Mining user actions with fuzzy related data security conviction in cloud computing

Alapati Naresh[1,*], Salman Ali Syed[2], B.V.V.S. Prasad[3]

1. *Vignan's Nirula Institute of Technology and Science for Women, palakaluru, Guntur, Andhra Pradesh, India*

2. *Department of Computer Science, College of arts and Sciences, Al-Jouf University, Tabarjal, Kingdom of Saudi Arabia*

3. *Department of CSE, DRK institute of Science & Technology, Bowrampet, Hyderabad, India*

alapatinaresh13@gmail.com

ABSTRACT. *Right now distributed computing assumes a fundamental job in various divisions. In spite of the fact that cloud is adaptable and savvy, it has a few testing issues to be tended to. A portion of the fundamental issues are cloud security and protection. The proposed fluffy based security instrument enhances the security level of information stockpiling in cloud by figuring cloud client's dependability relying upon their conduct. Reliability is assessed utilizing parameters that express client conduct, for example, number of bytes of information from specialist co-op to client, length of access to the cloud framework, timing of client visit, and IP address utilized by client for cloud get to. Cloud information is secured by encoding utilizing key produced dependent on trust level of clients and their successive access design. Visit get to design is identified by mining client's past conduct utilizing FP-Growth calculation. Investigation results demonstrate that the proposed plan withstands balckhole assault and offer higher bundle conveyance proportion.*

RÉSUMÉ. *À l'heure actuelle, l'informatique distribuée suppose un travail fondamental dans diverses divisions. Malgré le fait que le cloud est adaptable et intelligent, il a quelques problèmes de test à régler. Une partie des questions fondamentales sont la sécurité et la protection du Cloud. L'instrument de sécurité à base pelucheuse proposé améliore le niveau de sécurité du stockage d'informations dans le cloud en établissant que la fiabilité du client dans le cloud dépend de son comportement. La fiabilité est évaluée en utilisant des paramètres qui expriment le comportement du client, par exemple, le nombre d'octets d'informations d'une spécialiste à un client, la durée d'accès au cadre en nuage, le moment de la visite du client et l'adresse IP utilisée par le client pour accéder au cloud. Les informations sur le Cloud sont sécurisées par codage en utilisant la clé produite dépendant du niveau de confiance des clients et leur conception d'accès successive. La visite à la conception est identifiée par le comportement antérieur du client minier utilisant le calcul de FP. Les résultats de l'enquête montrent que le plan proposé résiste aux assauts du trou noir et offre une proportion plus élevée de moyens de transport.*

KEYWORDS: cloud computing, security, privacy, trust, fuzzy analysis, pattern mining.

MOTS-CLÉS: cloud computing, sécurité, confidentialité, confiance, analyse floue, exploration de modèles.

DOI:10.3166/ISI.23.5.201-212 © 2018 Lavoisier

1. Introduction

At present, several corporations are using cloud computing directly or indirectly to cut down the hardware and software deployment cost (Mell and Grance, 2009). One of the main cloud services is cloud storage in which data is maintained remotely and can be accessed by the users via a network. Cloud allows the user to work on any device without spending more money on hardware and servers. It also provides software as a service by allowing users to access application software and databases. The user can rent computing resources rather than spending money for their own installation.

Data accessed from remote virtual server are more vulnerable to security attacks. Hence a data security mechanism is essential for cloud environment. There is a substantial consideration in the development of cloud security mechanisms for data protection. Several researchers have proposed different encryption techniques for cloud computing such as hardware based encryption, software based encryption, transparent file encryption, full disk encryption, whole disk encryption etc. The primary problem is to investigate the security issues exist in cloud computing and related security mechanisms.

The proposed Fuzzy associated trust based data security mechanism withstands security attacks by encrypting data using key generated based on trustworthiness of user and the frequent access pattern. It deploys a fuzzy system to compute users trust degree based on their behavior in cloud. Frequent pattern mining algorithm is used to detect cloud user's frequent access pattern. The proposed framework satisfies security requirements such as confidentiality, integrity, availability, scalability, authentication, and authorization.

Most of the existing cryptography mechanisms (Jedeja and Modi, 2012) used for information sharing mainly focused on key generation based on identities of cloud users. Whereas in reality, identities can easily be affected by hacking attacks using which the attacker could gain access to cloud data. The proposed work computes CUs trustworthiness based on their behavior. It protects data by means of encryption using key generated based on user's trustworthiness and frequent behavior in cloud environment.

