ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

Enhancing Data Security during Transit in Public Cloud

[1]Amalraj Irudayasamy, [2]Arockiam L, [3]Veeraragavan N
[1]Research Scholar in Computer Science, Periyar University, Salem, Tamil Nadu, India
[2]Associate Professor in Computer Science, St. Joseph's College, Trichy, Tamil Nadu, India
[3]Research Scholar in Computer Science, St. Joseph's College, Trichy, Tamil Nadu, India

Abstract *– Cloud computing has developed as a widespread model in computing world in which resources of the computing infrastructure are provided as services over the Internet. A cloud computing scenario usually tailors the server as needed. Cloud computing pattern also brings out several new challenges in data security and access control when consumers outsource delicate data for sharing on cloud servers. Data security and admission control is one of the most challenging ongoing study in cloud computing. Therefore, it is essential that organizations regularly addresses risks such as location of data, industry compliance decrees and third party duties surrounding the usage and processing of sensitive data. However, security and privacy are still termed by many organizations as the highest inhibitors of cloud services adoption, which leads to introduce cloud encryption systems through anonymization.***

Keywords **–Cloud Computing, security, privacy, encryption, Data anonymization.**

I. INTRODUCTION

Many organizations compel their business counterparts to make use of cloud technologies in their overall IT strategies. However, when moving to cloud, the queries of data privacy and security are usually raised. Business procedures and the governing jurisdictions impact the way sensitive data is managed. The location, type of data and the data access, habitually decide the degree to which organizations can understand the worth of cloud computing. One of the frequently asked questions is if information is really protected in the cloud. It is a very scary aspect that our valuable information will get affected by malwares and hackers, even though it is safeguarded by firewalls or stored in on-premises servers [1]. But keeping all valuable company information in one location in physical servers is like having all eggs in one container. If one virus enters, if there is a letdown of the firewall, if someone hacks into the infrastructure, information is compromised. Best practices of cloud security and its related techniques are the requirements for cloud data security. By applying appropriate techniques in cloud data security it becomes more effective [2]. These measures provide a strong base which protects data at rest as well as in motion. Currently the practice of having encryption through anonymizationis a key constituent for cloud security. Customer or tenant control over these techniques will differ on the service and the deployment model. The remainder of this paper is organized into different sections in which the cloud computing security in general is presented in section II. In section III, the background and related work has been discussed. In section IV, overviews of data at rest and in transit are provided. In section V, a layered methodology is proposed. The overview of proposed system architecture based on the layered methodology is given in section VI. In section VII and VIII performance evaluation and conclusion are discussed respectively.

II. CLOUD COMPUTING SECURITY

Cloud computing makes everything flexible and easier but, there is an added phase concerning the security of data. It is doubt that, cloud computing provides confidentiality, integrity and being regulated by different compliances [3]. Resources are centralized in cloud computing. So, the disclosure as pectin creases proportionally which results to risk and it becomes important to have a countermeasure to mitigate the risk. The characterization of Cloud computing does not mention any security notion of the data that is stored in the Cloud. Therefore it is understood that the Cloud Computing is lacking security, confidentiality and visibility. To Provide Infrastructure (IaaS), Platform Service (PaaS) or Software (SaaS) as a Service is not sufficient if the cloud provider does not guaranty a better security and confidentiality of customer's data [4]. By convention, it is considered as Cloud computing is any treatment or storage of personal or professional information which are realized outside the concerned structure i.e. outside the company. To secure the Cloud means secure the treatments, storage and transition of data hosted by the Cloud provider. Cloud providers mostly use the virtualization on their Cloud platform and on the same server can co-exist a virtualized storage and treatment space that belong to concurrent enterprises [5]. The aspect of security and confidentiality must intervene to protect the data from each of the enterprises. Secure storage and data transmission requires using a modern aspect of encrypting that has the criteria for treatment such as, the necessary time to respond to any request sent from the client and the size of an encrypted data which will be stored and transmitted on the Cloud server. Security has emerged as the most important barrier to faster and widespread adoption of virtualization as well as cloud computing. Security depends from person to person as well as industry to industry how they analyze the concept of security in Cloud Computing [6]. Many questions arises while shifting to cloud about the security of data, location of data, data access, trustworthiness of the third party, confidentiality and security concerning the different clients data separated and inaccessible from other clients. Since the data is in the cloud, different companies and countries have different requirements, as well as controls placed on access, companies who adopt cloud may not realize that the data must reside in some physical location. Every cloud provider should have all the agreements in writing to provide maximum transparency to provide the different level of security required by different customers [7]. Each cloud provider must have fixed service level agreements regarding various things such as data privacy, limit of third party access to the confidential data etc.Access control is a key concern as insider attacks also possess a huge risk [8]. Persons entrusted with proper authentication to the cloud could be a potential hacker. Standards have been defined to ensure that third parties have sufficient control in handling data and ensure maximum cloud security.

