 Phases
(Shah and Malik, 2014) [18]	An approach towards digital forensic framework for cloud	4 Phases
(Rani and Geethakumari, 2015) [19]	An efficient approach to forensic investigation in cloud using VM snapshots	4 Phases
(Martini and Choo, 2012) [14]	An integrated conceptual digital forensic framework for cloud computing	4 Phases
(Quick and Choo, 2014b) [20]	Google drive: Forensic analysis of data remnants	4 Phases
(Pichan et al., 2015) [21]	Cloud forensics: Technical challenges, solutions and comparative analysis	6 Phases
(Easwaramoorthy et al., 2016) [22]	Digital forensic evidence collection of cloud storage data for investigation	4 Phases
(Khan et al., 2016) [23]	A Comprehensive Review on Adaptability of Network Forensics Frameworks for Mobile Cloud Computing A Comprehensive Review on Adaptability of Network Forensics Frameworks for Mobile Cloud Computing	4 Phases
(Ahmed Khan and Ullah, 2017) [24]	A log aggregation forensic analysis framework for cloud computing environments	5 Phases
(Almulla et al., 2014) [25]	a State-of-the-Art Review of Cloud	6 Phases

ISSN 2636-9680
eISSN 2682-9266

Author & Year	Research Title	Number of Phases
(Delport et al., 2011) [26]	Isolating a cloud instance for a digital forensic investigation	7 Phases
(Damshenas et al., 2012) [27]	Forensics investigation challenges in cloud computing environments	4 Phases
(Birk and Wegener, 2011) [9]	Technical Issues of Forensic Investigations in Cloud Computing Environments	3 Phases
(Simou et al., 2016) [28]	A survey on cloud forensics challenges and solutions	4 Phases
(Horsman, 2018) [29]	Framework for Reliable Experimental Design (FRED): A research framework to ensure the dependable interpretation of digital data for digital forensics	3 Phases
(Ho et al., 2018) [30]	Following the breadcrumbs: Timestamp pattern identification for cloud forensics	3 Phases
(Quick and Choo, 2014a) [20]	Impacts of increasing volume of digital forensic data: A survey and future research challenges	10 Phases
(Zhao, 2017) [31]	Study and Realization of Digital Forensics Key Technology Based on Cloud Computing	5 Phases

Based on the table above, there are different numbers of phases proposed by the authors. The cloud forensic investigation starts with three phases and one of them proposed until 10 phases. Basically, the main phases in cloud forensics are identification, collection, examination, analysis and reporting. The majority of the authors proposed four or five phases in cloud forensics. However, some authors have separated the tasks inside the main phase to become another different phase such as [13,20,26,25]. Following the majority of the authors, the main phases for cloud forensic might be four main phases and the analysis of the phases is shown in **TABLE 3**. **TABLE 3** shows a comparative analysis cloud forensic layers based on the previous authors in this field. Based on **TABLE 3** which is a comparative analysis on previous cloud forensic layers above, we analyze every phase to choose the best phase of cloud forensic investigation. In phase 1, identification has been used for almost every authors. Identification means to identify the scope of action before conducting any cloud forensic investigation that identifies the key players and custodians as well as the best sources of potential electronic evidence that need to be accessed for collection. In Phase 2, collection of data is the most preferred phase after identification phase. Collection means collecting digital information that may be relevant to the investigation.

TABLE 3: Comparative Analysis on Previous Cloud Forensic Layer

Author	Title	Phase 1	Phase 2	Phase 3	Phase 4
(Alex and Kishore, 2017) [5]	Forensics framework for cloud computing	Identification	Collection	Organization	Presentation
(Martini and Choo, 2012) [14]	An integrated conceptual digital forensic framework for cloud computing, Digital Investigation	Identification & Preservation	Collection	Examination & Analysis	Reporting & Presentation

Author	Title	Phase 1	Phase 2	Phase 3	Phase 4
(Raju and Geethakumari, 2017) [15]	An advanced forensic readiness model for the cloud environment	Identification	Collection	Examination	Analysis & Presentation
(Martini and Choo, 2013) [16]	Cloud storage forensics: ownCloud as a case study	Identification & Preservation	Collection	Examination & Analysis	Reporting & Presentation
(Shah and Malik, 2014) [18]	An approach towards digital forensic framework for cloud	Identification	Data Extraction, Preservation & Collection	Analysis/ Examination	Presentation
(Rani and Geethakumari, 2015) [19]	An efficient approach to forensic investigation in cloud using VM snapshots	Identification	Collection	Examination/ Analysis	Reporting/ Presentation
(Quick and Choo, 2014b) [20]	Google drive: Forensic analysis of data remnants	Prepare	Identify & Collect	Preserve (Forensic Copy)	Analysis
(Easwaramoorthy et al., 2016) [22]	Digital forensic evidence collection of cloud storage data for investigation	Identification & Preservation	Collection	Examination & Analysis	Reporting & Presentation
(Khan et al., 2016) [23]	A Comprehensive Review on Adaptability of Network Forensics Frameworks for Mobile Cloud Computing A Comprehensive Review on Adaptability of Network Forensics Frameworks for Mobile Cloud Computing	Collection	Examination	Analysis	Reporting
(Damshenas et al., 2012) [27]	Forensics investigation challenges in cloud	Identification	Collection	Preservation	Reconstruction

ISSN 2636-9680
eISSN 2682-9266

Author	Title	Phase 1	Phase 2	Phase 3	Phase 4
	computing environments				
(Simou et al., 2016) [28]	A survey on cloud forensics challenges and solutions	Identification	(Collection) Preservation	(Analysis) Examination	Presentation
(Rani and Sravani, 2016) [3]	Challenges of digital forensics in cloud computing environment	Identification	Collection & Preservation	Examination & Analysis	Reporting & Presentation

It involves removing the electronic device from the crime or incident scene and then imaging, copying or printing out the content. In Phase 3, examination and analysis are two different tasks, but can be combined in the same phase as the processes have similar objectives. This phase involves a systematic search of evidence related to the incident being investigated. The outputs of the examination are data objects found in the collected information; they may include system- and user-generated files. Analysis aims to draw conclusions based on the evidence found. Lastly, the fourth phase is reporting and presentation phase. The reports are based on proven techniques and methodology and the other competent forensic examiners should be able to duplicate and reproduce the same results. The results are then presented either in the court or not in the presence of a judge and juries. In conclusion, it can simplify that the major phases of cloud forensic investigation based on the majority of authors are identification, collection, examination and analysis, as well as reporting and presentation.

III. CHALLENGES IN CLOUD FORENSIC

Several challenges have been identified in the first phase of investigation. [3] described that it is difficult to access evidence in the logs when investigating in a cloud computing environment since it has several factors that need to be done. Consequently, [3] give a solution through accessing the logs in the eucalyptus cloud environment will ease the investigator to access logs in a cloud environment. With the lack of control in cloud system and lack of customer awareness, the investigator will face difficulties in identifying which cloud that has been affected. Thus, synchronization of volatile data and third-party member need to supply the logging information to cloud service provider and cloud user [3]. [32] also explained jurisdictional issues which are important for investigator before doing an investigation since it relates to who has the jurisdiction to investigate an international incident in cloud system.

The collection phase is one of the crucial parts of the cloud forensic investigation because data evidence that has been collected need to be secure from tampering or any external factor. [3] explained one of the challenges are data integrity which is important in order to maintain the chain of custody. [3] proposed a solution called Trust Platform Module (TPM) which preserves the integrity and confidentiality of the data in the cloud and using trained and qualified personnel will maintain chain of custody. [32] explained that investigator also has a minimum control and access to client side which is one of the possible data evidences that need to be collected. [32] also give a solution of using remote and control log server will shorten the process of digital

forensic investigation. [18] also said physical seizure is difficult to obtain for data collection since cloud does not have physical server. By using static data acquisition via virtual snapshot technique with fuzzy clustering method, it can be used for determining whether the VM is under safe or unsafe mode [18].

The third phase of cloud forensic investigation, which is an examination and analysis focusing on the analysis of the data evidence and what sort of tools that need to be considered as the investigation is going through. However, the lack of specific tools for cloud forensic [3] and lack of tested and certified tools [33] make the investigation is difficult to conduct. [3] proposed OWADE (Offline Windows Analysis and Data) which an open source software specifically for cloud forensic tools. Encase and FTK software are also available which are commercial digital forensic tools [3].

Finally, the last phase of cloud forensic investigation, which is reporting and presentation. [14] addressed the issues of metadata and logs can be modified to remove the traces of unauthorized access and malicious activities. This issue has been given a solution by [14] which stressed on the importance of keeping the data secure and does not break the chain.

Besides previous cloud forensic challenges and their solutions, **TABLE 4** shows challenges of cloud forensic based on category. It focuses on three phases, which are Identification, Collection, besides Examination and Analysis with the authors recommend to be the best method to encounter the problem in the cloud forensic investigation.

TABLE 4: Challenges of Cloud Forensic based on Category

Phase	Author	Challenges (Category)	Description	Recommendation
Identification	(Hay et al., 2011) [34]	Physical location	Unknown location	CSPs must ensure the flexibility and availability of the sources reserved
	(Alhamad et al., 2010) [35]	SLA issue	Lack of formal SLA terms	Must have forensic request in SLA from CSPs
	(Ruan and Carthy, 2013) [36]	System level logs	Lack of information on logs	Should contain all information such as access, created and deletion of system logs.
	(Sang, 2013) [37]	Decentralize log	Issue of hypervisor level logs in forensic process	Must have framework
	(Ruan and Carthy, 2013) (Alhamad et al., 2010) (Pichan et al., 2015) [36,35,21]	SLA issue	Lack of SLA focus on forensic requirement	Should have SLA that contain flexibility and server availability and accessibility of the resource in CSPs
		Data issue	Data duplication	Must have unique identification
			Data encryption	Must have guideline or process for cloud investigation and legal activity

ISSN 2636-9680
eISSN 2682-9266

Phase	Author	Challenges (Category)	Description	Recommendation
Collection	(Liu et al., 2010) [[38]	Lack of trust	Issue of hypervisor platform, virtual environment and cloud platform	Should have proposed mechanism between hypervisor platform, virtual environment and cloud platform
	(Delport et al., 2011) [6]	Cloud infrastructure isolation issue	Vendor control isolation process	Need a standard isolation process which accepted by forensic manor
		Lack of specialized cloud forensic issue	Lack of commercialize on specific tools	The tools which accepted by the jurisdiction
Examination and analysis	(Dykstra and Sherman, 2012) (Zawoad and Hasan, 2013) [39,40]	Logging issue	Log from cloud	Logging framework
			Evidence log resources	Proper resources of log
	(Pichan et al., 2015) [21]	No encrypted data facility	Current technology has no encrypted data facility	Password and key management infrastructure
		Issue of acquisition log	More focus on hardware integration and evidence finding	Correlations of evidence

IV. PREVIOUS RESEARCH RECOMMENDATIONS

Normally, digital forensics require investigators to do data acquisition, especially live acquisition by seizing physical hardware such as servers, computers or smart devices. However, in the cloud, acquiring the data by seizing equipment might be impossible as the data are diverse and classified across multiple regions and multiple countries with different service models. Hence, the investigator needs to require another permission if the case involves another country which makes acquisition highly challenging. [41] proposed an approach using VM snapshots in a cloud environment. It consists of Intrusion Detection System into VMM to monitor and detect malicious activity between VMs. The process of the approach is CSP stores snapshots of a VM whose activities are identified as malicious by an intrusion detection system. CSP is then require to provide log files of the suspected VM for investigator to acquire the evidence.

Suspected VM also needs to be isolated so other uninvolved instances does not interfere with digital investigation process. [26] proposed seven isolation technique which are Instance Relocation, Server Farming, Failover, Address Relocation, Sandboxing, Man in the Middle (MITM) and Let's Hope for the Best (LHFTB). When doing live forensics analysis, preventing the instance of tampering with evidence is the highest priority for investigators. Also, instances must be protected from the of the external factor such as power outage if the investigator choose to do dead analysis.

[39] addressed technical and trust issues in cloud that are constantly challenging to

tackle when acquiring evidence from the cloud service model, especially Infrastructure-as-a-Service (IaaS). It provides a model layer of trust in the cloud layer, presenting cloud forensic examination and analyzing the available method for investigators. Also, it describes forensic tools which are currently available and know how to use it in each cloud layer.

Various threats such as data hijacking, data loss or leakage are more common in cloud computing thus, decreasing the trust of potential customers to invest their business into cloud computing. [42] proposed a solution name TrustCloud which is a framework for accountability and trust in Cloud Computing. It classifies the main component into four which are security, privacy, accountability and audibility. TrustCloud consists of three components in abstraction layer which are system layer, data layer and workflow layer. These layers have each their own different role and set of sub-components for each context that simplifies the problem and makes accountability more achievable.

Service Level Agreement or SLA is an agreement between the CSP and the client that describe service terms such as policies, performance, availability, billing and other important items. The reason SLA is important because actions can be taken in instances such violation or breach of contract involving either side. [35] explained factors or elements that need to be considered when designing an SLA in cloud computing. The paper proposed a method to maintain the trust and reliability between each party involved during the negotiation process after investigating the negotiation strategies between CSP and client.

Additionally, [14] proposed an integrated conceptual digital forensic framework, emphasizing the differences in the preservation of forensic data and the collection of cloud computing data for forensic purposes. The framework is based on NIST framework and it is considered as one of the most widely used and accepted in forensic frameworks.

V. CONCLUSION

Cloud forensic has been recognized by the previous researchers. From cloud forensic layers or process to a solution and recommendation has been proposed, but they are not conclusive for investigators to use as a guide. With the comparative analysis, previous solution and possible types of evidence that can be found in cloud environments, this review can contribute to becoming a guide for investigators in cloud forensic investigation. The solutions and recommendations that have been proposed by the previous researchers are important contribution which can assist investigators to solve the issues in each phase of forensic investigation.

VI. REFERENCES

[1] S.K.A. Manoj and D.L. Bhaskari, "Cloud forensics-A framework for investigating cyber attacks in cloud environment," *Procedia Computer Science*, 85 (Cms), pp.149–154, 2016.

[2] H. Chung, J. Park, S. Lee, and C. Kang, "Digital forensic investigation of cloud storage services," *Digital Investigation*, 9 (2), pp.81–95, 2012.

[3] D.R. Rani, and P.L. Sravani, "Challenges of digital forensics in cloud computing environment," *Indian Journal of Science and Technology*, 9 (17), 2016.

[4] M. Galvan, Cloud Computing : Incident response and digital forensics, A Capstone Project Submitted to the Faculty of Utica College December 2013.

[5] M.E. Alex, and R. Kishore, "Forensics framework for cloud computing," *Computers and Electrical Engineering*, 60, pp.193–205, 2017,

[6] W. Delport, Forensic evidence isolation in clouds, submitted to the Faculty of Engineering, Built Environment and Information Technology University of Pretoria, November 2013.

[7] S. Sharma, "Expanded cloud plumes hiding Big Data ecosystem," *Future*

ISSN 2636-9680
eISSN 2682-9266

Generation Computer Systems, 59, pp.63–92, 2016.

[8] S. Park, Y. Kim, G. Park, O. Na, and H. Chang, "Research on digital forensic readiness design in a cloud computing-based smart work environment," *Sustainability (Switzerland)*, 10 (4), pp.1–24, 2018.

[9] D. Birk, and C. Wegener, "Technical issues of forensic investigations in cloud computing environments," in *2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*, Oakland, CA, USA, 2011, pp.1–10,

[10] M.D. Doran, A forensic look at Bitcoin cryptocurrency, A Capstone Project Submitted to the Faculty of Utica College, in Partial Fulfillment of the Requirements for the Degree of Master of Science in Cybersecurity, 2014.

[11] V. Roussev, I. Ahmed, A. Barreto, S. McCulley, and V. Shanmughan, "Cloud forensics–Tool development studies & future outlook," *Digital Investigation*, Vol. 18, pp.79–95, 2016.

[12] S.L. Garfinkel, "Digital forensics research : The next 10 years", *Digital Investigation*, Vol 7, 2010, pp. 64-73.

[13] D. Quick, and K.K.R. Choo, "Google drive: Forensic analysis of data remnants," *Journal of Network and Computer Applications*, Vol. 40, pp.179–193, 2014a.

[14] B. Martini, and K.K.R. Choo, "An integrated conceptual digital forensic framework for cloud computing," *Digital Investigation*, 9 (2), pp.71–80, 2012,

[15] B.K.S.P.K. Raju and G. Geethakumari, "An advanced forensic readines. model for the cloud environment," in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, pp.765–771, 2017

[16] B. Martini and K.K.R. Choo, "Cloud storage forensics: OwnCloud as a case study," *Digital Investigation*, 10 (4), pp.287–299, 2013.

[17] D. Quick and K.K.R. Choo, "Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata?," *Digital Investigation*, 10 (3), pp.266–277, 2013.

[18] J.J. Shah, and L.G. Malik, "An approach towards digital forensic framework for cloud," in *2014 IEEE International Advance Computing Conference (IACC)*, 2014, pp.798–801.

[19] D.R. Rani and G. Geethakumari, "An efficient approach to forensic investigation in cloud using VM snapshots.\," in *2015 International Conference on Pervasive Computing: Advance Communication Technology and Application for Society, ICPC 2015*, 00 (c), 2015.

[20] D. Quick and K.K.R. Choo, "Impacts of increasing volume of digital forensic data: A survey and future research challenges," *Digital Investigation*, Vol. 11, Issue 4, pp.273–294, 2014b.

[21] A. Pichan, M. Lazarescu, and S.T. Soh, "Cloud forensics: Technical challenges, solutions and comparative analysis," *Digital Investigation*, 13, 2015, pp.38–57.

[22] S. Easwaramoorthy, S. Thamburasa, G. Samy, S.B. Bhushan and K. Aravind, "Digital forensic evidence collection of cloud storage data for investigation," in *2016 International Conference on Recent Trends in Information Technology, ICRTIT 2016*.,2016.

[23] S. Khan, M. Shiraz, A. W Abdul Wahab, A. Gani, Q. Han, Z. Abdul Rahman, "A comprehensive review on adaptability of network forensics frameworks for mobile cloud computing, *The Scientific World Journal*, 2014

[24] M.N. Ahmed Khan and S.W. Ullah, "A log aggregation forensic analysis framework for cloud computing environments," *Computer Fraud and Security*, Issue 7, July 2017, pp.11–16.

[25] S. Almulla, Y. Iraqi and A. Jones, "A State-of-the-art review of cloud," Journal of Digital Forensics, Security and Law (February 2015). 2014.

[26] W. Delport, M.S. Oliver and M.D. Kohn, "Isolating a cloud instance for a digital forensic investigation," in *Information Security for South Africa (ISSA2011) Conference*, (September), 2011, pp.145–153,

[27] M. Damshenas, A. Dehghantanha, R. Mahmoud and S. Bin Shamsuddin, "Forensics investigation challenges in cloud computing environments", in

Proceedings 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic, CyberSec 2012, 2012, pp.190–194.

[28] S. Simou, C. Kalloniatis, S. Gritzalis, and H. Mouratidis, "A survey on cloud forensics challenges and solutions," *Security and Communication Networks*, 9 (18), pp.6285–6314, 2016.

[29] G. Horsman, "Framework for Reliable Experimental Design (FRED): A research framework to ensure the dependable interpretation of digital data for digital forensics," *Computers and Security*, 73, pp.294–306, 2018.

[30] S.M. Ho, D. Kao and W.Y. Wu, "Following the breadcrumbs: Timestamp pattern identification for cloud forensics," *Digital Investigation*, 24, pp.79–94, 2018.

[31] B. Zhao, "Study and Realization of Digital Forensics Key Technology Based on Cloud Computing," Revista de la Facultad de Ingenieria, 32, pp.53–57, 2017.

[32] P.M. Trenwith, Digital Forensic Readiness in the Cloud, 2013.

[33] G. Grispos, T. Storer and W.B. Glisson, "Calm before the storm: the challenges of cloud computing in digital forensics," *International Journal of Digital Crime and Forensics*, 4 (2), pp.28–48, 2012.

[34] B. Hay, K. Nance and M. Bishop, "Storm clouds rising: Security challenges for IaaS cloud computing," *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2011, pp.1–7.

[35] M. Alhamad, T. Dillon, and E. Chang, "Conceptual SLA framework for cloud computing," in *4th IEEE International Conference on Digital Ecosystems and Technologies - Conference Proceedings of IEEE-DEST 2010, DEST 2010*, 2010 , pp.606–610.

[36] K. Ruan and J. Carthy, Cloud Computing Reference Architecture and Its Forensic Implications: A Preliminary Analysis, pp.1–21, 2013

[37] T. Sang, "A log-based approach to make digital forensics easier on cloud computing, " in *Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications, ISDEA 2013*, 2013, pp.91–94.

[38] D. Liu, J. Lee, J. Jang, and J. Zic, "A cloud architecture of virtual trusted platform modules", *2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, Hong Kong, 2010, pp. 804-811.

[39] J. Dykstra and A.T. Sherman, "Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques," *Digital Investigation*, 9 (SUPPL.), pp.S90–S98, 2012.

[40] S. Zawoad and R. Hasan, Cloud Forensics: A Meta-Study of Challenges, Approaches, and Open Problems, 2013.

[41] R. Poisel, E. Malzer, and S. Tjoa, Evidence and cloud computing : The virtual machine introspection approach. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 4 (1), pp.135–152, 2012.

[42] R.K.L. Ko, P. Jagadpramana, M. Mowbray, S. Pearson, M. Kirchberg, Q. Liang and B.S. Lee, "TrustCloud: A framework for accountability and trust in cloud computing," *Proceedings - 2011 IEEE World Congress on Services, SERVICES 2011*, 2011, pp.584–588.

ISSN 2636-9680
eISSN 2682-9266

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
31 – 42

Digital Certificate's Level of Assurance Development with Information Value and Sensitivity Measurement

Nikson Badua Putra[1], and Arry A. Arman[2]
[1] Government CSIRT, Badan Siber dan Sandi Negara, Jakarta, Indonesia
[2] Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung, Bandung, Indonesia
[1]**nikson.badua@bssn.go.id,** [2]**arry.arman@yahoo.com**

ARTICLE INFO

Article History
Received 28 Jul 2019
Received in revised form 15 Aug 2019
Accepted 25 Sep 2019

Keywords:
the level of assurance, digital certificates, information sensitivity, synthesize, AHP

ABSTRACT

This paper presents the research to develop the digital certificate's level of assurance. The level of assurance (LoA) in this paper is a level of assurance which reflects the authenticity degree of digital certificate's ownership. This LoA has a three-level, which define in four LoA standards such as ISO 29115: 2013, NIST SP 800-63-3, STORK, and KANTARA. From the previous researches and initial interview with digital certificate provider from Indonesia Government, this research concludes that information sensitivity measurement should be assessed to select the appropriate LoA. The related works and standards so far were not given any solution to this problem. This paper tries to solve it by offering LoA and its determination guidance model. This solution is achieved by synthesizing the four LoA along with information value and sensitivity measurement, which indicators determined by prioritization with the analytical hierarchy process (AHP). The proposed model-simulated and discussed so the information sensitivity measurement might assist in getting the suitable LoA level of the sensitive information been protected by a digital certificate.

I. INTRODUCTION

A Certificate Authority provides digital certificate services with a Level of Assurance for each level; one may have three levels as described below [NIST SP 800-63-3]:

- Level 3: High assurance of the authenticity and ownership of the digital certificate
- Level 2: Medium assurance of the authenticity and ownership of the digital certificate
- Level 1: Low assurance of the authenticity and ownership of the digital certificate

This level of assurance reflects authenticity degree of digital certificate's ownership that includes [3]:

- assurance degree in identity's checking or the entity that the certificate is issued; and
- assurance degree that the person uses the certificate is indeed the person that has the correct certificate. If the level is higher, so the degree of confidence in the ownership of electronic certificates to the owner is higher.

Because this level of assurance describes the degree of trust or the degree of confidence in an electronic system uses electronic certificates as the identity of a legal entity accordingly by using electronic certificates. The provider of electronic certificates directly gives the guarantee of trust and confidence, so this study does not discuss the trust and confidence of the

ISSN 2636-9680
eISSN 2682-9266

digital certificate being assured or not assured.

The selection of this level of assurance should base on risk assessment of digital transactions or services in the authentication of the entity. By mapping the impact level to the level of assurance, the entity of the digital transactions or services may determine the level of assurance that they needed, can use the digital services or do the electronic transactions, and ensure their identity's safety. An example of the assessment of the level of impact can be seen in **TABLE 1** [1].

