Secure Data Aggregation Scheme for IOT Applications in Cloud

Mallareddy A, Sridevi R, Prasad Ch G V N

Abstract: While Internet of Things (IoT) technology comprises of nodes that are self-configuring and intelligent which are interconnected in a dynamic network, utilization of shared resources has been revolutionized by the cloud computing effectively reducing the cost overheadamong the cloud users.The major concerns of IoT infrastructure are reliability, performance, security and privacy. Cloud computing is popular for its unlimited storage and processing power. Cloud computing is much more matured with the capability to resolve most of the issues in IoT technology. A suitable way to address most of the issues in IoT technology is by integrating IoTparadigm into the Cloud technology.In this regard, we propose a methodology of applying our EPAS scheme for IoT applications. In our previous work[2] , we have proposed an Enhanced Privacy preserving gene based data Aggregation Scheme (EPAS) for private data transmission and storage by utilizing Enhanced P-Gene erasable data hiding approach. Enhanced P-Gene scheme ensures secure transmission and storage of private data by relying on a data aggregation scheme fully dependent on erasable data hiding technique. In the current work we analyse the applicability of the EPAS scheme for IoT applications. Experimental results show the suitability of the proposed scheme for application involving numeric data and also demonstrates performance improvement with existing proposals for data aggregation in cloud.

Index Terms: Cloud Computing, IoT, Data Hiding, Data Security, Data Aggregation, Enhanced P-Gene.

I. INTRODUCTION

Cloud computing has become very popular due to the efficient utilization and accessability of resources by viewing them as a service [11, 12]. The benefits of cloud resources can be utilized without realizing the exact location of such resources to predominently achieve data access and storage [9, 10, 3].The cloud computing environment also supports applications that require location independent data storage and periodic exchange of data in the cloud infrastructure[14].

Data security is a crucial issue that needs to be addressed in cloud data storage and transmission since the communication media used for transmission are insecure thereby increasing the risk of data misuse. An efficient method for achieving data privacy and security is the need of the hour to fully utilize the power of cloud data services.

Several solutions to address the issues of security and privacy in cloud computing have been proposed in the literaturein terms of monitoring cloud server services, data privacy protection , computation ofinformation accuracy, protection against modifications and deletions, prevention of malicious data and ensuring uninterrupted access to data [8]. However, majority of proposed approaches suffer from the overhead incurred by the encryption techniques invovled.

Hence an efficient methodology for data transmission, storage and aggregation is required to reduce the inherent computational complexity of encryption techniques.

To alleviate this problem, in [2]we proposed an Enhanced P-Gene technique for secured data transmission and storage in cloud (EPAS scheme).The focus of the proposal is on applications having a group of user nodes requiring a secured storage facility for aggregate data in the cloud. The application comes with a constraint that private data of user nodes are invisible to the aggregator in the cloud server, which makes the problem more challenging. However, data aggregation and data privacy protection contradict each other since to achieve data aggregation, any aggregator must access the original data. Several proposals have been made using End-to-end data encryption andsteganography. Our proposed EPAS method in [2] specifically addresses such contradiction without the usage of encryption techniques, thereby improving the overall performance.The scheme adopts data hiding to achieve private data aggregation.

The main contributions of the scheme proposed in [2]are as follows:

1. Enhanced P-Gene, a data-hiding scheme is constructed based on which an erasable data hiding technique is proposed. Here each user node hides its private data using simple mathematical calculations and the hidden data is then sent to cloud where it is aggregated with hidden data sent from other nodes in the cloud server. The aggregator in the cloud does not have access to the original private data of the user nodes connected in the group.
2. A method for secret Enhanced P-Gene generation is proposed wherein each user node independently and dynamically generates its Enhanced P-Gene through simple mathematical operations.

In this paper as an extension of the previous work in[2], we work on the applicability of the proposed method in applications that utilize a group of nodes and require aggregation of data sent from various nodes. One such emerging area that has received great attention and is applied in various domains currently is Internet of things(IoT).