III. BACKGROUND

Balakarishnan.S et al.[9]has introduced TPA (Third Party Auditor) between client and cloud service provider, which acts as external auditor to audit the user outsource data. This scheme has provided secured and efficient dynamic operations like data updates, deletion and append on data blocks stored in the cloud. But the method to secure client's resources reside on cloud server is beyond the scope of this paper. Joshi Ashay.M et al.[10] argued that Asymmetric Cryptography Algorithms and Digital Signature techniques are reliable and efficient to provide more security user's data in Cloud Computing. The potentiality of the paper was that, the authors have seeded the idea of using two different keys algorithms but failed to give the model or methodology for implementations. Richard Chow et al.[11] described a framework for supporting authentication decision, which is named as Trust Cube. A proposal of a high-level architecture of authentication flows was made. The architecture has four participants: client devices, data aggregators, an authentication engine, and authentication consumers. Client device, data aggregators and authentication consumers must be authenticated themselves through authentication engine before exchange of data. The strength of that paper was a model of cloud authentication. However this article only focuses on one threat (Authentication), facing Cloud Computing. Other threats in Cloud environment such as Repudiation, Denial of Services and Spoofing identity are probably ignored by the authors. Hongwei Li et al. [12] presented a Hierarchical Architecture for Cloud computing and proposed Identity-Based Encryption and Identity-Based Signature for that Hierarchical Architecture. Finally the author proposed Authentication Protocol for Cloud Computing (APCC). In the end a conclusion were made that, APCC is more light weight and efficient as compared to SSL Authentication Protocol, on the basis of performance analysis. The merits of this research works are the Cloud computing model along with Authentication Protocol. Moreover, the authors have given the simulation results to support the proposed APCC. The major drawback of this paper is the least preference to security element in their protocol has been assumed. Dai Yuefa et al.[13] analyzed the basic problem of Cloud Computing that is data security. The author has got data security requirement of Cloud Computing and has given a mathematical model on the basis of these requirements. The data security model proposed, is a worth addition in world of Cloud Computing security. However, writers are not able to give a comprehensive solution for security of Cloud Computing. Qiu-Xin.F et al.[14] proposed a multi-layered and multi-level secured architecture for Cloud Computing according to the characteristics of mobile user. The author has proposed the idea of SaaS (Security as a Service). The advantage of this proposal is that when implemented, is flexible to different scaling system to different requirements and can integrate different operating system and heterogeneous network. While researchers neither discussed the component of secured architecture in detail nor gave the prototype system for verification of theory. LatanyaSweeney [15] proposed a prescribed safety model termed k-anonymity and a set of rules for distribution. One of the versions shows that k-anonymity for each person cannot be eminent from at least k-1 entities whose information also appears in this version. This paper also inspects re-identification occurrences that can be comprehended on releases that follow k- anonymity unless all the other related rules are valued. The k-anonymity safety model is significant because it forms the base on which the real-world systems known as µ-Argus and k-Similar guarantees the safety of the private data. AshwinMachanavajjhala [16] proposed an innovative privacy classification known as ℓ-diversity. It is shown in an experimental evaluation that ℓ-diversity is hands-on technique and can be executed proficiently.ℓ-Diversity requires that there are ℓ-different delicate standards for every group of quasi-identifiers.

