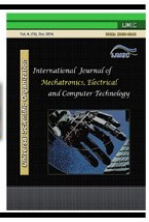
Vol. 7(25), Jul. 2017, PP. 3552-3562

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543



A New Authentication Method in Cloud-based Big Data in Order to Increase Data Security

Maryam Azimian[1*] and Rasoul Rustaei[2]

[1]*M.Sc., Computer Department, Malayer Branch, Islamic Azad University, Malayer, Iran*

[2]*Ph.D., Computer Department, Malayer Branch, Islamic Azad University, Malayer, Iran*

***Corresponding Author's E-mail:** m.azimian1983@gmail.com*

Abstract

Today, data are on a raise in many organizations. Big data is referred to a mechanism which is used by many organizations. Due to growth of cloud computing in recent years, the issue of big data has received very much attention. Many of organizations agree with use of big data as a solution to increase service quality to people and customers. Use of big data, namely, analysis, search and data mining of big data results in extraction of templates which leads to improve quality considered service or product. Concerns about breaching the people' privacy is a great challenge in the field of data mining and pattern extraction in big data. This paper investigates methods to protect people's privacy during data mining on big data and provides a new multi-factor authentication method. Consequently, it has been showed that the proposed method has a better performance compared to previous research.

Keywords: *cloud computing, security, big data, authentication.*

1. Introduction

Since databases have been progressed in recent years, therefore, their control and management is almost difficult and impossible in many cases. In these locations, databases are big enough that cannot more be controlled or analyzed sufficiently. Any operation which can be exerted in small databases with no problem, might face problem when addressing big data. This issue can be referred as a problem in big data [1]. Policies related to privacy and security is a challenge. Because, it protects privacy pertinent to information which are more likely to be accessed through internet. In other words, individuals who intend to access this data, will access via internet to private information of other users who don't want anybody access their information. Despite privacy protection is very important for these users, but most of them don't care about exposing these information when using social networks and increase the risk making the data visible for other users in various ways [2]. Each user has a working profile in big database in which saves his/her data in the profile. In databases in which data are exchanged very much, security is also very important. Many hackers and profiteer individuals lie in ambush for these data. User authentication is one of the most important parts which prevents unauthorized people to enter in order to keep data secure [2].

Article History:
Received Date: Mar. 23, 2017
Accepted Date: May. 03, 2017
Available Online: Jul. 04, 2017

2. Authentication

Authentication is referred to a process or operation during which individual's identity is investigated and validated or rejected, for example, when a service receiver (user) wants to connect to a service provider, the user's code (e.g. user name and password) is sent toward service provider so that service provider compares the data with information existent in its database, and if the data is correct, the user will enter into the next step [3].

Of disadvantage of the system are that password can be stolen, exposed or forgotten. For this reason, online businesses, banking transactions, and other important activities on internet and network will need other processes (other than authentication). Use of digital certification which is measured by Certificate Authority (CA) is a part of structure of a public key that has changed to an authentication standard on internet today.

As it was mentioned, the data bank can be implemented within equipment used in low capacity networks. In relatively high capacity networks in which a specific authentication system is required, active equipment of the network are configured in way that authentication is performed using databases established on specific systems for this process.

3. Previous works

In [4], the author has proposed two-factor authentication (2FA) method. In the first step, user is measured by username and password and smart card, and then, is validated by Out-Of-Band (OOB). The problem of this method is that smart card is prone to theft. Since message sending is possible only through saved data in the smart car, attacker can introduce himself as the legal user by robbery. Other disadvantage of the method is that one-time password might lead to phishing attacks, also, device clock of the user always must be accorded with service provider's clock.

Another method [5] is use of Elliptic Curve Cryptography (ECC) in which an n-rate curve equation and a base point is selected and private and public keys are produced. The size of curve determines intricacy of the problem. More intricate and more difficult implementation is of characteristics of the method which results in development of performance errors and consequently, reduces algorithm security.