Cloud computing and Internet of Things (IoT) are two very different technologies that are extensively used in many real-world applications [6]. The Internet of Things (IoT) paradigm is an interconnected dynamic network of nodes which are intelligent and self-configuring[7]. Characteristics of IoT include real world smart devices with limited storage and processing capacity that are widely distributed. The major concerns of IoT infrastructure are reliability, performance, security and privacy. On the other hand, Cloud computing has virtuallyunlimited capabilities in terms of storage and processing power.

Revised Manuscript Received on March 18, 2020.

Mallareddy A, AssociateProfessor, CVR College of Engineering and Research Scholar, JNTUH Hyderabad

Dr. Sridevi R, Professor and HOD, Department of CSE,JNTUH College of Engineering, Hyderabad

Dr. PrasadCh G V N, Professor and HOD, Department of CSE,
Sri Indu College of Engg & Tech., Hyderabad

Cloud technology ismuch more matured and has the capability to resolve most of the IoT issues. Most of the issues caused by limitations of IoT devices such as low processing ability, energy consumption can be solved by integrating IoT infrastructure with the Cloud environment.The Cloud can also benefitfrom the integration with IoT by providing new services for billions of devices in variousreal-life scenarios. The Cloud computing environment also simplifies the flow of the IoT based device generated datacollection and processing thereby providing low latency, low-costinstallation and simplified integration for complex data processing[7].A novel technology that merges Cloud and IoT together is the future since it supports a variety of real worldapplication [6].

At the outset Cloud can deal with more real life applications. Data sensed by various smart devices like mobile phones, sensors, vehicles, etc can be sent to cloud for faster processing and storage in various applications. However, though offloading the computation and storage tasks to the cloud may look attractive, the latency sensitive, security driven IoT applications need techniques that are efficient and secure in the cloud environment[5]. In this regard, we have conducted experiments to emphasise the applicability of our proposed method in [2] for such applications.

The paper is organized as follows : Section II discusses the various data hiding techniques and data aggregation methods applied for cloud data security. We briefly summarize our proposed Enhanced Privacy preserving gene based data Aggregation Scheme(EPAS)[2] in Section III. The experimental results are discussed in section IV. We conclude in Section V.

II. RELATED WORK

In [3] a scheme for privacy preserving data aggregation is proposed by Xiong Li et al. for mobile computing edge assisted IoT applications , where there is a necessity of privacy of data at terminal level. This approach involves three devices namely terminal device, edge server and the public cloud center. In this approach, each terminal device encrypts the data and sent to the edge server. The edge server acts as a bridge between the terminal device and cloud center where it's key role is to aggregate the data and forward the cipher text to the cloud center. The public cloud center uses its private key to recover the plain text from aggregated cipher text. This approach guarantees the data privacy at terminal devices, provides source authentication and integrity.

The proposed scheme follows Boneh-Goh-Nissim homomorphic crypto system using the following phases. Privacy preserving is done using bilinear map of composite order groups.

1. The public cloud center generates public/private keys for itself and the edge servers in addition to generating the system parameters during initialization.
2. During the registration phase, the terminal devices register themselves with the private key generated using their identity.
3. The plain text is encrypted at the terminal device and the resultant cipher text is sent to the edge server.
4. The edge server validates and aggregates the cipher texts received from terminal devices. After adding it's own signature to the aggregated cipher text, each edge server reports it to the cloud center.
5. Finally, the public cloud center validates and recovers the plain text from aggregated cipher text received from various edge servers.

Huaqun Wang et al. proposed anonymous and secure aggregation scheme in fog-based public cloud computing[4]. In this approach, each fog node is responsible for aggregating the data received from the terminal devices and forwarding the data to public cloud server.The main objective of this proposed architecture is to reduce the bandwidth utilization between fog node and the public cloud server. This approach protects the identities of the terminal devices by using pseudonyms and maintains the secrecy of data by applying homomorphic encryption technique. This scheme involves cryptographic techniques, which includes signature, encryption and aggregation. This scheme uses elliptic curve public key cryptography.