TABLE 1: Potential impact at each level of assurance [1]

Potential impact	Level of Assurance		
	1	2	3
Impact on standing, reputation, status	Low	Med	High
Money loss or agency accountability	Low	Med	High
Impact on organization capability or asset, and public concern	N/A	Low/Med	High
Illegal sensitive information disclosure	N/A	Low/Med	High
Personal physical damage	N/A	Low	Med High
Law and regulation violations	N/A	Low/Med	High

As concluded in NIST SP 800-30 Guide for Conducting Risk Assessments [2], for each risk that analyzed, the appropriate control is needed to be able to respond to the risk. Digital signatures are present as one of the cryptographic controls that can be applied against the security risk in information systems as described in ISO/IEC 27001:2013 Information Security Management Systems.

Previous research, as in [3], recommends that Guideline in determining the Level of Assurance should be there to build the certificate policy with its description in the Introduction section in the certificate policy. This recommendation is because of the LoA impacts, ie, 1) mechanism of registration verification; 2) the protection of crypto-key; and 3) certificate management and protection in Certificate Authority. Another research [4], investigated the previous efforts in defining the Level of Assurance. From their investigation, they found that the LoA may assist in getting the proper access control of the sensitive resources, for example is the information. From this research, the results find that the Level of Assurance which measures the information's sensitivity is required. Another study [5] has seen the differences of ISO/IEC, NIST, STORK, and KANTARA approaches with their historical order linked to each other, their summary of the four LoA then may be used later in the synthesis of these four LoA approaches.

However, the selection criteria of digital certificate's assurance level by considering the sensitivity of the information or data that transmitted secured by a digital certificate has not found. Therefore, this study will carry out the analysis and design of these required criteria. One research [6] found to apply the information's sensitivity measurement to help determine the Level of Assurance. This research provides the quantitative classification of information method by calculating the information value and information sensitivity.

This research carried out by analysis and synthesis on the four LoA standards, weighing the LoA criteria indicator with Analytic Hierarchy Process (AHP), and then simulate and discuss the results of the proposed LoA selection criteria.

II. LOA DEVELOPMENT

As described in the previous section that improved LoA is developed by analyzing and synthesizing the current four LoA standards with a gap analysis of the indicators and each level characteristics, later then the indicator's weighting is analyzed with AHP and combine it with information's sensitivity measurement to achieve the design of the improved LoA.

A. Analyze and synthesis of ISO 29115:2013, NIST SP 800-63-3, STORK, KANTARA, Information's Classification

The synthesis

In this subsection, the synthesis carried out on the four LoA and information classification indicators based on [6], comparing the linkage of impact characteristics and information classification's characteristics which describes in **TABLE 2-8**. A two-way arrow illustrates a powerful link between NIST, ISO, and KANTARA which can be used to construct the criteria of the indicator. While the straight-line sign on the indicator illustrates that the two do not have interrelations on the indicator.

TABLE 2: LOA Indicator Characteristics Synthesis

Indicator Characteristics	Level of Assurance			
	NIST	ISO, KANTARA	STORK	Info Class
1. Impact on standing, reputation, status	↔	↔	—	—
2. Money loss or agency accountability	↔	↔	—	—
3. Impact on organization capability or asset, and public concern	↔	↔	—	—
4. Illegal sensitive information disclosure	↔	↔	—	—
5. Personal physical damage	↔	↔	—	—
6. Law and regulation violations	↔	↔	—	—
7. Access to private data consideration	↔	—	—	—
8. Federated systems consideration	↔	—	—	—
9. Information's reliability	—	—	—	↔
10. Information's increment	—	—	—	↔
11. Information's timeliness	—	—	—	↔
12. Information's availability	—	—	—	↔
13. Information's users' ability	—	—	—	↔
14. Opportunity costs	—	—	—	↔
15. Degree of dependence	—	—	—	↔
16. Regeneration costs	—	—	—	↔
17. Regeneration time	—	—	—	↔

The table above illustrates the synthesis process for determining the LoA indicators. The 2-way arrow represents a deep connection, while the straight lines represent the lack of linkage. For example, the two arrows on the indicator "impact on standing, reputation, status" in the column of NIST, ISO, KANTARA illustrates that this indicator has a significant linkage on the NIST, ISO, and KANTARA. The straight line on the same indicator in the STORK, and the Info Class column describes the lack of linkages between STORK and Class Info on this indicator.

ISSN 2636-9680
eISSN 2682-9266

From the table, NIST has the most accurate indicator, besides that ISO has the same indicator as NIST. Information classification indicator doesn't match four LoA, so the result is the information classification indicator along with impact indicator from four LoA to develop the improved LoA based on this synthesis.

Next step is analyzing the characteristic of each LoA level as described below to determine the number of the LoA and the detailed characteristics.

TABLE 3: LEVEL 1 Characteristics Synthesis

Characteristics	ISO, KANTARA	NIST 800-63-3	STORK
Usage	◄———	———►	———
Assurance on the identity which is claimed or asserted	◄———	———►	———
Authentication method	———	◄———	———
Credential strength	———	◄———	———
Authentication protocol	———	◄———	———
Attack that prevented	———	———	———
Security of the credentials	———	◄———	———

From the synthesis above, for usage and assurance characteristics, ISO; KANTARA; and NIST have the strong bond to build the Level 1 LoA, while on four other characteristics only NIST that have the details. And for characteristic, which is the attack that desired to prevent, all of the standards do not have the details.

TABLE 4: LEVEL 2 Characteristics Synthesis

Characteristics	ISO, KANTARA	NIST 800-63-3	STORK
Usage for	◄———	———►	———
Assurance on the identity which is claimed or asserted	◄———	———►	———
Authentication method	◄———	———►	———
Credential strength	———	◄———	———►
Authentication protocol	◄———	———►	———
Attack that prevented	◄———	———►	———
Security of the credentials	◄———	———►	———

From the synthesis of level 2, the characteristic of usage and assurance from NIST, ISO, and KANTARA obtained, which have a strong bond with each other. The authentication method and authentication protocol characteristic obtained from all of the four standards. Credential strength characteristic from NIST and STORK, attack characteristics from ISO, KANTARA, and NIST, the security of the credentials characteristic from ISO, KANTARA, and NIST.

TABLE 5: LEVEL 3 Characteristics Synthesis

Characteristics	ISO, KANTARA	NIST 800-63-3	STORK
Usage for	◄———	———	———►
Assurance on the identity which is claimed or asserted	◄———	———	———►
Authentication method	◄———	———	———►
Credential strength	———	◄———	———►
Authentication protocol	◄———	———	———
Attack that prevented	———	———	◄———
Security of the credentials	———	◄———	———

All of the standards have a strong bond in usage, assurance, and authentication method characteristics, but only NIST that has high impact mitigation, while the others just substantial impact mitigation usage. Besides that, only NIST have difference details about assurance, which is a very high measurement, and authentication method which uses two separate authentication factor. In credential strength characteristic, NIST and STORK have a strong bond, but only NIST have a requirement which is hardware-based authenticator which resistant to verification forgery. And last for level 3 analysis, only NIST has a strong bond with the security of the credentials characteristics.

TABLE 6: LEVEL 4 Characteristics Synthesis

Characteristic	*ISO, KANTARA*	*STORK*	*NIST 800-63-3*
Usage for	◄———	———►	———
Assurance on the identity which is claimed or asserted	◄———	———►	———
Authentication method	◄———	———►	———
Credential strength	◄———	———►	———
Authentication protocol	◄———	———►	———
Attack that prevented	◄———	———►	———
Security of the credentials	◄———	———►	———

To get the characteristic of level 4, only the NIST standard which not has bonded with all of the characteristic, because NIST SP 800-63-3 doesn't have level 4.

Results of the synthesis

The result of the synthesis achieved by analysis in the previous subsection describes below. First, define the low, medium, and high levels of indicators.

ISSN 2636-9680
eISSN 2682-9266

TABLE 7: Synthesis Results of Indicator Characteristics

Indicator Characteristic	*Low*	*Medium*	*High*
Impact on standing, reputation, status	Impact on standing, reputation, status occur in short or limited term.	Impact on standing, reputation, status occur in short or long limited term.	Impact on standing, reputation, status occur in the long term severe and affect many parties.
Money loss or agency accountability	Financial loss or agency accountability, not significant/ not so serious.	Financial loss or agency accountability is significant/ serious.	Financial loss or agency accountability is severe at a catastrophic level.
Impact on organization capability or asset, and public concern	Declining organizational capabilities, increasing the time spent by the organization to perform tasks and functions at recognizable levels, low damage to organizational assets or the public interest.	Declining organizational capabilities, increasing the time spent by the organization to complete tasks and services, and damage to organizational assets or public interest at a significant level.	Declining organizational capabilities, increasing the time spent by the organization to perform tasks and functions, and damage to organizational assets or public interest at a very high level.
Illegal sensitive information disclosure	Disclosure of confidential personal/government/trade information on low-level secrecy impacts, in short, and limited scope and time.	Disclosure of sensitive personal/government/trade information on medium-level secrecy impacts.	Disclosure of sensitive personal/government/trade information on high-level secrecy impacts.
Personal physical damage	Injuries at a mild level, and do not require medical action.	Injuries at a mild level, and need medical action.	Severe injury or death, caused by a service access error.
Law and regulation violations	A violation caused by service access error with the possibility of not being subject to law enforcement.	A violation caused by service access error with the possibility of being subject to law enforcement such as KUHP, KUHAP, UU ITE.	A violation caused by service access error which is of particular interest in law enforcement such as corruption, terrorism, drugs.
Private data access	It doesn't need private information to do the transaction.	Information System Services makes personal data accessible.	
Information's reliability	The information does not describe the actual situation, and information may be obscured or misdirected, information may lose the minimum requirements, and may lose the content of the information.	The information describes some of the actual situations; information's correctness may be obscured or partially deviated; information may lose the minimum requirements/part of the main contents.	The information describes the actual situation; information's correctness should not be obscured or deviated; the information should not lose the minimum requirements/part of the main contents at all.
Information's increment	Information system service users do not obtain new information or benefits from information system services.	Information system users gain some new information or benefits from information systems services.	Users of information systems get a lot of new information or benefits from information systems services.
Information's timeliness	Information in information systems services doesn't have expired time.	Information in information systems services has relatively long expired time.	Information in information systems services has relatively short expired time.

Indicator Characteristic	*Low*	*Medium*	*High*
Information's availability	No difficulty is found/doesn't need to pay attention to the difficulty in creating, processing, storing, and entering information in information systems service.	Find some difficulty in creating, processing, storing, and entering information in information systems service.	Find many difficulties in creating, processing, storing, and entering information in information systems service.
Information's users' ability	Information system service users do not easily understand and use information obtained from information system service.	Information system service users can understand and use some of the information obtained from information system service.	Information system service users easily understand and use all of the information obtained from information system service.
Opportunity costs	There is no or very little interest in any particular party in using the information on the Information System Service.	There are several interests of certain parties in using the information on Information Systems Service.	There are many interests of certain parties in using the information on Information Systems Service.
Degree of dependence	Loss of access control of the information systems services does not or slightly affect/harm the owner of the information.	Loss of access control of the information systems services affects/harm the owner of the information on the middle level.	Loss of access control of the information systems services affects/harms the owner of the information on the high or entire level.
Regeneration costs	Only a small amount of cost is required to correct any missing, damaged, or leaked information in the Information System service.	Only a medium amount of cost to correct any missing, damaged or leaked information in the Information System service.	Needed many costs to correct any missing, damaged or leaked information in the Information System service.
Regeneration time	It doesn't need time to correct any missing, damaged or leaked information in the Information System service.	Needed a short time to correct any missing, damaged, or leaked information in the Information System service.	Needed a long time to correct any missing, damaged, or leaked information in the Information System service.

Next, the synthesis results of LoA level characteristics illustrate in **TABLE 8**.

TABLE 8: Level Characteristics Determination

Level Characteristic	*Level 1*	*Level 2*	*Level 3*
Usage	Minimum risk authentication errors, low impact identity abuse, information sensitivity has little/no value.	Medium-risk authentication errors, medium-impact identity abuse, information sensitivity has moderate value.	High-risk authentication errors, identity abuse, have a substantial impact; information sensitivity has a high value.
Assurance on the identity which is claimed or asserted	The claimant controls an authenticator registered to the subscriber	Verify identity which is claimed with identity in the real world, whether remote or physical presence.	High, verification of identities recognized by the government needs a physical presence.
Authentication method	*Single-factor authentication.*	Ownership proof and separate multi-factor authentication controls.	

ISSN 2636-9680
eISSN 2682-9266

Level Characteristic	*Level 1*	*Level 2*	*Level 3*
Credential strength	Secured only with non-cryptographic general authentication.	Crypto, authentication of key ownership secures the credential is secured.	Only hardware-based authenticator resistant to verifier forgery is allowed.
Authentication protocol	*Secure authentication protocol.*	*Secure authentication protocol.*	*Secure authentication protocol* cryptographic.
Attack that prevented	There are no specific requirements; minimal guarantee.	*Eavesdropping*, *online guessing attack.*	Protection focus on counterfeiting verifier forgery and MITM attacks.
Security of the credentials	No crypto method.	Encrypted using approved cryptography so that only RP (relying party) can decrypt it.	Cryptographically protected, the identity claim controller controls tied to the subscriber account.

B. Prioritization the LoA indicator with analytical hierarchy process

AHP is a measurement theory using pairwise comparisons which depends on expert judgment to obtain priority scales [7]. In this study, the AHP approach is carried out as a fair comparison on each LoA indicator which has synthesized, which aim that each indicator has weight, to develop the improved LoA. Furthermore, the analysis of flow chart determination of the assurance level uses this weight, which indicators are prioritized to be calculated compared to other indicators. This comparison is a person's subjective assessment of criteria based on several considerations. The AHP model uses individual perception, which is considered "expert" as the primary input. The "expert" criteria here refers to the individual who understands the problems correctly, feels the consequences of a problem or has an interest in the issue [8]. According to [6], the weights of utility and cost factors in information value measurement along with the comparison of pairs of sensitivity factors of the information can be obtained by asking expert opinions and then using the AHP method to calculate them. In this study, the assessment or weighting of AHP is carried out by the developer of the information systems which are secured by the digital certificate.

Some indicator variables to determine LoA are compared using pairwise comparisons matrices; that is, the comparison matrix contains the preference level of several alternatives for each criterion. The scale of preference used is a scale of 1 which shows the lowest importance level up to a range of 9 which shows the extreme importance. The comparison stage in AHP begins by comparing each alternative per criterion. The results of the AHP analysis for selecting the order of the flowchart improved LoA illustrates in **Fig. 1**.

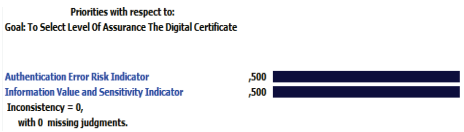

Fig. 1: AHP of Authentication and Information Sensitivity Indicator.

From **Fig. 1**, the authentication error indicator, with information value and sensitivity indicator, has equal relative importance, whose goal is to determine the LoA. The next step, selecting the priority order between six indicators in the authentication error risk indicator using AHP analysis as describes in **Fig. 2**.

ISSN 2636-9680
eISSN 2682-9266

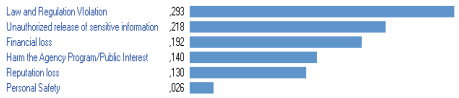

Fig. 2: AHP of Authentication Error Risk Sub Indicator

Concluded from **Fig. 2** that the relative importance between authentication error risk sub-indicator:

1. Law and regulation violation
2. Unauthorized release of sensitive information
3. Financial loss
4. Harm the agency program/public interest
5. Reputation loss
6. Personal safety

Finally, the final relative importance between all indicator orders from the top importance shows in Fig 3.

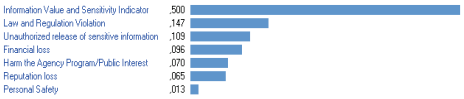

Fig. 3: AHP of All Indicator and Sub Indicator

C. Design of the improved LoA determination guidance

The design of the improved LoA determination guidance achieved by designing the flowchart matrix from the prioritization indicator resulted in the previous subsection. This flowchart, motivated by the measurement of information sensitivity needs to determine the desired LoA. The designed diagram illustrates in **Fig. 4**.

ISSN 2636-9680
eISSN 2682-9266

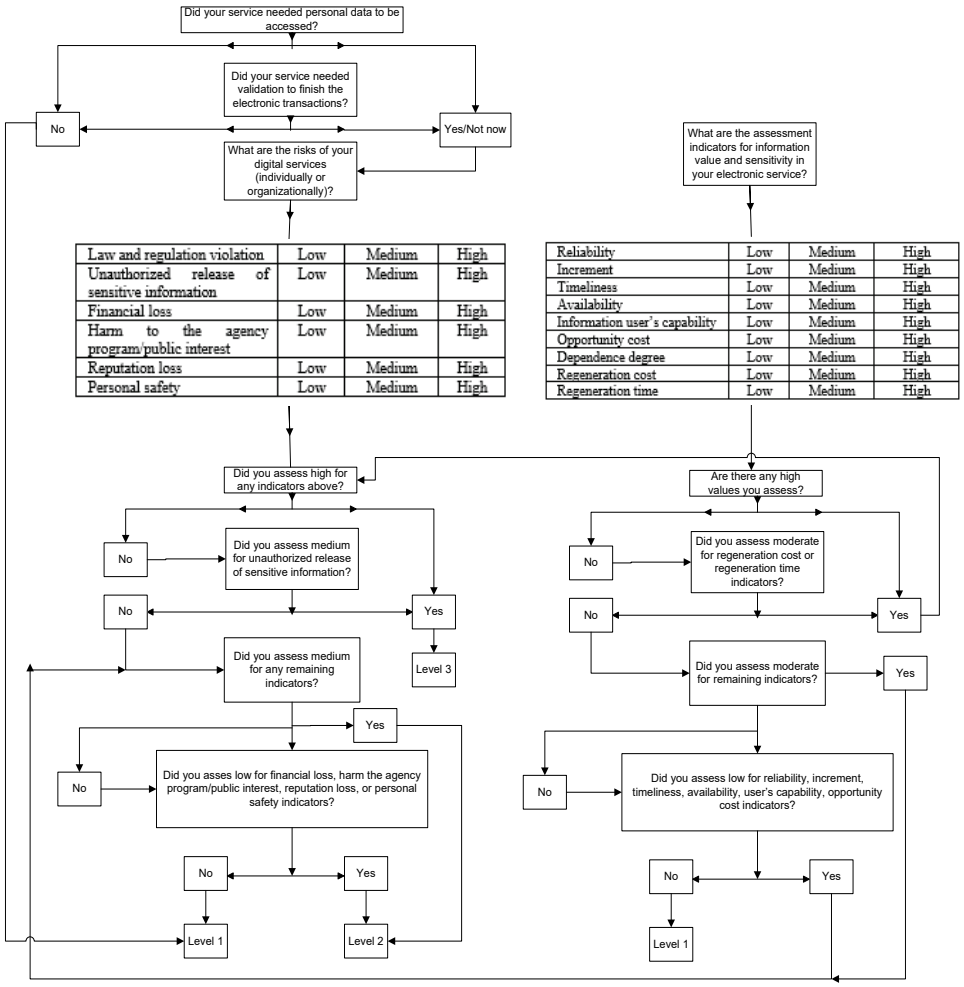

Fig. 4: Improved LoA Determination Flowchart

III. SIMULATION AND DISCUSSION

In this section, the same expert in AHP analysis simulating the improved LoA determination flowchart model. The simulation gives Level 3 LoA. The result of this simulation because the expert chooses the high rate for information indicators on the right side, and authentication error risk indicators on the left side, especially for dependence degree which reflects the high sensitivity of information, and high rate for financial loss indicator which indicates the agency accountability impact because the authentication risk error. The expert's digital service provides electronic processing of disbursement requests of the supported fund's cost for the infrastructure project from the executor bank. The digital signature implementation for these services is desired to assure the service's access control and the authenticity transmitted data. More advanced, this service very dependent on access account and its credential information, which to control the information and data being processed in the service, because these data and information have a high sensitivity which is private data of the debtors.

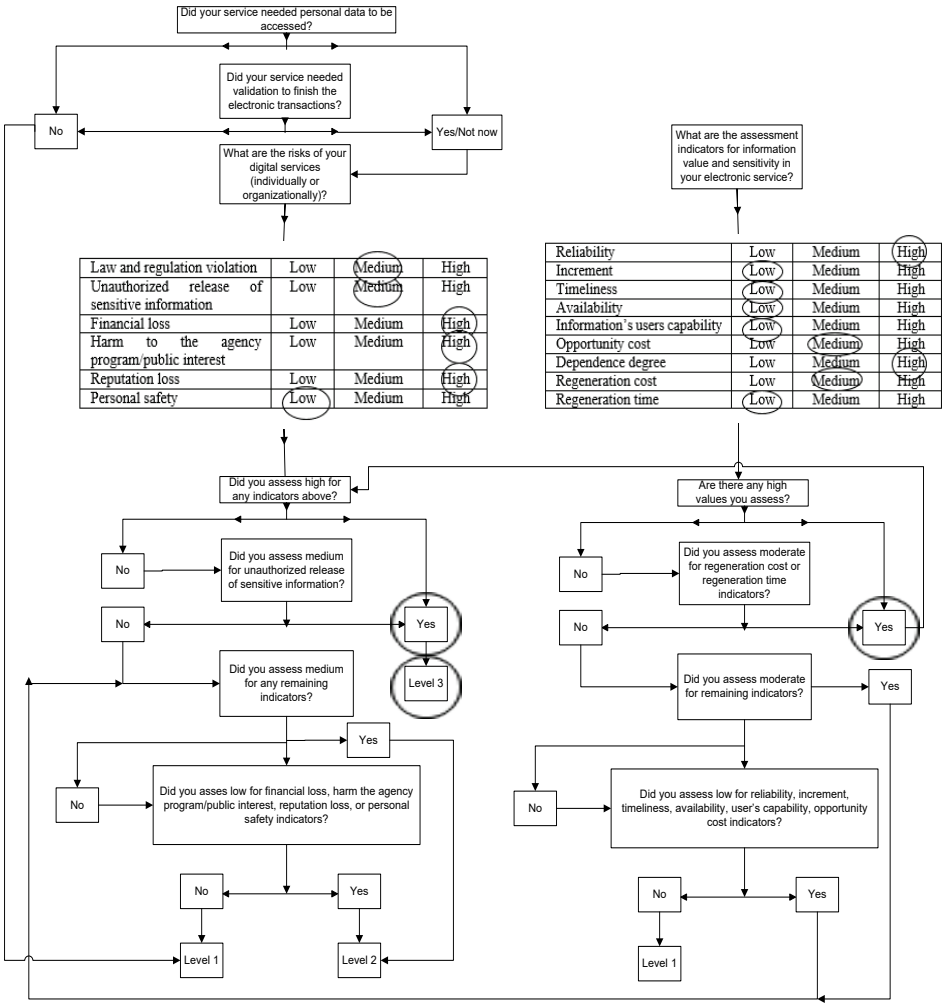

Fig. 5: Simulation Improved LoA Determination Flowchart

IV. CONCLUSION

This research proposes a model of Digital Certificate LoA and its determination guidance that improved by the synthesis of four current LoA standards and information value and sensitivity measurement. Concluded from the simulation and discussion with the expert of a digital service that being protected by a digital signature, this proposed model may assist in getting the suitable LoA level of the sensitive information being protected by a digital certificate which reflects in information value and sensitivity measurement assessed before the impact of authentication error risk.

V. REFERENCES

[1] M. E. Garcia and J. L. Fenton, "Digital identity guidelines."

[2] Joint Task Force Transformation Initiative, "Guide for conducting risk assessments," no. September, 2012.

[3] M. Endhy Aziz, "Certificate policy analysis and formulation of the

ISSN 2636-9680
eISSN 2682-9266

government public key infrastructure using SSM," Universitas Indonesia, 2016.

[4] A. Nenadic, N. Zhang, L. Yao, and T. Morrow, "Levels of authentication assurance : An investigation," pp. 155–158, 2007.

[5] T. Born and M. Peyrard, "Levels of Assurance," pp. 1–16.

[6] I. Sensitivity, "Supply chain information classification," 2007.

[7] T. L. Saaty, "Decision making with the analytic hierarchy process," vol. 1, no. 1, 2008.

[8] E. Helmud and S. Informasi, "Pemilihan paket internet android pada operator telepon gsm menggunakan metode analytical hierarchy process (AHP)," vol. 8, no. 1, pp. 918–927, 2016.