One-time password in [6] is produced by algorithm. This algorithm is different from common algorithms. Since, a password which is random and non-repeated must be produced. Each new password must be unique. The main item of this algorithm is a seed which is shared between user and service provider. During authentication process, service provider and user separately produce OTP. The user sends its OTP to the service provider for authentication and service provider compares it with its own produced OTP. If both are same, authentication is performed correctly. Software developers can use this algorithm for one-time password production. Password production software can be arranged in hardware. One-time password can be produced in two ways: 1) production based on code validation message via hash (Hash Message Authentication Code-HMAC), and 2) production based on Time-Based One-Time Password (TOTP) [6]. In HMAC, both user and service provider have a same primary seed. The user produces password by seed and any other input (e.g. QIN) and sends it via updating the seed. Service provider also produces the password in the same way. If both passwords are same, authentication will be successful. In TOTP, service provider and user use unified clock and an algorithm which produces password on the basis of unified clock. The rest of steps are same as those of HOTP.

Coordinating the clock of user's and service provider's devices and as it was mentioned earlier, phishing attacks are of problems in this method. Usually, phishing attackers guide the user to a fake

website similar to the main website. User sends its produced OTP password to the fake site, and now, the fake website can be logged in the main site using the obtained OTP.

Another proposed 2FA method is authentication using zero-knowledge proof [7]. At first, user's identity is measured using username and password. The second factor is validation file stored in the user's USB or phone. Of advantage of this method is that the password should not be saved in cloud and the defect of the method is that entrance will not possible if validation file is stolen or lost.

In [8], the author has provided a biometric model though recognizing the finger print and ear model in order to enhance protection and efficiency. This model can be used in various fields and programs such as authentication in banking transactions. In [9], the author has used the idea that physiological characteristics are unique. The author believes that these characteristics are less compromised compared to password, token, etc. The proposed model works by iris scan and helps data stream control. A sensor takes the iris image and then performs some processing such as noise reduction on the image. Euclidean distance method is used in order to match two images and according to the result of performed query, the decision to grant permission is taken.

Another framework provided in the literature review [10] is multi-user authentication. The proposed architecture has four phases of registration, entrance, authentication, and password change. The provided model has two advantages. The first is that it has an additional factor (OOB) which offers better security in two-factor authentication and the second is that attacks from two separate communication channels are prevented. General algorithm of the provided framework is that owner of smart card inserts it and then enters username and password. At first, a local system performs authentication. After authentication, entrance request is sent to cloud service provider. The service provider as soon as receiving the request, sends one-time password through HTTP/SMS port to the requesting user. Mobile network delivers the ley to the user and finally, user is authenticated based on smart ID card and one-time password. The provided framework in the first step, offers identity management and mutual authentication between user and service provider and then assigns user access rights.

In [11], also a mechanism similar to OTP, or more accurately, authentication by entrance permission and one-time password is provided. Of differences of this approach is that authentication is not more performed by local system in the first step. User must enter into legal provider's website through secured browser by https protocol and must enter the information, if the entered data is correct, one-time password will sent to his/her cell phone and user enters the password, service prodder compares the entered password and its own sent password and in case of being similar, will issue access permission.

A conventional and secure two-level solution based on hardware (USB) is provided in [12]. The proposed model is a solution for limitations in hardware model. The model is composed of two phases of SETUP and Login, and USB is of requirements in authentication and user recognition is not possible without it.

4. The proposed idea

The proposed method in this paper is a multi-factor authentication method to protect data privacy. The method include the following components.

4.1. Multi-factor encryption

A strong encryption has been used in order to send data and identity information in the proposed method as explained in the following.

Abbreviations used in this algorithm which uses multiple signatures are as follows:

C: user
S: service provider
ns, nc: new random number
SpecificationC: cryptic specification of C
SpecificationS: cryptic specification of S
SCS: a cryptic pre-master used to produce shared key
EPS [SCS]: cryptography of Scs with public key of S (PS) using IBE encryption algorithm
M: all of messages after ClientHello message
SigSC([M]): signature of M message with private key of C (SC) using IBS signature verification algorithm
Verpc (SigSc ([M])): verification of SigSC([M]) with the help of PC using IBS signature verification algorithm
DSs(Eps[Scs]): decoding the EPS [SCS] with the help of private key of C (SC) using IBE decoding algorithm

According to Figure 1, user C in the first step sends ClientHello message to service provide S. The message include nc (new random number), SID (session ID) and SpecificationC (cryptic specification of C).