IV. OVERVIEW OF DATA AT REST AND IN MOTION

Threats to data security in clouds are at two states. First, data at rest or stored in the cloud and the second, data in motion or moving into or out of the cloud. The security trio confidentiality, integrity, and availability along with hazard easiness drive the nature of data protection tools, measures, and practices [17].

A. Data at Rest

Data at rest denotes to any data in computer storage, including files of an individual, business files on a server, or replicas of these records on off-site tape as a standby. There is no much difference in protecting the data at rest or at transit. Same methodologies apply in safeguarding the data in both the cases [18]. Risks may arise when the data is not physically controlled by the data owning organizations. Effective security is an added advantage with on-premises data than in cloud.

B. Data in Motion

Data in transit denotes data in a moving state. Data can be moved from a stored state to a database or moving from a database to another. If data are being uploaded to a cloud, the interval at which the data is being uploaded is measured to be data in transfer [17]. The username and password that are used to access a web site or verifying to the cloud would be considered delicate sections of data in motion that are not really stored in an unencrypted manner. Data in transit is at transition between outlets, such as in memory or among end points. Safeguarding these data, focuses on avoiding the data from being interfered with, as well as making sure that it remains trustworthy. There is a great risk with a third party perceiving the data while it is in transit. Strange situation may occur when data is transferred between distant end points [19]. Data in bulk may be reserved on intermediate systems, or temporary files may be designed at both end points. There is no healthier protection approach for data in transit than encryption possibly with anonymization. It is reasonable that a cloud user is experiencing security concerns like storing and processing delicate data in a public cloud [20]. When compared with a private cloud, these concerns usually focus on two ranges:

- Less control by the individual organization when data are not managed within its sites.
- A multi-tenancy cloud naturally has risks to sensitive data.

The above mentioned concerns are not new to the field but the risk of data disclosure is actual organizations has huge control over their data that is being stored and processed in their sites and organizations implement strict data policies to achieve full and effective control of their data. Whereas, when a shift is made to a public cloud, data security is not assured. Added risks may arise when data is not kept within the organization. Having achieved security in their premises, organization requires promises of better security from the cloud providers when the organization has shift to cloud [21]. Most organizations are neither capable to be in the information security professional nor are they in that industry. Organizations simply use workstations and webs to get their work done. Even though a Safety computation is a desired quality in all organizations, data security expertise is not a core business nor is it collective in most organizations. So, moving data off sites not necessarily create new risks, but it may in fact increase the security [22].Entrusting the data to an outside up holder may end in enhanced security and may well be more cost operative. Some data are really more sensitive and when it is moved to the public cloud, the data exposure is very large.

V. METHODOLOGY

A methodology has been proposed in which an efficient security framework that incorporates the various securities in different layers is introduced. It is also proposed in the methodologies a framework for privacy preserving and auditing when the data is in transit. Security in cloud computing raises the anxiety over the confidentialitylinkedconcerns in data transit such that no critical information can be interrupted [23]. A suggestion is made to encrypt data by using anonymization techniques before transfer it to the cloud. Before downloading the data it must be decrypted. Until now it seems to be difficult to encrypt data and to trust a third party to keep safe and able to accomplish distant calculations [24].A new methodology has been proposed to allow the cloud provider to perform the operations on data security which are carried out with encryption by anonymization.

A. A. Proposed Data Security Model in Cloud Computing

The layered approach has been given in figure 1 where the first layer is responsible for user authentication. The second layer is responsible for user's data anonymization and protects the privacy of users through a certain way by using anonymization techniques and the third layer is responsible for the speedy recovery of data using decryption [25].