Zhitao Guan et al. proposed device oriented anonymous privacy preserving scheme (APPA) for data aggregation in fog enhanced IoT environment[5]. This scheme guarantees anonymity and authenticity of the smart device using pseudonym and pseudonym certificates.In the proposed scheme, asymmetric encryption technique called Paillier Cryptosystem is followed which involves key generation, encryption and decryption steps.

1. The smart devices collaborate with local certificate authorities and trusted certificate authorities for generation of pseudonym certificates to prevent themselves from certificate forgery.
2. Certificate update at smart device is done autonomously using on demand approach.
3. Reliable data aggregation is followed to guarantee data authentication, integrity verification and privacy preservation.

Murat Yesilyurt and Yildiray Yalman proposed a data hiding scheme based on watermarking in order to secure the user access by the process of hiding users information by defining a coverage file that results in a highly secure covered audit [15]. This data hiding scheme was consideredcapable of securing the services by preventing external attacks through the encryption process enabled by the cloud architecture and deployment models used for security. This data hiding scheme was proved to be superior in accessing security and robust during the process of storing data and access them in the time of usage.

A reversible data hiding technique was proposed for hiding the data by applying the process of multiple encryption [16]. This reversible data hidingscheme confirmed superior recovery of hidden databy utilizing the method of histogram shuffling process. This reversible data hiding scheme was also proved to recover hidden data andcover images without any kind of errors that are possible under data exchange in cloud computing processes. The computation in during the process of data hiding was considered to be phenomenally minimized than most of the works of the literature.

III. EPAS SCHEME SUMMARY

Inour previous work we proposed a new private data aggregation scheme based on the Enhanced P- gene scheme for secured storage of private data in cloud. The scheme adopts data hiding to achieve private data aggregation. A brief summary of the EPAS scheme [2] is discussed in this section.

A Enhanced P-Gene (Privacy Preserving Gene)

The scheme assumes working with a group of user nodes along with the cloud server $G_a$of size , each having a unique group ID from {1,..,*n*}. G_a^1 refers to a subset of $G_a$that comprises of the user nodes participating in data aggregation. An Enhanced P-Gene methodology is proposed for data hiding to achieve private data aggregation.The scheme uses a mathematical model for data hiding and the key terms are explained below:

1. *P-list :*P-list refers to the list of integers generated by every user node b ($b\in G_a^l$) for enhanced P-Gene generation. The P-list $\{P_c^b, c\in G_a^l\}$satisfies

$$\left(\sum_{c\in G_a^l} P_c^b\right) modU=0 \qquad (1)$$

where $U\geq d_{max}\times n$, d_{max} being the upper bound of user data, n being the maximum group size that includes the cloud server and l being the number of bits in U.

2. *P-Seed :*Each $P_c^b, c\in G_a^l$ is called a P-seed and is only shared between user nodes *b* and *c*.

3. *Enhanced P-Gene :*

The enhanced P-Gene of any random node $b(b\in G_a^l)$ is denoted as R^b

$$R^b=\left(\sum_{c\in G_a^l} P_b^c\right) modU \qquad (2)$$

The values in the nodes satisfies the following property - Let d^b denote data in user node b, D^b denote the hidden data

$$D^b=\left(d^b+R^b\right)\ modU$$

For any group G_a^l,

$$\left(\sum_{b\in G_a^l} R^b\right) modU=0 \qquad (3)$$

$$if\left(\sum_{b\in G_a^l} d^b\right)\leq(U-1)\ then$$

$$\left(\sum_{b\in G_a^l}(d^b+R^b)\right) \text{mod}\quad U = \sum_{b\in G_a^l} d^b \qquad (4)$$

This property assures that for any random G_a^l, the sum of all hidden data $D^b(b\in G_a^l)$ is equivalent to that of all original data d^b under the modular addition operation.

