User Prediction in a Role for Secure Data Sharing Through Cloud

Pattabhi Mary Jyosthna, Konala Thammi Reddy

Abstract: Nowadays cloud is being used by both individuals and organizations to store and share the data without establishing their own data center. The outsourcings of these data are becoming a major security issue for businesses. Searchable encryption is one of the prominent techniques which allow the data owner to securely store the data and then share the data for their growth in business. With this technique, Cloud Service Provider can process the user request by searching on encrypted stored data without decrypting the data. In this paper we analyze different searchable encryption techniques for secure data sharing and their preventive attacks. We also proposed a method named "User Prediction in Role" to reduce the insider attack possibility in Role Based Data Sharing (RBDS), which is based on user priority levels in a role. Priorities will be decided by role manager based on the roles of a user in the organization and also predict the role assignment for new user. The proposed method can helps the organization to avoid the unauthorized data leakage.

Index Terms: cloud store, Data sharing, insider attack, Role based data sharing, Searchable Encryption.

I. INTRODUCTION

Every day, users in every small to large scale organizations, are generating massive amount of data. That data need to be maintained for the future purpose and also share with the users of their organization. The emerging technology of today's internet world to achieve the organizations requirement is Cloud computing. It has huge datacenter, equipped with thousands of servers to process the user's request. Cloud storage system is one of the services provided by the cloud service providers (CSPs) where people can store large amount of data with less investment instead of spending more money for building their own data center. People are very familiar in utilizing these cloud services based on rental basis. Due to the elasticity characteristic of cloud computing, data owners can scale up and down the storage space utilization as per their requirement. The benefits of utilizing Storage-as-a-Service are low maintenance cost, easy data sharing, and easy data accessing. Organizations can share their data with employees to collaboratively work together from any location as long as the internet is available. Similarly users can access data from the cloud repository at any time, at any place and from any device with internet as the basic requirement. Every sector like IT, Education, Medical, Business, and Banking etc. can share their data through cloud storage service.

Manuscript published on 30 August 2019.
***Correspondence Author(s)**

Mrs. P. Mary Jyosthna, Department of CSE, GITAM Institute of Technology, Visakhapatnam, INDIA.
Dr. Konala Thammi Reddy, Department of CSE, GITAM Institute of Technology, Visakhapatnam, INDIA.

Generally, data owner places encrypted data in the cloud servers. As per the traditional data sharing approach, Cloud service provider (CSP) used to decrypt the data to process the user request on the data. This architecture may leads to many security issues. One, the data has revealed to CSP. Though CSP is trusted but he/she is curious to know about the data, so he may utilize that information for his intended use and sometimes he may delete some of the files which are never/rarely used by the users. Second, the outside hacker who hacks the public network can use tools to gain some plaintext or ciphertext or secrete key which is used for encryption/decryption. Third, data owner has no control over the outsourced data. These security issues are treated as possible attacks at cloud storage and that are named as Known-ciphertext attack, Known-Plaintext attack and keyword-Guessing attack.

Therefore data owners might think about the data center location and data security related issues. It may prevent users from utilizing the cloud services. So there is a need to establish trust between data owner and service provider (CSP). Many researchers have come up with new techniques for availing cloud services like data storage, data computing, and data sharing etc. securely. Searchable encryption techniques are the one to achieve secure data sharing through cloud storage. Fig.1. shows the general architecture to share data with multiple users by using searchable encryption methods.

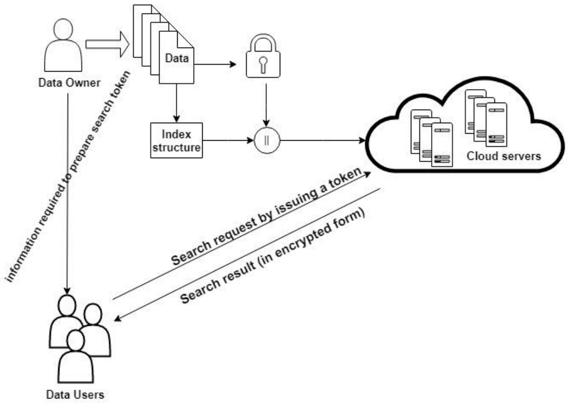
Data Owner
Data
Index structure
Cloud servers
information required to prepare search token
Search request by issuing a token
Search result (in encrypted form)
Data Users

Fig. 1. Architecture for secure data sharing through Cloud with searchable encryption

Searchable encryption technique allows the users to recover only required data and also prevent CSP from decrypting the entire data. i.e., CSP can perform search operation on encrypted data for the received trapdoor from user. Therefore Cloud Service Provider (CSP) cannot learn any information about the stored data.

