Improved Privacy Multi-Keyword Based Secure Retrieval Scheme For Cloud Data

Sengathir Janakiraman, M. Deva Priya, A. Christy Jeba Malar

Abstract : Secured retrieval of data stored on the cloud has attracted the research community in the recent years. Individuals and business enterprises utilize the cloud for storing and managing potential data due to the significance of high economic savings and comfort. İn this context, data need to be secured before it is outsourced in order to achieve confidentiality in the cloud data retrieval process. In this paper, a secure retrieval process over cloud data using Improved Privacy Multi-Keyword Retrieval Scheme (IPMKRS) is proposed. This secure retrieval process includes a recursive procedure that uses single integer operations which rapidly converge to the collection of points that aid in constructing the scrambled data propagated through the clouds. IPMKRS includes Cloud set-up, initialization and retrieval phases for securing the data shared on the cloud. It plays a vital role in encrypting the cloud data using the concept of Multi-Keyword based Ranking (MKR) derived based on the technique of co-ordinate matching. The cloud servers are responsible for ranking the files that are stored on the server. The results prove that IPMKRS reduces the cost of communication during the identification of top secure files on the file server.

Keywords : Public Cloud, Privacy Multi-Keyword Retrieval Process, Multi-Keyword based Ranking, Cloud servers

1 INTRODUCTION

Cloud computing involves massive, scalable IT enabled capabilities delivered 'As-a-Service' to external customers using internet technologies. It offers several beneficiary aspects to users such as fast deployment of services in the user's environment, easy access, cost effectiveness, rapid provisioning, elasticity of services and sharing of resources in a ubiquitous network. The cloud providers provide resilient access services, mitigation against network vulnerabilities, disaster recovery, demand based storage, security control services and rapid re-composition of services. The main challenge in Cloud computing is the risk involved in understanding the security aspects rightly and finding the responsibility (provider or user) in case of any crisis. Apart from these issues, the cloud business also faces a challenge of addressing privacy issue raised in the new way of computing. The application software and data of an organization are transferred to its data centre. The cloud data centre does not guarantee 100% reliability of data and service management. Transferring an application to the cloud imposes new security issues like unauthorized access of data, attacks based on virtualization, Structured Query Language (SQL) injection attack, cross-site scripting attack, privacy and other control issues raised by the owner of the application software and data, identification based issues in handling data, various issues related to data verification, data interfering and integrity, message confidentiality, data loss, challenges related to authentication of an authorized user of the application and data/message spoofing. Though cloud computing provides better service in sharing of resources, it suffers from high level of security risks.

The cloud deployment models can be categorized into four types' viz., Public cloud, Private cloud, Community cloud and Hybrid cloud.

- ✓ Public cloud: Public cloud provides cost-effective solutions for deployment, management and secures the infrastructure used in the cloud service providing environment. Organizations can use the cloud resources by means of public deployment model on demand basis and pay for their usage. The public cloud computing environment has numerous advantages which makes it dominant over the private and community deployment models. It requires high level of security, since services, applications, data centres and other infrastructures are openly accessed. As the public cloud is benficial, this significant issue is to be considered and demands enhanced security aspects.
- ✓ Private cloud: In a private deployment model, the aspect of security is more significant as the cloud resources are used by an organization after a contract based agreement. When compared to the public cloud deployment model, the private model has more number of security features offered by the cloud service providers. Verification of the cloud resource user, authorization process, deployment and usage of the resource within an organization are the more prominent security aspects of the private deployment model.
- ✓ Community cloud: The community cloud includes framework like environment which allows sharing of resources among the agreed group of users in an organization or multiple organizations. Since this deployment model ensures the concepts of multi-tenancy, the security aspects seem to be the biggest challenge. For the known group of users, this deployment model acts as an open framework which makes security a crucial issue.