B EPAS (Enhanced Privacy Preserving Gene Based Aggregation Scheme)

The EPAS Scheme proposed**[2]**ensures data privacy in the additive data aggregation process based on the erasable data hiding technique.

Table 1 presents the basic notations used in the scheme.

Table 1: Basic notations used in EPAS.

Notation	Meaning
G_a	Group of cloud server and ($n-1$)user nodes
G_a^l	Group of cloud server and *m* participating user nodes $m\leq(n-1)$
P_c^b	Secret P-Seed generated by user *b*for P-Gene generation and shared between user nodes *b* and *c*.
r_c^b	Secret seed for P-Seed generation that is generated by user *b* and shared between user nodes *b* and *c*.
d^b	Original data in user node *b*.
R^b	Enhanced P-Gene in user node *b*
D^b	Hidden data in user node *b* where $D^b=(d^b+R^b)\ modU$

Fig1(a) shows how each data hiding is done in each user node using the enhanced P-Gene and how the hidden data is sent to the cloud server(CS). This ensures data privacy during the communication process.

$$\left(\sum_{a\in G_a^l}(d^a+R^a)\right) modU=\sum_{a\in G_a^l} d^a \qquad (5)$$

The cloud server receives the data from the users in the group G_a^l, erases the P-Genes from the hidden data and obtains the aggregation result as shown in Fig1(b).

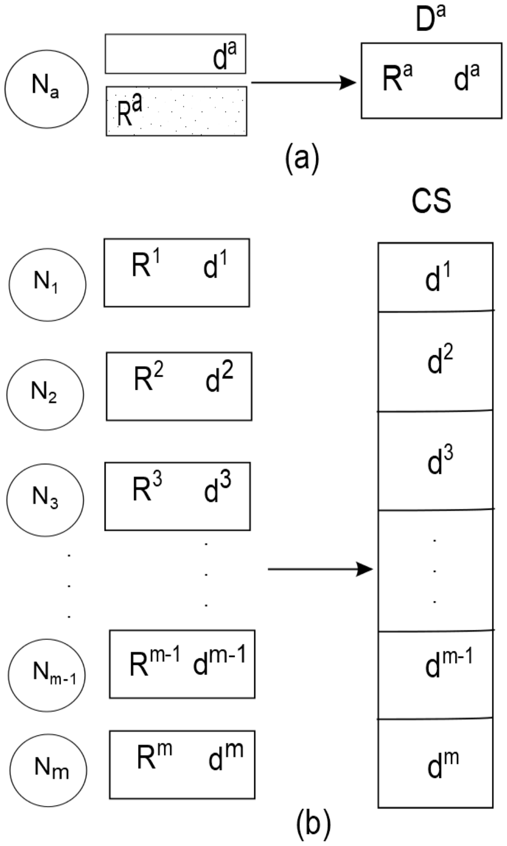
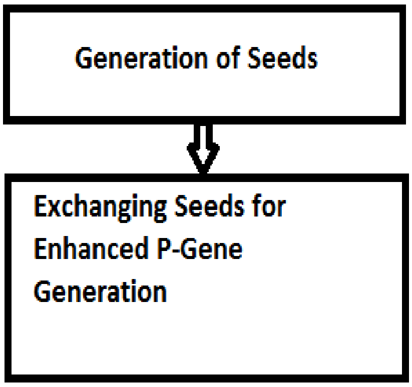
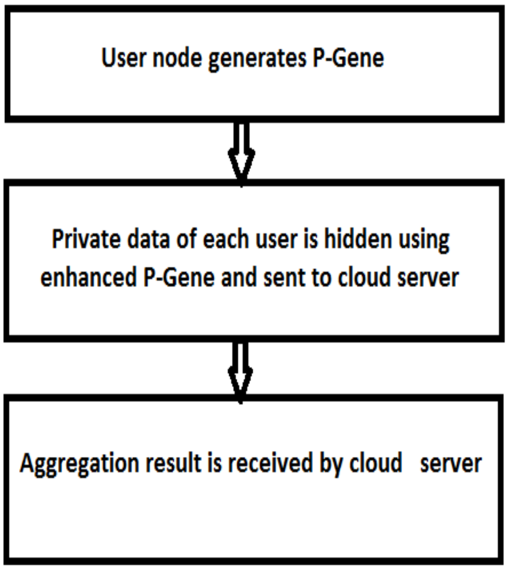