OIC-CERT Computer Emergency Response Team

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
43 – 55

Identity-Division Multiplexing Technique for Enhancing Privacy of Paging Procedure in LTE

Abdulrahman Muthana[1], and Abdulraqeb Al-Samei[2]
[1,2]Smart Security Solutions, Sana'a, Yemen
[1]**ab.muthana@smartsecurity-y.com**, [2]**abdu.alsamee@gmail.com**

ARTICLE INFO

Article History
Received 5 Jul 2019
Received in revised form 15 Aug 2019
Accepted 15 Aug 2019

Keywords:
LTE (Long Term Evolution), TMSI (Temporary Mobile Subscriber Identifier), identity-division multiplexing, linkability, traceability, paging

ABSTRACT

Despite efforts have been made by Long Term Evolution (LTE) toward enhancing privacy preserving capabilities, LTE is still vulnerable to privacy attacks. This paper evaluates the privacy issues of paging procedure in LTE and suggests a solution for enhancing the privacy of paging procedure in LTE. The solution introduces the Identity-Division Multiplexing (IDM) technique, in which the total sequence of temporary mobile subscriber identifiers M-TMSIs (2^{32} unique M-TMSI identities) is divided into a series of overlapping M-TMSIs ranges, each of which is allocated to one or more user equipment (UE). The solution guarantees the use of frequently changing unrelated TMSIs for identification; and thus, providing unlinkability and untraceability of the user. The solution provides an effective identity management that protects privacy of LTE users during paging process. The solution is formally verified using proVerif and proved to protect user privacy adequately.

I. INTRODUCTION

Long Term Evolution (LTE) cellular network technology [1] enhances the security of its predecessors: Global System Mobile Communication (GSM) and Universal Mobile Telecommunications System (UMTS) and offers a range of new security features.

LTE protects user identity privacy by allocating each user equipment (UE) various different temporary identities such as Global User Temporary Identifier (GUTI), temporary mobile subscriber identifier (TMSI), and cell radio network temporary identifier (C-RNTI) at different levels of LTE network architecture for different services. The UE can use these temporary identities instead of the International Mobile Subscriber Identifier (IMSI) to identity itself. This strategy reduces IMSI exposure risk and mitigates user identity privacy attacks.

Despite this strategy, LTE is still vulnerable to privacy attacks [2-6]. Temporary identifiers remain unchanged for amount of time sufficient for hacker to track the user and are transmitted in clear. For example, TMSI will not be changed within certain tracking area and that the paging messages are not encrypted [3].

This paper evaluates the privacy issues of paging procedure in LTE and also presents a solution for enhancing the paging privacy. The solution provides a high level of user unlinkability and anonymity within LTE cellular networks by using Identity-Division Multiplexing (IDM) technique, in which the total M-TMSIs sequence (2^{32} unique M-TMSI values) is divided into a series of overlapping ranges, each of which is allocated to one user equipment (UE). The solution guarantees the use of frequently changing unrelated temporary mobile subscriber identifiers (TMSI) for identification; and thus, providing unlinkability and untraceability of the user.

ISSN 2636-9680
eISSN 2682-9266

The proposed solution preserves privacy of user identity during paging procedure with minimal modifications at network architecture. The design strategy of the proposed solution aims at keeping LTE messaging system away as much as possible from the changes and modifications. We believe that this solution could be easily fit in current LTE cellular network architecture.

The main contribution of this paper is proposing a security solution that substantially enhances LTE capabilities in preserving user identity privacy during paging procedure. The privacy is preserved with minimal modifications on the network entities (i.e., Mobile Management Entity MME and UE) and with no modifications in the message system. The second contribution is an extensive theoretical study on privacy of paging procedure in LTE.

The rest of this paper is organized as follows: Section 2 reviews the related work and Section 3 describes privacy issues of paging procedure in LTE. Section 4 presents the proposed solution, Section 5 analyzes its security, and Section 6 concludes.

II. RELATED WORK

Many research works have discussed privacy in LTE and suggested solutions for protecting privacy in LTE. A number of researches have focused on privacy of user identity in LTE [2-13, 16, 19]. Paging and location privacy have also been highlighted in [3- 6, 15, 17,18, 20,21]. In this paper, we restrict ourselves to the closest related works (i.e., the research works addressing paging privacy issue in LTE).

Several researches investigate security issues of paging procedure in LTE. The research work [9] proposed encrypting the paging request using a shared session key, called as unlinkability key. The key is maintained for privacy preserving purpose only and is generated by applying a new one-way keyed function f to the long-term shared key KIMSI, and a random number rand included in the paging request. Furthermore, the solution requires that the encrypted request message should include a sequence number SQN and a random challenge chall.

The network stores the random challenge and checks it against the one received from the UE in the paging response. The aim of the sequence number SQN is to ensure freshness of the paging request and avoid replay attacks. A UE that receives a legitimate IMSI paging request should discard the request if the SQN is not in the correct range. The use of this procedure should still be kept minimal to avoid burdening the signaling communication with cryptographic operations. Each UE must decrypt all the received IMSI paging requests to determine whether it is the intended recipient or not.

The research work [21] studied the information leakage problem in the paging procedure and provided a solution by using a physical layer identification scheme. The scheme is a complementary and does not eliminate the need for enhanced privacy in the other signaling procedures. The scheme proposed a function that takes the UE's temporary ID as an input and has a tag as an outcome. During the paging period of a UE, instead of transmitting TMSI, the corresponding tag would be inserted. The scheme requires that no correlation should exist among the tags for different users. An interesting point is that in this case, the transmission power of the signal need not to be at such a level that the receiver could decode it. The receiver should only be able to detect the signal to be able to ensure if the user has been paged or not. This results in saving energy. The drawback of the scheme is that it requires changes in the physical layer procedure that would lead to changing the hardware, which might be costly.

The research work [3] suggested a solution to mitigating paging procedure privacy attacks through sending a hashed value of TMSI identifier allocated to the UE. The security solution requires that a random value like a nonce or a time stamp should be utilized as input to the hash

function in order to change the hashed value of the pseudonym after each calculation.

III. PAGING PROCEDURE ISSUES IN LTE

The LTE network locates an idle UE using paging procedure in order to deliver a service to it (e.g, an incoming call, SMS message). The serving network MME locates an idle UE as per tracking area TA basis and sends the paging request message to every evolved eNodeB (eNB) within a particular tracking area. The transmitted paging request may contain the identity of one or more UEs. The UEs targeted by the paging request identified by temporary identities (TMSIs) [3].

Once a UE finds its TMSI identifier in the paging request message, it establishes a dedicated channel to allow the delivery of the service (responding to incoming call or receiving the SMS). It should be noted that TMSI is not changed within a certain tracking area TA and that the paging message are sent in clear text.

The possibility of initiating a paging request for a specific TMSI allows an attacker to check for the presence of a particular UE within a specific area. Assume that an attacker initiates several calls to a specific user within the user's tracking area and monitors the paging channel to obtain several sets of TMSIs that have been paged by the eNB. The attacker could reveal the identity of the concerned user by intersecting the sets of TMSIs identifiers.

IV. METHODOLOGY

The proposed solution replaces the fixed M-TMSI identifier with frequently changing M-TMSI identifiers selected from a range of M-TMSI identifiers. More than one UE can share M-TMSI values in one M-TMSI range. During paging process MME implements identity-division multiplexing IDM technique in order to select proper M-TMSI values for each UE. It is worthy to mention that, while it introduces the concept of identity ranges overlapping that allows the ranges of identities allocated to the UEs to be overlapped, the solution allows the MME to uniquely identify a specific UE using Identity-Division Multiplexing IDM technique and ensuring the unlinkability and the untracebility of the UE.

The MME first allocates a range of M-TMSI identifiers to the UE and delivers the range to it. We use R to refer to the M-TMSI range. Each range R is a pair(S, L), where S is 32bit value representing the starting point of R (i.e., the first M-TMSI value in R) while L is 16bit value representing the length of the range R (thus, the maximum length of R is 65536 values). However, it is up to the network operator to determine the length L. The UE, interprets the allocated range R as follows: S is the smallest M-TMSI value in R while the value (S+L) is the largest M-TMSI in R. The UE also understands that the valid M-TMSI used for paging UE should lie between S and S+L. Next, when the MME wishes to page the UE, it generates a random fresh M-TMSI value between S and S+L, embeds it within the paging request message, and transmits the message to the UE. Once the UE receives the paging request message, it checks whether the received M-TMSI lies between S and S+L. If the received M-TMSI is within the correct range, the UE responds to the request by initiating a service request procedure; otherwise it discards the request.

The proposed solution changes the way of creating and allocating M-TMSI identifiers to ensure subscribers unlinkability. It enhances the characteristics of M-TMSI identifiers allocated to UEs as follows: (1) each allocated M-TMSI is independent from other identifiers such as IMSI, GUTI, and from any previous allocated M-TMSIs. An attacker, who is monitoring the paging channel, cannot correlate the intercepted

ISSN 2636-9680
eISSN 2682-9266

M-TMSIs with each other nor correlate them with a particular UE, (2) the allocated M-TMSI is random and computationally unpredictable, (3) the allocated M-TMSI is used only once for paging a specific UE, (4) allocated M-TMSI is changed frequently, (5) allocated M-TMSI is not reused, (6) there are no collisions in the allocation areas, and (7) the concerned UE can easily check M-TMSIs in the paging message to find whether it is intended by the paging request or not.

The proposed method enhances the privacy of paging procedure in LTE and ensures the unlinkability of UEs with minimal modifications at the two network nodes (i.e., the MME and the UE) with no modifications on any other network node. It also considers the computational power and storage capabilities of both the MME and the UE. It introduces a negligible computation overhead at the UE and an affordable computation overhead at the MME. The method protects against paging-related attacks at a minimal cost and thus it can be easily integrated with the current mobile technology. Moreover, the proposed solution requires no changes on the messaging system. The boundary values (*S* and *L*) of M-TMSI range allocated to the UE can be included in the normal messages that serving network SN sends to the UE during communications between UE and SN

A. The MME

S	*L*	*STATUS*
S_1	L_1	1
…	…	…
Si	L_i	1
…	…	…
SK	*LK*	0
…	…	…
Sn	*Ln*	0

(a) M-pool

IMSI	*S*	*L*	*T*	*V*
$IMSI_1$	S_1	L_1	T_1	V_1
…	…	…	…	…
…	…	…	…	…
IMSIi	*Si*	*Li*	*Ti*	*Vi*
…	…	…	…	…
…	…	…	…	…
$IMSI_K$	S_K	L_K	T_K	V_K

(b) M-table

Fig. 1: (a) M-pool (b) M-table

The solution extends the MME storage with two tables: M-pool and M-table **(Fig. 1**). M-pool table stores a list of all available M-TMSI ranges from which MME can allocate ranges for the UEs in its service area while M-table stores information details of the ranges that are allocated to the UEs in the MME's service area. The M-pool table maintains three values for each M-TMSI range: *S*, *L* and *STATUS* values. *S* and *L* store respectively the start and the length of the range while *STATUS* is a binary value indicating whether the range is free for the use or not. The range with value 0 in *STATUS* is free for the use. The value 1 in *STATUS* indicates that the corresponding range is allocated.

Each record of M-table stores M-TMSI information details of one UE. Each record comprises a set of fields including: IMSI, *S*, *L*, *T* and *V*. The IMSI holds the IMSI identifier of the UE. *S* and *L* denote the start and the length of M-TMSI range allocated to the UE. *T* denotes the last M-TMSI value used by the MME for paging the UE while the *V* denotes the last M-TMSI value used by the UE for initiating service request procedure.

The proposed solution implements a set of algorithms at MME side:

*1) M-TMSI Ranges Creation Algorithm:*initializes M-pool table with the boundaries of all M-TMSI ranges available for the MME to use. **Fig. 2** shows the major steps of the algorithm and the details are as follows:

1. The total sequence of 2^{32} unique M-TMSI values is partitioned into a set of overlapping partitions each of which is a range *R*.

2. The boundaries of each range *R* are stored in *S* and *L* fields of a particular record at M-pool. The field *S* stores the first M-TMSI value in *R* whereas the field *L* stores the length of *R*. Initially, all created M-TMSI ranges are marked as free for the use (i.e., *STATUS* is set to 0 for all M-TMSI ranges).

The M-TMSI ranges creation algorithm ensures that every M-TMSI range overlaps with its previous range in some partition. This is done by selecting the starting point *S* of the next M-TMSI range from within current M-TMSI range. The algorithm computes the value of the starting point *S* of next range as the 2/3 of *L* of the current range (refer to step 10 at **Fig. 2**). By following this partitioning strategy, the boundaries of the consecutive ranges interleave with each other and each M-TMSI range has two types of partitions: shared and unshared partitions. Shared partition is basically a range of M-TMSI values that belong to two overlapping ranges while the unshared partition is a range of M-TMSI values that belong to one range only. The shared partition among two overlapping ranges say R_w and R_z is denoted by $R_{w,z}$; the unshared partition that belongs to range say R_x is denoted by R_x*.

```
Input: limits of range length min and max
1:    Let Avail← 2^32
2:    Let Stop← 0
3:    Let S ← 0
4:    While (Avail ≥ min) do
5:        generate a random L (min ≤ L ≤ max)
6:        if (Avail < min ) then
7:            L= Avail
8:        end if
9:        S = Stop;
10:       Stop = Stop + (2/3) L
11:       create an empty record at M-pool
12:       insert into the new record a tuple (S, L)
13:       Avail =Avail – L
14:   end while
```

Fig. 2: M-pool initialization algorithm

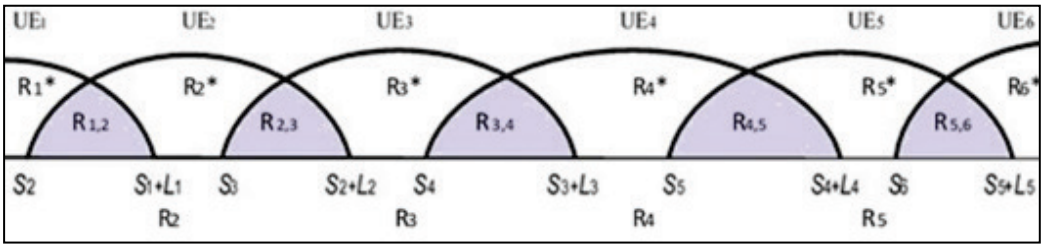

Fig. 3: Overlapping M-TMSI ranges

As **Fig. 3** shows, the M-TMSI range is divided into three partitions: two partitions shared with the previous and the next neighbor ranges respectively (in gray color) and one unshared partition (in white color). This partitioning strategy has several advantages: (1) It allows for controlled partitions overlapping, (2) MME is always able to uniquely identify a particular UE in its service area, (3) It is possible for MME to simultaneously identify two UEs using only one M-TMSI value. This can be done when two UEs, whose M-TMSI ranges overlapping with each other, are intended in the same paging message. The MME can select one T-MSI value from the shared range and include it in the paging message for both UEs.

2) M-TMSI Ranges Assignment Algorithm: selects a free M-TMSI range R from M-pool for the purpose of allocating it to the UE and delivers R information to the UE. **Fig. 4** shows the major steps of the allocation algorithm that MME follows for allocating M-TMSI range for a new UE that enters the MME's service area.

ISSN 2636-9680
eISSN 2682-9266

```
Input: UE's IMSI identifier (IMSI_UE)
1.    Check if M-pool has free M-TMSI
      ranges if there exist free ranges then
2.       select a free range i (S,L)
3.          set STATUS(i) = 1
4.          allocate range i (Si, Li) to UE
5.    else
6.       select an arbitrary allocated M-TMSI
7.    range X (Sx, Lx) such that:
8.    (Sx, Lx) is allocated to UE_X
9.    AND TAL(UEX) ∩ TAL(UE) = φ
10.         allocate range X (SX, LX) to UE
11.   End if
12.   create an entry at M-table and insert the
13.   tuple (IMSI_UE, S_UE L_UE, 0, 0) into it
```

Fig. 4: M-TMSI ranges allocation algorithm

1. The MME selects a fresh not-in-use M-TMSI range *R* from the M-pool, updates its *STATUS* value in M-pool to 1 and associates *R* with IMSI of the requesting UE. If M-pool has no free M-TMSI range to be allocated to the UE, the MME arbitrarily selects for the new UE an allocated M-TMSI range that would not cause M-TMSI collision in the tracking area the UE is in. The MME selects an M-TMSI range of any UE whose, Tracking Area List (TAL) does not overlap with the TAL of the concerned UE and reuses M-TMSI range for the concerned UE.
2. A new tuple with IMSI identifier, *S*, *L*, initial value of *T*, initial value of *R* is inserted into the M-table.

The information of the first allocated range *R* is delivered to the concerned UE in two stages. First, the length *L* is delivered during the attachment procedure included in the authentication vector AV, and then the starting value *S* of M-TMSI range is delivered during GUTI procedure.

-Delivery of *L value*: The main steps of the delivery of the length *L* of the allocated range *R* to the UE are shown in **Fig. 5** and explained below:

1. Upon receiving an attachment request issued by a UE, the MME allocates M-TMSI range *R* (*S, L*) to the UE.
2. The MME computes *L'* as a result of XORing *L* and first half of *S*, and forwards *L'* along with the attachment request to Home Subscriber Station HSS.
3. The HSS generates random token *RAND* and embeds *L'* into *RAND*. The calculation of authentication vector AV proceeds as in normal AKA (authentication and key agreement) procedure and transmitted to the MME.
4. The MME forwards the authentication request to the UE and completes with the UE the AKA procedure steps
5. If authentication succeeds, the UE extracts *L'* from *RAND* and get back *L* by XORing *L'* with first half of *S*, which is received in GUTI message.
6. Finally, the UE stores S and L for the purpose of paging procedure.

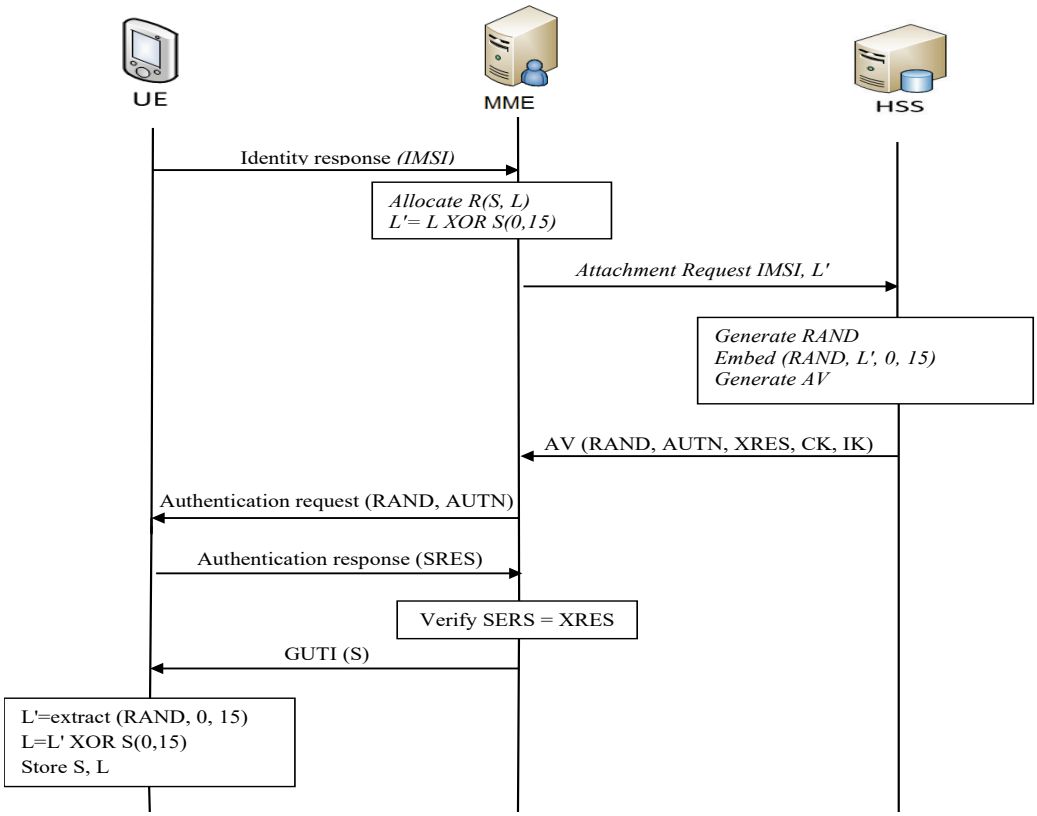

Fig. 5: The main steps of allocating and delivery of M-TMSI range to the UE

-Delivery of *S value*: the UE receives *L* value within *RAND* token during AKA procedure. The UE is then supplied with the starting point value *S* of the allocated M-TMSI range within the GUTI messages after a successful run of AKA procedure. Later, the UE may receive the *S* value included in the GUTI message in the following occasions:

- After Inter-MME handover request
- After a successful TAU request
- The serving network can be scheduled to send GUTI messages including *S* values to the UE at regular time intervals.
- The UE can be provided with the capability to request a fresh *S* value at arbitrary times.

3) M-TMSI Ranges De-Allocation Algorithm: the MME de-allocates the M-TMSI range R allocated to the UE by deleting R from M-table and frees up R if possible (i.e.; if R is non-sharable). The de-allocation algorithm is run whenever an existing UE is leaving the MME's service area. **Fig. 5** shows the main steps the MME runs in de-allocation algorithm and the steps details are given below:

1. Searches M-table for the UE's entry that includes the range *R* using the IMSI of the concerned UE as a key and removes it.
2. Locates *R* at the M-table and verifies whether *R* is currently in use by another UE or not. If *R* is not found, the MME sets 0 value in the *STATUS* field corresponding to *R* in M-pool table and frees up *R*.

4) M-TMSI Ranges Re-Allocation Algorithm: replaces an M-TMSI range allocated to a UE with a different range. The Re-Allocation algorithm is run during the UE's movement within the service area when the UE moves into a new tracking

ISSN 2636-9680
eISSN 2682-9266

area which is not in the tracking area list TAL registered in the UE. MME can also run Re-Allocation algorithm at arbitrary time intervals. It is worth mentioning that the newly allocated range has the same length as the currently allocated R. This is because the MME sends only S to the UE during the Reallocation procedure. The major steps of the algorithm are:
1. The MME allocate a new range *R* whose *L* is the same as existing range *R* currently allocated to the UE.
2. The MME initiates GUTI relocation procedure and sends new *S* to the UE, which will replace its *S* with the newly received *S* and recalculate the boundary value (*S*+*L*).

5) Paging UE Algorithm: generates a fresh M-TMSI value and includes it in the paging request message transmitted to the UE. The MME runs the algorithm for page an idle UEs. **Fig. 6** demonstrates the major steps of the algorithm:
1. Searches M-table for the UE's entry using UE's IMSI identifier and generates a random fresh M-TMSI value M_{MME} such that: (i) M_{MME} is within the range *R* allocated to the UE, (ii) M_{MME} is different from the last M-TMSI value sent to the UE, and (iii) M_{MME} is different from the last M-TMSI received from the UE.
2. Proceeds with normal paging procedure steps with M_{MME} as M-TMSI value.

Input: UE's IMSI identifier
1. Search M-table using UE's MSI as a key
2. get *S*UE, *L*UE, *T*UE, and *V*UE of the UE
3. generate a fresh M-TMSI value(*M*MME) such that:
4. *S*UE ≤ *M*MME ≤ (*S*UE + *L*UE) and
5. *M*MME ≠ *T*UE and
6. *M*MME ≠ *V*UE
7. Update M-table at the UE's entry
8. *set T*UE = *M*MME
9. embed *M*MME within the paging request message
10. proceeds with normal paging steps

Fig. 6: Algorithm for Paging UE

TABLE 1 shows the possible scenarios for paging three UEs with three overlapping ranges. It also shows the number of M-TMSI identities required for paging and the source ranges from which M-TMSI used in the paging can be selected. The following facts can be derived from the table:

- Two UEs having shared range can be identified in the same paging message using only one M-TMSI value selected from a shared range. For example, one M-TMSI value selected from shared range R1,2 can simultaneously identify UE1 and UE2 (Scenario 4(1)). Similarly, one M-TMSI value selected from shared range R1,2 can simultaneously identify UE2 and UE3 can be simultaneously identified using only one M-TMSI value selected from shared range R2,3 (Scenario 6(1)) .
- Three UEs (e.g., UE1, UE2, and UE3) allocated three consecutive ranges (e.g., R1, R2, and R3) can be targeted in the same paging message using two M-TMSI values selected from one range or two ranges or using three M-TMSI values selected from three ranges (Scenarios 7(1), 7(2), and7(3)).