1.1. Cloud security issues

Cloud services and models face several security threats. It is essential to assure the security requirements such as confidentiality, integrity, authenticity, accountability, non- repudiation. Confidentiality ensures that the data are accessed only by authorized parties. Trustworthiness plays a vital role in cloud computing

environment as it is commonly accessible by different users. Trusting on-line services is harder than trusting off-line services because of lack of centralized control (Beat *et al*., 2008). The distrust of on-line service will cause a negative effect on the truthness of the organization (Jaeger and Fleischmann, 2007). Security issues associated with cloud computing are: lack of user control, unauthorized resource usage, data flow restrictions, litigation and legal uncertainty. Top Threats stated in cloud environment are as follows:

(1) Multi-tenancy in cloud allows sharing of resources across multiple cloud users. Resource sharing has an impact on data confidentiality and results in data loss and increased number of attacks.

(2) Scalability of cloud allows cloud users to scale up and down as needed. The resources assigned for one user may later be assigned to other. This causes confidentiality issues.

(3) Due to absence of hiring standards for cloud employees, malicious insiders can easily hack the organization's data and sell it to competitors.

(4) Cloud APIs can be exploited by outsider hackers and attackers.

(5) Abuse and miscreant use of cloud resources. The proposed work mainly focuses on securing data in cloud environment. It ensures data confidentiality, integrity, authentication, and non-repudiation (Behl and Behl, 2012).

1.2. Background techniques

1.2.1. Data mining

Data Mining (Zhang *et al*., 2009) techniques are generally utilized for data analysis in various fields such as marketing, finance, supply chain management (SCM), and customer relationship management (CRM) (Gaurha and Shrivastava, 2012).

1.2.2. Data mining in cloud security

The proposed work assures basic security requirements by using encryption in data transmission. Data may be transmitted between CSP and CU, or between CU and CU. The session key used in encrypting data is obtained from frequent access pattern exists between the end parties. GenMax, a Backtracking search based frequent pattern mining algorithm is applied to derive frequent access pattern by mining data collected from past communications (Gouda and Zaki, 2005).

1.2.3. Fuzzy logic

Fuzzy logic system (FLS) is a rule based system that have the ability to express uncertainty (Wang *et al*., 2011; Takagi and Sugeno, 1985). The proposed work applies a sugeno based fuzzy system to evaluate the trust degree of cloud users. Trust Degree indicate the degree of trustworthiness of the cloud users computed based on their behavior.

2. Previous work

Cloud security issues are handled and data are protected using various existing security algorithms such as DES, AES, RSA, Blowfish. The symmetric key algorithms DES, AES, and Blowfish provide security for both providers and users. Whereas RSA provides security only for users (Jachak *et al*., 2012). Kaur and Mahajan proposed a security mechanism that selects any of the required algorithms from DES, AES, Blowfish, and RSA (Kaur and Mahjan, 2012). It analyzes technical privacy and encryption controls. Security mechanism proposed in (Kalpana and Singaraju, 2012) is based on RSA algorithm. Only the authorised users are allowed to access the data. But it involves high complexity in key management. The research work proposed by Tebaa *et al*. (2012) uses homomorphic encryption method. It involves high implementation cost. The method proposed in (El-etriby and Mohamed, 2012) is suitable for applications that concentrate more on encryption duration. It uses different algorithms such as RC4, RC6, MARS, AES, DES, 3 DES, Twofish, and Blowfish. Craig Gentry's homomorphic scheme and bootstrapping suggested by Meissen (2012) have unlimited usability of cipher text. Due to the limitation of the circuit, it cannot properly decrypt the cipher text. Key policy and cipher text policy introduced in (Chung *et al.*, 2014) involves less computation overhead. The sensitive information cannot be revealed. But the method cannot fails on collusion attack. The mechanism discussed by song *et al*., uses chaos block encryption algorithm and homomorphic signature algorithm. It improves encryption/decryption speed and maintains high level of data confidentiality. But in this method, the complexity involved in implementation is high. Gaurha and Shrivastava (Gaurha and Shrivastava, 2012) presented an enhanced compete alu sequence algorithm that increases cloud efficiency. Its limitations involve less security and less data encryption process.