SpecificationC uses improved TTLS which has been explained. For example, cryptic specification might be in the form of Advanced Encryption Standard (AES). IBS, IBS are used as provider of communication security. MID5 is a sample mixing function. AES is a symmetrical encryption algorithm. The message of ClientHelloDone means end of the first step. In the second step, service provider S answers with ServerHello message which include ns (new random number), SID (session ID) and SpecificationS (cryptic specification of S). SpecificationS is the collection of encryption supported by service provider S. The message of ServerHello means end of the second step.

In the third step, user C firstly selects pre-master SCS and encrypts it using public key of service provider S (PS) and using IBE encryption algorithm. Cryptic text as ClientKeyExchange and ClientKeyExchange messages are sent to service provider S. Then, user C creates a signature (SigSC([M])) and sends it as IdentityVerify message to service provider S. finally, message of ClientFinished means end of the third step.

In the fourth step, service provider S firstly obtains public key of user C (PC) with the help of IDc and investigates SigSC([M]) signature using PC in IBS signature validation algorithm. User C will be validated if only possess valid IDC. This operation completes validation of C by S. Then, service provider S decrypts Eps[Scs] with its private key SS. due to novelty of SSC, correct decryption reveals that S possesses a valid IDS. This step validates validation of S. The message of ServerFinished means end of the fourth step. Finally, a shared cryptic key is calculated between S and C by KCS=PRF(SCS, nC, nS). PRF is a semi-random function (a PRF is a definitive function which maps two separate collection).

$$
\begin{aligned}
&(1)\ C \rightarrow S: \textbf{ClientHello}\ (n_C, SID, specification_C)\\
&\qquad\qquad \textbf{ClientHelloDone}\\
&(2)\ S \rightarrow C: \textbf{ServerHello}\ (n_S, SID, specification_S)\\
&\qquad\qquad \textbf{ServerHelloDone}\\
&(3)\ C \rightarrow S: \textbf{ClientKeyExchange}\ (E_{P_S}[S_{CS}])\\
&\qquad\qquad \textbf{IdentityVerify}\ (Sig_{S_C}[M])\\
&\qquad\qquad \textbf{ClientFinished}\\
&(4)\ S \rightarrow C: \textbf{ServerFinished}(Ver_{ID_C}(Sig_{S_C}[M]),\\
&\qquad\qquad\qquad\qquad D_{S_S}(E_{P_S}[S_{CS}]))
\end{aligned}
$$

Figure 1: Encryption operation and digital signature

4.2. Using the pattern password

One of the factors of password in the proposed method is in the form of pattern. Firstly, user validation pattern is saved in the registration phase. It is supposed that processing phase of validation pattern registration is secure and protocol of registered pattern is not revealed.

Validation pattern includes 3 lines of 4 pixels to be selected from matrix showed in Figure 2. This type is the first type of validation pattern which requires some limitations in neighbor areas.

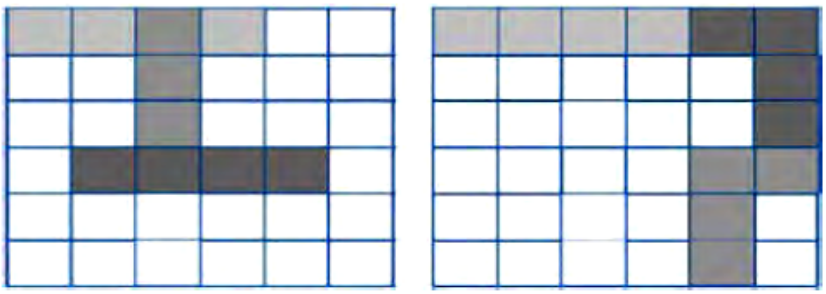


Figure 2: Graphical validation pattern

4.3. Typing pattern password

In this method, user selects a text password during registration so that the shape of keys together on the keyboard is saved in the system as a pattern. For example. User selects “567ygvbn” as password that its typing pattern is showed in Figure 3.