In the Cloud Service Provider (CSP), the user's data will be transmitted and stored in the provider's data centre, wherein all types of operations can be carried out. The users

- *Sengathir Janakiraman is working as Associate Professor in the Department of Information Technology, CVR College of Engineering, Mangalpally, Vastunagar, Hyderabad, Telangana, India. E-mail: j.sengathir@gmail.com*
- *M. Deva Priya is working as Associate Professor in the Department of Computer Science & Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India. E-mail: m.devapriya@skct.edu.in*
- *Christy Jeba Malar is working as Associate Professor in the Department of Information Technology, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India. E-mail: a.christyjebamalar@skct.edu.in*

need not deploy a separate data centre model, instead share the resources available on the cloud, which is considered to be a most significant aspect of the cloud computing. However, the security aspects related to data transfer and storage remains crucial in this type of implementation. The organization applies some standard classifications on the data and also follows some cryptographic strategies which needs to be accommodated in the cloud based processing.

Apart from the merits of cloud computing, there are also certain security issues with greater significance demanding attention.

- ✓ Lack of governance
- ✓ Ambiguity in responsibility
- ✓ Authentication and authorization
- ✓ Failure separation
- ✓ Compliance and legal risks
- ✓ Protection of application in cloud
- ✓ Protection of data
- ✓ Dependency on proprietary services
- ✓ Insecure or incomplete data deletion
- ✓ Visibility and audit

Cloud computing not only creates new security risks but it also possess inherent security services in order to solve most of the security issues raised in resources sharing. CSPs offer enhanced security and privacy facilities that influence their scale and skills at computerized infrastructure management tasks. This potentiality in cloud computing is more beneficial to customers with limited security factors in the organization.

2 RELATED WORK

It is essential to maintain confidentiality of identity and support secured access of resources. In the literature, very few works are focussed on the issues in distributed application environments. Bertino et al. (2001) [1] have proposed an identity management system for achieving exhaustive authentication and privacy in cloud infrastructure. The authors have dealt with the issues of sharing the “Master Key Secret” and have used an entity based system to handle diverse entities in the cloud infrastructure. Goyal et al (2006) [2] and Mon & Naing (2011) [4] have designed a hierarchy based encryption scheme in which encryption is done based on data attributes. A weighted parameter based encryption scheme with cipher policy is proposed so as to ensure privacy and integrity in data transformation. The client with weighted parameter set matching is used in shared information decryption. Sanka et al. (2010) [3] have propounded SPICE for identity management to deal with non-linking, delegable authentication and other properties to support access control and privacy. A hybrid scheme is proposed for combining securities at multiple levels along with access control. Chen et al. (2011) [5] have propounded a scheme to safegaurd perceptible information. Information is extracted from the user’s identification information using data mining. Yang et al. (2012) [6] have combined batch verification with identity confirmation and control. Information is broadcast to support authentication and verification. The identity management system does not involve a centralized agent. Zhang et al. (2013) [7] have proposed a trusted identification scheme to ensure secured communication and anonymous authentication of participants. The authors have applied it on a network to handle user’s privacy. Shashidhara & Jaini (2014) [8] have focused on a proof dependent token issuing scheme to deal with identity authentication without any reliable central authority. The propounded architecture of the coordinative authentication scheme is applied to ensure cloud service security. A flexible, vibrant confidential preserving key management system is propounded to ensure client’s privacy in a cloud environment. The above methods deal with security based on data and not on cloud infrastructure, services such as resources, platforms and interfaces. In addition, Almutairi et al. (2012) [9] have proposed a trusted distributed architecture to support access control in distributed systems. These methods deal with policies capable of handling issues related to security. Blind signature is used to authenticate the third party in a secured distributed environment and ensure identification, authentication, accountability and differentiated access control method. More recently, Yang et al. (2012) [10] have designed parameter based signature to ensure authentication and identity confidentiality. A decentralized privacy preserving access control approach is proposed in which replay attacks and Denial of Service (DoS) attacks are dealt with. Chen et al (2014) [11] have proposed vCNSMS, a cooperative network security prototype used in multi-tenant data center. It is demonstrated using a central cooperative method and deep packet review with an open source UTM system. A protection policy is propounded to ensure security rule management for vCNSMS. Diverse security levels have varying packet inspection schemes, and diverse security plugins are imposed. A smart packet decision method is incorporated into vCNSMS for processing flow to defend the system from network attacks at a data center. Luna et al (2015) [12] have designed assessment schemes for performing quantifiable calculation and investigation of the secSLA based security offered by providers based on a set of customer security demands. These schemes aid in improving the specifications of security demands by presenting a simple scheme that permits customers to recognize and signify their security needs precisely. Quantitative Policy Trees (QPT) and Quantitative Hierarchical Process (QHP) are used independently and also together using 2 usecase scenarios and a prototype. Hu & Qiao (2016) [13] have proposed expert knowledge and quantifiable data called Cloud Belief Rule Base (CBRB). This model uses a cloud model to define the belief rule’s referential point for detailing the expert knowledge. Constraint Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm is proposed to obtain optimal parameters. Yang et al (2018) [14] have explored the choice of multi-granularity level of trust, the users’ liking computation and have designed an algorithm for service selection. Initially, trust is evaluated among diverse entities in the human society, and the multi-granularity levels of trust based on Gaussian cloud transformation is built. The user preferences are considered to build a model based on cloud analytic hierarchy. An algorithm is proposed to select service based on fuzzy comprehensive assessment. Wu et al (2019) [15] have dealt with the security of cloud storage providers and third-party intermediaries using equilibrium analysis. A set of game models including providers and users are considered and the results are analyzed for diverse service scenarios. The providers can choose approaches to improve the trustworthiness of their services.