Figure 1: EPAS Scheme

The process of EPAS is shown in Fig 2(a) and 2 (b)



(a) Initialization Process


User node generates P-Gene
Private data of each user is hidden using
enhanced P-Gene and sent to cloud server
Aggregation result is received by cloud server

(b) Data Reporting Process

Figure 2: Process of EPAS

IV. EXPERIMENTAL RESULTS

The effectiveness of Enhanced Privacy preserving gene based data Aggregation Scheme (EPAS) method is shown by conducting two sets of experiments:

1. Performance Analysis of our EPAS approach.
2. Performance comparison with existing approaches.

Experimental Setup

To evaluate the performance of EPAS, asdiscussed in section III, we conducted experiments on CloudSim [1], a simulation framework for cloud computing.CloudSim is the most popular and widely used simulator from CLOUDS Laboratory, at the Department of Computer Science and Software Engineering Department, University of Melbourne. It provides a generalized and extensible simulation framework using which we can model, simulate, and experiment of research proposals in emerging Cloud computing infrastructures and application services.

We ran our experiments on a 1Intel Core (TM) i3-5005U CPU @2.00GHz (4 CPUs), ~2.0GHz machine with 4 GB RAM running Microsoft Windows 10. Our methodology was implemented with JDK 1.8, Eclipse **3.6.0** and CloudSim 3.0.3. In all the experiments, the values of U were taken as 12626 in all the test cases. The experiments were initially conducted for 4 user nodes and then repeated for a varying number of user nodes from 5 to 100.

1. Performance Analysis of EPAS Approach

Here, we analyse the performance of our EPASin terms of varying number of user nodes sending data, varying number of VMs (Virtual Machines) connected to the Data Center in CloudSim and with varying configurations of the VMs (Virtual Machines) connected to the Data Center in CloudSim. The time taken for P-Gene generation for varying number of user nodes is also analysed.

Seed generation and P-Gene computation is an important phase of our proposal and is done initially when the group of user nodes working on the same application try to store the data in the cloud server and these values are used for the group for further data storage too. We analysed the time taken for this entire process by varying the number of nodes as shown in Fig 3. It can be seen that the time taken although increases with increasing number of nodes, the time taken for P-Gene generation is very less (in microsecs), even when the number of nodes is high. Hence this time can be ignored since it's a one time process for a given user node group.