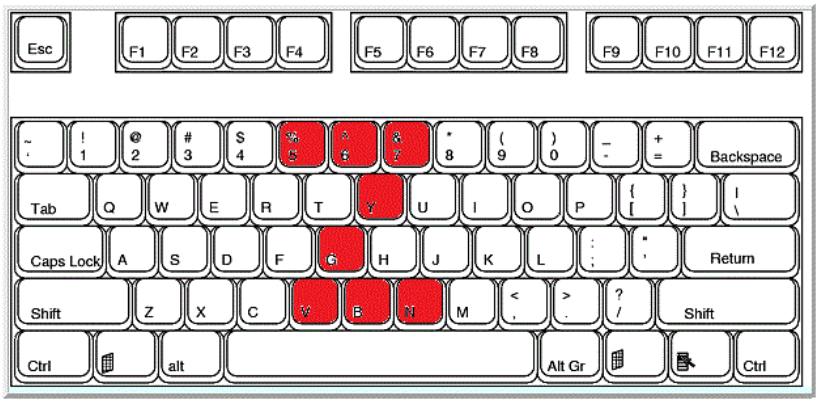


Figure 3: Typing pattern

Hereafter, user must only observe the typing pattern. For example, user can enter the system by typing "345rdxcv" which is showed in Figure 4.

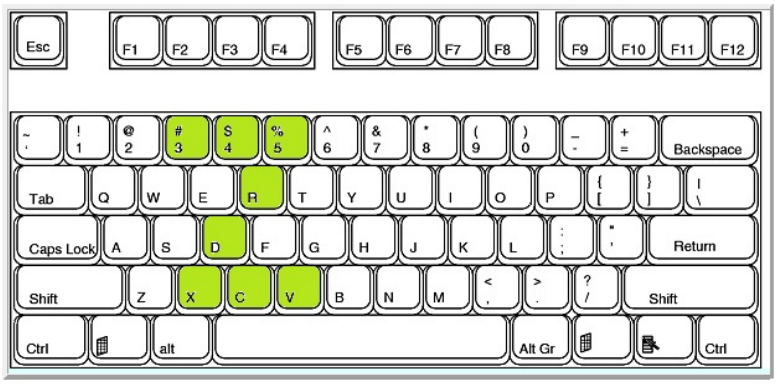


Figure 4: Typing pattern

Features of this pattern are as follows:

- Typing arrangement of pattern components is not important and only the final result will be investigated.
- The password is dynamic. This means that when user enters a pattern correctly, cannot enter the last pattern in the next entrance and must use another letters on keyboard.

4.4. Time limitation

Time limitation in entering the password is one of security factors considered for authentication in the proposed method. In this method, user at registration phase determines a time period for password entering and must enter his/her password within the determined limitation to obtain entrance permission.

This time limitation enhances security of password for authentication and can include the following items:

- Considering a time period for entering the whole password
- Considering time period between characters of password

5. Simulation and comparison

Parameters presented in Table 1 have been used in order to simulate the proposed method.

Table 1: Simulation parameters

Parameter	Value or type
Number of users	Variable from 5 to 20
Number of routers	Variable from 5 to 20
Number of service providers	1
Number of simulation	10 times for each step

5.2. Comparison method

In this paper, the proposed method has been compared with different ways which are explained in the following.

5.2.1. Elliptic Curve Digital Signature Algorithm: ECDSA is an analog elliptic curve of digital signature algorithm (DSA) and has been standardized by many standard organizations of the world such as ANSI, IEEE, NIST, and ISO. In case of ECDSA, two different signatures are produced by two production function which are used to produce points on curve and an extractor. The first function produces the points from private key and the other produces new points from public key. The process of signature include selecting a random password, producing a point using the password, and creation of two signatures. ECDSA firstly creates appropriate parameters to ensure about authentication of this domain. Although encryption using public key method (RSA) does not provide domain parameters. Also, receiver provides required parameters and public key through digital certifications. In order to produce ECDSA, domain parameters, public key and main data are created respectively, and the mentioned hash algorithm and random number producer are used, and the same domain parameters and the same hash algorithm are used to investigate signature [13].

5.2.2. RSA method: Security of RSA is based on vigor of hash function and standard length. For signature, sender signs hash message by private key and then, both the message and signature are sent to the other part. The second part uses the received signature by sender's public key and investigates the signature. If the results were the same, the signature is successfully authenticated. In simpler terms, if a text is released from a person to others, the text include the main but encrypted text by private key of the same person. Now, if the encrypted text is decrypted by public key of the person which you know it, accordance of the resulted and the main texts reveals accuracy of the sender and thereby the person's signature is authenticated. People who do not know this person's private key, are not able to create the primary text by decrypting the encrypted text by this person's public key. An attacker simply can produce a great RSA signature key by low power, therefore, key assessment needs great power so that the other part consumes great calculation sources in order to investigate signature; hence, public key must have an acceptable power value [14].