Initially, Searchable Symmetric (SSE) [1] and Asymmetric (PEKS) [2] encryption algorithms were proposed to satisfy the data owner's requirement.

In both the algorithms, data owner prepares an index as a set of keywords for his data. In SSE both data and index encrypts with secret key and the same key will also be used for decryption. This is suitable when both the owner and user will be same. Public key encryption (PEKS) is well appropriate approach to share data with others because it encrypts the data with user's public key. So the intended user can only decrypt the data with his private key. In the above two methods user need to submit a trapdoor to the CSP to access the required block of data from the cloud. Data user will get that trapdoor or trapdoor related information from the data owner. Data owner need to prepare more number of tokens/trapdoor when they have multiple users to share their data. It might be a burden to the owner and mainly the security concern regarding data sharing with a group of users through cloud servers is lack of owner control on the outsourced data. In this regard, Attribute Based Encryption (ABE) and Role Based Encryption (RBE) are suitable and efficient techniques to achieve owner-controlled data sharing based on user access policy structures [19]. The access policy structure in ABE prepares based on the user attributes and ciphertext attributes where as in RBE prepares based on the role hierarchies in the organization. The main motivation of RBE is to share data with a group of user in an enterprise instead of individual users.

The remaining part of this paper is organized as follows. Related work is in section 2, threat analysis in RBE is in section 3, proposed method is in section 4, and conclusion &future work is in section 5.

II. RELATED WORK

Towards searchable encryption methods, researchers have introduced new approaches to achieve secure data sharing with multiple users. Alok Kumbhare *et al.* [5] used Broad Cast Encryption for distributing encryption keys to the users. Data owner in this approach performs symmetric key encryption on the outsourced data. There is a secure key distribution overhead in the symmetric key encryption and in the cloud environment multiple data owners stores the data and different users wants to access the data. To support multi-owner and multi-user scenario, authors in [6], [7] & [11] have introduced multi-party searchable encryption scheme (MPSE). To prevent Keyword Guessing Attack (KGA) in Public key encryption with keyword search (PEKS), [8] introduced a framework that uses two servers for verification. [9] Proposed two levels of security for secure data sharing. Key aggregate scheme in [10] allows the data owner to supply single key to share data with users and user can only submit single trapdoor to the CSP for searching data in the cloud servers.

Confidentiality with access control service [3] is essential to share data with multiple users. Attribute Based Encryption (ABE) [25] was initiated to share data with a set of users. ABE allows the users whose private key attributes are matched with the attributes of ciphertext to decrypt the data. The two variants Key-Policy ABE [4] and Ciphertext-Policy ABE [26] of ABE are differentiated based on the association among private key attributes, ciphertext attributes and access policy structures. To prevent chosen-ciphertext attack, Zhu, H. *et al.* [21] were introduced KP-ABE with equality Test (ET) to find whether two ciphertexts contains the same information or not without decryption. [22] Wang, Q. *et al.* were introduced CP-ABE with equality test. Shen.z *et al.* [12] introduced Hierarchical Predicate Encryption (HPE) that includes access policy, keywords and symmetric key in the index information.

RBE allows the organizations to share data based on the roles of the users that they play in the organization. Researchers of [13-16] proposed RBE scheme by enforcing role based access control (RBAC) models [20] on encrypted data. Roles in the organization are arranged as a hierarchy of roles (RH) to provide access control on the data. The underlying principle of RBE is users whose role match with the role which specifies in the role hierarchy can only decrypt the data. It also preserves the principle of Searchable encryption