3 PROPOSED SYSTEM

In this paper, Improved Privacy Multi-Keyword Retrieval Scheme (IPMKRS) is presented for analysing its significance in improving the security in the retrieval process. The steps involved in the implementation of IPMKRS and the inferences on simulation results are discussed.

3.1 Improved Privacy Multi-Keyword Retrieval Scheme (IPMKRS)

IPMKRS is an attempt to improve the security of the cloud data by deriving the benefits of co-ordinate mating that aids in realizing the key features of the data documents based on search query. Specifically, inner product similarity is used for estimating the similarity of the data documents. The search potential for securing the cloud data is optimized. IPMKRS includes three steps - Cloud set-up, initialization and retrieval. IPMKRS is mainly propounded for facilitating the encryption of the cloud data using the concept of Multi-Keyword based Ranking (MKR). MKR is based on the technique of co-ordinate matching and the cloud servers are responsible for ranking the files that are stored on the server. This technique of ranking reduces the cost of communication that is incurred during the identification of top secure files on the file server. It also uses searchable key encryption for determining the optimally secure file on the server.

Creation of Cloud Environment

The cloud environment is modelled and the behaviour of resource policies, data centres and virtual machine are visualized using simulation toolkit. It aids in adapting the generic applications to the environment and enables them to be deployed with comfort and ease. It also supports federated and single cloud environments, and the concept of instance type is used in IPMKRS for enabling the categorization of virtual machines with the underlying hardware characteristics. In IPMKRS, the data centres, virtual machines, cloud users and the cloud application policies are implemented using the simulation toolkit.

Initialization Phase of IPMKRS

In the ınitialization phase of IPMKRS, the secure parameters that facilitate feasible search in the network are enabled. It includes two steps namely, set-up and build index steps respectively. The set-up phase is responsible for initializing security and the latter step enforces plaintext operation. For security concerns, much work is imposed on the data owner.

- ✓ Set-up phase of IPMKRS: The owner of the data invokes the operation Key-Generation (λ) for obtaining the secret and the public key respectively for enabling homographic encryption. Further, the owner of the data fixes the specific secret key to the data users who are authenticated in the security process. Furthermore, the owner of the data elucidates the set of keywords, $W = \{kw_1, kw_2, \ldots . kw_n\}$ with the computed İnverse Document Frequency (IDF) and Term Frequency (TF) values from the set of files, $F = \{f_1, f_2, \ldots . f_n\}$. In addition, the data owner creates a 'i+1' dimensional vector, $D_v = \{T_{i:1}, T_{t:2}, \ldots ., T_{i:n}\}$ for each of the file 'F' under the constraint, $T_{i:1} = TF_i - IDF_i$ with $1 \leq j \leq n$. In this context, the search index is estimated as $SI = \{S_{ij} / \ 1 \leq j \leq n\}$.
- ✓ Build Index phase of IPMKRS: The owner encrypts the Search Index (SI) to form Secure Searchable Index (SSI) using $SSI = \{S_i / \ 1 \leq j \leq n\}$ and encrypts the set of files $F = \{f_1, f_2, \ldots . f_n\}$ into $F^1 = \{f_1^1, f_2^1, \ldots . . f_n^1\}$ based on the classical cryptographic techniques. The estimated SSI and F1 are outsourced to the cloud server of the network.