The methodology proposed by Kumar and Venkateswarlu uses attribute based encryption, full homomorphic encryption, and linear programming. It satisfies basic security requirements. The security mechanism suggested by Wang *et al.*, addresses security issues like modification attacks, byzantine failures, and other attacks from cloud server (Wang *et al*., 2011). It maintains high efficiency and resilience. But it does not focus on dynamic data operations. Jachak *et al*. (2012) proposed a framework of a very light-weight and provably secure provable data possession scheme. It supports dynamic operations on data. But it does not guarantee about privacy. The encryption technique discussed in (Jain and Kaur, 2012) analyzes the data security risk, its requirements, and deployment of security functions. It provides acceptable level of security. Kamara and Raykova (2013) suggested a scheme suitable for massive datasets. But it is limited by large-scale clusters. Bouti and Keller (2012) followed homomorphic properties of AES to improve the security level. The method lacks in data confidentiality. Singh and Maini (2011) used Blowfish, DES, and AES algorithm in their security mechanism. Inbarani *et al*. (2013) suggest proxy re-encryption mechanism that is secure against chosen cipher text attack. Its performance is limited by collision problem and plaintext attack. Compared with other fully homomorphic encryption schemes, the scheme suggested by Zhang *et al*. (2012) handles practical message attack. The suggested method is

incapable of detecting the attack. None of the above has considered the user behavior in their encryption mechanism.

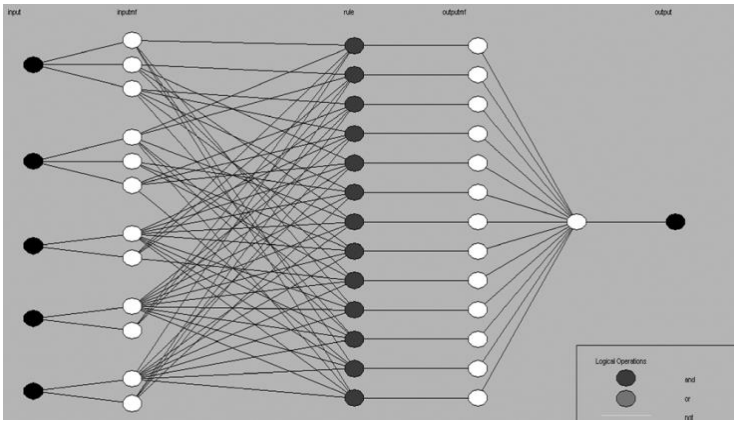

Figure 1. Architecture of fuzzy system

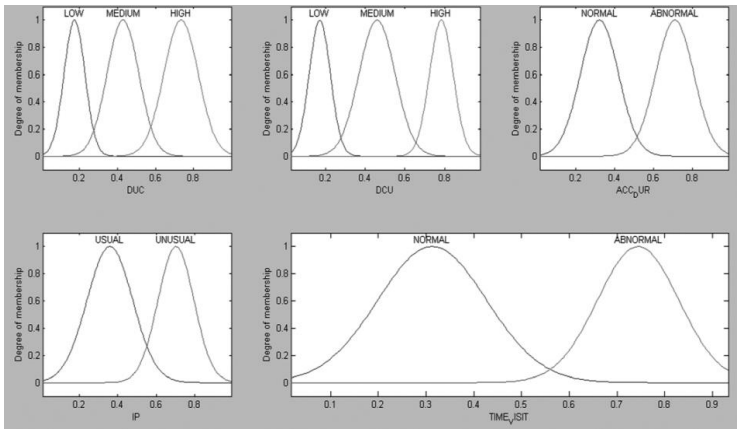

Figure 2. Membership functions for inputs: DUC, DCU, ACC_DUR, IP, TIME_VISIT

3. Proposed work

3.1. Fuzzy system for evaluating trust degree

The trustworthiness of CUs is computed based on the following parameters: number of bytes of data from user to CSP, number of bytes of data from CSP to user, duration of access to the system, whether IP address is unusual, and whether timing

of visit is abnormal. Figure 1 represent the architecture of 3 input and 1 output sugeno based fuzzy system. This section describes the development of fuzzy logic controller for trust degree evaluation.