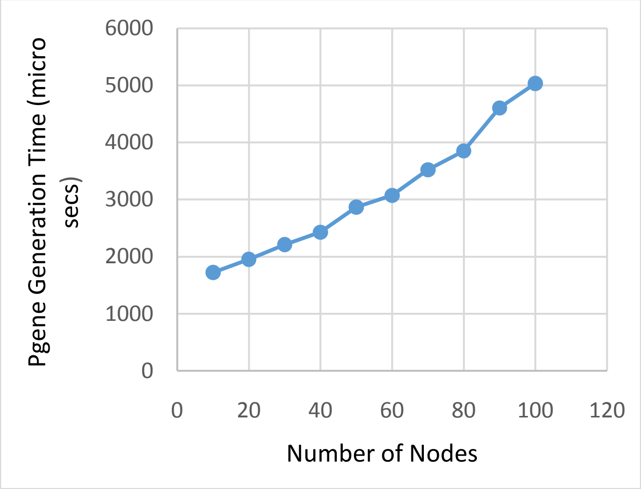


Fig 3. Number of Nodes v/s PgeneGeneration Time

We next analysed the time taken for the entire P-Gene data aggregation which includes the time taken for P-Gene generation, time for data hiding in the user nodes, transferring hidden data from the many user nodes to the Data Center in CloudSim via the VM and the data aggregation time in the Data Center. Initially we tested the time taken keeping only a single VM for the DataCenter and varied the number of nodes from 4 to 100. The performance of our proposed method is plotted as a graph in Fig 4. It can be noticed that the time for data aggregation increases with increasing number of nodes and is linearly proportional to the number of nodes. However, the time taken for the entire process is just a few seconds even for high number of user nodes.

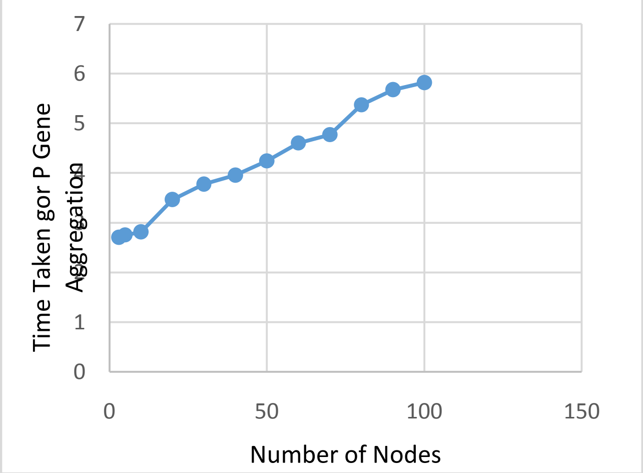


Fig 4. Time Taken for PGene Data Aggregation V/S Number of Nodes

We then varied the number of VMs connected to the DataCenter in CloudSim, all VMs having the same configuration, and analysed the performance of our methodology, the results are as seen in Fig 5. From the results obtained it can be inferred that there is an improvement in the performance with an increase in number of VMs in the cloud, especially for greater number of user nodes. Hence it is suggested to have more VMs for a higher number of user nodes for efficient performance of the approach.

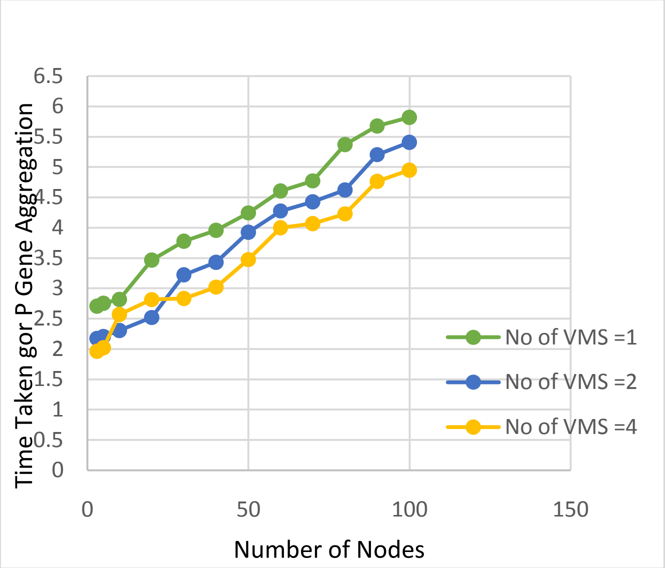


Fig 5. Time Taken for PGene Aggregation V/S Number of Nodes

2. Performance comparison with existing approaches.

We have compared the efficiency of our proposed EPAS scheme with few existing approaches**[2,3]**.These approaches also proposes data aggregation methods for cloud environment and have been discussed in section II. The total time taken for the approaches compared includes time taken for encryption in user nodes, data transfer timebetween the user nodes and cloudsim and the time taken for decryption and aggregation in Data Center. The total time taken by our proposed approach includes time taken for P-Gene generation, time for data hiding in the user nodes, data transfer time from the many user nodes to the Data Center and the data aggregation time in the Data Center.