-UE Confusion Prevention Principle*: For any two UEs: UEx and UEz having overlapping ranges Rx and Rz with a shared range Rx,z. If, at any time, the MME wishes to page either UEx or UEz and uses M-TMSI identifier from Rx,z, then the unintended one will respond to the paging message thinking that it is intended. To prevent such situation, MME refrains from selecting M-TMSI identifier from Rx,z when paging either UEx or UEz. However, a selection from Rx,z can be made when both UEx and UEz are intended by the paging message. The principle could be relaxed and MME can select and use M-TMSI identifier from Rx,z for paging either UEx or UEz if and only if it is knows with certainty that UEx and UEz are in different tracking areas TAs.*

6) M-TMSI Validation Algorithm: upon receiving a service request initiated by the UE, the MME validates the M-TMSI value, M_{UE}, included in the service request. Depending on validation results the MME

may respond to the request or not. If TRUE is returned from the validation algorithm after validating the request, the MME is assured that the request came from a legitimate UE and thus responds to the incoming request; if FALSE is returned the incoming request is discarded. **Fig. 7** shows the major steps of the algorithm to validate M_{UE} value:

TABLE 1: Scenarios for paging 3 UEs {UE1,UE2,UE3} with 3 overlapping M-TMSI ranges {R1,R2,R3}

Scenario (case)	UEs intended by the paging message			No. of M-TMSIs required for Paging	No. of Ranges from which M-TMSI(s) can be selected	Possible Source Range(s) for selecting M-TMSI(s)
	UE1	UE2	UE3			
1	X			1	1	R1*
2		X		1	1	R2*
3			X	1	1	R3*
4 (1)	X	X		1	1	R1,2
4 (2)				2	2	(R1*, R2*) **Or** (R1*, R1,2**) Or (**R1,2, R2*)
5 (1)	X		X	2	1	(R1,2, R2,3)
5 (2)				2	2	(R1,2, R3*) **Or** (R1*,R3*) **Or** (R1*, R2,3)
6 (1)		X	X	1	1	R2,3
6 (2)				2	2	(R2*, R3*) **Or** (R2*, R2,3) **Or** (R2,3, R3*)
7 (1)	X	X	X	2	1	(R1,2 ,R2,3)
7 (2)				2	2	(R1*,R2,3) **Or** (R1,2, R3*)
7 (3)				3	3	R1*,R2*,R3*

1. Searches M-table looking for any M-TMSI range that contains the M_{UE} value. If no M-TMSI range is found, then discards the request; otherwise proceeds with the next step.
2. Verify that M_{UE} differs from the last M-TMSI sent to the UE, and from the last M-TMSI received from the UE.

```
Input: UE's IMSI and MUE included in the request
Output: TRUE or FALSE
1. Searches M-table for M-TMSI range where:
2.      SUE ≤ MUE < 2/3(SUE + LUE)
3. If a particular range is found
4.      If ((MUE ≠ TUE ) and (MUE ≠VUE ))
5.          return TRUE
6.      End if
7. End if
8. return FALSE
```

Fig. 7: Service request validation algorithm

B. The UE

The solution extends the UE with four 32 bit fields to store M-TMSI values: S_{UE}, L_{UE}, T_{UE} and V_{UE}. The S_{UE} and L_{UE} fields store respectively the values of the start and the length of the M-TMSI range supplied by the MME. The last M-TMSI value transmitted by the UE and the last M-TMSI value received by the UE are stored in T_{UE} and the V_{UE} fields respectively. For successful operation of the proposed solution, the functionalities of the UE with respect to GUTI relocation, paging, and service request procedures are modified.

The proposed solution implements a number of algorithms at UE side:

1) Receive Paging Request Message: Once a paging message request is received from the MME, the UE verifies that the incoming M-TMSI value M_{MME} included within the paging request message is within the correct range and is also different from the T_{UE} and V_{UE} that are stored at the UE.

ISSN 2636-9680
eISSN 2682-9266

If so, the UE updates its VUE to the newly arrived M_{MME} identity and initiates a service request; otherwise the paging request is ignored. **Fig. 8** presents the M-TMSI validation algorithm.

```
Input: paging request from MME including MMME
1:    if ( SUE ≤ MMME ≤ SUE +LUE )
2:      if ((MMME≠TUE) and ( MMME≠VUE))
3:            set VUE = MMME
4:            initiate a service request
5:      else ignore the paging request
6:      end if
7:    else ignore the paging request
8:    end if
```

Fig. 8: Paging request validation algorithm

2) Initiate a Service Request If the UE is confirmed that it is intended by the paging request message received from the MME, it initiate a service request. First, it UE first generates a random fresh M-TMSI value M_{UE}, embeds M_{UE} within the service request message, and updates TUE to M_{UE}. **Fig. 9** presents the steps of service request algorithm.

The condition in step 2 is to comply with MME Confusion Avoidance Principle. It allows the MME to uniquely identify the UE, while the conditions (in steps 3 and 4) are to ensure that the fresh M-TMSI is different from the last M-TMSIs values exchanged with the MME. The conditions (in steps 3 and 4) aim to eliminate the possibility of replay attack.

```
1:  Generate a fresh MUE value such that:
2:        SUE ≤ MUE < (2/3) (SUE + LUE),
3:          MUE  ≠ TUE, and
4:          MUE  ≠ VUE
5:  update  TUE = MUE
6:  initiate service request
7:
```

Fig. 9: Service Request algorithm

-MME Confusion Prevention Principle: *To prevent MME getting confused with the M-TMSI identifier M_{UE} transmitted by the paged UE, the UE must select a fresh M_{UE} that satisfies the following condition:*

$$S_{UE} \leq M_{UE} < (2/3)\,(S_{UE} + L_{UE}) \quad (1)$$

S_{UE} is the starting point of the UE's M-TMSI range and L_{UE} is the range's length.

V. ANALYSIS AND RESULTS

LTE architecture assigns each UE a unique TMSI identifier in order to identify the user during the paging process. During the paging process, the MME includes the UE's TMSI within the paging request message and sends it to the UE. The security issue in using TMSI is that it remains for a period that is enough for an attacker to link the TMSI included in the paging requests with the user's permanent identity IMSI. Therefore, the existing LTE paging procedure is vulnerable to user linkability attack. The proposed solution enhances the characteristics of TMSI identifiers and their allocation procedure and improves the capabilities of LTE in preventing linkability attacks. Paging the UE with random TMSI identifier each time guarantees that an attacker cannot link the paging requests with the same user.

A. The key features

1) Minimal Computation Overhead: The majority of computation overhead is placed on the MME since its computation power is unlimited while a minimal computation is placed on the UE. We can claim that the overhead is negligible at both the MME and the UE.

2) Minimal system Impact: The solution does not change the messages and the messaging system, which makes it transparent to the intermediary networks.

3) Compatibility with LTE architecture: The solution can be easily integrated in the current LTE architecture with minimal modifications on the network parties.

B. Security analysis

This section analyzes the unlinkability and untraceability of the proposed solution.

1) User Unlinkability

The capability of an observer to linking between permanent identity and temporary identities of users is known as the linkability. The proposed solution mitigates the linkability attack that can be caused by using fixed TMSI in LTE paging procedure. The proposed solution provides unlinkability of LTE network subscribers by assigning each UE frequently changing M-TMSIs identifiers instead of a fixed M-TMSI that can be tracked and linked to a specific UE. Furthermore, since M-TMSI ranges overlap, it is possible for one M-TMSI from shared range used for identifying one UE to be safely reused for identifying another UE in some events (as discussed earlier in paging algorithm). Thus, ranges overlapping makes it harder for an adversary to track a specific UE.

2) User Untraceability

Traceability refers to the possibility of identifying past of identity requests and responses of the same subscriber. The proposed solution eliminates user traceability and protects user against tracking attack through enhancing the characteristics of, and the allocation procedures of, the pseudonyms (TMSIs). The allocation procedure of TMSI pseudonyms adopted by the presented solution prevents tracking of the user. The user is assigned a range of TMSIs and upon each request message, a random pseudonym TMSI is selected from within the range. Moreover, each pseudonym is utilized only once by respective network parties. Besides that, the same pseudonym can be reused by different UEs. This complicates the task of an observer to identify the requests and the responses that destined the same user as the M-TMSI exchanged in the network, from the observer's viewpoint are unrelated. As a result, the observer cannot identify the past identity requests and responses of the same user and cannot track the user.

VI. CONCLUSION

This paper presents a solution for enhancing the privacy of paging procedure in LTE network. It introduces identity ranges overlapping concept that allows the ranges of identities allocated to the UEs to be overlapped while it allows the MME to uniquely identify the UE using Identity-Division Multiplexing IDM technique. The UE is identified every time with a fresh M-TMSI identifier selected from a range of M-TMSI values allocated for the UE. The solution preserves paging procedure privacy through a secure identification system that allows a user to remain anonymous and be uniquely identified within the network and prevents attackers from being able to track the user. The solution is compatible with recent standards of LTE cellular network technology. The solution enhances the privacy of paging procedure in LTE and ensures user untraceability and unlinkability with minimal changes at the network and the UE and with low computation overhead on the part of the network and the UE.

VII. APPENDIX

The main result of this section is that the proposed solution indeed enhances the privacy of paging protocol in LTE and preserves unlinkability. The main idea of the proof is that an outside observer (attacker) sees no difference in the output of two runs of paging protocol that they differ only in user identifiers. The proVerif [14] is used for verifying that proposed solution enhances the privacy of paging protocol. The proof proceeds using the notion of observational equivalence.

ISSN 2636-9680
eISSN 2682-9266

```
Enhanced_Paging_Protocol:
event accept TMSI (bitstring, bitstring).
free net: channel.
free A:bitstring.
free B:bitstring.
(* constants *)
const PAGING_REQUEST:bitstring.
const PAGING_RSPONSE bitstring.
let UE(id: bitstring, out_tmsi: bitstring) =
   in( net, (=PAGING_REQEST, in_tmsi:bitstring));
   out(net, (PAGING_RSPONSE, out_tmsi)).
let MME(id: bitstring, in_tmsi:bitstring) =
       (*new in_tmsi: bitstring;*)
       out(net, (PAGING_REQUEST, in_tmsi));
       in(net,(=PAGING_RSPONSE,
out_tmsi:bitstring)).
process
      ((! (new out_tmsi1a: bitstring;  new out_tmsi1b:
bitstring;
      new in_tmsi1a: bitstring;  new in_tmsi1b:
bitstring;
(!  ((MME(choice[A, B], choice[in_tmsi1a,
in_tmsi1b])) | (UE(choice[A, B], choice[out_tmsi1a,
out_tmsi1b])))).
```

VIII. REFERENCES

[1] 3GPP, 3GPP System Architecture Evolution (SAE); Security architecture. 3GPP, TS 33.401, 2013.

[2] H. Choudhury, B. Roychoudhury and D. K. Saikia, "Enhancing user identity privacy in LTE". In IEEE 11th International Conference on Security and Privacy in Computing and Communications (TrustCom), 2012. p. 949–957.

[3] H. Ghafghazi, A. El-Mougy, H. T. Mouftah, "Enhancing the privacy of LTE-based public safety networks". In 13th Annual IEEE Workshop on Wireless Local Networks, Edmonton, Canada 2014.

[4] A. Bikos and N. Sklavos, "LTE/SAE security issues on 4g wireless networks". IEEE Security and Privacy, 11(2):p. 55–62, 2013.

[5] N. Seddigh, B. Nandy, R. Makkar and J. F. Beaumont, "Security advances and challenges in 4g wireless networks". In Eighth Annual International Conference on Privacy Security and Trust (PST), 2010. p. 62-71.

[6] I. Bilogrevic, M. Jadliwala and J. P. Hubaux, "Security and privacy in next generation mobile networks: LTE and femtocells". In 2nd International Femtocell Workshop, Luton, UK. Citeseer, 2010.

[7] A. J. Bou, H. Chaouchi and M. Aoude, "Ensured confidentiality authentication and key agreement protocol for EPS". In 3rd Symposium on Broadband Networks and Fast Internet, May 2012, 28-29.

[8] Li Xiehua, and Y, Wang, "Security Enhanced authentication and key agreement protocol for LTE/SAE network", 2011, In 7th International Conference on Wireless Communications, Networking and Mobile Computing, IEEE.

[9] M. Arapinis, et al., "New privacy issues in mobile telephony: fix and verification". In ACM Conference on Computer and Communications Security, 2012, p. 205–216.

[10] Z. Muxing, F. Yuguang, "Security analysis and enhancements of 3gpp authentication and key agreement protocol," IEEE Trans, vol. 4, 2005.p. 734-742.

[11] G. M. Køien, "Mutual entity authentication for LTE". In 7th International Wireless Communications and Mobile Computing Conference, 2011, IEEE.

[12] G. M. Køien, "Privacy enhanced mutual authentication in LTE". In IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE, 2013. p 614–621.

[13] F. Broek, R. Verdult and J. Ruiter, "Defeating IMSI catchers". In CCS '15 Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, ACM New York, NY, USA.

[14] B. Blanchet. "Proverif: Cryptographic protocol verifier in the formal model". http://www.proverif.ens.fr/.

[15] K. Shubber for Wired magazine. "Tracking devices hidden in London's recycling bins are stalking your smartphone". http://www.wired.co.uk/news/archive/2013-08/09/recycling-bins-are-watching-you. Last accessed May 2015.

[16] M. Arapinis, L. I. Mancini, E. Ritter, and M. Ryan. "Privacy through pseudonymity

in mobile telephony systems". In NDSS, 2014.

[17] S. Datoo for The Guardian. "How tracking customers in-store will soon be the norm". http://gu.com/p/3ym4v/sbl. Last accessed May 2015.

[18] N. Balasaheb, B. N. Gawande,"Hybrid model for location privacy in wireless ad-hoc networks", IJCNIS, vol.5, no.1, pp.14-23,.DOI: 10.5815/ijcnis.2013.01.02, 2013.

[19] A. Muthana, M. Saeed, "Analysis of user identity privacy in LTE and proposed solution", I. J. Computer Network and Information Security, 1, 54-63 Published Online January 2017 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijcnis.2017.01.07, 2017.

[20] D. Forsberg, H. Leping, K. Tsuyoshi, and S. Alanara, "Enhancing security and privacy in 3gpp e-utran radio interface," in Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on. IEEE, pp. 1–5.

[21] T. Tuan, and J. Baras, "Enhancing Privacy in LTE Paging System Using Physical Layer Identification", Data Privacy Management and Autonomous Spontaneous Security. pp. 15-28. Springer Berlin Heidelberg, 2013.

ISSN 2636-9680
eISSN 2682-9266

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Computer Emergency Response Team

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
57 – 67

Knowledge Impact on Information Quality, Service Quality and System Quality for Security of 1GovUC

Rossly Salleh[1], and Azni Ab Halim [2]
[1,2]Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Malaysia
[1]**rosslysalleh@yahoo.com,** [2]**ahazni@usim.edu.my**

ARTICLE INFO

Article History
Received 24 Jun 2019
Received in revised form 20 Sep 2019
Accepted 25 Sep 2019

Keywords:
knowledge, service quality, system quality, information quality, 1GovUC

ABSTRACT

The users play a crucial role in information system (IS) implementation. They are the ones who learned, utilized, and experienced the system. Studying and understanding user satisfaction are important since user satisfaction – along with the quality of the system, the utilization by the users, and the support the users received during the system implementation – are all important aspects that influenced the IS success. Despite the governments' growing investment in electronic services, e-government services do not always meet the expectations of the users. Therefore, measuring the success of management systems within an organisation is crucial to understand how the systems should be built and implemented in practice. Therefore, in this study, a MyGovUC implementation was developed to expose the essential implementation factors that may influence the usage of 1GovUC implementation. The relationship between their knowledge and the information, service, and system quality will be measured statistically using PLS-SEM.

I. INTRODUCTION

In today's style of work, everything moves very fast which include meetings as one of the communications ways between officers and staffs. Today's business communication requires business workers to communicate with others (either with other employees or customers) almost instantaneously, irrespective of their geographic locations [1]. It has been realized that speed in meetings is essential to achieve growth and profitability. The emerging of next-generation applications enable productivity gains in the government sectors and create new agility in communications among the government agencies.

In Malaysia, the use of ICT in public started in 2003 by the launching of Public Sector ICT Strategic Plan through the MyGovernment services as the main partner. The Malaysian Administrative Modernizations and Management Planning Unit (MAMPU) has been given the mandate to lead the MyGoverment project. Kerravala [2] has mentioned that the organizations' competitive advantage is no longer based on a single core competency but must stay ahead if their competitors can make critical decisions in the shortest time. In today's global business competition, people has changed their nature of work. People now are working differently where the requirement for faster decision-making dictates that project teams need to be smaller and nimbler.

MAMPU has launched 1GovUC project which can benefit all government departments and agencies in improving the productivity and reduce management costs. 1GovUC project currently used by 186

ISSN 2636-9680
eISSN 2682-9266

agencies with 296,802 active users. 1GovUC is a Unified Communication and Collaboration services that are centrally managed by MAMPU. The service combines channels of communication via e-mail, video, audio conferencing, and instant messaging. In addition to the above, 1GovUC enables sharing of information through the Collaborative Portal and 1GovUC portal. 1GovUC implemented as cost-saving measures through integrated collaborative communications which all public sector in Malaysia can liaise and undertake projects in communication systems offered by 1GovUC. There are 5 services of 1GovUC such as email (application that allows users to communicate with each other via e-mail), unified communication (UC) (allows users to communicate directly through text, voice call and sharing files online), portal and social media (consisting of 1GovUC portal, collaborative Portal and IMPS) and add-on value (1GovUC provide add-on value such as e-mail archiving, big mail transfer (BMT), Active Directory Right Management Service (ADRMS), e-mail relay (Simple Mail Transfer Protocol - SMTP) and secured e-mail).

II. LITERATURE REVIEW

The users play a crucial role in information system (IS) implementation. They are the ones who learn the system, utilize it, and experience the impact of the IS. Studying and understanding users' satisfaction are important since users' satisfaction – along with the quality of the system, the utilization of the system by the user, and the support user receive during the system implementation – are all important aspects that influence IS success. [3]-[7].

"The review of literature on IS has indicated that the majority of prior studies has attempted to develop summative evaluation frameworks rather than implementation frameworks [8], [9]. [10] and [11] also advocate that by assessing an existing system, the factors leading to implementation could be found. Nonetheless, it is more important to have all the necessary constituents leading to implementation success prior to evaluation." [12] and other researchers developed service quality measurement models, but these models have been developed for assessing private organization's service performance. It has been found that the area of service quality and measurement in the public sector has been less considered and the introduction of the service quality in the public sector is a more recent phenomenon. Despite the governments' growing investment in electronic services, e-government services do not always meet the expectations of users [13]. However, measuring the success of management systems within an organisation is crucial to understand how the systems should be built and implemented in practice [14]. Therefore, in this study, a MyGovUC implementation was developed to expose the essential implementation factors that may influence the usage of 1GovUC implementation.

With regard to IS implementation factors, numerous IS studies have been sought to identify factors related to IS implementation successes and failures, among others [15]- [17]. The importance of these factors cannot be ignored as they guide practitioners and researchers to focus on key areas during implementation. "Some researchers have observed that the factors research approach has little practicality in coping with IS problems [18]; the approach emphasizes the factors and their associated outcomes without much information to structure or implement them [19]. Also, it seems that most prior studies have discussed and identified factors only for successes and failures and the effect of these implementation factors are rarely tested [20]. Similarly, prior studies tend to list the implementation factors without giving any empirical evidence to support their findings."

The study conducted in 2000 by MAMPU found the majority of users were not aware of 1GovUC services and

Mohamed [21] argued that level of knowledge awareness and the use of e-Government services among citizens in this country are still low. This was in-line with Noraidah et al. [22] that the knowledge awareness among the Selangor people in the e-Government was still low. Erika [23] said that UC tools often go unused by employees. The reason he said that, because the irony for organization implementing UC services isn't very "unified" at all. The reality for organization turning to UC services to boost employee communication and collaboration is that these tools often sit dormant and unused on a worker's tablet, smart phone or laptop. He also added the employees must use different applications for different functions to make the video conferencing, messaging, presence and others service available. Other than that, the UC services should be designed to improve teamwork by bringing together individuals and workgroups and increase reachability through single applications access to communicate services. The apps need to be run on mobile devices that allow for employees to work from anywhere and on any network.

MAMPU conducted a survey about 1GovUC Impact on 2014 and it has been found that majority of users are still not knowledgeable with the 1GovUC services, especially in terms of using "Telepresence". The past research also showed that the level of awareness and the use of eGovernment services among the citizens are still low [24]-[27]. Although 1GovUC is the only integrated government communications projects that have been developed so far, there are still issues of global concern as expressed by Chon [29] where he found there are some loop holes in UCaaS, such as slow connection of email and teleconferences when the enterprises connect all UC applications to an integrated framework. Ashaari et al. [22] added that the use and awareness of e-Government among the people of Selangor who are not Internet users (56%) and non e-Government users (74%) was still low. Such statement was also supported by [21],[26] and [27]. Therefore, in this study the attempt is to overcome this limitation by conducting an empirical study that tests the knowledge factor influencing user satisfaction (system quality, information quality, service quality). The study not only adapts the factors research approach, but also combines it with the DeLone and McLean IS model measurements to further examine and evaluate whether the implementation factors influence success or otherwise. The objective of this study is to determine the knowledge influence information quality, service quality and system quality of 1GovUC among government user. Therefore, this study develops a conceptual model of user satisfaction as shown in **Fig. 1**.

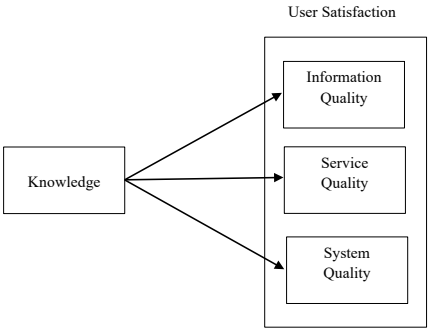

Fig. 1: Conceptual Model of the Study

Based on **Fig. 1**, three (3) hypotheses have been developed. There are:

H1: Knowledge is significant influence on information quality;
H2: Knowledge is significant influence on service quality; and
H3: Knowledge is significant influence on system quality.

III. METHODOLOGY

The target population in this study is government employees in 15 Ministries and 5 Government Agencies. The sampling method for this study was a stratified random sampling. To determine the number of participants that can represent an

ISSN 2636-9680
eISSN 2682-9266

adequate sample size, this study used the data on the total target population that is 2892. The recommended sample size 341 for the 2892 government employees was using the sample size as provided by [29]. Data was collected via, structured questionnaire using web application survey (*www.surveymonkey.com*). The questionnaires were distributed to selected sample of public organizations with the assistance of the information technology (IT) officers who provided the list of user email address in their organization.

A. Population and procedure

In this section the demographic profile of the research participants is presented. **TABLE 1** exhibits the respondents' profile. The respondents are classified according to six distinct categories; gender, age, education, nation, job grade and job experience. The survey results demonstrate that female respondents (34.4%) and the male respondents (65.6%), and that 68.9% of the respondents are 30 years to 39 years of age. This indicates that respondents are mostly from Generation Y where technology should not be an unknown topic to them. In terms of experience, 44% of respondents are have experience between 6 to 10 years. Regarding education level, 63.3% have at least a bachelor's degree. 72% of the respondents hold a 41-48 job grade which most of them are officers' position and most respondents (98.6%) are Malay.

TABLE 1: Demographic Characteristics of the Respondents

Items	Details	Frequency	Percentage
Gender	Male	279	65.6
	Female	146	34.4
Age	Below 20 years	6	1.4
	20 years – 29 years	37	11.1
	30 years – 39 years	293	68.9
	40 years – 49 years	72	16.9
	50 years and above	7	1.6
Education	Primary	6	1.4
	Secondary	17	4.0
	Diploma	82	19.3
	Bachelor	269	63.3
	Master/PhD	51	12.0
Nation	Malay	419	98.6
	Chinese	0	0
	Indian	0	0
	Others	6	1.4
Job Grade	Jusa/Turus	6	1.4
	52 – 54	6	1.4
	41 – 48	306	72.0
	17 – 40	104	24.5
	1 – 16	3	0.7
Job Experience	Below 1 years	12	2.8
	1 – 5 years	7	1.6
	6 – 10 years	187	44.0
	11 – 15 years	141	33.2
	16 years and above	78	18.4

B. Instrument

Quantitative data will be collected based on structured closed-ended questions. Six measures have been used in this study to test the proposed hypotheses. In total, the questionnaire comprised of seven (7) parts: Section A: Demographic Background, Section B: Individual Impact, Section C: Information Quality, Section D: Service Quality, Section E: System Quality, Section F: Knowledge and Section G: Usefulness. The questionnaire is using a five-point Likert-type scale, where 5= Strongly Agree, 4 = Agree, 3= neutral, 2= disagree, and 1= strongly disagree. The information quality, Service Quality and system quality developed by [4],[30]-[31]. Meanwhile knowledge developed by [32] and consists 4 items.