Retrieval Phase of IPMKRS

The retrieval phase of IPMKRS incorporates trapdoor generation, score calculation and rank generation using cloud server and data user. Mostly, the computations are performed on the cloud server as there is limited computation capability in the cloud user perspective. The trapdoor generates the user request. Vector 'T' is constructed from the multi-keyword user request and then encrypted based on the secured trapdoor using the public key generated in the set-up and build index phases. The score is calculated using the determined secure trapdoor and returns the data file to the user who initiated the request. Finally, the user decrypts the vector generated during score calculation using the secret key and updates the top scores of the cloud data files stored in the cloud space. The ranking of the secure data files is done as shown in the following algorithm.

Algorithm - TOPRANKING (Source files, n)
Input - Cloud data to be selected for ranking purpose
n - Number of files to be ranked
Initialize TOP(N) = Φ and TOP(NID)= Φ
Iterate steps 1 to k do
for 'n' number of source files on the cloud server do
ADD (top-n, Index-ITEM-n) and generate Secret key and Public key set
Build index value based on the generated Secret key and public key set
End for
for the entire element in the list of source files
Determine the TOP (NID)
Add tuples into the TOP (NID) based on retrieval when the source keywords possess similar index of categorization
End for
End

4 RESULT INVESTIGATION

In Experiment 1, the performance of IPMKRS is analysed by varying the number of files. The plots of query time based on different thresholds are portrayed in Figure 1 & Figure 2. Table 1 shows the query time for a threshold of 0.3.

TABLE 1.
QUERY TIME FOR THRESHOLD=0.3

No.of files	IPMKRS	MKRS	SRKS
5	1680	1760	1880
10	1780	1850	2010
15	2480	2520	2650
20	3520	3700	3800
25	3590	3620	3970
30	3810	3890	4030
35	5530	5620	5800
40	5650	5820	6810
45	5650	6570	7450

From Figure 1, it is evident that the query time of IPMKRS, Multi-Keyword Retrieval System (MKRS) and Single Keyword Retrieval System (SKRS) decreases marginally with increase in the number of files. The increase in number of files introduces additional time for processing the query. IPMKRS offers a phenomenal query processing rate of 27% greater than the MKRS and SKRS as it uses multi-keyword ranked search encryption. It is transparent that IPMKRS improves the query time by 8%-12% over MKRS and 18%-21% over SKRS. In addition, it is also clear that IPMKRS improves the query time on an average by 17% in contrast to the benchmark security approaches considered for study.

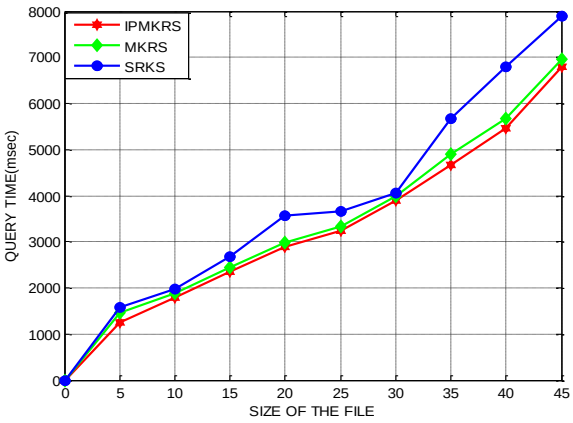

***Fig. 1**. IPMKRS - Query Time (Threshold=0.3)*

Figure 2 shows the query times of IPMKRS, MKRS and SKRS under a threshold of 0.6. This increase in threshold reduces the query processing time by a rapid processing rate of 32% greater than MKRS and SKRS. It is transparent that IPMKRS improves the query time by 10%-15% over MKRS and 21-24% over SKRS. It is also clear that IPMKRS improves the query time on an average by 21% in contrast to the benchmark security approaches considered for study.