In Table 2, the proposed method has been compared to four other protocols in the field of security.

Table 2: Comparison between the proposed method and other methods

Type of attack	The proposed method	Sood method	Li method	Wen method	Choudhury method
Insider attack	*	-	-	-	*
Key security meeting	*	*	*	*	*
Attack to guess passwords	*	-	*	*	*
Impersonation attack	*	-	-	-	-
Man-in-the-middle attack	*	*	*	*	*
Denial of Service Attack	*	-	-	-	*
Denial of Service Attack	*	*	-	*	*

5.2.3. Execution time: One of problems of EAP-TTLS is great time of execution which results in high rates of data transactions and high overload related to security calculations. Using the new encryption method and architecture for EAP-TTLS, execution time will be improved greatly. Therefore, this algorithm can be used in networks which have weak sets. In addition, efficiency of the new method has been enhanced.

In this condition, number of clients has been considered to be fixed and in each step, number of routers between AP and the main service provider is increased. Supposing equal conditions for both new encryption method and ECDSA in EAP-TTLS, execution time can be seen in Figure 5.

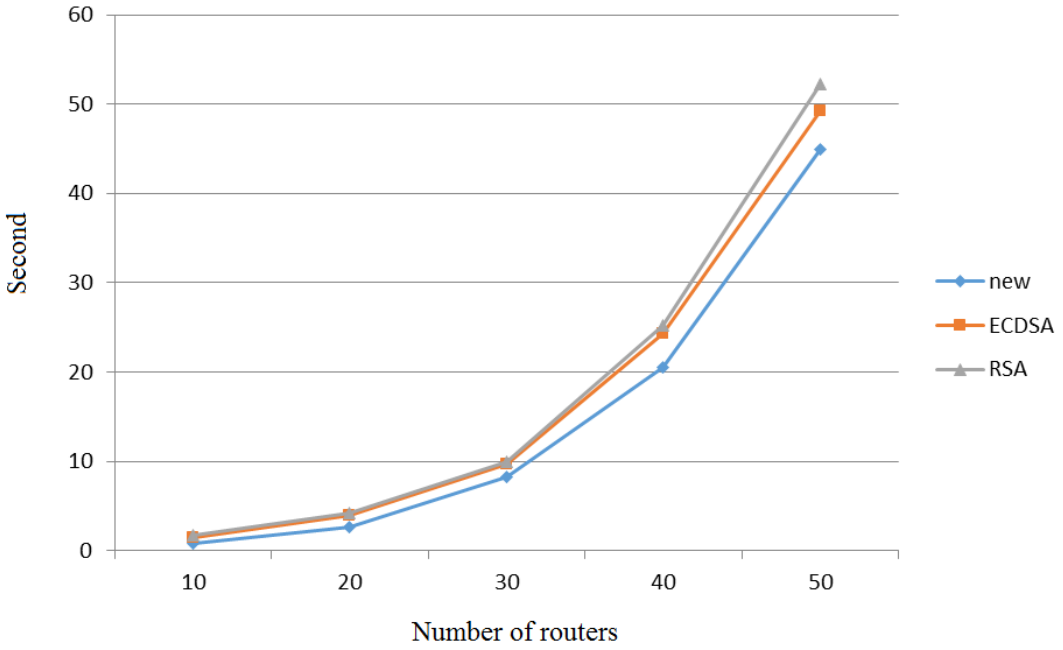


Figure 6: Comparison of execution time between the new method and previous methods as number of routers is increased

Figure 5 indicates that execution time in the new authentication method has improved %22 compared to previous methods.

Figure 6 shows authentication time for the condition in which number of users is variable and number of routers is fixed.