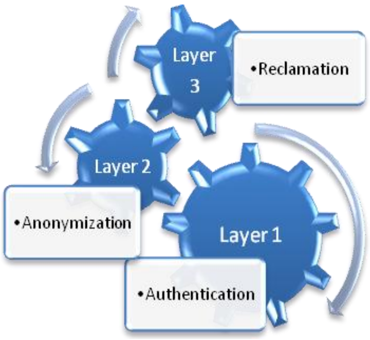

Fig 1: Architecture Block Diagram

In three layered security scheme, when client uploads the data over the cloud classifies it on the basis of its security level. The data is classified based on three parameters Confidentiality, Integrity & Availability (CIA) [26]. Confidentiality means up to what level data should be kept secured on the cloud. Integrity provides assurance that data is not altered but is accurate. Reliability is providing correct data as output to valid users. These parameters decide the level of security provided on a particular data.

B. B. One Time Password Authentication

A one-time password *(OTP)* is a password that is only valid for a single login session or transaction. Using a multi-factor authentication with *OTP's* reduces the risks associated with logging into your system from insecure workstations [27]. It is as a validation scheme which provides an additional layer of security for delicate data and information by demanding a password that is only valid for one login. This password, created nearly every period of seconds, is provided to the customer by a hardware authenticator method and is obligatory with a *user name* and *PIN*. One Time Passwords are not susceptible to malicious users finding the user name and password to access the resource. Nothing can be done to get in to the cloud without the proper combination of *user name*, *PIN*, and *one-time password*. In order to secure the system in a more effective manner, the generated *OTP* must be difficult to estimate, regain, or trace by hackers. Therefore, it is very significant to progress a secure *OTP* producing procedure [9]. International Mobile Equipment Identity *(IMEI)* number, International Mobile Subscriber Identity *(IMSI)*, user name, pin, hour minute, day, year/month/date are the various parameters used in *OTP* generation.

C. Selection of Anonymization Technique

In cloud computing where resources are shared and provided to the users, security plays an important role in cloud paradigm. In case of IT infrastructure, public cloud leads to the sharing of computing resources with other companies as well. Here is the risk of data or any other important asset, the risk of seizure. Cloud computing makes use of virtualization where data and resources are stored in a virtual environment [15]. Users will not know exact location of data or other source of data. To ensure data storage a safe confidentiality, integrity and availability should be provided. To extend further safeguards of data it is important to encrypt it by using anonymization schema along with backup and auditing [28]. Anonymization is a technique that enterprises can use to increase the security of data in Public cloud while still allowing the data to be analyzed and used. Data anonymization is the process of changing data that will be used or published in a way that prevents the identification of key information. Using data anonymization, key pieces of confidential data are obscured in a way that maintains data privacy. The data can still be processed to gain useful information [16]. Anonymized data can be transmitted in a cloud and processed without the concern of the data owner. Later, the results can be collected and mapped to the original data in a secure area.

VI. PROPOSED SYSTEM ARCHITECTURE

Based on the methodology mentioned in section V,a related architecture has been developed and the proposed system architecture is given in figure 2. In this system three layered architecture is provide in which one time password authentication is made for the access into the cloud. Basic aim is to create a private cloud in which an internal cloud is to be created for the authentication purpose. Multiple users can connect to cloud using different web services. Once the access is allowed the overall operation is provided by the web server in the cloud. Separate servers are used for specified purposes such as Audit server for auditing in the cloud and data anonymization algorithm execution [9]. For ultra secure download, two factor authentications is used in which *OTP'*s are generated and send to client's E-Mail or Mobile and then client is authorized to download the data.

VII. PERFORMANCE EVALUATION

In this section, performance is scaled only for the anonymization in which the well-known problem of using generalization to publish a single view *(R*)* of a single base relation *(R)*, while limiting the risk of a linking attack is considered. It is assumed that, as in the majority of previous works, each attribute in *R* can be uniquely characterized by at most one of the following types based on knowledge of the application domain:

- ***Identifier:*** Unique identifiers are removed entirely from the published data.
- ***Quasi-Identifier:*** The quasi-identifier is a set of attributes $Q_1, ..., Q_d$ that can potentially be used to re-identify individuals when combined with other public data.
- ***Sensitive Attribute:*** An attribute which are highly confidential is considered sensitive, if an adversary should not be permitted to uniquely associate its value with a unique identifier.