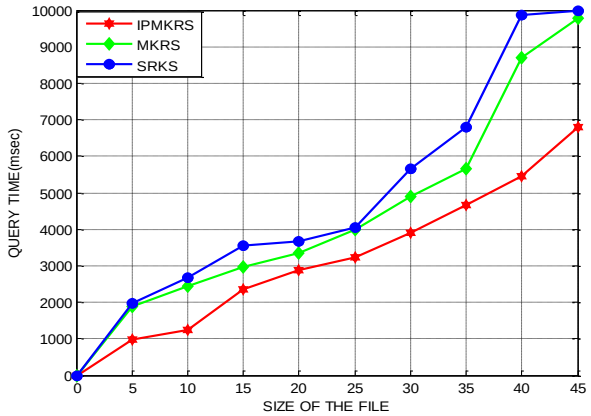

***Fig. 2**. IPMKRS - Query Time (Threshold=0.6)*

Likewise, Figure 3 shows the query time of IPMKRS, MKRS and SKRS for a threshold of 0.8. The increase in threshold reduces an additional time for processing the query by ensuing a rapid processing rate of 35% greater than the considered traditional approaches like MKRS and SKRS. It is transparent that IPMKRS improves the query time by 13%-18% over MKRS and 24%-27% over SKRS. In addition, it is also clear that IPMKRS improves the query time on an average by 25% when compared to the benchmark security approaches.

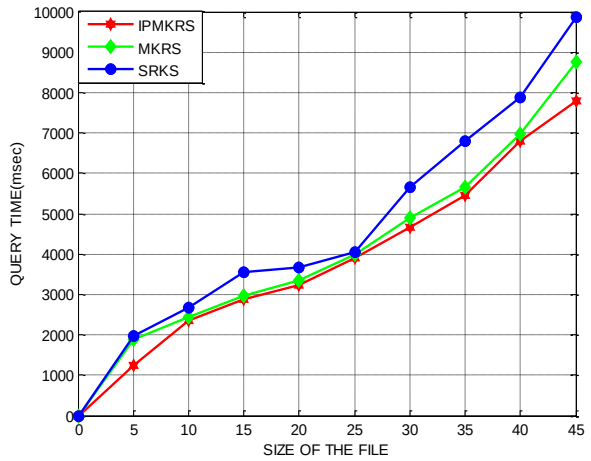

***Fig. 3** IPMKRS - Query Time (Threshold=0.8)*

In Experiment 2, the performance of IPMKRS is analysed based on computation overhead by varying the number of nodes. The plots of computation overhead on different thresholds are shown in Figure 4 & Figure 5. From Figure 4, it is evident that the control overheads of IPMKRS, MKRS and SKRS decrease with increase in the number of nodes. IPMKRS handles this scenario by reducing the computation overhead by 15% in contrast to MKRS and SKRS. It is transparent that IPMKRS reduces computation overhead by 12%-14% over MKRS and 17%-23% over SKRS. IPMKRS reduces the computation overhead on an average by 13% when compared to the baseline approaches.