I. ***summary of mechanisms for secure data sharing through cloud***

Method	**Encryption**	**Access control**	**revocation**	**Verifiability**	**Attack Resistant**
Tang, Q. [6]	SE	ASE	No	Yes	Collusion attack
Chen, R *et al.* [8]	ASE	ASE	No	No	Inside Keyword Guessing
Liu, J. K. [9]	ASE	IBE	Yes	Yes	Device Guessing
Cui, B [10]	SE	Aggregate key	No	No	Known Aggregate key
Sun, W *et al.* [11]	SE	ABAC	Yes	Yes	Chosen Keyword
Shen.z *et al.* [12]	SE	ABAC	Yes	No	Access violation
Zhu, Y. *et al.* [13]	ASE	RBAC	Yes	No	Hierarchical Collusion
Zhou, L. *et al.* [14]	BE	RBAC	Yes	No	Selective ID, Revocable ID
Zhou, L. *et al.*[15]	ASE	RBAC	Yes	Yes	Chosen cipher Text, Chosen Plain Text

Researchers of [13-16] use different variants of RBE to provide user revocation and verifiability in addition to the secure data sharing. The author of [18] enforced RBAC model to prevent the unauthorized access of personal electronic health records (PEHR) before outsourcing the EHRs into the cloud. [15] Proposed hybrid cloud architecture for Role Based Encryption.

It allows organization to place access policy structure in private cloud while the actual data stores in public cloud.

Works done till now are used to prevent the outsiders from unauthorized access. About the insider attack, the researchers in [17] address the issue of trust between data owners and roles and between roles and users. This approach allows data owners and roles to determine the trustworthiness of their users before sharing the data. Here, data owner determines the trust factor of a user based on the feedbacks from other roles in which the user involved and history of number of failures in which the user accessed just before the failure.

The summary of searchable encryption mechanisms and their preventive attacks is shown in the table I. As per the observations, there is a possibility of insider attack in the RBE scheme when data are sharing according to the roles. In the next section we explained the analysis of insider threat possibility in the organization.

III. Threat Analysis

All the Previous works have done based on either user attributes or user roles for secure data sharing. RBE allows data owners to restrict the accessing of outsourced data by a predefined access policy structure. The access policy structure includes role hierarchy which constructs as a partial ordered relation. Roles in the role hierarchy contain set of users and permissions are assigned to roles. According to the NIST model for RBAC [27], a single user may have more than one role. If a user belongs to more number of roles then he may have excess access privileges because roles have inheritance property. A role can acquire its own permissions along with its successor role's permissions but it cannot acquire its predecessor's access permissions. The sample role hierarchy of an organization is shown in the Fig.2.

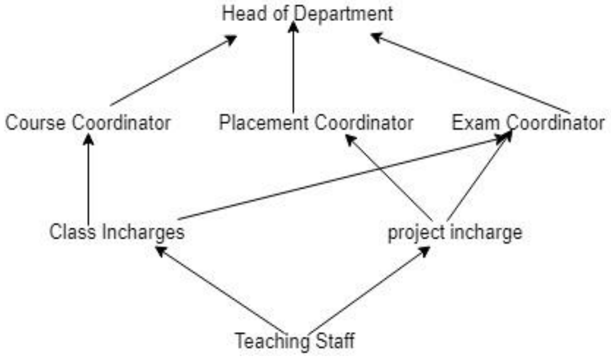

Fig.2. Role Hierarchy

The access rights assigned to the teaching staff are inherited by the all the in-charges and transitively by the coordinators and Head of the Department whereas access privileges assigned to in-charges are not inherited by the teaching staff. Organizations points of view, there are some levels of priority to get access of the data. If the organizations have people/users of different roles then the possibility of insider attack is directly proportional to the total number of roles in which a particular user acts as a member to avail access rights. Most of the previous works addressed only chosen plaintext attack, chosen ciphertext and honest but curious attacks. None of them have discussed about insider attack within the organization. As per the survey report [24] on insider threat, the cause of insider attack at organizations is excess access rights assigned to the users. Insider may be a user in the single role or multiple roles. Insider attacks may happen unintentionally because of the employee's negligence or may happen intentionally to gain the data.

According to the Binomial Distribution theorem, consider an event as accessing data by a user with right access privileges. The possibility of getting success on this event is ½ and possibility of failure is ½. The sum of success rate and failure rate is one.

Let,

R is the set of roles in the role hierarchy.