K-anonymity provides a simple and intuitive means for protecting individual identity with respect to linking attacks [15], [16]. It stipulates that, no individual record should be uniquely identifiable from a group of less than *k* on the basis of its quasi-identifier values. The values will be referred to each group of tuples in R* with identical quasi-identifier values as an equivalence class. K-anonymityR* is k-anonymous with respect to quasi-identifier attributes $Q_1, ..., Q_d$,if every unique tuple *(q1,….,qd)* in the projection of R* on $Q_1, ..., Q_d$occurs at least *k* times. It is often natural to extend the *k*-anonymity model to protect a known sensitive attribute *S*[16]. One of the important components in the design of the proposed algorithm is choosing an appropriate means of checking each anonymity requirement using a sample. It is important to have a reasonable procedure in order to avoid excessive pruning. For each experiment, we used an input of 100,000 tuples, and varied the sample size. The results are given in Table 1, in which each entry indicates the total number of nodes that were pruned during the algorithm's entire execution. The numbers in parentheses indicate the number of nodes that are pruned. An "*x*" indicates that the resulting partitioning is (potentially) non-minimal. There are two important things to note from these results. First and foremost, the estimates are reasonably well-behaved, and do not lead to an excessive amount of pruning. Secondly, they provide for much cleaner execution. As expected, the incidence of both non-minimality and pruning decreases with increased sample size.

TABLE 1: Pruning & Non –Minimality In K-Anonymity

k

n	10	100	1000	10000
100	68 (1384)	x(x)	x(x)	x(x)
250	30 (1110)	7(97)	x(x)	x(x)
500	11 (419)	12(55)	x(x)	x(x)
1000	0 (0)	5 (6)	x(x)	x(x)
2500	0 (0)	2 (1)	1 (3)	x(x)
5000	0 (0)	0 (0)	1 (4)	x(x)
10000	0 (0)	0 (0)	0 (0)	x(x)
25000	0 (0)	0 (0)	0 (0)	0 (0)

VIII. CONCLUSION

This paper considered scaling of anonymization for pruning of data sets that are much larger. The output of this technique is guaranteed to satisfy all given anonymity requirements and the minimal partitioning. End users are secured by the way of accessing the network and benefit out of it. The proposed architecture will give way for secured access and identity using anonymization. This method ensures an isolated and secure execution environment at the cloud by providing a set of security protocols. The proposed scheme empowers the data owner to outsource the security enforcement process on the outsourced data files without losing control over the process. Future extensions will include enhancement in design decisions like inclusion of a trusted third party auditor which will have capabilities of assessing and exposing cloud service risks, key management and distribution scenarios, and formal security proofs of proposed security protocols.

REFERENCES

[1] Jansen, Wayne and GranceTimothy, “Guidelines on Security and Privacy in Public Cloud Computing”, National Institute of Standards and Technology, Vol. 12, January 2011.

[2] J. Brodkin. “Gartner: Seven cloud-computing security risks” InfoWorld, Available: http://www.infoworld.com/d/security-central/ gartner -seven-loudcomputingsecurity-risks-853, March 13, 2010.

[3] Weichao Wang, Zhiwei Li, Rodney Owens, Bharat Bhargava. “Secure and Efficient Access to Outsourced Data”, ACM workshop on Cloud computing security, November 2009, pp. 55-65.

[4] R. K. Balachandra, P. V. Ramakrishna and A. Rakshit, “Cloud Security Issues”,IEEE International Conference on Services Computing,2009, pp. 517-520,

[5] P. Kresimir and H. Zeljko. "Cloud computing security issues and challenges”, Third International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services,2010, pp. 344-349.

[6] S. Ramgovind, M. M. Eloff, E. Smith,” The Management of Security in Cloud Computing”, IEEE International Conference on Cloud Computing, 2010.

[7] Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing”, National Institute of Standards and Technology (NIST), Information Technology Laboratory, Version 15, October 2009.

[8] M. A. Morsy, J. Grundy and Müller I. “An Analysis of the Cloud Computing Security Problem”, APSEC 2010 Cloud Workshop, 2010.