The performance analysis was done by executing all the 3 approaches in CloudSim, with varying number of user nodes. For all the 3 approaches a single VM was connected to the Data Center. The results obtained are as shown in Fig 6. It can be noticed that our proposed approach outperforms both the compared approaches. The performance improvement obtained is mainly due to the simple mathematical model used in our proposed approach. This approach does not involve the overhead time for encryption and decryption. Hence it can be concluded that the proposed EPAS approach is efficient and it is scalable for any number of user nodes. Also, there is a significant performance improvement with increasing number of VMs for higher number of user nodes.

Retrieval Number: D10170394S220/2020©BEIESP
DOI: 10.35940/ijitee.D1017.0394S220

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

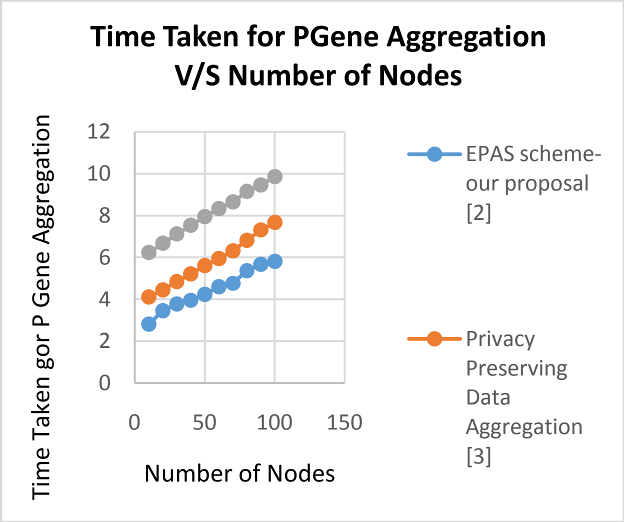


Fig 6. Comparison of PGene Aggregation method V/S Other methods

V. CONCLUSION

The major concerns of IoT infrastructure are reliability, performance, security and privacy. Cloud computing is popular for its unlimited storage and processing power. Cloud computing is much more matured with the capability to resolve most of the IoT issues.The integration of the IoT into the Cloud is certainly the most suitable way to resolve most of the issues in IoT.The Cloud can also benefitfrom the integration with IoT by providing new services for billions of devices in variousreal-life scenarios. A novel technology that merges Cloud and IoT together is the future since it supports a variety of real-worldapplication [6].

At the outset Cloud can deal with more real life applications. Data sensed by various smart devices like mobile phones, sensors, vehicles, etc can be sent to cloud for faster processing and storage in various applications. However, offloading the computation and storage tasks to the cloud may look attractive, the latency sensitive, security driven IoT applications need techniques that are efficient and secure in the cloud environment[5].

In this regard, we have conducted experiments to emphasise the applicability of our proposed method in [2] for such applications. In the current work we analyse the applicability of the EPAS scheme for IoT applications. We conducted 2 sets of experiments – for analysing the performance of our proposed approach and for comparing the efficiency of our approach with the existing approaches. The experiments were conducted in cloudsim, a standard and widely used simulation for cloud computing. From the results we can infer that our proposed work is efficient and also scalable for varying number of user nodes. Experimental results show the suitability of the proposed scheme for application involving numeric data and also demonstrates performance improvement with existing proposals.