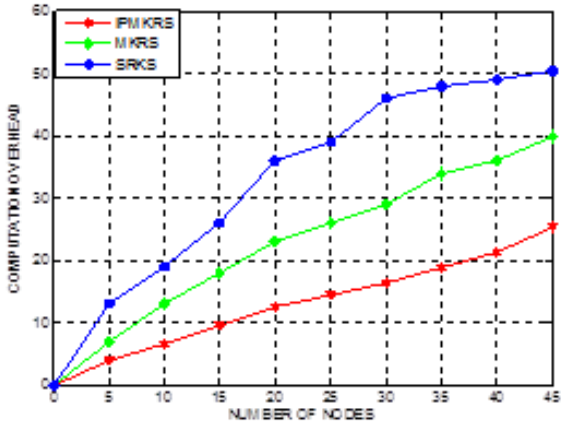

Fig. 4 *IPMKRS - Computation Overhead (Threshold=0.3)*

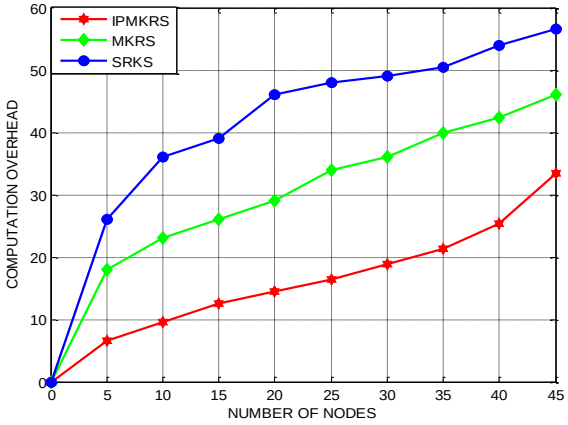

Fig. 5 *IPMKRS -Computation Overhead (Threshold=0.6)*

Figure 5 shows the computation overhead of IPMKRS, MKRS and SKRS under the threshold of 0.6. IPMKRS reduces the computation overhead by 6%-9% over MKRS and 13%-17% in contrast to SKRS. IPMKRS minimizes the computation overhead on an average by 12% when compared to the benchmarked approaches.

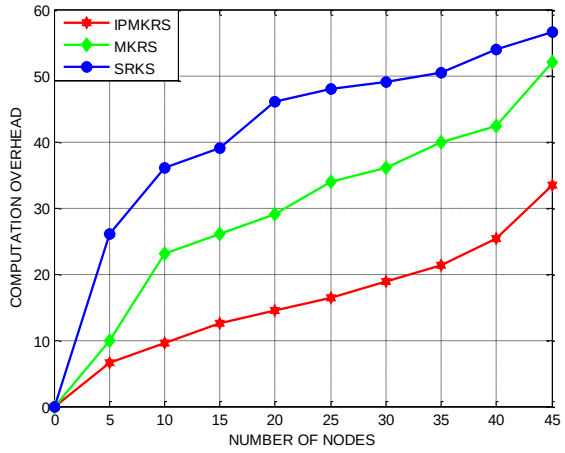

Fig. 6 *IPMKRS - Computation Overhead (Threshold=0.8)*

Figure 6 presents the computation overhead of IPMKRS, MKRS and SKRS for a threshold of 0.8. This increase in threshold reduces the computation overhead, thus facilitating faster data delivery. It is found that IPMKRS improves the query time by 11%-16% over MKRS and 18%-22% over SKRS. In addition, it is also clear that IPMKRS minimizes the computation overhead by 17% in contrast to the benchmark security approaches. In Experiment 3, the performance of IPMKRS is analysed by varying the number of nodes based on storage overhead. The plots of storage overhead on different thresholds are portrayed in Figure 7 & Figure 8.

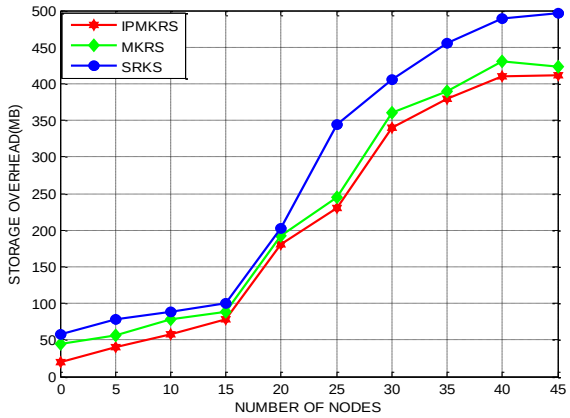

Fig. 7 *IPMRKS Query Time - Storage Overhead (Threshold=0.3)*

From Figure 7, it is evident that the storage overheads of IPMKRS, MKRS and SKRS decrease marginally with increase in the number of nodes. IPMKRS involves a reduced storage overhead of 13% in contrast to the considered traditional approaches like MKRS and SKRS. It is transparent that IPMKRS reduces the computation overhead by 11%-13% over MKRS and 15%-18% over SKRS.