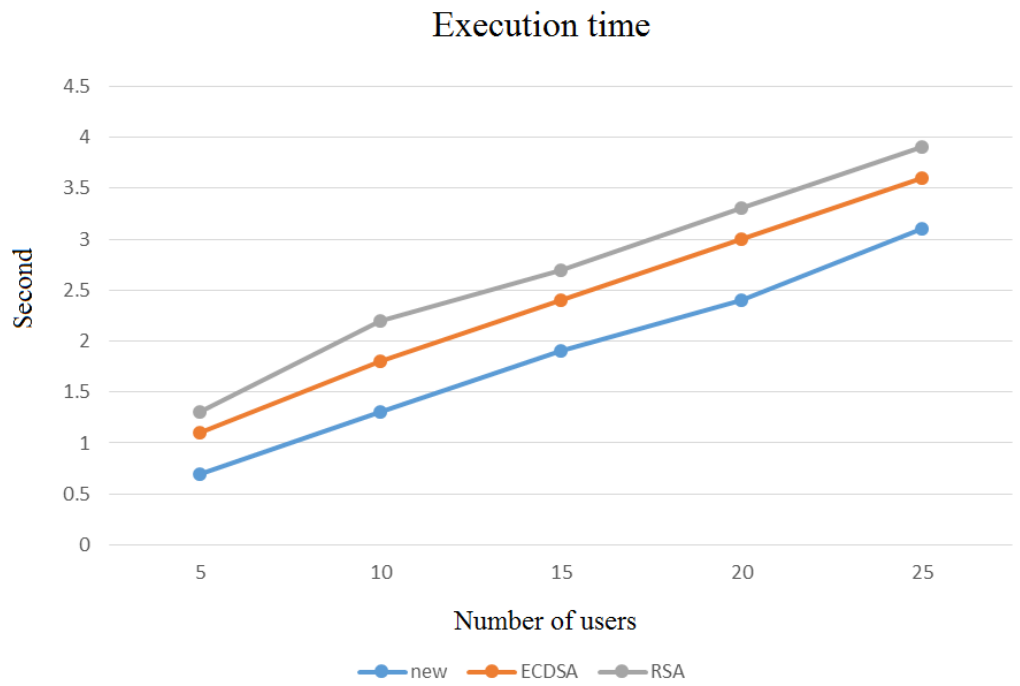


Figure 6: Comparison of execution time between the new method and previous methods as number of users is increased

5.2.4. Used memory: In this condition, number of users is fixed and number of routers is variable and the result is showed in Figure 7.