IV. DATA ANALYSIS

The researchers had tested the proposed conceptual model using PLS-SEM. The results were then interpreted. **Fig. 2** indicates the results of the proposed model used in this study. Knowledge is modelled as a reflective construct together with the information quality, service quality and system quality. The measured items loadings, composite reliability (CR), and average variance extracted (AVE) of all reflective constructs are presented in **TABLE 2.**

TABLE 2: Measurement of the Reflective Constructs

Construct & Measured Items	Factor Loading (< 0.50)	Composite Reliability (> 0.70)	Average Variance Extracted (> 0.50)
Knowledge		0.885	0.659
KN1	0.827		
KN2	0.850		
KN3	0.752		
KN4	0.814		
Information Quality		0.936	0.830
IQ3	0.873		
IQ4	0.915		
IQ12	0.944		
Service Quality		0.897	0.500
SQ10	0.760		
SQ2	0.647		
SQ3	0.730		
SQ4	0.752		
SQ5	0.674		
SQ6	0.707		
SQ7	0.692		
SQ8	0.666		
SQ9	0.690		
System Quality		0.923	0.548
SYQ10	0.784		
SYQ11	0.710		
SYQ12	0.753		
SYQ2	0.695		
SYQ3	0.623		
SYQ5	0.759		
SYQ6	0.806		
SYQ7	0.668		
SYQ8	0.753		
SYQ9	0.823		

ISSN 2636-9680
eISSN 2682-9266

Factor Loadings. In PLS-SEM, the loading estimates should be 0.50 or higher [33] and it is a good rule thumb. Analyses have shown that, some items are low loadings which is below then 0.50 and it has been deleted. Then the convergence validity of the items has been measured.

Composite Reliability. In PLS-SEM, the composite reliability (CR) should be equal or greater than 0.70 [34]. This study found that all CR values are greater than 0.70 thus indicating an acceptable range of reliability.

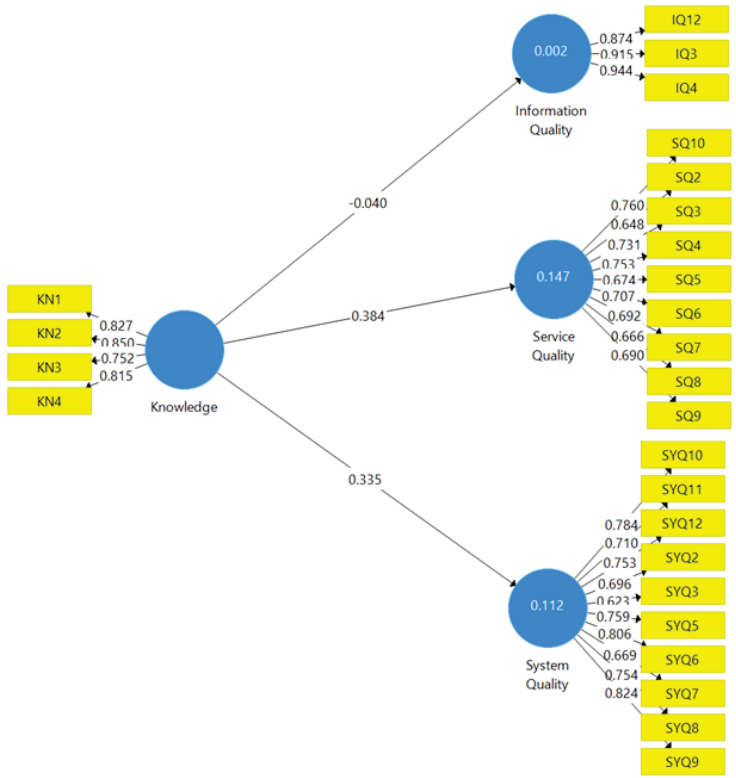

Fig. 2: Result of the Proposed Conceptual model Using SEM PLS

Average Variance Extracted (AVE). The rule of thumb of AVE is 0.50 or higher to indicate adequate convergence of each construct [35]. This study achieved that all AVE values are greater than 0.50, suggesting the convergence validity of the construct.

TABLE 3: Construct Correlations

	Information Quality	Knowledge	Service Quality	System Quality
Information Quality	**0.915**			
Knowledge	-0.040	**0.812**		
Service Quality	-0.044	0.382	**0.703**	
System Quality	-0.008	0.334	0.489	**0.740**

Discriminant Validity. The discriminant validity is to identify the correlation between each construct. It can be seen by comparing the square root of a given construct AVE with the correlation of each construct. As shown in **TABLE 3**, the square root of each AVE is greater than the construct correlations. It indicates acceptable discriminant validity for all constructs.

Fig. 2 indicates the results of the SEM-PLS path analysis and all three hypotheses (H1, H2, H3) were measured. The first hypothesis, H1 assumed that the knowledge has a significant effect on information quality. **TABLE 4** indicates that a no significant effect can be traced between the knowledge and information quality (PC = -0.040, T-statistic = 0.668 and p = 0.5040). Therefore, H1 is not accepted. The second hypothesis, H2 assumed that the knowledge significantly affects the service quality. Analyses have shown that, the path coefficient and T-statistic above 1.96, indicate significant values (PC = 383, T-statistic = 8.651, and p = 0.00). The values demonstrate that knowledge has significant effect on the service quality. Hence, H2 can be accepted. The third hypothesis, H3 presumed that the knowledge has a significant effect on system quality. Based on the analysis, H3 can be described the T-statistic value is below 1.96 (PC = 0.334, T-statistic = 7.712 and p = 0.00). The values demonstrate that knowledge has significant positive effect on the system quality Therefore, H3 is accepted within the context Malaysian.

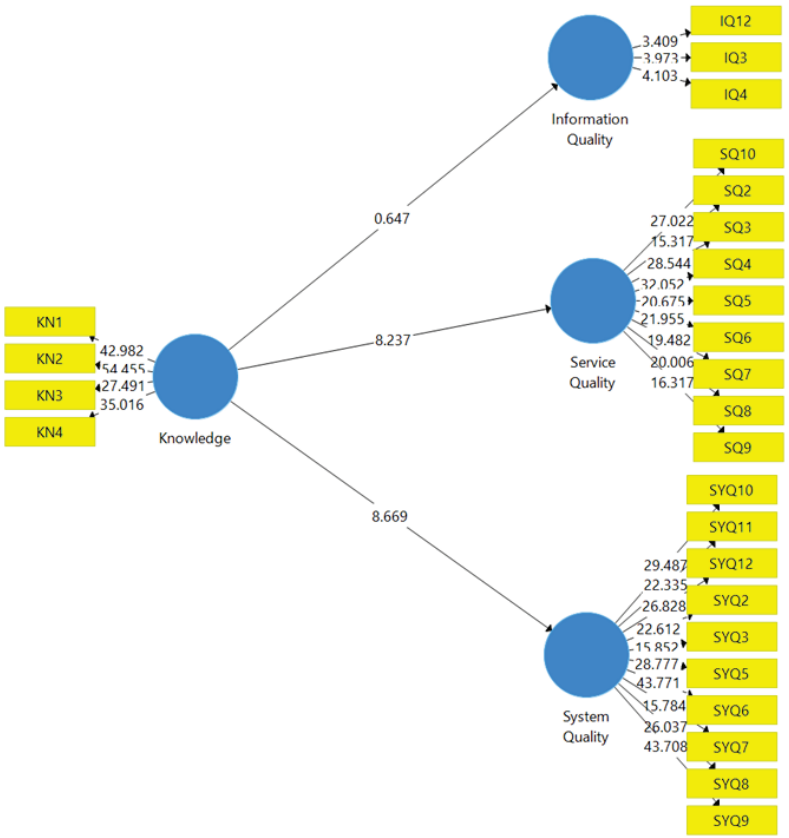

Fig. 3: Bootstrapping Result

TABLE 4: Structural Estimates of the Model

Hypothesis	Path	Path Coefficient	T-stat (>1,96)	P
H1	Knowledge → Information Quality	-0.040	0.668	0.50
H2	Knowledge → Service Quality	0.383	8.651	0.00
H3	Knowledge → System Quality	0.334	7.712	0.00

ISSN 2636-9680
eISSN 2682-9266

V. DISCUSSION

The importance of knowledge towards an information system and IS has been asserted in many studies. Several studies found a significant relationship between knowledge and IS [36]-[39]. However, the quantitative findings of this study suggested *knowledge* as an antecedent for *system quality*. In addition, *knowledge* is statistically found to have a significant positive relationship with *service quality*. This finding confirmed the earlier study about the importance of knowledge within the phenomenon of the 1GovUC effectiveness. Practically, knowledge needs to be applied in order to get an outcome from it. Knowledge is often discussed as having an impact on system usage [40] suggested that the enhancement of system user's knowledge is important for better utilisation of the 1GovUC. However, insufficient knowledge among the IS users is one of the challenges in the IS field [41]. In practice, the 1GovUC users have to be equipped with adequate knowledge in order to gain benefits from the system [42]. The impact of knowledge on the system usage is also mentioned by [43], in which an inadequacy of knowledge may cause inefficiency of the 1GovUC usage. It is found that the knowledge is influence towards service quality and system quality. Nevertheless, the result reported the insignificance of knowledge to influence the information quality. Knowledge is also important for security matters in 1GovUC as mentioned by [44]. The cloud computing security has become a basic necessity nowadays. In is important to have the knowledge about vulnerabilities, attack, activities of attacker and tools to secure it.

VI. CONCLUSION

The advancement of technology in the last decades has enabled the capability of the 1GovUC to offer various outcomes beyond its traditional purpose (i.e. providing information to support the decision-making process). The findings of this study provide substantial contributions to the understanding about the phenomenon of the factors of 1GovUC effectiveness. Quantitative methods used in this study allowed quantitative findings, which have enlightened the understanding of the phenomenon currently happening. As a result, this study presented a comprehensive knowledge on the factors influencing user satisfaction in measuring the 1GovUC effectiveness that is constituted by three dimensions: system quality; information quality; and service quality of the system and the effect on individual impact. In spite of thorough methods applied in this study and its detailed analysis, there are some limitations that should be noted. First, this study applied a quantitative methods approach, in which the research model and the survey instrument are developed based on the review. Thus, the proposed variables and its measurements are limited to the findings from the quantitative fieldwork and the literature review. There might be other variables not discovered in the literature, due to the context of the no interviewees are conducted. Therefore, future studies may wish to consider a comprehensive qualitative method that includes the accounting office and the responsibility centre.

Second, the data of this study were solely obtained from individuals' opinion. As such, bias in opinions might be present because opinions are easily influenced by other factors, such as experience, background and environment. However, totally neglecting their opinions might not reflect the real phenomenon. In addition, individuals' opinion and rating on the 1GovUC effectiveness and other factors in the research model are depending on personal judgement, which may or may not be accurately disclosed by the respondents. Thus, the findings of this study, which is subject to individuals' opinion towards the system, may reflect situational bias. In reality, individual bias is impossible to eliminate. Nevertheless, critical

consideration has been taken during the selection of targeted sample for this study in order to minimise the irrelevant opinions. Therefore, future studies may consider mixing opinions and technical evaluations, using a mixed methods approach. For example, knowledge can be evaluated technically through a test of the related questions or user commitment can be assessed through an observation of the system's user on how he or she uses the system.

Third, the context of this study is limited to the Malaysian Federal Government. In addition, it focused on user for its participants. Non-user involved in data recording or personnel that develop or technically maintain the technology related to the system, such as those from the IT department, might have different opinions towards the studied phenomenon. Moreover, other contexts such as the private sector, Small and Medium-sized Enterprises (SME), public listed companies and so on, might lead to different findings due to the different nature of their business and environment. As such, future studies may wish to apply the methods, research model and survey instrument of this study to other contexts of study.

Fourth, system usage in the context of this study is mandatory. Voluntary usage of the system may discover other variables or result in different findings from this study. In the case of voluntary use, the users' willingness to use the system depends on what they thought as the best practice. Thus, the antecedents for their commitment might be different from the findings of this study. Hence, selecting a voluntary usage context in a future study by using the method and instrument of this study can be a basis that permits a comparison between studies of mandatory and voluntary usage of the system.

As a conclusion, it is believed that the findings of this study could foster a strategic plan for the achievement of 1GovUC effectiveness. Additionally, this study sheds light on the issue of various measures for 1GovUC effectiveness, as well as too many factors influencing the system's effectiveness in the literature. Therefore, the researchers are encouraged to further examine the relationships within the presented phenomenon based on the research model and the survey instrument of this study, in other contexts, to enable comparison between studies.

VII. REFERENCES

[1] S. J. Kowalewski and M. E. Halasz, "Why are written communication skills important for business students?," *Arch. Bus. Res.*, vol. 7, no. 2, pp. 95–102, 2019.

[2] Z. Kerravala, "Today's workers require a new way to work with their teams," 2014.

[3] W. H. Moon, "Development and evaluation of NRMIS (Nursing Resources Management Information System) for managing healthcare resources," *Technol. Heal. Care*, vol. 27, no. 5, pp. 557–565, 2019.

[4] W. H. Delone and E. R. Mclean, "Information systems success: The quest for the dependent variable," *Inf. Syst. Res.*, vol. 3, no. 1, pp. 60–95, 1992.

[5] W. H. Delone and E. R. Mclean, "Information systems success revisited," in *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002, pp. 2966–2976.

[6] N. Urbach and B. Müller, "The updated DeLone and McLean model of information systems success," in *Information Systems Theory*, New York, NY: Spinger, 2012, pp. 1–18.

[7] A. Afnan and C. Ranganathan, "Factors associated with EHR user satisfaction in small clinic settings," in *Thirty Sixth International Conference on Information Systems*, 2015, pp. 1–10.

[8] A. Stylianides, J. Mantas, Z. Roupa, and E. N. Yamasaki, "Development of an evaluation framework for health information systems (DIPSA)," *Acta Inform. Medica*, vol. 26, no. 4, pp. 230–234, 2018.

[9] M. M. Yusof, J. Kuljis, A. Papazafeiropoulou, and L. K. Stergioulas, "An evaluation framework for Health Information Systems: human,

ISSN 2636-9680
eISSN 2682-9266

organization and technology-fit factors (HOT-fit)," *Int. J. Med. Inform.*, vol. 77, no. 6, pp. 386–398, Jun. 2008.

[10] E. Ammenwerth *et al.*, "Visions and strategies to improve evaluation of health information systems. Reflections and lessons based on the HIS-EVAL workshop in Innsbruck," *Int. J. Med. Inform.*, vol. 73, no. 6, pp. 479–491, 2004.

[11] M. J. Van Der Meijden, H. J. Tange, J. Troost, and A. Hasman, "Determinants of success of inpatient clinical information systems: A literature review," *J. Am. Med. Informatics Assoc.*, vol. 10, no. 3, pp. 235–243, May 2003.

[12] O. H. Petersen, U. Hjelmar, and K. Vrangbaek, "Is contracting out of public services still the great panacea? A systematic review of studies on economic and quality effects from 2000 to 2014," *Soc. Policy Adm.*, vol. 52, no. 1, pp. 130–157, Jan. 2018.

[13] S. F. H. Zaidi and M. K. Qteishat, "Assessing e-government service delivery (government to citizen)," *Int. J. Ebus. eGovernment Stud.*, vol. 4, no. 1, pp. 45–54, 2012.

[14] J. P. Vijai, "Examining the relationship between system quality, knowledge quality and user satisfaction in the success of knowledge management system: Àn empirical study," *Int. J. Knowl. Manag. Stud.*, vol. 9, no. 3, pp. 203–221, 2018.

[15] C. Granja, W. Janssen, and M. A. Johansen, "Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature," *J. Med. Internet Res.*, vol. 20, no. 5, p. e10235, May 2018.

[16] S. Y. Hung, W. H. Hung, C. A. Tsai, and S. C. Jiang, "Critical factors of hospital adoption on CRM system: Organizational and information system perspectives," *Decis. Support Syst.*, vol. 48, no. 4, pp. 592–603, Mar. 2010.

[17] B. Kaplan and K. D. Harris-Salamone, "Health IT success and failure: Recommendations from literature and an AMIA workshop," *J. Am. Med. Informatics Assoc.*, vol. 16, no. 3, pp. 291–299, May 2009.

[18] S. Kaushal and D. Hussain, "Key future challenges of indian banking information system," *Int. J. Virtual Communities Soc. Netw.*, vol. 10, no. 1, pp. 65–74, Dec. 2018.

[19] D. Bergemann and S. Morris, "Information design: A unified perspective," *J. Econ. Lit.*, vol. 57, no. 1, pp. 44–95, Mar. 2019.

[20] J. Wolters, U. Y. Eseryel, and D. Eseryel, "Identifying the Critical Success Factors for Low Customized ERP System Implementations in SMEs," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018, pp. 4662–4671.

[21] N. Mohamed, H. Husnayati, and H. Ramlah, "Measuring users' satisfaction with Malaysia's electronic government systems," *Electron. J. e-Government*, vol. 7, no. 3, pp. 283–294, 2009.

[22] N. Sahari, N. Zainal Abidin, H. Kasimin, and H. Mohd Idris, "Malaysian e-government application: Factors of actual use," *Aust. J. Basic Appl. Sci.*, vol. 6, no. 12, pp. 325–334, 2012.

[23] E. Van Noort, "Crafting effective UC training," *Network Computing*, 2015. [Online]. Available: https://www.networkcomputing.com/networking/crafting-effective-uc-training. [Accessed: 09-Jan-2020].

[24] S. H. M. Ali and M. E. A. De Vigal Capuno, "Assessing electronic banking services implementation in Khartoum, Sudan using electronic services dimensions," *Asian J. Manag. Sci. Educ.*, vol. 8, no. 2, pp. 37–41, 2019.

[25] L. Shuib, E. Yadegaridehkordi, and S. Ainin, "Malaysian urban poor adoption of e-government applications and their satisfaction," *Cogent Soc. Sci.*, vol. 5, no. 1, pp. 1–18, Jan. 2019.

[26] M. Dorasamy, M. Raman, and M. Kaliannan, "Knowledge management systems in support of disasters management: A two decade review," *Technol. Forecast. Soc. Change*, vol. 80, no. 9, pp. 1834–1853, Nov. 2013.

[27] M. Kaliannan and H. Awang, "Adoption and use of e-government services: A case study on e-procurement in Malaysia," *WSEAS Trans. Bus. Econ.*, vol. 7, no. 1, pp. 1–10, 2010.

[28] M.-G. Chon, "Government public relations when trouble hits: exploring political dispositions, situational

variables, and government–public relationships to predict communicative action of publics," *Asian J. Commun.*, vol. 29, no. 5, pp. 424–440, Sep. 2019.

[29] R. V. Krejcie and D. W. Morgan, "Determining sample size for research activities," *Educ. Psychol. Meas.*, vol. 30, no. 3, pp. 607–610, Sep. 1970.

[30] W. H. Delone and E. R. Mclean, "The DeLone and McLean model of information systems success: A ten-year update," *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, 2003.

[31] P. Seddon and M.-Y. Kiew, "A partial test and development of DeLone and McLean's model of IS success," *Australas. J. Inf. Syst.*, vol. 4, no. 1, pp. 90–109, Nov. 1996.

[32] S. Puasa, J. Smith, and S. M. Amirul, "Perceptions of accounting information effectiveness: preliminary findings from the Malaysian federal government," *Labu. e-Journal Muamalat Soc.*, vol. S1, pp. 48–59, Jun. 2019.

[33] P. B. Lowry and J. Gaskin, "Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it," *IEEE Trans. Prof. Commun.*, vol. 57, no. 2, pp. 123–146, 2014.

[34] G. D. Garson, *Partial least squares: Regression and structural equation models*. Asheboro, NC: Statistical Associates Publishers, 2016.

[35] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*, 2nd ed. Thousand Oaks: Sage, 2017.

[36] A. R. Komala, "The influence of the accounting managers' knowledge and the top managements' support on the accounting information system and its impact on the quality of accounting information: A case of zakat institutions in Bandung," *J. Glob. Manag.*, vol. 4, no. 1, pp. 53–73, 2012.

[37] L. Wiechetek, "Effectiveness of information systems implementation: The case of the Polish small and medium enterprises," in *Management, Knowledge and Learning International Conference 2012*, 2012, pp. 193–202.

[38] R. Kouser, G. e Rana, and F. A. Shahzad, "Determinants of AIS Effectiveness: Assessment thereof in Pakistan," *Int. J. Contemp. Bus. Stud.*, vol. 2, no. 12, pp. 6–21, 2011.

[39] N. A. Ismail and R. M. Zin, "Usage of accounting information among Malaysian bumiputra small and medium non-manufacturing firms," *J. Enterp. Resour. Plan. Stud.*, vol. 2009, pp. 1–7, 2009.

[40] O. J. Awosejo, M. Kekwaletswe, R, P. Pretorius, and T. Zuva, "The effect of accounting information systems in accounting," *Int. J. Adv. Comput. Res.*, vol. 3, no. 12, pp. 142–150, 2013.

[41] K. O. Appiah, F. Agyemang, Y. F. R. Agyei, S. Nketiah, and B. J. Mensah, "Computerised accounting information systems : Lessons in state-owned enterprise in developing economies," *J. Financ. Manag. Public Serv.*, vol. 12, no. 1, pp. 1–23, 2014.

[42] M. Agung, "Accounting information system and improvement on financial reporting," *Int. J. Recent Adv. Multidiscip. Res.*, vol. 2, no. 11, pp. 950–957, 2015.

[43] A.-K. Pierre, G. Khalil, K. Marwan, G. Nivine, and A. Tarek, "The tendency for using accounting information systems in Lebanese firms," *Int. J. Comput. Theory Eng.*, vol. 5, no. 6, pp. 895–899, 2013.

[44] S. Chaimae and C. Habiba, "Cloud computing security using IDS-AM-Clust, Honeyd, Honeywall and Honeycomb," *Int. Conf. Comput. Model. Secur.*, vol. 5, no. 85, pp. 433–442, 2016.

ISSN 2636-9680
eISSN 2682-9266

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Computer Emergency Response Team

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
69 – 76

Malware Discovery using Lebahnet Technology

Fathi Kamil Mohad Zainudin[1], Izzatul Hazirah Ishak[2], Sharifuddin Sulaman[3], Farah Ramlee[4], Nur Sarah Jamaludin[5], and Shuaib Chantando[6]
[1,2,4,5,6] Malaysia Computer Emergency Response Team, CyberSecurity Malaysia, Cyberjaya, Malaysia
[3] International Engagement, CyberSecurity Malaysia, Cyberjaya, Malaysia
fathi.kamil@cybersecurity.my, izzatul.hazirah@cybersecurity.my, sharifuddin@cybersecurity.my, farah.ramlee@cybersecurity.my, nursarah.jamaludin@cybersecurity.my, shuaib@cybersecurity.my

ARTICLE INFO

Article History
Received 4 Jul 2019
Received in revised form 15 Aug 2019
Accepted 25 Sep 2019

Keywords:
malware discovery, lebahnet, honeypot

ABSTRACT

Recent trends indicate that the cyber-crimes caused by the malware is increasing as these malicious tools are authored to spread through multiple platform and affecting the millions of users. In order to explore new attack and exploitation trends, virtual honeypot is used to simulate the virtual computer systems at the network level. This paper presents the Lebahnet technology, an improved version of virtual honeypots which consists of simulated the networking stack of different operating systems, data analytics and visualisation platform and also the sandboxing technology to examine the code samples behaviour. This paper also discusses the Lebahnet architecture and shows how the Lebahnet framework helps to explore new attack trends and provide insight for early warning mechanism.

I. INTRODUCTION

Cyber-crimes are crimes that occurs via the Internet that is increasing from previous years and shows no sign of decreasing in the near future. It has already been considered in alarming stage since the Internet of Things (IoT) nowadays are crucial in day-to-day activities to help human life easier. This eventually opens up an opportunity for adversaries to gain financially and keep on evolving the technique, tactics and procedures of cyber-attacks to potential targets every day.

In recent news, most of the cyber threats' goals are targeting for profitable gains. Based on a study conducted by Centre for Risk Studies, University of Cambridge, the WannaCry ransomware cryptoworm resulted in an estimated US$4 billion in losses globally where NotPetya's wiper cryptoworm caused an estimated US$10 billion in losses [1]. Surprisingly, the victims especially businesses would be willing to forked out the expenses and pay upwards of close to a million dollar to decrypt their data [2] without knowing whether they would receive the decryption keys from the attacker. This fact will only lead the attacker to gain more financially and is deemed intolerable for the victim's business returns.

NTT Communication in 2018 Global Threat Intelligence Report (GTIR) highlighted that ransomware attacks are growing more than 350 percent compared to attacks in 2016 [3]. Well known malware such as ransomware, crypto jacking and data breaches are foreseen as top common cybercrime listed nowadays. These attacks

ISSN 2636-9680
eISSN 2682-9266

are focusing on the adversary's intent, capability and opportunity to compromise an organization to achieve their desired objectives. Threat actors are then motivated to persistently attack their targets and would eventually formulate behavioural profiles based on the trends of intrusion used that is known as campaign. As a result, a harmless computer can turnaround into a powerful device for illegal activity based on the actor's preferences.