The input parameters of the fuzzy logic controller are as follows:

-number of bytes of data from user to CSP (DUC)

-number of bytes of data from CSP to user (DCU)

-duration of access to the system (ACC_DUR)

-whether IP address is unusual (IP)

-whether timing of visit is abnormal (TIME_VISIT)

The number of membership functions used for 5 inputs are 3, 3, 2, 2, and 2. The output uses 5 membership functions. The output of the Sugeno based fuzzy controller is a constant (fuzzy singleton). The Gaussian membership functions for the inputs are shown in Fig. 2. Some of the fuzzy rules used for trust evaluation are as follows.

If (DUC is LOW) and (DCU is LOW) and (ACC_DUR is ABNORMAL) and (IP is UNUSUAL) and (TIME_VISIT is NORMAL) then (TD is VERYLOW) (1)

If (DUC is HIGH) and (DCU is HIGH) and (ACC_DUR is NORMAL) and (IP is USUAL) and (TIME_VISIT is ABNORMAL) then (TD is LOW) (1)

If (DUC is MEDIUM) and (DCU is LOW) and (ACC_DUR is NORMAL) and (IP is USUAL) and (TIME_VISIT is NORMAL) then (TD is MEDIUM) (1)

If (DUC is LOW) and (DCU is LOW) and (ACC_DUR is NORMAL) and (IP is USUAL) and (TIME_VISIT is NORMAL) then (TD is HIGH) (1)

.

.

.

If (DUC is LOW) and (DCU is LOW) and (ACC_DUR is NORMAL) and (IP is USUAL) and (TIME_VISIT is NORMAL) then (TD is VERYHIGH) (1)

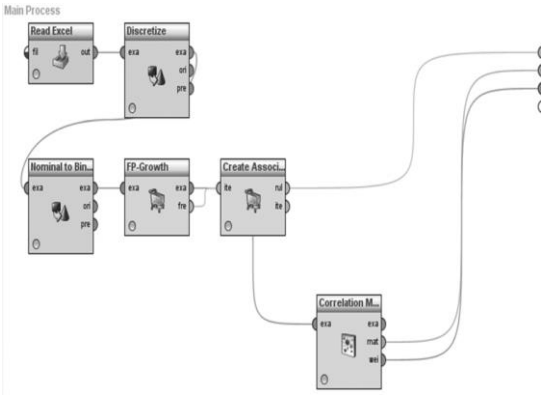

Figure 3. Generating association rules using FP-growth

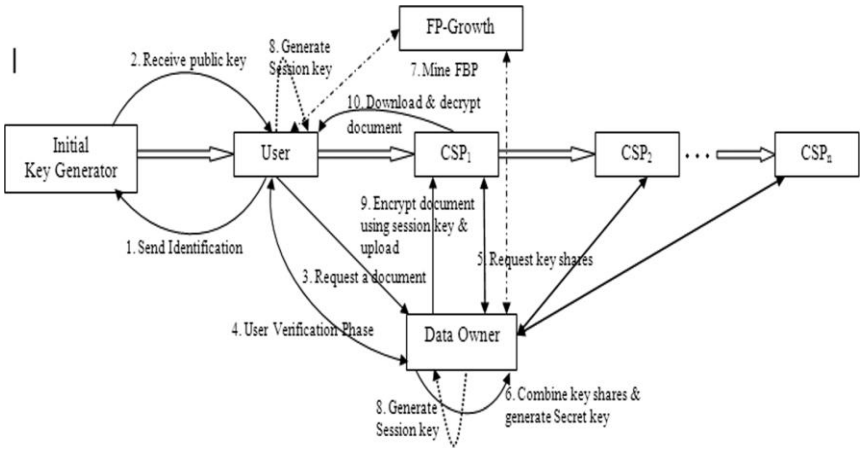

Figure 4. Generating association rules using FP-growth

3.2. Data security in cloud using key management scheme based on user behavior mining

FP-Growth, a frequent pattern mining algorithm is applied to derive CUs behavior and the hidden relationship exists between CUs and CSPs. The proposed work applies the idea of Key Distribution through Traffic Mining (KDTM) (Stalling, 2003). It uses user behavior mining to generate the session key and to satisfy all basic security requirements. The FP-growth (Hunyadi, 2011) pattern mining algorithm has been used in frequent pattern identification and is applied later to generate session keys. Apriori, the most commonly used frequent pattern mining algorithm lists every single frequent itemset using bottom up, breadth first search. Using Apriori listing all possible subsets of length pattern for dense data with long patterns involves high computational overhead. Generation of candidate itemset using Apriori is expensive interms of space and time. Whereas FP- Growth allows discovery of frequent itemset without generating candidate itemset. It is efficient and scalable for mining both long and short frequent patterns. It is faster than the Apriori algorithm. Hence, the proposed method applies FP- growth algorithm to identify frequent behavior pattern of CUs. The process implemented using RapidMiner is presented in Fig. 3. The basic idea is derived from previous work proposed in (Lakshmi and Kumar, 2015).