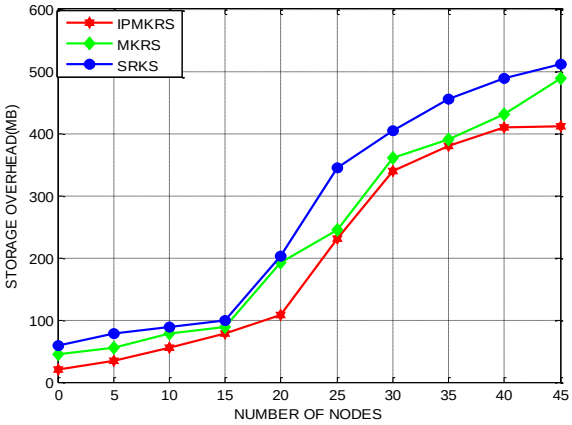

Fig. 8 *IPMRKS Query Time - Storage Overhead (Threshold=0.6)*

Figure 8 shows the storage overhead of IPMKRS, MKRS and SKRS for a threshold of 0.6. IPMKRS reduces the storage overhead by 8%-11% over MKRS and 14%-19% over SKRS. IPMKRS minimizes the storage overhead on an average by 14% when compared to the benchmark security approaches.

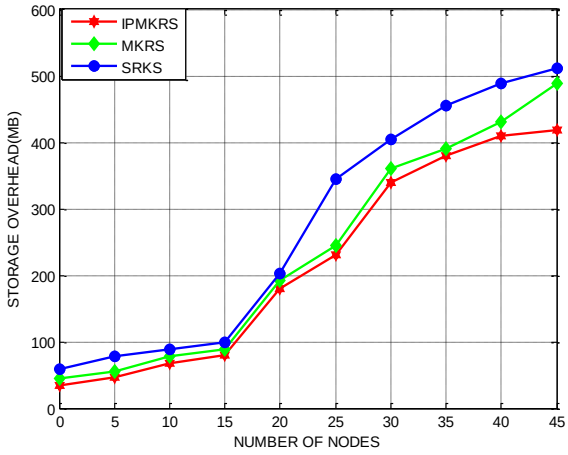

***Fig. 9**. IPMRKS Query Time - Storage Overhead (Threshold=0.8)*

Figure 9 shows the storage overhead of IPMKRS, MKRS and SKRS for a threshold of 0.8. This increase in threshold reduces the storage overhead, thus facilitating faster data delivery. It is found that IPMKRS improves the query time by 13%-18% over MKRS and 19%-24% over SKRS. It is also clear that IPMKRS minimizes the computation overhead by 21% when compared to the benchmark security approaches.

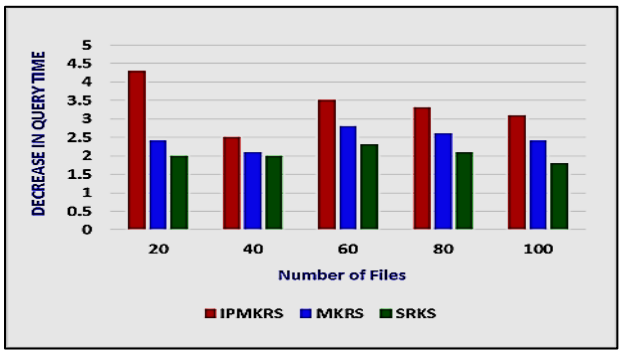

***Fig. 10** IPQUERY Time - Decrease in Query time*

From Figure 10 presents the decrease in query processing time of IPMKRS, MKRS and SKRS for different file sizes. IPMKRS reduces the query processing time by 3%-6% over MKRS and 8%-11% over SKRS.