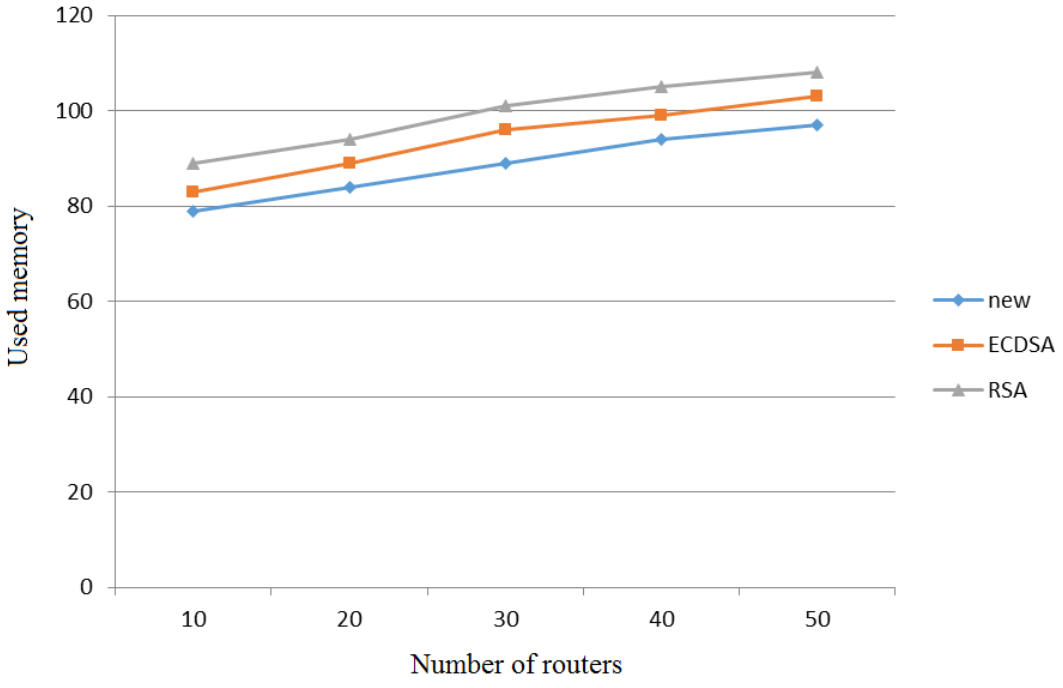


Figure 7: Comparison of used memory with increase in number of routers

As it can be seen in the chart above, used memory is reduced in the new architecture of authentication because fewer messages have been exchanged in EAP-TTLS.

5.2.5. Authentication time: In this model, time duration of authentication operation has been compared and the results are presented in Figure 8.

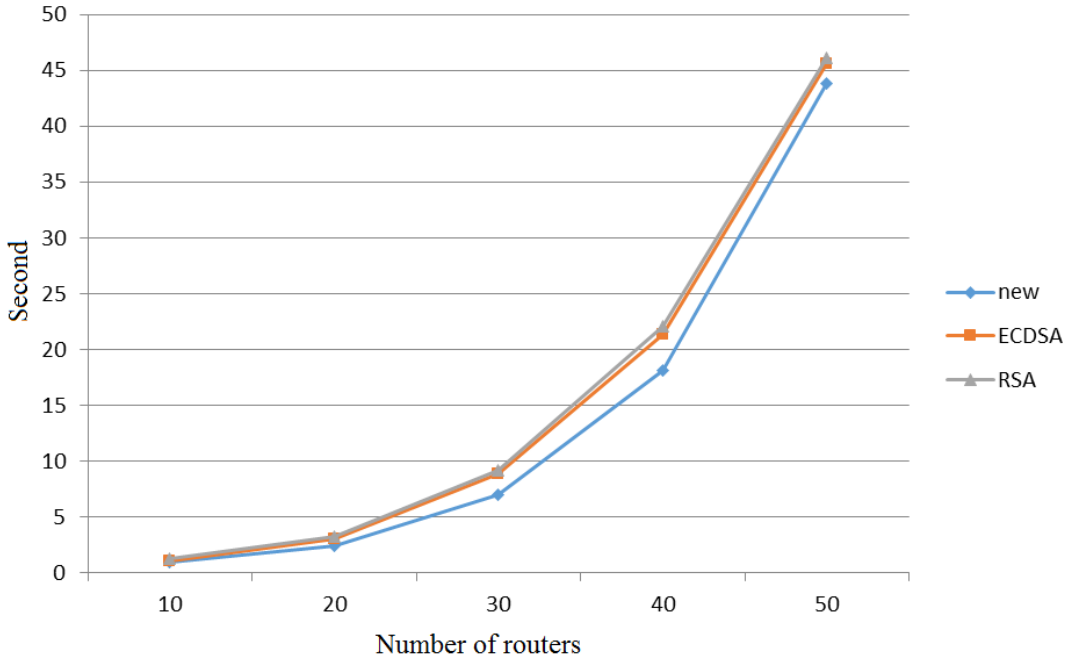


Figure 8: comparison of authentication time duration

Results of simulation shows that the proposed method operates faster than previous methods.

5.2.6. Delay: Delay is referred to the time that authentication operation takes to connect user to network. Delay time is showed in Figure 9.