$R= \{R_1, R_2, R_3 \ldots R_n\}$

U_{Ri} is the set of users of a role R_i

$U_{Ri}= \{u_1, u_2 \ldots . u_m\}$

T_{Ui} is the total number of roles in which a particular user u_i is a member with roles access privileges.

A is the insider threat possibility by a user of a role.

When a user belongs to multiple roles, he has many access privileges and it leads to high possibility of insider threat.

The cumulative distribution of insider threat when a user belongs to r number of roles out of n roles is

$P(A) = \sum n_{Cr}\, p^{r} (1 - p)^{n-r}, r = 0, 1, 2, \ldots, n,$

Where n is a positive integer and $0 \leq p \leq 1$.

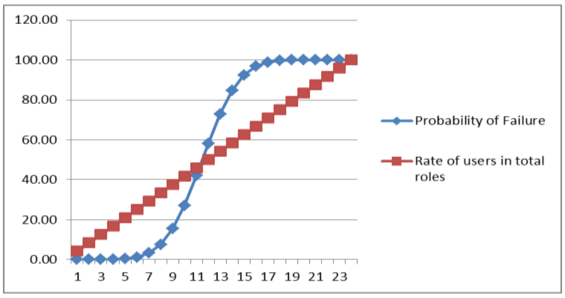

Fig.3. Probability of failure

We can say that P(A) is directly proportional to total number of roles of a user T_{Ui}.

$P(A)\ \alpha\ T_{Ui}$ The above observation saying that, we should restrict the excess permissions to the users. Therefore, user priorities need to be evaluated before distributing the data to user in the roles.

IV. PROPOSED METHOD

The intentional idea is diagrammatically represented in fig.3.

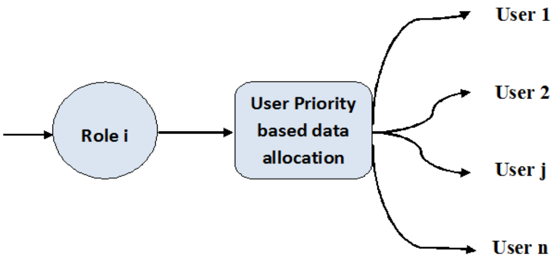

Fig.4. Sharing data to the users in a particular Role

As per the fig.4 every role may have one or more users. All users in a single role may or may not have the same priority to allocate the data. So, according to the user priority levels data should be allocated to the users to avoid the insider attacks. The user priority is evaluated based on their behavior in the roles in which he belongs to. Different parameters that are needed to be considered to evaluate the user behavior are user's expertise, resource consumption, spending time and data accessing.

When the data are from the different sources, they should be combined using information fusion technique to get the accurate priority level. We use the Bayesian method to combine the information and produce the priority value between [0, 1] interval. Based on the threshold value the priority levels can be defined as low, medium and high.

A. Evaluation and Results

The Bayesian inference method produces the value based on the probability rules. The rule is

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$

For statistics of a user to be a insider attacker, the Bayesian theorem can be taken as

$$P(\boldsymbol{\Theta}|data) = \frac{P(data|\boldsymbol{\Theta}) \times P(\boldsymbol{\Theta})}{P(data)}.$$

Here Θ is the target that we wanted to know
Data D is the parameters that we have about Θ
D= {H, E, U, R}
P(Θ) is the prior distribution
P(Θ| data) posterior distribution. This is the distribution representing our belief about the parameter values
To assess the user's priorities, we have taken sample data about 100 users and predict the user based on the probability of a user to be an insider attacker. Table II contains the sample values of the parameters H, E, U, R.
H indicated number of hours spent on the system per month.
E indicates User's expertise on the software and hardware.
U indicates utilization of resources that may be or may not be available to that particular user.
R indicates total number of roles that the respective user has in the organization.
As we discussed in the previous section, the possibility of an attack by insider might be more if he/she has high rate of access privileges on data. A user can get more privileges if he/she has more roles. So, we should consider total number of roles R in which he/she is a member, to predict the user.