3.3. Session key generation

For data transmission the source and the destination users or source CU and destination CSP will generate a session key. The concept of behavior mining is applied to attain frequent user's behavior Pattern (FBP) that exists between the CU's, or CU and CSP. Each CU maintains details about last few traffic carried on with

other CU in the cloud.

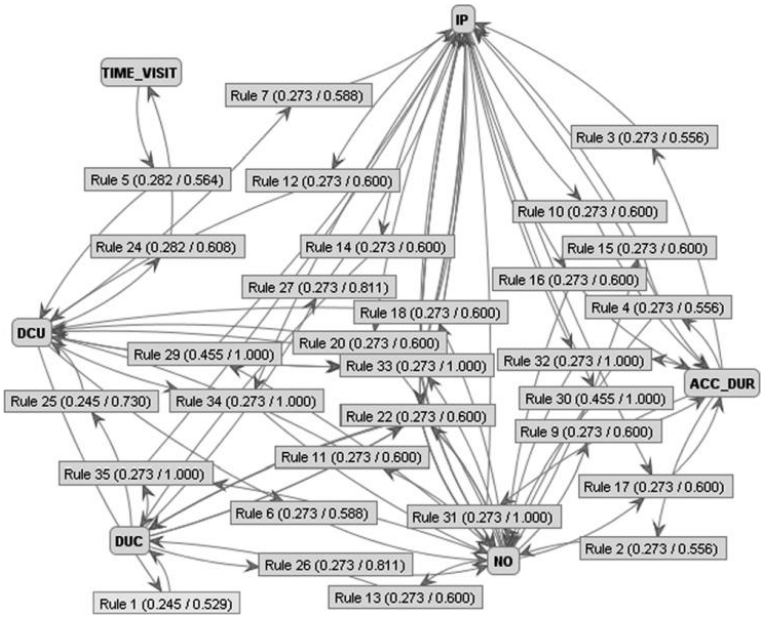

Figure 5. Association rules

The users CU1 and CU2 generate their session key SCU1CU2.

Session key, $SK_{CU\_data\ owner} = H\ (FBP \mid\mid ID_{CU} \mid\mid$

4. Experimental results

Cloud environment is simulated using CloudSim and the CU's trust is evaluated using the fuzzy system designed using MATLAB. The fuzzy system is generated with 5 inputs and 1 output. The inputs DUC and DCU have 3 Gaussian membership functions, ACC_DUR, IP, TIME_VISIT have 2 Gaussian membership functions, and output has 5 membership functions. The Sugeno based fuzzy system is trained using 250 data sets and its architecture is shown in Figure 1. FBP analysis is carried out with the help of RAPIDMINER tool.

Then the performance of the proposed scheme is analyzed using the metric packet delivery ratio (PDR).The experiment is conducted by introducing blackhole attack in the network and by varying the number of malicious CUs in the network. Figure 8 show that the PDR achieved using proposed scheme is almost 10% more compared to the PDR achieved without using the proposed scheme.

$ID_{data\ owner} \mid\mid (r_{CU} P_{CU} \mid\mid r_{data\ owner} P_{data\ owner}))$

Attributes	NO	DUC	DCU	ACC_DUR	IP	TIME_VISIT
NO	1	0.259	0.062	0.040	1	0.033
DUC	0.259	1	0.144	0.005	0.259	0.003
DCU	0.062	0.144	1	0.012	0.062	0.040
ACC_DUR	0.040	0.005	0.012	1	0.040	0.012
IP	1	0.259	0.062	0.040	1	0.033
TIME_VISIT	0.033	0.003	0.040	0.012	0.033	1