As an initiative to mitigate cyber-attacks, Malaysia Computer Emergency Response (MyCERT) developed a CyberSecurity Malaysia Honeynet Project, also known as Lebahnet. Lebahnet is developed based on honeypot technology. The purpose of this technology is to lure cyber attackers to invade into other valuable machine and obtain unauthorised access to information systems, allow in-depth analysis of techniques and approaches of adversaries during and after exploitation of the machine or provide early warning about new attack and exploitation trends[4]. MyCERT utilised Lebahnet to generate high value information that is focused on network trends and malicious activities to support advisory and incident handling activities. This paper will discuss on malware collected and analysis starting in 2017 until May 2019 using Lebahnet.

.

II. METHODOLOGY

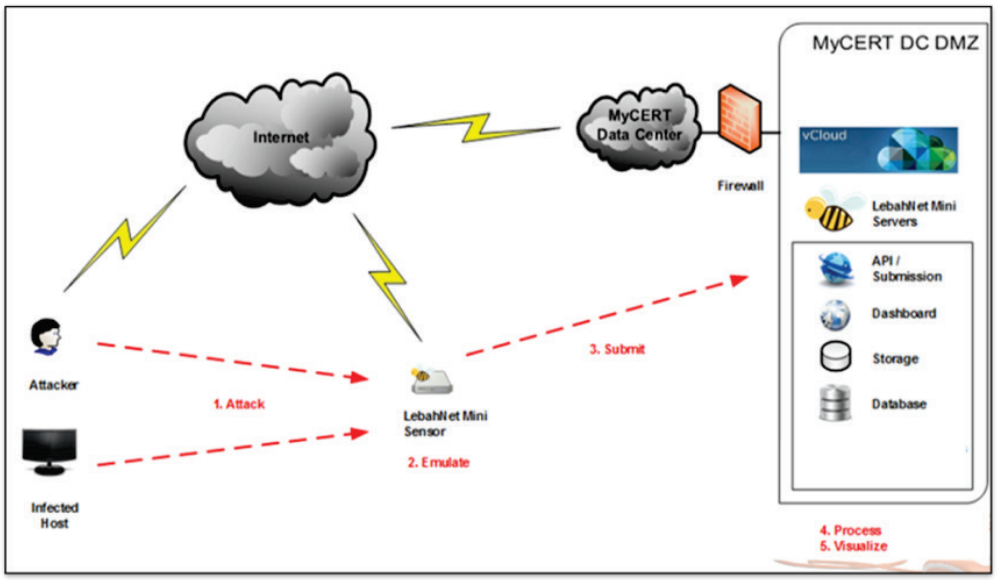

Fig. 1: Overview of Lebahnet

The current version of Lebahnet sensor consist of 2 main components for service emulations which known as Cowrie and Dionaea. These emulations work by luring attacker to think they have successfully compromised a real valuable machine when it is actually a disguise setup for them to run malicious activities. This is because the emulation services respond to each command and request from the attacker.

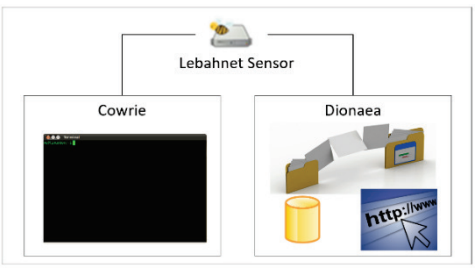

Fig. 2: Lebahnet component

ISSN 2636-9680
eISSN 2682-9266

A. Cowrie Honeypot

Cowrie is an open source project which is developed by Michel Oosterhof. It covers the medium interaction Secure Shell (SSH) and Telnet honeypot which can generate log brute force attacks and attacker's shell interaction. When accessing Lebahnet sensor using SSH, the attacker needs to guess all possible combination of username and password to have access into the server. This scenario gives the attacker to experience "real situation" before gaining access into the server. Upon success gaining access into the sensor, the attacker may proceed with the malicious activities which will be logged and used for further analysis in Lebahnet.

B. Dionaea Honeypot

Dionaea is intended to replace Nepenthes honeypot which is renowned as a malware capturing honeypot which was initially developed under the 2009 Google Summer of Code (GSoC) Honeynet Project's. By utilising Dionaea component in the sensor, the malicious payloads will be captured and used for deep analysis and investigation. Dionaea uses LibEmu to detect the shellcode in the malicious file which make it surpass the capability in Nepenthes honeypot. LibEmu is a library that can be used for x86 emulation and shellcode detection [5]. The small piece of executable binary code is called shellcode [6]. If there is existence of the shellcode in the payload, LibEmu has the capability to execute it in LibEmu VM for assessment or profiling. Each API calls and arguments are recorded and need to be allowed to act such as creating network connection at the first place. This will facilitate the evaluation of the shellcode. Shellcode execution is sufficient to profiling most shellcodes; but not for multi-stage shellcodes [7], [8].

C. Kibana

All data captured via Lebahnet sensors is then been analysed and visualised in Kibana, one of the open source analytics and visualisation tools available. Due to the large amount of Lebahnet data, using Kibana as a platform to analyse data is convenient and cost saving.

Fig. 3: Overview Lebahnet Data in Kibana

D. Cuckoo Sandbox

Once the malware details have been harvested and listed, the sample of the malware is being searched through open source repository or in the wild and will be analysed or re-analysed using MyCERT's very own Cuckoo Sandbox. The sandbox is chosen because it is the leading open source automated malware analysis system [9].

All data captured via Lebahnet sensors is then be analysed and visualised in Kibana. Due to the large amount of Lebahnet data, using Kibana as a platform to analyse data is convenient and cost saving.

ISSN 2636-9680
eISSN 2682-9266

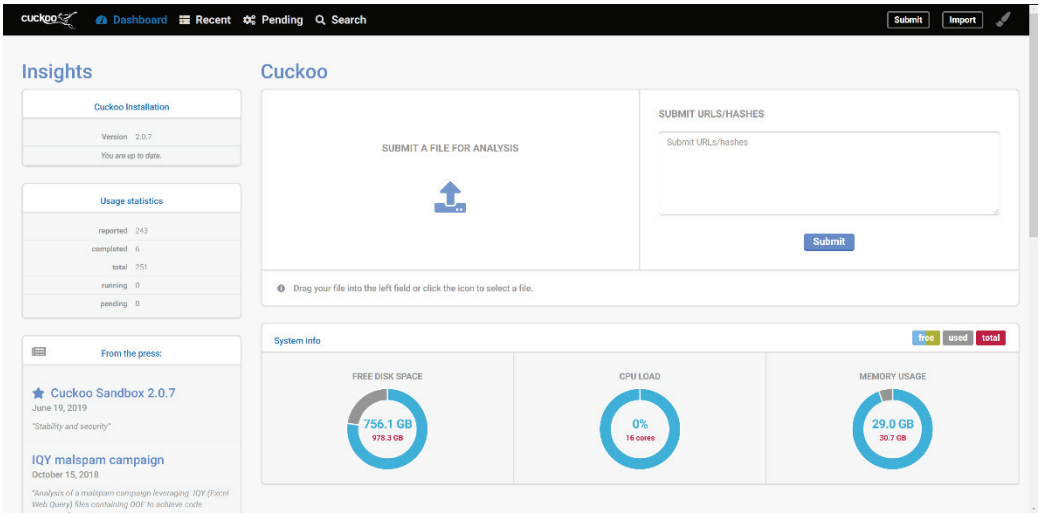

Fig. 4: MyCERT Cuckoo Sandbox

III. DISCUSSION

A typical IT Infrastructure that most small medium organisations used usually are vulnerable despite taking best practice measurement into consideration. The existing security element in a traditional infrastructure implemented may only be sufficient to block and filter any incoming attacks. However, attackers are often a step ahead and will always find ways to bypass the infrastructure and creating new attack patterns. It is important to understand how malware is operating in order to grasp the context, motivations, and the goals of an attack.

Hence, Lebahnet sensors are being placed in organisation's network. At the initial state of the attack, the sensor will respond by emulating according to the attack patterns and attackers' continuous attempts to drop the malware artefact. The payload is then captured and details of the malware such as binaries and hashes are being stored which later will be analyse and visualise in Kibana.

In Kibana, for malware visualisation, selecting term *metadata.md5.keyword* and sorting the result in descending order, the malware hashes later will be displayed where the highest count is displayed at the top. It is recommended to rename metadata.md5.keyword for a better display in data table type of visualisation.

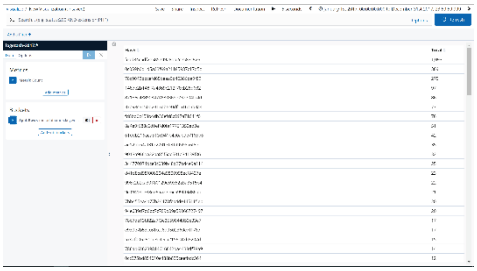

Fig. 5: Overview malware visualisation in Kibana

Illustrated in **Fig. 6** shows the general total malware captured within 2017 until May 2019. Starting from January 2017 until May 2019, a total of 15 858 of known malware binaries are been captured using Lebahnet technology. From 2703 in 2017, the total known malware captured is increased by 5582 in 2018. Even though the statistical data used on 2019 is until May, however, the total known malware captured is more than half of the statistic value in 2018.

ISSN 2636-9680
eISSN 2682-9266

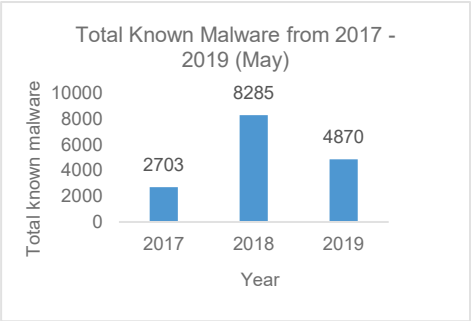

Fig. 6: Malware Trend From 2017 – 2019 (May) 1

Moving forward, **Fig. 7** shows the malware section classified by malware type. Ransomware attack in 2018 hit the highest total captured compare to the previous year, where Worm is the highest malware captured in 2017. Trojan Downloader and Trojan in 2018 also increased contrast with the previous year. Until May 2019, there is no sign for the Ransomware attack to be decline. In 2019, there is a new detection malware type which is known as Virus starts to peep in the Lebahnet technology. However, there is no sign that this malware type will be maintain in the future as the DDOS and Backdoor is not yet been found after 2017 and above. The pattern on the malware type is affected by the user utilisation which can be proved by the ransomware attack started to skyrocket from 2017 until today due to the big impact from all around the world.

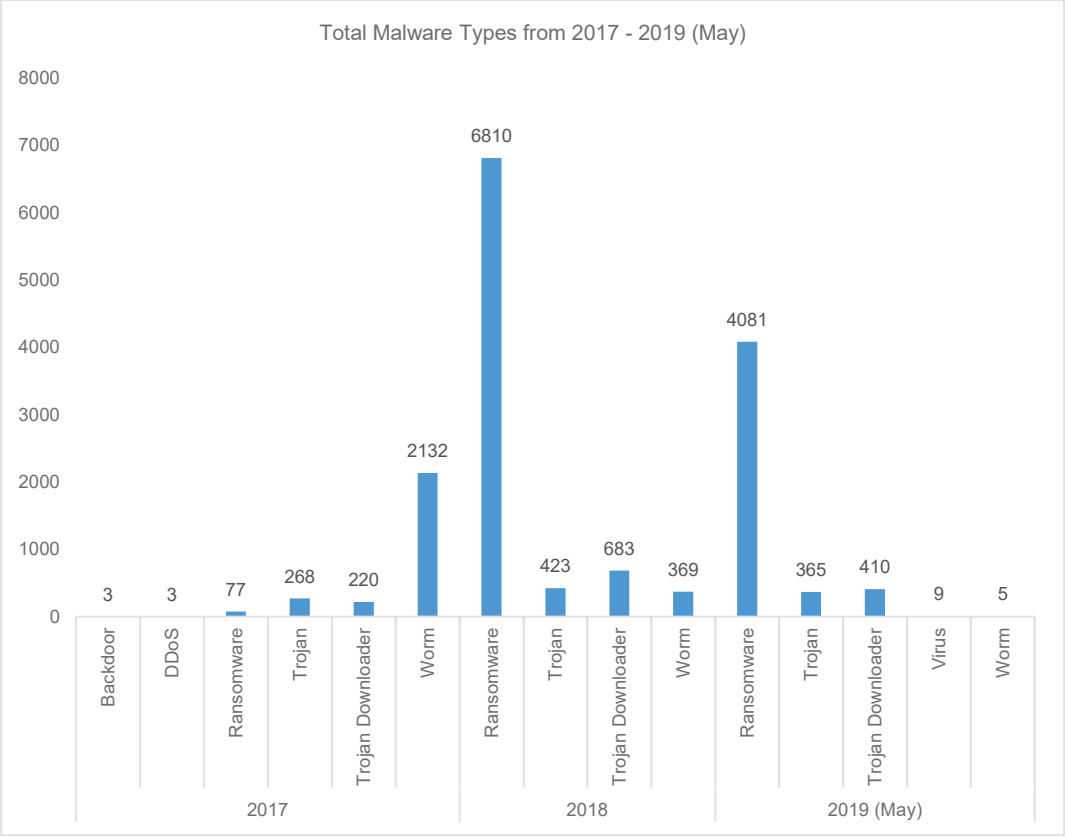

Fig. 7: Malware Trend From 2017 – 2019 (May) 2

From the results of the search, below are the highest malware hashes that have targeted our sensors.

ISSN 2636-9680
eISSN 2682-9266

TABLE 1: Data Extractions results from Kibana

Year	Total hits	Hash (MD5)	Malware
2017	1093	fead84c5df2e585749a8da2ce583c926	Conficker
2018	2266	ae12bb54af31227017feffd9598a6f5e	Wannacry Ransomware
2019 (January - May)	1186	ae12bb54af31227017feffd9598a6f5e	WannaCry Ransomware

In 2017, the highest hit to our sensors was a Conficker malware that was named after a vmware file. Details of the malware are as below:

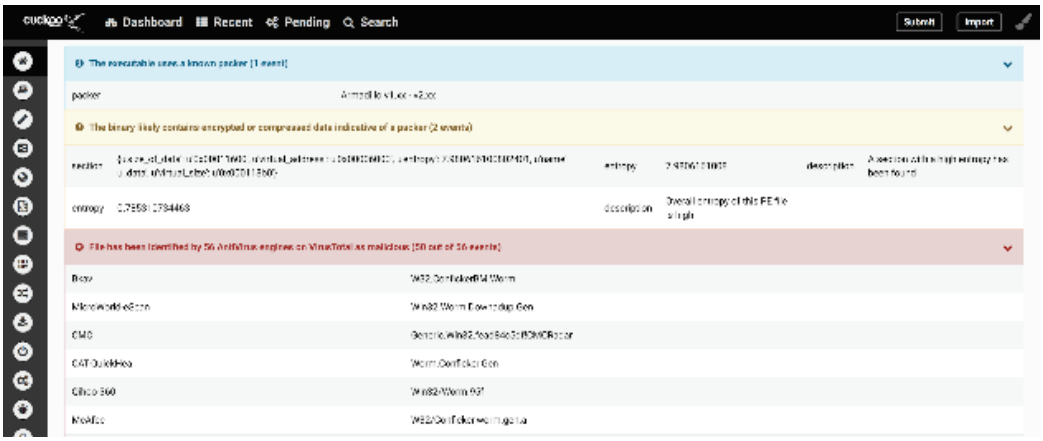

Fig. 8: Conficker result in MyCERT's Cuckoo Sandbox

Filename: jwgkvsq.vmx or jwgkvsq.dll
Md5:
fead84c5df2e585749a8da2ce583c926
File type: Win32 DLL
PE32 executable for MS Windows (DLL) (GUI) Intel 80386 32-bit
File Size: 166.51 KB (170505 bytes)

This malware also known as Downadup, Downadup and Kido, that was first detected in November 2008 and is a fast-spreading worm that targets a vulnerability (MS08-067) in Windows operating systems.

From the sandbox analysis result, the malware is seen using the .vmx file extension in naming the malicious file to lure victims to think they are opening a harmless virtual machine file. This file is a disguise and it is actually a DLL that could allow remote code execution if an affected system received a specially crafted RPC request and later run arbitrary codes without authentication.

This shows that malware is still at large even after many years of existing. This could be the reason that the operating system that is vulnerable to this exploit is still being used in the network where our sensors were being deployed. It proves that malwares can be reused and evolved as years gone by.

As for 2018 and mid-year of 2019, the top hit malware found in our sensors is known as WannaCry ransomware. Ransomware is one of malware types that infects computing platform and limits

ISSN 2636-9680
eISSN 2682-9266

user's access until the stated ransom amount is paid in order to access the encrypted file. WannaCry ransomware infects the victim computers via EternalBlue vulnerability that exist in the Windows Server Message Block (SMB) service which later patched by Microsoft in March (MS17-010) [10]. Details of the malware are as per below:

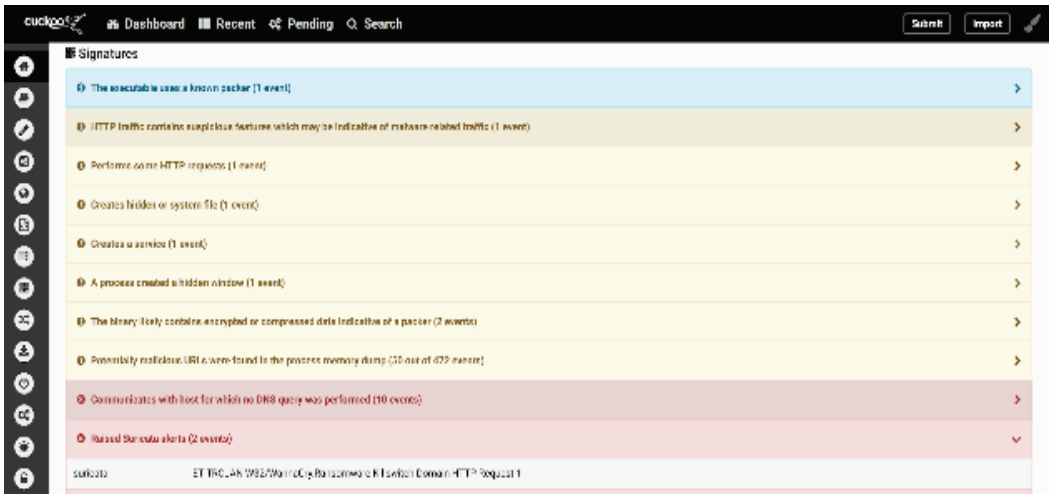

Fig. 9: Wannacry result in MyCERT's Cuckoo Sandbox

Filename: 1561866018332_hbtbr_dionaeajpn1_ae12bb54af31227017feffd9598a6f5e
Md5: ae12bb54af31227017feffd9598a6f5e
File type: PE32 executable (DLL) (GUI) Intel 80386, for MS Windows
File Size: 5.0 MB (5145000 bytes)

From the sandbox analysis, the malware is a DLL file. It is a dropper that is being used to download process that can impersonate a legitimate process. Through an export function called "PlayGame" in the DLL file, the process is then being spawned as "mssecsvc.exe" which is exploiting the "CVE-2017-0147" to compromise the neighbouring PCs in the same network.

WannaCry ransomware started to spread in May 2017 but Microsoft have announced the patch in March 2017. Meanwhile, Lebahnet sensors detected this malware binary on October 2017. Over the past two years, two hundred sixty (260) of different binaries and hashes has been detected which is related to WannaCry ransomware.

IV. CHALLENGES

From the past few years of Lebahnet technology consumption, the data captured only covered certain region and area based on the participation from organisation and institution. Thus, the statistic data limited to the covered region or area. Apart from that, Lebahnet technology need to be enhanced and keep maintaining the same pace with the current IoT device as IoT evolves for the past few years and even for the future. In order to detect new malware, the detection of malware technology needs to be enhanced. However, the deficiency of LibEmu library which is been used in Dionaea component is only capable to analyse and profile for single-stage shellcode. As a result, the profiling is insufficient for the multi-stage shellcode as the data from the second shellcode and upward cannot be recorded for further analysis. This might affect Lebahnet technology to miss certain valuable binary to be analysed.

ISSN 2636-9680
eISSN 2682-9266

V. CONCLUSION

Lebahnet technology is developed based on honeypot technology. The deployment of Lebahnet technology successfully serve it purpose. Not to mention that the data captured can be used for further analysis such as malware binaries and hashes. According to the malware binaries and hashes collected, this paper concludes the findings from 2017 until May 2019, the malware trend changes from Conficker malware in 2017 to Wannacry ransomware in early 2018. However, there is a possibility to have a change in the malware trend in the future based on the vulnerable exploits which will be discovered in the future. Apart from the evolving of malware, the number of operating system using the vulnerability to the exploit contributes to the growth of the malware within the network. This is because the attacker or attackers believes that the potential victims exist all around the world, thanks to the Internet.

VI. REFERENCES

[1] S.K.A. Manoj and D.L. Bhaskari, "Cloud forensics-A framework for investigating cyber attacks in cloud environment," *Procedia Computer Science*, 85 (Cms), pp.149–154, 2016.

[1] A. Coburn et al., "Cyber risk outlook", Risk Management Solutions, Inc., California, 2019.

[2] IBM Security, "IBM study: Businesses more likely to pay ransomware than consumers", 2016. [Online]. Available: https://www03.ibm.com/press/us/en/pressrelease/51230.wss. [Accessed: 14-Jun-2019].

[3] NTT Communications, "2018 global threat intelligence report", 2018.

[4] N. Provos, "A virtual honeypot framework", in *Proceedings of the 13th USENIX Security Symposium*, San Diego, CA, USA, 2004, pp. 1–14.

[5] D. Lukan, "Shellcode detection and emulation with libemu", 2014. [Online]. Available: https://resources.infosecinstitute.com/shellcode-detection-emulation-libemu/. [Accessed: 15-Jun-2019].

[6] T. Lu, L. Zhang, and Y. Fu, "A novel immune-inspired shellcode detection algorithm based on hyperellipsoid detectors", *Secur. Commun. Networks*, vol. 2018, p. 10, 2018.

[7] E. Tan, "Dionaea – A malware capturing honeypot", 2014. [Online]. Available: https://www.div0.sg/single-post/dionaea-malware-honeypot. [Accessed: 15-Jun-2019].

[8] S. Shahrivartehrani and Shadil Akimi Bin Zainal Abidin, "Dionaea honeypot implementation and malware analysis in cloud environment", *Journal of Computing Technologies and Creative Content,* vol. 1, pp. 1-5, 2016.

[9] S. Jamalpur, Y. S. Navya, P. Raja, G. Tagore, and G. R. K. Rao, "Dynamic malware analysis using cuckoo sandbox", in *2018 Second International Conference on Inventive Communication and Computational Technologies, Coimbatore, India, April 20-21 2018,* 2018, pp. 1058–1060.

[10] MyCERT, "MA-663.052017: MyCERT Advisory – Technical Detail: WannaCry Ransomware," 2017. [Online]. Available: https://www.mycert.org.my/en/services/advisories/mycert/2017/main/detail/1265/index.html. [Accessed: 10-Dec-2017].

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Computer Emergency Response Team

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
77 – 86

Securing the OLSR Routing Protocol

Amin Nurian Dehkordi[1], and Fazlollah Adibnia[2]
[1,2]Cert (APA) Center, Yazd University. Yazd, Iran
[1]**apa@offices.yazd.ac.ir,** [2]**fadib@yazd.ac.ir**

ARTICLE INFO

Article History
Received 6 Oct 2018
Received in revised form
15 Aug 2019
Accepted 25 Sep 2019

Keywords:
security, ad hoc networks, trust, OLSR

ABSTRACT

Mobile Ad-hoc networks are self-organized wireless mobile networks that do not rely on any fixed network infrastructure. Due to limited capability including battery, local memory, CPU cycle, any design of protocols in these networks has to consider these limitations. Trust always exists in protocols, which their running are based on cooperation, especially in routing operations between the nodes in these networks. Indeed, these networks can operate properly only when nodes cooperate with each other truly and in a routing operation, which is defined by a standard. In this paper, we propose a method to verify the trust of nodes by their neighbors, while its amount influences the decision about choosing the MPR nodes and causes an improvement in OLSR's routing protocol security. The proposed method brings few modifications and is still compatible with the bare OLSR. We perform an overall evaluation of our proposed method through simulations. Simulation's results indicate performance of our approach while providing effective security.

I. INTRODUCTION

Several routing protocols have been defined for the MANETs [1]. Generally, it is possible to classify routing protocols in MANETs into two classes: reactive and proactive. The reactive protocols such as DSR will find the shortest route by broadcasting a route request only when they require sending data [2], but in proactive protocols, for instance OLSR, each node holds an entire overview of the network topology [3]. Since the usage of MANETs have been increased, in order to benefit from their advantages and apply them in everyday life, and also because of self-organization in these networks, soon it was clarified that the routing protocols security such as OLSR protocol, are the problems in these networks which have never been considered in designs at all. These protocols have been designed on this fact that all of their nodes are correct.