II. Sample behavioral data of individual users

Users	H	Parameters		
		E	U	R
1	98	0.37	0.55	0.42
2	82	0.45	0.54	0.47
3	53	0.49	0.63	0.34
4	80	0.56	0.48	0.39
5	87	0.38	0.48	0.64
6	40	0.57	0.61	0.17
7	83	0.37	0.33	0.38
8	75	0.18	0.26	0.69
9	89	0.26	0.45	0.51
10	89	0.15	0.69	0.91
11	59	0.46	0.48	0.81
12	44	0.22	0.43	0.31
13	75	0.26	0.27	0.71
14	107	0.25	0.4	0.7
15	95	0.38	0.59	0.61
16	117	0.4	0.25	0.47
Average	79.56	0.36	0.47	0.53
Minimum	40	0.15	0.25	0.17
Maximum	117	0.57	0.69	0.91

III. Priority Levels

Likelihood	Posterior	probability	Priority level
1.173	0.077	0.072	0.5
1.564	0.103	0.093	0
1.631	0.107	0.097	0
0.047	0.724	0.045	1
1.127	0.074	0.069	0.5
1.592	0.105	0.095	0

Based on the probability values, we have proposed priority levels to the users for data distribution between [0, 1] interval. Table 3 shows the proposed priority levels. In this level 0 indicates least priority to access the data and level 1 indicates highest priority to access the data. Based on the requirement Medium level user can have the access rights on data. It will be decided by role manager. The data and results that I used here are the sample small data set.

V. CONCLUSION AND FUTURE WORK

Cloud storage service attracts the users from all sectors to store their data at cloud data centers. Low maintenance cost and easy data sharing are the benefits to the cloud stakeholders. These benefits proportionally increase the security challenges. RBAC is used to control the user access rights on shared data. But it neglects the effect of insider attack if a user is belongs to multiple roles. In this paper we have discussed the existing searchable encryption methods for secure data sharing through cloud and proposed a new data sharing method to reduce insider attack possibility within the organization. This new method allows the role manager to predict user priority and can share the data based on those user priorities instead of just a role. We also provided sample results of this new approach for secure data sharing in Role Based Data sharing methods. In future we will specify the complete implementation details of our method.

REFERENCES

1. D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," *Proceedings of the IEEE Symposium on Security and Privacy*, 2000, pp. 44-55.
2. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," *International Conference on the Theory and Applications of Cryptographic Technique*, Springer Berlin Heidelberg, May 2004, pp. 506-522.
3. D. Thilakanathan, S. Chen, S. Nepal, and R. A. Calvo, "Secure data sharing in the Cloud," *Security, Privacy and Trust in Cloud System,* 2014, pp. 45-72, Springer Berlin Heidelberg.,
4. V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine- grained access control of encrypted data," *Proceedings of the 13th ACM conference on Computer and communications security,* OCT. 2006, pp. 89-98.
5. A. Kumbhare, Y. Simmhan, and V. Prasanna, "Cryptonite: a secure and performant data repository on public clouds," *IEEE 5th International Conference on Cloud Computing, June 2012,* pp. 510-517.