5 CONCLUSION

In this paper, a secure retrieval process over cloud data using Improved Privacy Multi-Keyword Retrieval Process (IPMKRS) is proposed for securing data shared on the cloud. It incorporates Multi-Keyword based Ranking (MKR) for securing data shared on the cloud. IPMKRS offers 4.18%, 5.68% and 6.84% reduced computation overhead in contrast to the benchmarked schemes. It involves 5.12%, 6.84% and 7.18% less storage overhead when compared to the benchmarked schemes. In addition, the query time of the proposed IPMKRS is minimized on an average by 4.32%, 6.18% and 7.12% in contrast to the benchmarked schemes taken for evaluation.

REFERENCES

[1] Bertino, E., Bonatti P. A., Ferrari, E., "TRBAC: A Temporal Role-based Access Control Mode", ACM Transactions on Information and System Security, vol. 4, no. 3, pp. 91-233, 2001.

[2] Goyal, V., Pandey, O., Sahai, A., Waters, B., "Attribute-based encryption for fine-grained access control of encrypted data", In Proceedings of the 13th ACM conference on Computer and communications security, pp. 89-98, 2006.

[3] Sanka, S., Hota, C., Rajarajan, M., "Secure data access in cloud computing", In Proceedings of the 4th IEEE International Conference on Internet Multimedia Services Architecture and Application, pp. 1-6, 2010.

[4] Mon, E. E., Naing, T. T., "The privacy-aware access control system using attribute-and role-based access control in private cloud", In Proceedings of the IEEE International Conference on Broadband Network and Multimedia Technology, pp. 447-451, 2011.

[5] Chen, K. Y., Lin, C. Y., Hou, T. W., "The low-cost secure sessions of access control model for distributed applications by public personal smart cards", In Proceedings of the 17th IEEE International Conference on Parallel and Distributed Systems, pp. 894-899, 2011.

[6] Yang, K., Jia, X., "Attributed-based access control for multi-authority systems in cloud storage", In Proceedings of the 32nd IEEE International Conference on Distributed Computing Systems, pp. 536-545, 2012.

[7] Zhang, Y., Chen, X., Li, J., Wong, D. S., Li, H., "Anonymous attribute-based encryption supporting efficient decryption test", In Proceedings of the 8th ACM SIGSAC symposium on Information, Computer and Communications Security, pp. 511-516, 2013.

[8] Shashidhara, M. S., Jaini, C. P., "Privacy Preserving Third Party Auditing In Multi Cloud Storage Environment", In Proceedings of the IEEE International Conference on Cloud Computing in Emerging Markets, pp. 1-6, 2014.

[9] Almutairi, A., Sarfraz, M., Basalamah, S., Aref, W., & Ghafoor, A., "A distributed access control architecture for cloud computing", IEEE software, vol. 29, no. 2, pp. 36-44, 2012.

[10] Yang, K., Liu, Z., Cao, Z., Jia, X., Wong, D. S., Ren, K., "Taac: Temporal attribute-based access control for multi-authority cloud storage systems", IACR Cryptology EPrint Archive, pp. 651, 2012.

[11] Chen, Z., Dong, W., Li, H., Zhang, P., Chen, X., & Cao, J., Collaborative network security in multi-tenant data center for cloud computing, Tsinghua Science and Technology, vol. 19, no. 1, pp. 82-94, 2014.

[12] Luna, J., Taha, A., Trapero, R., & Suri, N., "Quantitative reasoning about cloud security using service level agreements", IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 457-471, 2015.

[13] Hu, G. Y., & Qiao, P. L., "Cloud belief rule base model for network security situation prediction", IEEE Communications Letters, vol. 20, no. 5, pp. 914-917, 2016.

[14] Yang, Y., Liu, R., Chen, Y., Li, T., & Tang, Y., "Normal cloud model-based algorithm for multi-attribute trusted cloud service selection", IEEE Access, vol. 6, pp. 37644-37652, 2018.

[15] Wu, Y., Lyu, Y., & Shi, Y., "Cloud storage security assessment through equilibrium analysis", Tsinghua Science and Technology, vol. 24, no. 6, pp. 738-749, 2019.