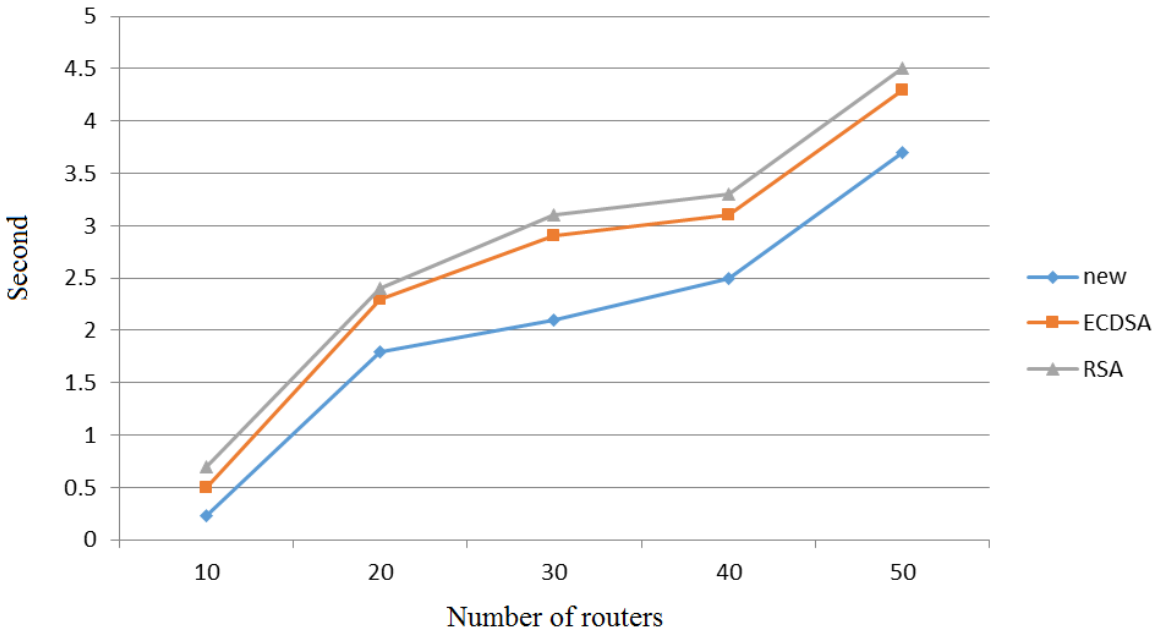


Figure 9: comparison of delay

Since operation is performed almost locally in the proposed method, there is a little delay.

Conclusion

Privacy is one of the most important factors of quality and reliability in big data and authentication is one of its main items which prevents unauthorized people to enter and increases reliability. Authentication is performed by password and the password can be possessed by attackers in various ways. In this paper, multi-factor encryption, pattern password, type pattern password, and time limitation were introduced that each one in its place has some advantages and disadvantages. Results of simulation showed that the proposed method has higher reliability and can be considered as a strong encryption in order to send data and identity information.

References

[1] A.-L. Abul-Ela, B. G. Greenberg, and D. G. Horvitz,"A multi-proportions randomized response model," J. Am. Stat. Assoc. 62, 319 (Sept.), 2011, pp. 990-1008.

[2] C. L.Philip Chen and Z. Chun-Yang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, 275, 2014, pp 314-347.

[3] J. C. W.Lin, T. Y. Wu,, P. Fournier-Viger, G. Lin, T. P. Hong and J. S. Pan, "A Sanitization Approach of Privacy Preserving Utility Mining," In International Conference on Genetic and Evolutionary Computing (pp. 47-57). Springer International Publishing, August 2015.

[4] J.Q. Anderson, "Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well," Ed, 2012.

[5] K. Martin, P. Chitalia, M. Pugalenthi and K. Raghava Rau, "Dell's Channel Transformation – Leveraging Operations Research to Unleash Potential across the Value Chain," Interfaces, Vols. Vol. 44, No. 1, 2014, pp. 55-69.

[6] L. Kamm, "Privacy-preserving statistical analysis using secure multi-party computation," (Doctoral dissertation), 2015.

[7] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski and M. Zaharia, "A view of cloud computing," Communications of the ACM, 53(4), 2010, pp- 50-58.

[8] M. Modak and R. Shaikh, "Privacy Preserving Distributed Association Rule Hiding Using Concept Hierarchy," Procedia Computer Science, 2016, 79, 993-1000.

[9] M. Wang, J. Wang, F. Tian, "City Intelligent Energy and Transportation Network Policy Based on the Big Data Analysis," 2014.

[10] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," In 21st International Conference on Data Engineering (ICDE'05) 2014, April, pp. 217-228. IEEE

[11] S. P. Ahuja and B. Moore, "State of Big Data Analysis in the Cloud," Network and Communication Technologies, 2(1), 2013.

[12] T. S. Lim, W. Y. Low and Y. S. Shih, "A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms," Machine Learning, 40, 2014, pp. 203-229.

[13] Vailaya, "What's All the Buzz Around "Big Data?"," IEEE Women in Engineering Magazine, December 2012, pp. 24-31.