6. Q. Tang, “Nothing is for free: security in searching shared and encrypted data,” *IEEE Transactions on Information Forensics and Security*, vol. *9, no.* 11, Nov. 2014, pp. 1943-1952.
7. Q. Zheng, S. Xu, and G. Ateniese, “VABKS: verifiable attribute-based keyword search over outsourced encrypted data,” *Infocom, 2014 proceedings IEEE*, April 2014, pp. 522-530.
8. R. Chen, Y. Mu, G. Yang, F. Guo, and X. Wang, “Dual-server public-key encryption with keyword search for secure cloud storage,” *IEEE transactions on information forensics and security*, vol. 11, no. 4, 2016, pp. 789-798.
9. J. K. Liu, K. Liang, W. Susilo, J. Liu, and Y. Xiang, “Two-factor data security protection mechanism for cloud storage system,” *IEEE Transactions on Computers*, vol. 65, no. 6, 2016, pp. 1992-2004.
10. B. Cui, Z. Liu, and L. Wang, “Key-aggregate searchable encryption (KASE) for group data sharing via cloud storage,” *IEEE Transactions on computers,* vol. 65, no. 8, 2016, pp. 2374-2385.
11. W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, “Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 4, 2016, pp. 1187-1198.
12. Z. Shen, J. Shu, and W. Xue, “Keyword Search with Access Control Over Encrypted Cloud Data,” *IEEE Sensors Journal*, vol. 17, no. 3, 2017, pp. 858-868.
13. Y. Zhu, H. X. Hu, G. J. Ahn, H. X. Wang, and S. B. Wang, “Provably secure role-based encryption with revocation mechanism,” *Journal of Computer Science and Technology*, vol. *26, no.* 4, May 2011, pp. 697-710.
14. L. Zhou, V. Varadharajan, and M. Hitchens, “Enforcing role-based access control for secure data storage in the cloud,” *The Computer Journal*, vol. *54, no.* 10, 2011, pp. 1675-1687.
15. L. Zhou, V. Varadharajan, and M. Hitchens, “Achieving secure role-based access control on encrypted data in cloud storage,” *IEEE transactions on information forensics and security*, vol. 8, no. 12, Dec. 2013, pp. 1947-1960.
16. Y. Zhu, G. J. Ahn, H. Hu, D. Ma, and S. Wang, “Role-based cryptosystem: A new cryptographic RBAC system based on role-key hierarchy,” *IEEE Transactions on Information Forensics and Security*, vol. *8, no.* 12, 2013, pp. 2138-2153.
17. L. Zhou, V. Varadharajan, and M. Hitchens, “Trust Enhanced Cryptographic Role-Based Access Control for Secure Cloud Data Storage,” *IEEE Trans. Information Forensics and Security*, vol.10, no. 11, Nov. 2015, pp. 2381-2395.
18. L. Zhou, V. Varadharajan, and K. Gopinath, “A secure role-based cloud storage system for encrypted patient-centric health records,” *The Computer Journal*, vol. 59, no. 11, 2016, pp. 1593-1611.
19. J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, “Ensuring security and privacy preservation for cloud data services,” *ACM Computing Surveys (CSUR)*, vol. 49, no. 1,2016, p. 13.
20. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, “Role-based access control models,” IEEE *Computer*, vol. *29, no.* 2, 1996, pp. 38-47.
21. H. Zhu, L. Wang, H. Ahmad, and X. Niu, “Key-Policy Attribute-Based Encryption with Equality Test in Cloud Computing,” *IEEE Access*, vol. 5, 2017, pp. 20428-20439.
22. Q. Wang, L. Peng, H. Xiong, J. Sun, and Z. Qin, “Ciphertext-Policy Attribute-Based Encryption With Delegated Equality Test in Cloud Computing,” *IEEE Access*, vol. *6*, 2018, pp. 760-771.
23. J. Sun, X. Zhu, and Y. Fang, “A privacy-preserving scheme for online social networks with efficient revocation,” *INFOCOM, 2010 Proceedings IEEE* , March 2010, pp. 1-9.
24. Holger Schulze. (2018 Report). Insider Threat [Online]. Available: https://www.ca.com/content/dam/ca/us/files/ebook/insider-threat-report.pdf
25. A. Sahai, and B. Waters, “Fuzzy identity-based encryption,” *Annual International Conference on the Theory and Applications of Cryptographic Techniques* , Springer, Berlin, Heidelberg, May 2005, pp. 457-473.
26. J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” *IEEE symposium on security and privacy (SP'07)*, May 2007, pp. 321-334.
27. R. Sandhu, D. Ferraiolo, and R. Kuhn,” The NIST model for role-based access control: towards a unified standard,” ACM *workshop on Role-based access control*, Vol. 10, July 2000, No. 344287.344301.

AUTHORS PROFILE

Mrs. P. Mary Jyosthna, received B.Tech. and M.Tech. from JNTUH. Currently she is pursuing Ph.D. in the Department of Computer Science and Engineering at GITAM University (Deemed to be University), Visakhapatnam, Andhra Pradesh. At present she is working as an Assistant Professor in Dept. of CSE at B.V.R.I.T, Narsapur, Hyd., Telangana and she has 14 years of teaching experience. Her research interests include Cloud Computing and Information Security. Life member of CSI.

Dr. K. Thammi Reddy, received M.Tech (CST) from Andhra University and Doctoral degree from JNTUH, in the area of Data mining. Having 25 years of Teaching & Research experience with an expertise in AI, Data Mining & Security. Published good number of papers in the indexed journals. He is Professor & HOD in the Dept. of CSE, Chairperson, Board of Studies, GITAM University. Life member of CSI, IETE, IE, ISTE.