Adnane et al. [7] have proposed a trust-based solution for securing the OLSR Ad hoc routing protocol in three phases. The first phase was the analysis of the implicit trust relations in OLSR protocol. This analysis highlights the possible measures to make OLSR more reliable by exploiting the operations and information already existing in the protocol. To detect misbehaving nodes, they have developed in the second phase, trust-based reasoning by correlating information provided in the OLSR messages received from the network. The integration of this reasoning allows each node to test the consistency of the behavior of other nodes and validate trust relationships established implicitly. Finally, the third phase complements the second by offering two complementary solutions: prevention to resolve certain vulnerabilities of OLSR protocol, and countermeasures to isolate malicious nodes. Trust reasoning in here is, validating neighbors' behavior and

ISSN 2636-9680
eISSN 2682-9266

performance according to OSLR protocol specification. In this model, trust of nodes, which carry out wrong or abnormal behavior and in general do not operate according to OLSR protocol specification, will be reduced. In this paper, we introduce a method to verify the trust of nodes by their neighbors, while its amount influences the decision about choosing the MPR nodes and causes an evolution in OLSR's routing protocol security. On the other hand, considering the networks MANETs nodes limitations in processing capacity, energy and bandwidth an incorrect operation from a node can't always result a malicious node. Because the node did not have enough processing capacity for directing the received package, thus the node has to be given another opportunity to proof its accuracy. The node's record security is calculated by its neighbors' sent packages and old relations in time series, and for computing its security for this moment a time decay function is used.

This paper is organized as follows: Section 2 presents a brief introduction of related works. In Section 3, we introduce the concept of trust management, trust specification language and we introduce the analysis of implicit trust in OLSR. In Section 4 we present the proposed method. In Section 5 the simulation's results will be described.

II. RELATED WORK

In [4], the authors have proposed a protocol, called CONFIDANT, for making misbehavior unattractive; it is based on selective altruism and utilitarianism. It aims at detecting and isolating misbehaving nodes, thus making it unattractive to deny cooperation. Trust relationships and routing decisions are based on experienced, observed, or reported routing and forwarding behavior of other nodes.

Meka et al. [5] have proposed trust-based reputation model for AODV. Reputation is calculated according to the degree of participation in the routing protocol and the information it provides about the network topology.

Ariadne [6] is another secured protocol based on DSR and TESLA: the authors assume that a shared secret key is distributed for a group of trusted nodes using TESLA and that the nodes are synchronized.

Adnane et al. [7] have proposed a trust-based solution for securing the OLSR Ad hoc routing protocol in three phases. OLSR protocol function occurs in three steps: neighborhood discovery, MPR selection, and routing table calculation.

Gadekar et al. [8] have proposed another secured protocol based on OSLR. To detect and mitigate this Node isolation attack, OLSR protocol is modified by improving its MPR selection procedure. This modified OLSR protocol gives better results than Fictitious Node Mechanism.

Madhvi et al. [9] have proposed another secured protocol based on OSLR. The proposed scheme is to observe malicious nodes misbehavior and stop their malicious activities. This protection scheme provides the protection against DoS attack and routing attack and provides secure communication in dynamic network. The infection just in case of security theme is totally removes in network.

III. TRUST MODEL

Trust, trust models and trust management are subjects which researchers have studied in decision-making in distributed and auto-organized applications. Actually, it has not been defined a specific definition for trust, and all researchers have used their own definition of it for their research area. In this work, Yahalom's [10] trust language and notations are used for indicating trust relationship between nodes. Based on this language, trust is some relations, which help to understand interactions between entities such as humans, network's nodes and organizations. When it is saying, we trust node A, it exhibits a confidence, that node

ISSN 2636-9680
eISSN 2682-9266

A will behave in a certain way and will perform some action under certain specific circumstances.

Yahalom's trust language includes two classes: direct trust relations and the derived trust relations as mentioned in [10], the latter being established on recommendations from other entities. Because of energy, bandwidth and processing limitations derived trust, which is established on recommendation from other nodes and aggregation of them, are not considered.

Therefore, the notations are used for the trust languages in this paper are obtained from Yahalom's trust language and as follow:

Each entity is shown by a capital letter A, B.

The trust between A and B is written by **A trust B**; this means that node A trusts node B and is sure that node B is not a malicious node.

When node A mistrusts node B, it is written by **A ¬ trust B**, this means that node A has detected that node B is a malicious node.

OLSR protocol is classified in proactive routing protocols. It discovers the links between network's nodes by using HELLO and TC messages and then it will broadcast the information in MANET. OLSR protocol consists of three steps: neighborhood discovery, MPR nodes selection and routing table calculation. HELLO messages help each node to detect its one-hop and two-hop neighbors. Then it, among its one-hop neighbors, can select its minimum number of nodes (MPRs) enabling it to reach all the two-hop neighbors. The selected neighbors are called MPRs and advertised in the HELLO message with ''mpr'' status.

A. OLSR protocols notations

An OLSR node stores different data about network's topology and its neighborhood:

- $MANET$: the set of the whole MANET nodes.
- $Asym_x$: includes all asymmetric neighbors of node x.
- N_x : the set of symmetric neighbors of node x.
- L_x : includes all neighbors of node x
- $2HN_x$: the set of two-hop neighbors of node x.
- MPR_x : the set of neighbors of node x which have been recognized as MPR, $MPR_x \subseteq N_x$
- $MPRS_x$: the set of nodes selected as MPRs by node x, which are in charge of routing and forwarding the packets sent by x, $MPRS_x \subseteq N_x$.
- RT_x : the routing table of node x.

In OLSR protocol, each node must advertise its presence by diffusing HELLO messages to its MPR and neighbors, periodically. The selected nodes as MPR also have to advertise all nodes, which have selected them as MPR, by TC messages in the whole network, periodically.

- $HELLO$ is the HELLO message generated by node x; it includes the set of the neighbors of x.
- TC_x is the TC message generated by node x. These messages are broadcasted only by MPR nodes in OLSR and in definite validity time inside the whole network.
- $x \xleftarrow{Hello_y} y$ and $x \xleftarrow{TC_y} y$ are the reception of HELLO and TC messages from y by node x, respectively.

OLSR protocol function occurs in three steps: neighborhood discovery, MPR selection, and routing table calculation [7].

Nodes in OLSR protocol advertise their neighbors by sending HELLO messages periodically, through these HELLO messages each node can detect all nodes, which are two-hop away from itself and control the set of $2HN_x$. In addition, each node must, among its one-hop neighbors, select minimum number of symmetrical neighbors nodes (MPR) enabling it to reach all $2HN_x$ nodes. A node can obtain $MPRS_x$ set by receiving HELLO messages, and by

ISSN 2636-9680
eISSN 2682-9266

observing the present MPR nodes in the messages.

The defined threshold for trust can take different levels in various situations according to the requested service and the mobility degree of the network. The level, which is shown by θ, $0 \leq \theta \leq 1$ is always and is used to detect malicious nodes. On the other hand, if a node evaluates the trust level of other node less than θ, this node will be known as malicious node and will be hold in trust table from the node. Particularly, the bigger value for θ indicates a more trustful network with low efficiency. It is possible to obtain a adequate status between the security and the efficiency of the network by changing the value of θ.

At first, by detecting the asymmetrical neighbors, trust on these nodes initializes less than the initialized threshold. Therefore, we will have $A \in Asym_B \Rightarrow B \neg trust\ A$, and after a node becomes a symmetrical neighbor, trust on it will initialize to minimum value of trust.

$$A \xleftarrow{Hello_B} B, A \in Asym_B \Longrightarrow A\ trust\ B,$$
$$N_A = N_A \cup B, 2HN_A = 2HN_A \bigcup (N_B - A) \quad (1)$$

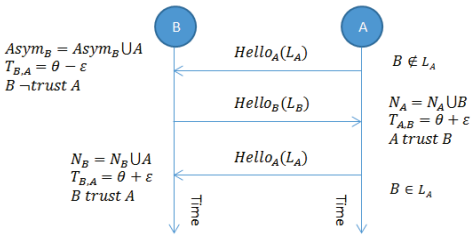

Fig. 1: Trust relationship creating in discovering neighbors.

In **Fig. 1**, at first while between node A and B is not any trust relation, node A sends a $HELLO_A$ message. After receiving this message by node B, a new simplex link is defined for B and the initialized trust to A will be equal to $\theta - \varepsilon$, while ε is much less than θ. After receiving the $HELLO_B$ message by node A and observing itself as a neighbor from node B, a symmetric link will be formed between A and B, because B and A have received $HELLO_A$ and $HELLO_B$, respectively. In this state, a trust relation will be established between these two nodes (A, B) by changing the link type to symmetric and initialized trust will be set θ.

B. MPR selection

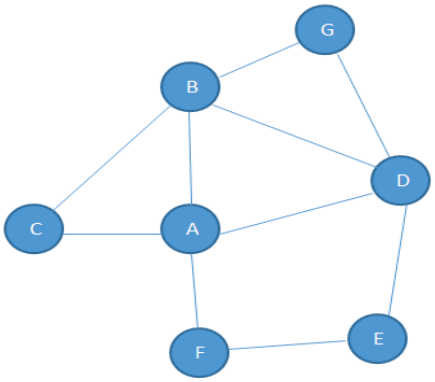

Fig. 2: Node A selects node D as its MPR.

After the detection of one-hop and two-hop neighbors, each node have to, among its one-hop neighbors, select the minimum number of nodes enabling it to reach all the two-hop neighbors. All selected nodes as MPR are advertised to neighbors by HELLO message. For instance, in **Fig. 2** it is only necessary that node A selects node D as its MPR.

When a node is selected as MPR, it will advertise the nodes, which have selected it as MPR, by TC messages in the whole network periodically. TC messages contain the necessary topological information for computing routes to the whole network and will only broadcast by MPRs in the whole network. A node can build its set of topologies and establish its routing table by receiving TC messages, and always the computed routes from one node to other one include MPR nodes across the route. In the presented trust model, choosing node D as MPR by node A means that node A trusts node D, which is to say:

$$\forall x \in MPR_A : A\ trust\ x \quad (2)$$

Selected MPR nodes have to choose their MPR nodes likewise, thus a chain of

ISSN 2636-9680
eISSN 2682-9266

trust between the MPR nodes will create. Surely, success in this approach does not only depend on selecting the correct local MPR nodes but it depends on selecting correct MPR nodes by its neighbor, as well. Therefore, each node has to trust its MPR's selection, either.

C. Routing table calculation

For computing the routing table, which is the result of the OLSR protocol, TC messages are broadcasted in the whole network. The TC message sender advertises that, which nodes have selected it as their MPR node. TC messages help each node to create the network topology from its own point of view and calculate its own routing table. Moreover, if a node is not chosen as an MPR node by its neighbors, its set of $MPRS_x$ will be empty and will not send any TC message. Every node in the network has topology information that is based on received TC messages, stores the information related to every MPR node in the network. Based on this information, the routing table will be calculated. RT routing table is shown as:

$$\forall z \in MANET\ \exists y \in MPR_x \Rightarrow \exists T \in RT_x, T = (z, y, N, I) \quad (3)$$

Each entry in RT table consists of (z, y, N, I), and specifies that the node identified by z is located N hops away from the local node. The symmetric neighbor node, which is identified by y, is the next hop node in the route to z. From the trust point of view, the computation of the shortest path between x and z through the MPR y means that x trusts y for the routing towards z. Hence, if $T = (z, y, N, I)$ is an entry in RT routing table than:

$$\forall T \in RT_x \Rightarrow x\ trust\ y \quad (4)$$

In addition, in routing table only a route is calculated towards each destination node which is the shortest path that starts from the MPR node. The inherent risk in choosing only one route towards any destination is to choose a misbehaving node as a router. Thus, an attacker can put itself between sender and receiver nodes and disrupts protocol operation by giving false information to the neighbors. Finally, in this model selecting y as x's MPR not only implies that x trusts y, but also trusts all y's MPR nodes and as result trusts all selected routes by y, as well.

D. Trust Reasoning

In this paper, it is assumed that when a packet is sent by a node, all of its neighbors will receive it correctly. Trust reasoning in here is, validating neighbors' behavior and performance according to OSLR protocol specification. In this model, trust of nodes, which carry out wrong or abnormal behavior and in general do not operate according to OLSR protocol specification, will be reduced. Therefore, by passing the time each node observes its neighbors' behavior and stores for each of them N_{neg} and N_{all} values, which are the numbers of observed misbehavior and total behavior respectively. If a node shows misbehavior, then it will increase its N_{neg}. It is obvious that $N_{all} - N_{neg}$ is the observed correct behavior. The definition of is:

$$T_{A,B}^{T_k} = \frac{N_{all}(t_k) - N_{neg}(t_k)}{N_{all}(t_k)} \quad (5)$$

IV. PROPOSED METHOD

Fig. 3 shows algorithm flowchart of the proposed method. This section introduces that how a node can detect malicious nodes by employing the received information from the network and reasoning between them. Detection of abnormal behavior includes verification of consistency between OLSR messages and trust-based reasoning that can be carried out by each node in the network. This is a continuous process which will start from reception of first HELLO and TC messages and will take part in calculating routing table. In other words, a continuous

ISSN 2636-9680
eISSN 2682-9266

and recursive checking of trust properties have to be performed, in order to validate all information received from the network. By using concept of trust, it is possible to change the mentioned properties in above to design a mistrust reasoning, so a node is able to protect itself from malicious nodes.

A. Validation in MPR selection

Selection of MPR nodes is the most important phase in OLSR protocol. Nodes get access to the network through their MPRs and this causes them to be known by the whole nodes of network. In OLSR protocol there is not any way to verify the MPR behavior. This vulnerability is exploited by attackers which try to be selected as MPR by a target node and through this, control the target node input–output messages [11]. The most serious reason of vulnerability in MPR selection is in selecting of nodes; because only degree of reachability to two-hop nodes is important and an attacker is able to give wrong information which cannot be verified [7]. Each node has to use the trust concept to have control over its MPR nodes. Based on OLSR protocol specifications, correct behavior of an MPR regarding to routing is definite by two operations: producing TC messages and forwarding data packets and TC messages of nodes which have selected it as their MPR. If a node can validate this and confirm these functions based on MPR's behavior, then a trust relationship will be created correctly, and the node can operate as MPR. Otherwise, if it is not possible for a node to confirm these two functions and does not obtain enough trust, it has to delete that node from its MPR nodes and find other one among other trustful nodes. Hence, to achieve this goal a blacklist is employed for the selected MPRs that the malicious node will be added to this blacklist. Nodes in the blacklist will not be selected as MPR in the future, but after an expiration time the node will be remove from the blacklist and will have a new chance to be selected as MPR again.

If an MPR node like y has generated false TC messages, and nodes, which use it as MPR node, do not advertise this, then trust to this will be decreased, it has to be deleted from MPR nodes set and has to be inserted to the black list and alternative node or nodes has to be selected as MPR nodes.

If a y's selected MPR does not forward data packets and TC messages, then trust to this node has to be decreased and be deleted from the MPR nodes set and be inserted to the blacklist. Also, alternative node or nodes have to be selected as MPR nodes.

ISSN 2636-9680
eISSN 2682-9266

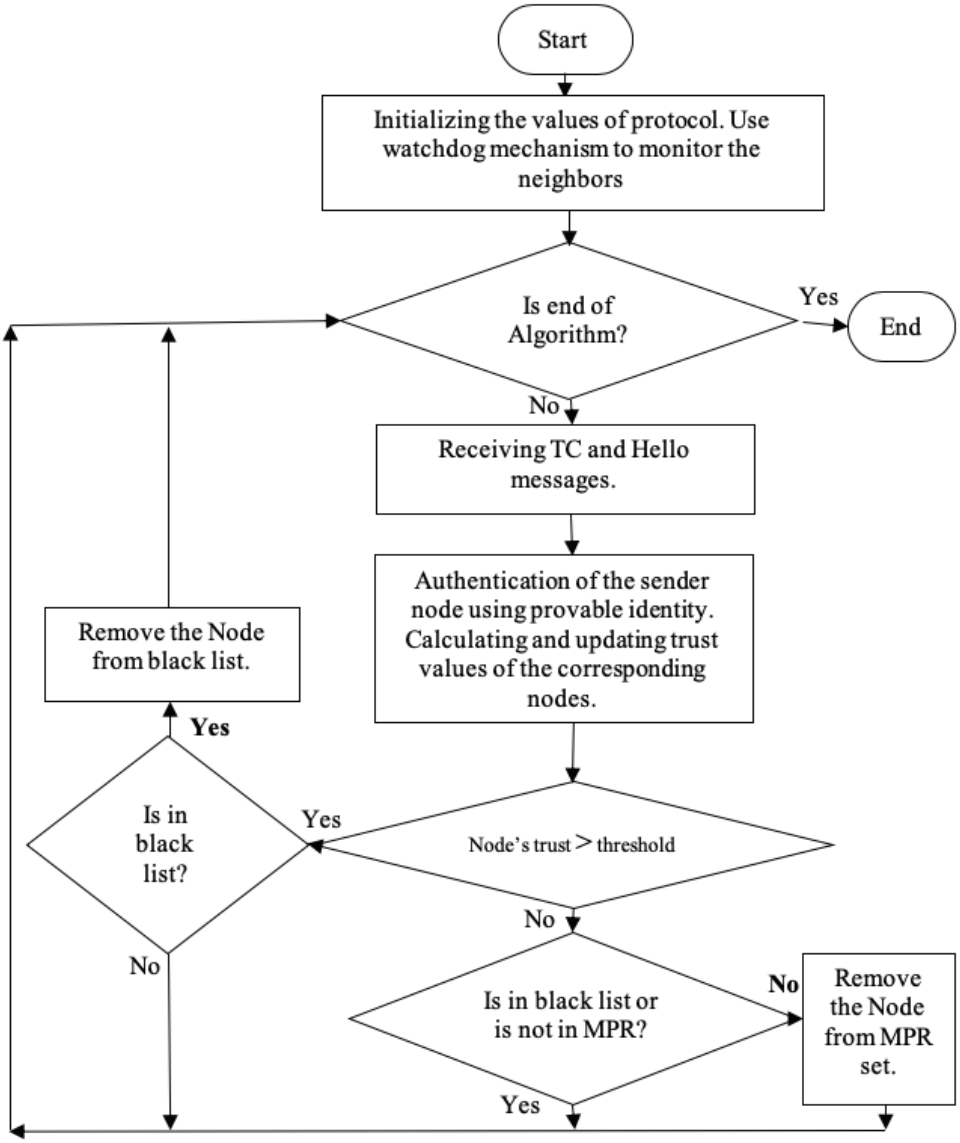

Fig. 3: The proposed method.

ISSN 2636-9680
eISSN 2682-9266

V. SIMULATIONS AND RESULTS

In **TABLE 1**, the proposed method are compared with OLSR protocol. In the proposed method, it is possible to define the required security level for each connection by changing the threshold of it. For example, a multimedia and secret connection needs a high security route and as result a higher threshold in comparison with a file sharing connection. On the other hand, routing overhead in the proposed is higher than OLSR protocol. Authentication of nodes and being confidence about their ID in both methods is performed by using provable identity mechanism. None of these methods needs a centralized entity for operating and they are still compatible with OLSR protocol. Suggestions approach, which is used in credit systems, includes receiving trust value as a suggestion from other nodes about a special node.

TABLE 1: Explanation of the Main KSA Descriptor Sections

	OLSR Protocol	Proposed Method
Security Mechanism	None	Based on trust
Trust history	None	Have
Routing Overhead	Minimum	High
Qualifying nodes origin	None	Provable identity and OLSR protocol
Protocol compatible with OLSR	Yes	Yes

For simulation, NS2 simulator is used. In the simulations, MANETs is composed from 50 nodes which are placed in a flat network with the dimensions 500x500 square meters randomly. In addition, a node's range which has been identified as a step is 250m. We have considered that nodes are not mobile, due to allow the attacker to do its attack over time and also the threshold value is initialized to 0.8.

The attacker node is selected randomly and then it will perform its attack's scenario. **Fig. 4** shows the rate that a malicious node is detected by its neighbors. While, one of the basic assumptions about OLSR protocol is that the whole nodes of network are correct, any security mechanism for protection against malicious nodes are used, and all attacks from malicious nodes will be done successfully. The OLSR method and proposed method have same results in Simulation number 5, because two methods have the same behavior in this number.

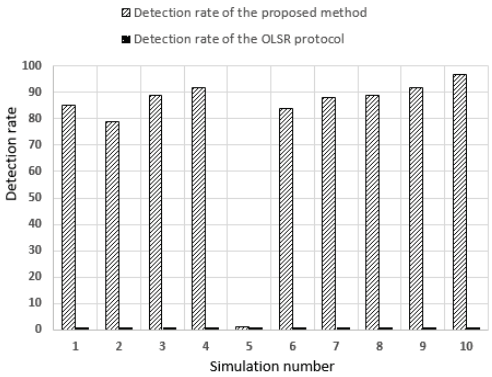

Fig. 4: Comparison the detection rate of the proposed method with OLSR protocol.

In **Fig. 5** is presented the evaluated trust value of the proposed method and Adnane's method [14]. The simulations, which have been done, are in an identical state for both methods and each curve shows the evaluated trust value by the first neighbor of the malicious node.

ISSN 2636-9680
eISSN 2682-9266

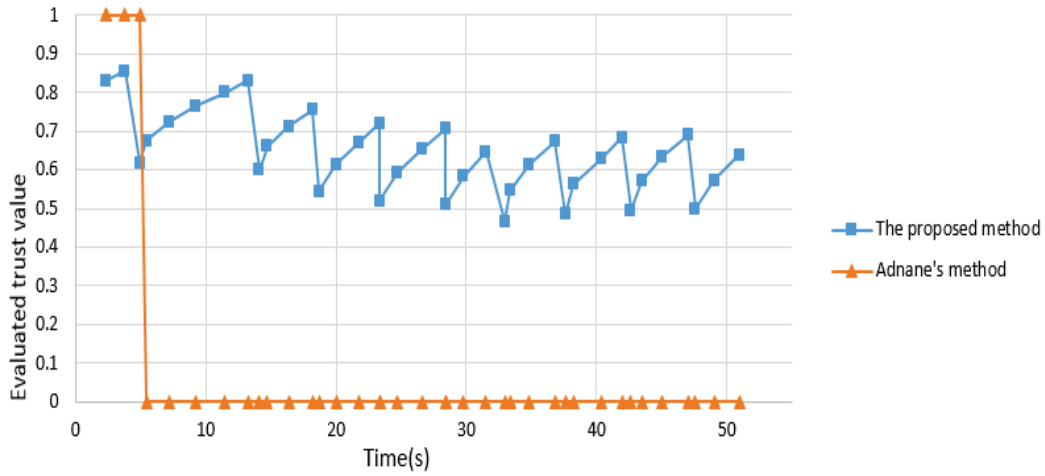

Fig. 5: Evaluated trust value by the first neighbor.

VI. CONCLUSION

It is possible to calculate the trust value of each node in the network by monitoring the network, analyzing received information from neighbors and filtering them. Moreover, by storing and comparing the messages, it is possible to detect false relations and malicious nodes from chain of messages. This solution gives the possibility to use a memorial method in OLSR routing protocol to calculate trust of a node.

The proposed method unlike Adnane's method creates its trust relations by calculating the trust history of a node continuously. The simulations, which have been done, are in an identical state for both methods and each curve shows the evaluated trust value by the first neighbor of the malicious node. The simulations have been presented effectiveness of the proposed method in detecting malicious nodes.

VII. REFERENCES

[1] A. Boukerche, et al., "Routing protocols in ad hoc networks: A survey", *Computer Networks*, 2011.

[2] D.B. Johnson, D.A. Maltz, J. Broch, "DSR: the dynamic source routing protocol for multi-hop wireless ad hoc networks," in: C.E. Perkins (Ed.), *Ad Hoc Networking*, Addison-Wesley, 139–172, 2001.

[3] T. Clausen, P. Jacquet, "Optimized Link State Routing Protocol OLSR," IETF RFC-3626, 2003.

[4] S. Buchegger, J-Y. Le Boudec, "Performance analysis of the confidant protocol: cooperation of nodes – fairness. Dynamic Ad-hoc networks," *Proceedings of IEEE/ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC)*, IEEE , 2002.

[5] K. Meka, M. Virendra, S. Upadhyaya, "Trust based routing decisions in mobile ad- hoc networks," In: *Workshop on Secure Knowledge Management (SKM)*, 2006.

[6] Y. Hu, YA. Perrig, D. Johnson, "Ariadne: a secure on-demand routing protocol for ad hoc networks," *Wireless Networks*, vol. 11, Kluwer Academic Publishers, 21–38, 2002.