REFERENCES

1. Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, Cesar A. F. De Rose, and Rajkumar Buyya, CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, Software: Practice and Experience (SPE), Volume 41, Number 1, Pages: 23-50, ISSN: 0038-0644, Wiley Press, New York, USA, January, 2011.
2. A Mallareddy, R Sridevi, Ch G V N Prasad, "Enhanced P-Gene based Data Hiding for Data Security in Cloud", IJRTE, Volume-8 Issue-1, May 2019, ISSN: 2277-3878
3. Li, Xiong, et al. "Privacy preserving data aggregation scheme for mobile edge computing assisted IoT applications." *IEEE Internet of Things Journal* 6.3 (2018): 4755-4763.
4. Wang, Huaqun, Zhiwei Wang, and Josep Domingo-Ferrer. "Anonymous and secure aggregation scheme in fog-based public cloud computing." *Future Generation Computer Systems* 78 (2018): 712-719.
5. Guan, Zhitao, et al. "APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT." *Journal of Network and Computer Applications* 125 (2019): 82-92.
6. Botta, Alessio, et al. "Integration of cloud computing and internet of things: a survey." *Future generation computer systems* 56 (2016): 684-700.
7. Atlam, Hany F., et al. "Integration of cloud computing with internet of things: challenges and open issues." *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2017.
8. Avizienis A, Laprie J C, Randell B and Landwehr C 2004 Basic concepts and taxonomy of dependable and secure computing. IEEE Trans. Depend. Secure Comput. 1(1): 11-3
9. Balu, A., & Kuppusamy, K. (2014). An expressive and provably secure Ciphertext-Policy Attribute-Based Encryption. Information Sciences, 276(2), 354-362.
10. Liang, X., Cao, Z., Lin, H., & Xing, D. (2009). Provably secure and efficient bounded ciphertext policy attribute based encryption. Proceedings of the 4th International Symposium on Information, Computer, and Communications Security - ASIACCS '09, 1(1), 56-87.
11. Sarhan, A. Y., & Carr, S. (2017). A Highly-Secure Self-Protection Data Scheme in Clouds Using Active Data Bundles and Agent-Based Secure Multi-party Computation. 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud). 2(1), 59-83.
12. Waters, B. (2011). Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization. Public Key Cryptography â€“ PKC 2011, 1(1), 53-70.
13. Yang, K., Jia, X., & Ren, K. (2013). Attribute-based fine-grained access control with efficient revocation in cloud storage systems. Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security - ASIA CCS '13, 1(2), 34-47.
14. Zhu, X., Liu, Q., & Wang, G. (2016). A Novel Verifiable and Dynamic Fuzzy Keyword Search Scheme over Encrypted Data in Cloud Computing. 2016 IEEE Trustcom/BigDataSE/ISPA, 2(1), 45-59.
15. Murat Yesilyurt & Yildiray Yalman. (2016). New approach for ensuring cloud computing security: using data hiding methods, Sadhana, 1(1), 36-47.
16. Chendulkar, N. N., & Mahajani, P. (2015). Reversible Data Hiding in Cloud Based Applications. 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 1(1), 24-32.

AUTHORS PROFILE

Mallareddy A, M.Tech(Ph.D) is working as Associate Professor in Department of Information Technology at CVR College of Engineering, Hyderabad, andalso pursuing Ph.D in Computer Science and Engineering from JNTUH Hyderabad. His research interests focus on Cloud Security, Cryptography and Network Security.

Email:mallareddyadudhodla@gmail.com

Dr. R. Sridevi is a Professor and heading Computer Science andEngineering Department at JNTUCEH, Hyderabad, Jawaharlal Technological University Hyderabad.She received her Ph.D in 2010. Her research interests includeData Structures, Steganography, Steganalysis, Network security and Cryptography, Computer Networks and Cloud Security. She has published more than twenty five research papers in reputed journals and eight international andnational conferences.

Email:sridevirangu@jntuh.ac.in

Dr. Ch GVN Prasad ,M.Tech. , Ph.D(Experience-- 20 years ; 12 years IT industry (8 years in National Informatics Centre, Govt. of India, as Scientist and Software Analyst in AT&T in US)and 19years Teaching as Professor and HOD of CSE dept). He is currently working as Professor in Department of Computer Science & Engineering in Sri Indu College of Engg& Tech. Hyderabad. His research interests include Network security and Cryptography, Data mining, Cloud Security. He has published more than twenty research papers in reputed journals , international and national conferences.
Email: prasadch204@gmail.com