Figure 6. Correlation matrix

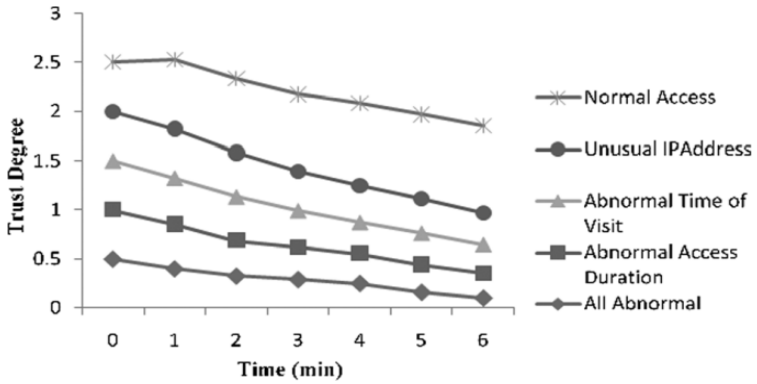

Figure 7. Trust degree Vs. Time

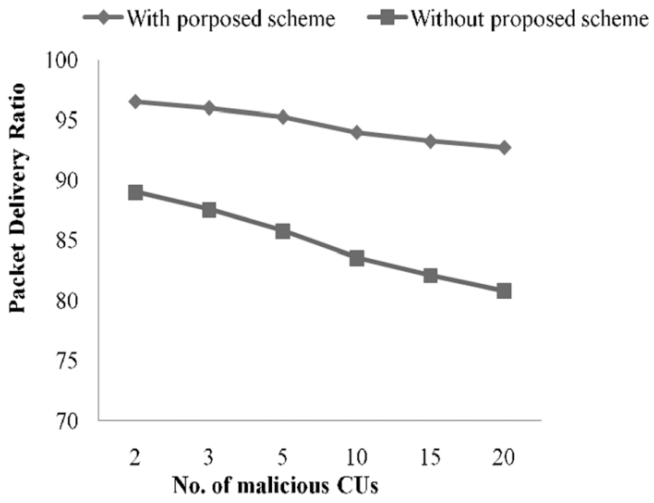

Figure 8. PDR Vs No. of malicious CUs

5. Conclusion & future work

This paper presents a fuzzy associated trust based data security scheme by mining user behavior for cloud environment. The proposed approach ensures authentication and confidentiality of CUs by applying encryption using the key generated based on CU's behavior. It computes direct trust of cloud users with fuzzy system using user behavior parameters such as access duration, time of visit, IP address, and data transfer rate. Frequent access pattern of CUs is identified by mining their past cloud access pattern. The effect of PDR has been analyzed with respect to blackhole attack in cloud network. Simulation results show that the proposed scheme satisfies all security requirements, withstand balckhole attack, and yields higher PDR. The primary focus of future work includes analyzing the performance of proposed scheme under distributed denial of service attack.

References

Behl A., Behl K. (2012). An analysis of cloud computing security issues. *Proceedings of IEEE World Congress on Information and Communication Technologies*, pp.109-114.

Best S. J., Kreuger B. S., Ladewig J. (2008). The effect of risk perceptionson online political participatory decisions. *Journal of Information Technology & Politics*, Vol. 4, No. 1, pp. 5-17. http://dx.doi.org/10.1300/J516v04n01_02

Bouti A., Keller J. (2012). Securing cloud-based computations against malicious providers. *ACM SIGOPS Operating System Review*, Vol. 46, No. 2, pp. 38-42. http://dx.doi.org/10.1145/2331576.2331583

Chung P., Liu C., Hwang M. (2014). A study of attribute-based proxy re-encryption scheme in cloud environments. *International Journal of Network Security*, Vol. 16, No. 1, pp. 1-13.

El-etriby S., Mohamed E. (2012). Modern encryption techniques for cloud computing randomness and performance testing. *3rd International Conference on Communications and Information Technology (ICCIT)*, pp. 800-805. http://dx.doi.org/10.13140/2.1.4685.8880

Gaurha N., Shrivastava M. (2012). Data security in cloud computing using linear programming. *Int. J. Emerging Technology. Adv. Eng.*, Vol. 2, No. 7, pp. 28-30.

Gouda K., Zaki M. J. (2005). GenMax: An efficient algorithm for mining maximal frequent itemsets. *Data Mining and Knowledge Discovery*, Vol. 11, No. 3, pp. 1-20. https://doi.org/10.1007/s10618-005-0002-x

Hunyadi D. (2011). Performance comparison of Apriori and FP- growth algorithms in generating association rules. *Proceedings of the 5th European conference on European computing conference*, pp. 376-381.