[7] A. Adnane, C. Bidan, CR. Timóteo, "Trust-based security for the OLSR routing protocol," *Computer Communication*, 2013.

[8] S. Gadekar, S. Kadam, "Secure optimized link state routing (OLSR) protocol against node isolation attack," *IEEE International Conf. on Power, Control, Signals and Instrumentation Engineering (ICPCSI),* Chennai, India, 2017.

ISSN 2636-9680
eISSN 2682-9266

[9] C. Mahdvi, P. Bhanu, "Prevention of DOS and routing attack in OLSR under MANET," *International Journal of Engineering Science and Computing*, Vol. 7, No. 4, 2017.

[10] R. Yahalom, B. Klein, T. Beth, "Trust relationships in secure systems – a distributed authentication perspective," In: *SP'93: Proceedings of the 1993 IEEE Symposium on Security and Privacy, IEEE Computer Society*, Washington, USA 150–164, 1993.

[11] L. Buttyan, J-P. Hubaux, "Nuglets: A virtual currency to stimulate cooperation in self-organized mobile ad hoc networks," *Proc. of Technical Report DSC/2001/001, Swiss Federal Institute of Technology – Lausanne, Department of Communication Systems*, 2001.

.

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Journal of Cyber Security
Volume 2, Issue 1 (February 2020)
87 – 96

The Development of Constraints in Role-based Access Control: A Systematic Review

Nazirah Abd Hamid [1], Rabiah Ahmad [2], and Siti Rahayu Selamat [3]
[1,2,3] Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia
[3]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Besut, Terengganu, Malaysia
[1]**owenira@gmail.com**

ARTICLE INFO

Article History
Received 22 May 2019
Received in revised form 15 Aug 2019
Accepted 25 Sep 2019

Keywords:
access control, role-based, role mining, user-permission assignment, constraints

ABSTRACT

Role-based access control (RBAC) model attracts many organizations to transform their traditional access control method to RBAC model mainly because of the security features in the RBAC. The RBAC model is generated to achieve security objectives and to do so, RBAC enforces various constraints to accomplish those objectives. Up to now, very limited studies provide systematic review of constraints in roles mining and in this study, we focus on the publications that published during 2011–2018. The main objective of this study is to recommend conceptual understanding through a systematic review by classifying the constraints and its proposed solutions. The analysis offered variety of areas that can be explored in leveraging constraints in the role mining development hence providing an improvement to roles-based access control growth.

I. INTRODUCTION

Access control model is very needed by an organization to administer their security policies and the resources. Role engineering or also known as role mining in RBAC can be described as a process to discover an appropriate set of roles that could perform as a control mechanism to access the organization's resources [1].

Nevertheless, to ensure that the generated roles are signifying the organizations security policies and user requirements persist a difficult challenge. Thus, there is a need to introduce the constraints that could govern the implementation of those policies and requirements especially in designing and developing role mining algorithms [1].

The development of constraints in RBAC are still quite limited so the key contribution of this paper is to present conceptual understanding through a systematic review by classifying the constraints and its proposed solution. The analysis offered variety of areas that can be explored in leveraging constraints in the role mining development.

This paper is adapted from the work presented by [2]. The remainder of this paper is organized as the following: section II provides the preliminaries study on the various constraints. Then in the section III shows the methodology of systematic literature review process. In section IV explains the data extraction and analysis process and, in the section, presents and classifies the constraints in role mining. Lastly, section VI and VII describe the discussion and conclusion of this study.

ISSN 2636-9680
eISSN 2682-9266

II. PRELIMINARIES

In this section, we would examine the notions of constraints in role-based access control specifically in role mining.

The concept of constraints in RBAC models was introduced by the authors in [3] and they classified the constraints into several categories namely mutually exclusive roles, cardinality constraints and prerequisite roles. Moreover, the authors also characterized the constraints into several general classes: cardinality constraints and prerequisite constraints [4].

Furthermore, according to [5] constraints could be described as an important set of rules that governed the architectural structure of RBAC and the constraints were significant to be used as a control and protection mechanism in RBAC because of the nature of a RBAC model that heavily relied on the flow of the security such as who should has the permission to the objects or resources and so on.

The research study by the authors in [5] also have discovered that in the RBAC environment, cardinality constraints administrated the organization security policies. In the real scenario, in the beginning, the chief security officer is needed to list the minimum and maximum requirements of the security officers and users that would involve in the RBAC system.

III. SYSTEMATIC LITERATURE REVIEW

In order to do the survey and analysis on the current state-of-the-art of constraints in role-based access control, we have initiated a systematic literature review based on these research questions:

Question 1: What is the relationship between role-based access control and constraints?

Question 2: Which constraints that involved in existing role mining algorithms?

A. Exploration process

The exploration process has been commenced with the searching activity for the relevant research studies using search engines through digital libraries and databases as illustrated in **TABLE 1**. The keywords or search strings that have been applied were "constraints in role-based access control" and "constraints in role mining".

B. Inclusion and Exclusion Criteria

We have defined some inclusion and exclusion criteria to select the articles for the review process:

a) Inclusion criteria

Studies that have been published during 2011 to 2018 and related to the constraints in role-based access control and role mining. More articles have been discovered through examining the reference list of each one of these articles.

b) Exclusion criteria

The articles that unrelated to the research questions.

TABLE 1 provides the information on the number of research studies that have been ascertained during the exploration process of the digital libraries and databases.

TABLE 1: Number of Articles and Databases

Databases	Number of Articles		
	Based on keywords	Based on titles	Based on abstracts
ACM Digital Library	10	3	10
IEEE Xplore	7	3	11
Science Direct	4	2	4
Scopus	19	22	27
Google Scholar	19	19	26

ISSN 2636-9680
eISSN 2682-9266

Total	**59**	**49**	**78**

IV. DATA EXTRACTION

For this data extraction section, there were 14 final articles that have been selected based on the inclusion and exclusion criteria as mentioned in section III. The final articles were focusing on "constraints in role-based access control" and "constraints in role mining" and each article has been summarized in **TABLE 2** in the term of the aim of the study and the methodology that involved, and the articles have been ordered by most recent year of publication.

V. DATA ANALYSIS

From the data that has been extracted as in **TABLE 2**, there were two main findings that could be concluded as followed:

a) There were five different classes of cardinality constraints and
b) The algorithms from those articles, have been studied and sorted into the classes of abovementioned cardinality constraints

TABLE 2: Summary of Articles

No	Year	Title	Aim & Methodology	Type of Constraints
1	2018	Dynamic User-Oriented Role Based Access Control Model (DUO-RBAC)	This study discussed a dynamic user-oriented role-based access control model that could accommodate a way of new user-permission assignment (UPA)to be included into the existing one.	The number of role assignments for each user (R-U)
			The model involved three main processes including a preprocessing phase by eliminating users with the same permission, a candidate roles generation phase using a combination formula and lastly in role selection and assignment phase, dynamically choose candidate roles form the uncovered permission.	
2	2017	PRUCC-RM: Permission-Role-Usage Cardinality Constrained Role Mining	This study deliberated on two heuristics algorithms that could be used to execute the two type of constraints by implementing a pre-processing process in one algorithm where role was split and assigned based on the mentioned constraints and for the second algorithm, the conflicting roles were eliminated.	The number of permissions included in a role (P-R)
				The number of roles a user can own (R-U)
3	2016	Performance Evaluation of a Role Based Access Control Constraints in Role Mining Using Cardinality	This paper proposed a new concept of objective that intended to limit the most number of permissions that can be incorporated into the role by designing and developing a Matrix Based Role Assignment (MBRA) algorithm and role miner algorithm.	The number of permissions that can be incorporated in a role (P-R)
			The algorithms were designed by employing the visual method that could represent the UPA in a better manner and facilitated the way of doing analysis rapidly.	

ISSN 2636-9680
eISSN 2682-9266

No	Year	Title	Aim & Methodology	Type of Constraints
4	2016	Role Mining Using Answer Set Programming	This article discussed an innovative way to leverage multiple constraints with numerous optimization objectives. Multiple constraints could cause conflicts if they were not being handled properly and the authors introduced role mining method using answer set programming (ASP) named constrained role miner (CRM).	Multiple constraints
			Fundamentally, ASP is a declarative problem-solving technique that enable a computer intelligently to propose a solution based on a problem such as in this research the conflicts that happen between constraints.	
5	2015	Meeting Cardinality Constraints in Role Mining	The discoursed two strategies in this article were to solve multiple constraints problems simultaneously and in this case, the multiple constraints were known as Multiple Cardinality Constraint Problem (MCP).	The number of roles to which an individual user can own (R-U)
			The two strategies specifically called as the postprocessing strategy and the concurrent processing strategy. The difference between those approaches was the way the user-assignment (UA) and permission-assignment (PA) matrices were obtained where the postprocessing was built without deliberating the constraints meanwhile in concurrent processing both constraints were simultaneously developed.	The number of roles that can include in a permission (R-P)
6	2015	Towards a General Framework for Optimal Role Mining: A Constraint Satisfaction Approach	The authors suggested a technique to convert the role mining problem (RMP) into a constraint satisfaction problem using satisfiability modulo theories (SMT) solvers that permitted RBAC model to be represented into multiple constraints. The transformation enabled to get an optimal RBAC model based on customized optimization metrics.	Multiple constraints
7	2015	Role Mining based on Cardinality Constraints	In this article, a role mining algorithm with the consideration of two constraints namely the number of roles to which an individual user can belong should be limited and the number of roles to which a permission can be assigned should also be restricted.	Multiple constraints

ISSN 2636-9680
eISSN 2682-9266

No	Year	Title	Aim & Methodology	Type of Constraints
			The algorithm was developed by employing an improved graph optimization theory and it consisted of three major phases; generating the initial role set, selecting role pair for role update algorithm and updating the initial role state.	
8	2015	Towards User-oriented RBAC Model	This paper proposed role mining algorithms that complied with the constraint of the number of roles a user can own. The algorithm was designed and developed to solve four different problems specifically i) the user RMP and its approximate ii) the personalized RMP and approximate personalized RMP. The user RMP would allow all the users with the same maximum role assignment while the personalized RMP permitted different values for each user. The approximate versions could be defined as to place a threshold value that enabled a little bit deviation from the complete reconstruction.	The maximum number of roles each user can have (R-U)
			The heuristic algorithms consisted of i) candidate role generation phase and ii) role selection and assignment phase.	
9	2015	Role Mining Based on Permission Cardinality Constraint and User Cardinality Constraint	The constraints that involved in this study were permission cardinality constraint and user cardinality constraint and the authors developed a role mining algorithm to leverage those constraints while minimize the assignment cost. The algorithm comprised of three phases namely initial role set generation phase, role selection phase, and role state generation phase.	Multiple constraints
10	2014	Visual Approach to Role Mining with Permission Usage Cardinality Constraint	In this research, a graphic or visual approach was introduced to illustrate the UPA and eventually an analysis and optimal roles with constraint could be extracted rapidly. The authors built two heuristics algorithms particularly ADVISER and t-SMAR that could produce a UPA with the intended constraint.	The number of permissions that can be included in a role (P-R)

ISSN 2636-9680
eISSN 2682-9266

No	Year	Title	Aim & Methodology	Type of Constraints
11	2013	Towards User-Oriented RBAC Model	A heuristic solution was proposed to enforce the intended constraint and the solution utilizing a dynamic role generation that used an iterative technique to uncover optimum roles. The author also mentioned that the algorithm was integrated with the end-user standpoint. In this study, two main role mining algorithms were designed i) user-oriented exact and user-oriented approximate and the difference was the threshold value of the reconstruction of the UPA where exact algorithm had to completely mimic the UPA whilst the approximate allowed some deviation.	The number of roles each user can have (R-U)
12	2012	Constrained Role Mining	The authors described two heuristic techniques to enforce the permission cardinality constraint, namely t-SMAR and t-SMAC to reconstruct a complete UPA and the they claimed that the algorithms could be expended to other types of cardinality constraints.	The number of permissions included in each role (P-R)
13	2012	Role Mining under Role-Usage Cardinality Constraint	Study by this article discovered two techniques to solve the role-user cardinality constraint problem in building the UPA. The first technique called as Role Priority based Approach (RPA) involved the process of ranking the roles based on the sizes and then enforced the constraint using the produced ranking. The Coverage of Permissions based Approach (CPA) as the second technique worked by choosing a role based on the largest uncovered number of permissions to impose the intended constraint.	The number of roles any user can have (R-U)
14	2011	Towards Role Mining with Restricted User-Role Assignment	Three algorithms were introduced by the authors to implement the number of users assigned to any role constraint and those algorithms could accommodate different requirements. The algorithms were designed by considering the problem of finding smallest biclique cover of the edges of a bipartite graph.	The number of users assigned to any role (U-R)

ISSN 2636-9680
eISSN 2682-9266

A. Role-user cardinality constraint

The role-user cardinality constraint of role could be defined as the maximum number of roles that can be assigned to a user (R-U) or in the other words, the users could only perform a task based on the privileges that have been granted to them only as mentioned in [5][6].

The authors highlighted the need of a dynamic user-oriented role-based access control algorithm that could accommodate a way of new user-permission assignment (UPA) to be included into the existing one while the constraints were being maintained [7].

According to [8], a heuristic algorithm was designed and developed to solve four different problems specifically i) user RMP, ii) approximate user RMP, iii) personalized RMP and iv) approximate personalized RMP. The proposed algorithms were complied with the constraint of the number of roles a user can own and consisted of i) candidate role generation phase and ii) role selection and assignment phase.

A heuristic solution was proposed to enforce the intended constraint and the solution utilizing a dynamic role generation that used an iterative technique to uncover optimum roles. This paper [9] also mentioned that the algorithm was integrated with the end-user standpoint. Two role mining algorithms were designed, i) user-oriented exact and user-oriented approximate where exact algorithm had to completely mimic the UPA whilst the approximate allowed some deviation.

In this study, the authors discovered two techniques and the first technique called as Role Priority based Approach (RPA) involved the process of ranking the roles based on the sizes and then enforced the constraint using the produced ranking. The Coverage of Permissions based Approach (CPA) as the second technique worked by choosing a role based on the largest uncovered number of permissions to impose the intended constraint [10].

B. User cardinality constraint

The user cardinality constraint of user can be described as a constraint that restrict the number of users to which a role can have (U-R). For example, if a RBAC model allocates a huge number of users to a specific role, then the security officer would have difficulty to administrate the RBAC system.

So, to resolve the abovementioned constraint, three algorithms were introduced by the authors [11] to implement the number of users assigned to any role constraint and those algorithms could accommodate different requirements. The algorithms were designed by considering the problem of finding smallest biclique cover of the edges of a bipartite graph. Fundamentally, biclique problem could be expressed as a problem to discover a node that could stimulate a thorough subgraph [12].

C. Role-permission cardinality constraint

The role-permission cardinality constraint can be expressed as the maximum number of roles to which a permission can belong (R-P). One of the study [13] designed and developed two strategies known as the postprocessing strategy and the concurrent processing strategy. Moreover, the algorithm was developed by employing an improved graph optimization theory and it consisted of three major phases [14]. Both articles involved in the multiple constraints research.

D. Permission cardinality constraint

This permission cardinality constraint is applied to determine the maximum number of permissions that can be present in a role (P-R) and in [15] a new concept of objective was proposed by designing and developing algorithms that have employed the visual method that could represent the UPA in a better manner and facilitated the way of doing analysis rapidly.

In this research, a graphic or visual approach was used to illustrate the UPA and eventually an analysis and optimal roles with constraint could be extracted rapidly. The two heuristics algorithms were produced particularly ADVISER and t-SMAR that could construct a UPA with the intended constraint [16].

ISSN 2636-9680
eISSN 2682-9266

The two heuristic techniques could be applied to enforce the permission cardinality constraint, namely, t-SMAR and t-SMAC to reconstruct a complete UPA and the authors claimed that the algorithms could be expended to other types of cardinality constraints [17].

E. Multiple constraints

This study deliberated on two heuristics algorithms that could be used to execute the two type of constraints specifically constraints on the number of permissions included in a role (P-R) and the number of roles a user can own (R-U) by implementing an algorithm with a pre-processing process and another algorithm without the pre-processing stage [1].

The authors investigated two strategies to solve multiple constraints problems simultaneously known as Multiple Cardinality Constraint Problem (MCP). The multiple constraints that involved were known as dual of each other specifically a role-user cardinality constraint (R-U) and its dual namely role-permission cardinality constraint (R-P). The two strategies were called as the postprocessing strategy and the concurrent processing strategy. The difference between those approaches was the way the user-assignment (UA) and permission-assignment (PA) matrices were obtained, where the postprocessing was built without deliberating the constraints meanwhile, in concurrent processing both constraints were simultaneously developed. [13]

In this article [14], a role mining algorithm with the consideration of two constraints namely the number of roles to an individual user can belong (R-U) and the number of roles can be assigned in a permission (R-P). The algorithm was developed by employing an improved graph optimization theory [18] and it consisted of three major phases.

The constraints that involved in this study were permission cardinality constraint (P-R) and user cardinality constraint (U-R) and the authors [19] developed a role mining algorithm to leverage those constraints while minimize the assignment cost. The algorithm comprised of three phases namely initial role set generation phase, role selection phase, and role state generation phase.

According to the authors in [20], they discovered an innovative way to leverage the conflicts between multiple constraints with numerous optimization objectives using answer set programming (ASP) named constrained role miner (CRM) that enable a computer intelligently to propose a solution based on a problem.

The authors [21] suggested a technique to convert the role mining problem (RMP) into a constraint satisfaction problem using satisfiability modulo theories (SMT) solvers [22] that permitted RBAC model to be represented into multiple constraints. The transformation enabled to get an optimal RBAC model based on customized optimization metrics.

VI. DISCUSSION

From the exploration process that has been described in the section III, there were 78 publications have been published from 2011-2018 and after the inclusion and exclusion criteria, 14 related studies have been selected for further analysis.

In the data analysis sections as in section V, there are two main findings that needed to be deliberated. Firstly, there are five different classes of cardinality constraints that can be discovered as the following:

a) Role-user cardinality constraint (R-U) - the number of roles a user can own.
b) User cardinality constraint (U-R) - the number of users to which a role can have.
c) Role-permission cardinality constraint (R-P) - the number of roles that can include in a permission.
d) Permission cardinality constraint (P-R) - the number of permissions that can be incorporated in a role.

ISSN 2636-9680
eISSN 2682-9266

e) Multiple constraints – involve the enforcement of two constraits as above.

Secondly, over the years, the research on constraints in the RBAC have attracted many researchers but the resources are quite inadequate. Based on the reviews, there are some discussion that could be summarized:

a) Most of the researchers have concentrated on findings the solutions for a single constraint and the most discussed constraint is the role-user cardinality constraint.
b) The researchers have designed and developed heuristics algorithms because they have showed that the role mining problem of constraints are NP-hard.
c) For the multiple constraints' techniques, most of them have developed two separate strategies to enforce the constraints.

Furthermore, all the selected articles have been using some or all the nine real-world datasets as portrayed in **TABLE 3**. The datasets have been introduced by [23] and the elements in **TABLE 3** can be denoted as |U| as user and |P| as permission.

TABLE 3: Elements in the Datasets

Dataset	|U|	|P|
americas_large	3485	10127
americas_small	3477	1587
apj	2044	1164
emea	35	3046
healthcare	46	46
domino	79	231
customer	10021	277
firewall1	365	709
firewall2	325	590

VII. CONCLUSION

The development of constraints in RBAC are still quite limited and the main contribution of this paper is to present a systematic review and the analysis offered variety of areas that can be explored in the research of constraints in the role mining development.

There are some potential paths that can be explored by the researchers in the future. The first path is to propose a dynamic solution to find more interesting roles that can resultant more accurate results and for the second direction in the multiple constraints, the researchers could explore some techniques to enforce the constraints to be executed simultaneously.

VIII. REFERENCES

[1] C. Blundo, S. Cimato, and L. Siniscalchi, "PRUCC-RM: Permission-Role-Usage Cardinality Constrained Role Mining," in *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, 2017, pp. 149–154.

[2] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared, "Approaches to cross-domain sentiment analysis: A systematic literature review," *IEEE Access*, vol. 5, pp. 16173–16192, 2017.

[3] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *IEEE Comput.*, vol. 29, no. 2, pp. 38–47, 1996.

[4] Q. Jiong and M. Chen-hua, "Detecting and resolving constraint conflicts in role-based access control," in *2011 International Conference on Electrical and Control Engineering*, 2011, pp. 5845–5848.

[5] X. Ma, R. Li, Z. Lu, and W. Wang, "Mining constraints in role-based access control," *Math. Comput. Model.*, vol. 55, no. 1–2, pp. 87–96, Jan. 2012.

[6] B. Mitra, S. Sural, J. Vaidya, and V. Atluri, "A survey of role mining," *ACM Comput. Surv.*, vol. 48, no. 4, pp. 1–37, 2016.

[7] H. Kiwan and R. Jayousi, "Dynamic User-Oriented Role Based Access

ISSN 2636-9680
eISSN 2682-9266

Control Model (DUO-RBC)," in *EDA*, 2018, pp. 281–290.

[8] H. Lu, Y. Hong, Y. Yang, L. Duan, and N. Badar, "Towards user-oriented RBAC model," *J. Comput. Secur.*, vol. 23, no. 1, pp. 107–129, 2015.

[9] H. Lu, Y. Hong, Y. Yang, L. Duan, and N. Badar, "Towards user-oriented RBAC model," *Lect. Notes Comput. Sci.*, vol. 7964, pp. 81–96, Mar. 2013.

[10] J. C. John, S. Sural, V. Atluri, and J. S. Vaidya, "Role mining under role-usage cardinality constraint," in *IFIP Advances in Information and Communication Technology*, Springer Berlin Heidelberg, 2012, pp. 150–161.

[11] M. Hingankar and S. Sural, "Towards role mining with restricted user-role assignment," in *2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE)*, 2011, pp. 1–5.

[12] D. S. Hochbaum, "Approximating clique and biclique problems," *J. Algorithms*, vol. 29, no. 1, pp. 174–200, Oct. 1998.

[13] P. Harika, M. Nagajyothi, J. C. John, S. Sural, J. Vaidya, and V. Atluri, "Meeting cardinality constraints in role mining," *IEEE Trans. Dependable Secur. Comput.*, vol. 12, no. 1, pp. 71–84, 2015.

[14] R. Li, H. Li, X. Gu, Y. Li, W. Ye, and X. Ma, "Role Mining based on Cardinality Constraints," *Concurr. Comput. Pract. Exp.*, vol. 27, pp. 3126–3144, 2015.

[15] Y. R. More and S. V Gumaste, "Performance evaluation of a role based access control constraints in role mining using cardinality," *Int. J. Adv. Res. Sci. Manag. Technol.*, vol. 2, no. 7, pp. 1–7, 2016.

[16] P. Sarana, A. Roy, S. Sural, J. Vaidya, and V. Atluri, "Role mining in the presence of separation of duty constraints," in *International Conference on Information Systems Security*, 2015, pp. 98–117.

[17] C. Blundo and S. Cimato, "Constrained role mining," in *International Workshop on Security and Trust Management*, 2012, pp. 289–304.

[18] D. Zhang, K. Ramamohanarao, and T. Ebringer, "Role engineering using graph optimisation," in *Proceedings of the 12th ACM symposium on Access control models and technologies - SACMAT '07*, 2007, p. 139.

[19] X. Ma, R. Li, H. Wang, and H. Li, "Role mining based on permission cardinality constraint and user cardinality constraint," *Secur. Commun. Networks*, vol. 8, no. 13, pp. 2317–2328, 2015.

[20] W. Ye, R. Li, X. Gu, Y. Li, and K. Wen, "Role mining using answer set programming," *Futur. Gener. Comput. Syst.*, vol. 55, pp. 336–343, Feb. 2016.

[21] J. H. Jafarian, H. Takabi, H. Touati, E. Hesamifard, and M. Shehab, "Towards a general framework for optimal role mining," in *Proceedings of the 20th ACM Symposium on Access Control Models and Technologies - SACMAT '15*, 2015, pp. 211–220.

[22] L. De Moura and N. Bjørner, "Satisfiability modulo theories: introduction and applications," *Commun. ACM*, vol. 54, no. 9, p. 69, Sep. 2011.

[23] A. Ene, W. Horne, N. Milosavljevic, P. Rao, R. Schreiber, and R. E. Tarjan, "Fast exact and heuristic methods for role minimization problems," in *Proceedings of the 13th ACM Symposium on Access Control Models and Technologies*, 2008, pp. 1–10.

ISSN 2636-9680
eISSN 2682-9266

OIC-CERT Permanent Secretariat:
CyberSecurity Malaysia
Level 7, Tower 1, Menara Cyber Axis,
Jalan Impact, 63000 Cyberjaya,
Selangor Darul Ehsan,
Malaysia.

secretariat@oic-cert.org
www.oic-cert.org