Inbarani W. S., Shenbagamoorthy G., Paul C. K. C. (2013). Proxy re- encryption schemes for data storage security in cloud- a survey. *International Journal of Engineering Research & Technology*, Vol. 2, No. 1, pp. 1-5.

Jachak K. B., Korde S. K., Ghorpade P. P., Gagare G. J. (2012). Homomorphic authentication with random masking technique ensuring privacy and security in cloud computing. *J. Bioinfo Secur. Inform.*, Vol. 2, No. 2, pp. 49-52.

Jadeja Y., Modi K. (2012). Cloud computing - concepts, architecture and challenges. *Proc International Conference on Computing, Electronics and Electrical Technologies*, pp. 877-880. https://doi.org/10.1109/ICCEET.2012.6203873

Jaeger P. T., Fleischmann K. R. (2007). Public libraries, values, trust, and egovernment. *Information Technology and Libraries*, Vol. 26, No. 4, pp. 35-43.

Jain N., Kaur G. (2012). Implementing DES algorithm in cloud for data security. *International Journal of Computer Science & Information Technology*, Vol. 2, No. 4, pp. 316-321.

Kalpana P., Singaraju S. (2012). Data security in cloud computing using RSA. *International Journal of Research in Computer and Communication Technology*, Vol. 1, No. 4, pp. 143-146.

Kamara S., Raykova M. (2013). Parallel homomorphic encryption. *Workshop on Applied Homomorphic Encryption (WAHC '13)*, pp. 213-225. http://dx.doi.org/10.1007%2F978-3-642-41320-9_15

Kaur M., Mahajan M. (2012). Implementing various encryption algorithms to enhance the data security of cloud in cloud computing. VSRD. *International Journal of Computer Science & Information Technology,* Vol. 2, No. 10, pp. 831-835.

Kumar S. K., Venkateswarlu S. (2013). Efficiently providing data security and linear programming in cloud computing. *International Journal Of Computer Science & Technology*, Vol. 4, No. 2, pp. 1569-1570.

Lakshmi R. P., Kumar A. V. A. (2015). Parallel key management scheme for mobile ad hoc network based on traffic mining. *IET Information Security*, Vol. 9, No. 1, pp. 14-23. https://doi.org/10.1049/iet-ifs.2013.0076

Liaw S. H., Su P. C., Chang H. K., Lu E. H., Pon S. F. (2005). Secured key exchange protocol in wireless mobile ad hoc networks. *Proc. 39th Annual Int. Carnahan Conf. (CCST '05)*, pp. 171-173. https://doi.org/10.1109/CCST.2005.1594839

Meiseen R. (2012). A mathematical approach to fully homomorphic encryption. *Project Report, WPI.*

Mell P., Grance T. (2009). A NIST definition of cloud computing, National Institute of Standards and Technology. *NIST Special Publication*, pp. 800-145.

Singh S. P., Maini R. (2011). Comparison of data encryption algorithms. *International Journal of Computer Science and Communication*, Vol. 2, No. 1, pp. 125-127.

Stallings W. (2003). Cryptography & network security: Principles & practices. *(3rd edition) Prentice Hall.*

Takagi T., Sugeno M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on System, Man and Cybernetics*, Vol. 15, No. 1, pp. 116-132. http://dx.doi.org/10.1016/B978-1-4832-1450-4.50045-6

Tebaa M., Hajji S. L., Ghazi A. E. (2012). Homomorphic encryption applied to the cloud computing security. *Proc. on Engineerin*g.

Wang G., Liu Q., Wu J., Guo M. (2011). Hierarchical attribute- based encryption and scalable user revocation for sharing data in cloud servers. *Computers & Security*, Vol. 30, No. 5, pp. 320-331. http://dx.doi.org/10.1016/j.cose.2011.05.006

Zhang C., Tiwari R., Chen W. (2009). A data mining method to extract and rank papers describing coexpression predicates semantically. *IEEE International Conference on Data Mining Workshops*, pp. 483-488. https://doi.org/10.1109/WICT.2012.6409059

Zhang Z., Plantard T., Susilo W. (2012). Reaction attack on outsourced computing with fully homomorphic encryption schemes. *Information Security and Cryptology, Springer Berlin Heidelberg*, pp. 419-436. https://doi.org/10.1007/978-3-642-31912-9_28