[9] Balakarishnan.S, Saranya.G, Shobana.S, Karthikeyan.S,“Introducing Effective Third Party Auditing (TPA) for Data Storage Security in Cloud”, IJCST Vol. 2, Issue 2, June 2011.

[10] Joshi Ashay.M et al., “Enhancing Security in Cloud Computing”, ISSN 2224-5758 (Paper) ISSN 2224-896X (Online) Vol. 1, 2011.

[11] Richard Chow, Markus Jakobsson, RyusukeMasuoka, Jesus Molina, Yuan Niu, Elaine Shi, Zhexuan Song,“Authentication in the Clouds: A Framework and its Application to Mobile Users”,IACR, October 2010.

[12] Hongwei Li, Yuanshun Dai, Bo Yang, “Identity-Based Cryptography for Cloud Security”, IACR, 2011, pp.169.

[13] Dai Yuefa, Wu Bo, GuYaqiang, Zhang Quan, Tang Chaojing, “Data Security Model for Cloud Computing”, ISBN 978-952-5726-06-0, Qingdao, China, November 2009.

[14] QiuXiu-feng, Liu Jian-Wei, Zhao Peng-Chuan, “Secure Cloud Computing Architecture on Mobile Internet”, Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), IEEE 2nd International Conference, August 2011, ISBN: 978-1-4577-0535-9.

[15] L. Sweeney. K-anonymity. “A model for protecting privacy”, International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems, Volume II, May 2011, ISSN: 557-570

[16] AshwinMachanavajjhala, Johannes Gehrke, D.Kifer, and M. Venkita Subramanian, “ℓ-diversity: Privacy beyond k-anonymity”, http://www.cs.cornell.edu/_mvnak, 2009.

[17] N. Leavitt. “Is Cloud Computing Really Ready for Prime Time?”,Vol. 42, May 2009, pp. 15-20.

[18] Wang, Lizhe von Laszewski. “Cloud computing: A Perspective study”, Grid Computing Environments workshop, November 16, 2008.

[19] M. Jensen, J. Schwenk, N. Gruschka and L. L. Iacono, "On Technical Security Issues in Cloud Computing", IEEE ICCC, 2009, pp. 109-116.

[20] S. Arnold. “Cloud computing and the issue of privacy”, KM World, www.kmworld.com,August 19, 2009, pp. 14-22.

[21] M. Klems, A. Lenk, J. Nimis, T. Sandholm and S. Tai. “What’s Inside the Cloud? An Architectural Map of the Cloud Landscape”, IEEEXplore, June 2009, pp. 23-31.

[22] N. Gruschka, L. L. Iancono, M. Jensen and J. Schwenk, “On Technical Security Issues in Cloud Computing”, IEEE International Conference on Cloud Computing, July 2009, pp. 110-112.

[23] Tim Mather, SubraKumaraswamy, ShahedLatif, “Cloud Security and Privacy: An Enterprise perspective of Risks and Compliance”, O'Reilly Media, Inc., February 2009.

[24] Jinpeng Wei, Glenn Ammons, VasanthBala, PengNing. “Managing security of virtual machine images in a cloud environment”, CCSW ACM workshop on Cloud computing security, November 2009, pp. 91-96.

[25] Flavio Lombardi, Roberto Di Pietro, “Transparent Security for Cloud”, ACM Symposium on Applied Computing, March 2010, pp. 414-415.

[26] Richard Chow, Philippe Golle, Markus Jakobsson, Elaine Shi. “Controlling Data in the Cloud Outsourcing Computation without Outsourcing Control”, ACM workshop on Cloud computing security, November 2009, pp. 85-90.

[27] Xinwen Zhang, Joshua Schiffman, Simon Gibbs. “Securing Elastic Applications on Mobile Devices for Cloud Computing”, ACM workshop on Cloud computing security, November 2009, pp. 127-134.

[28] R. J. Bayardo and R. Agrawal, “Data privacy through optimal k-anonymization”, International Conference on Data Engineering”, Volume I, January 2005.

APPENDIX

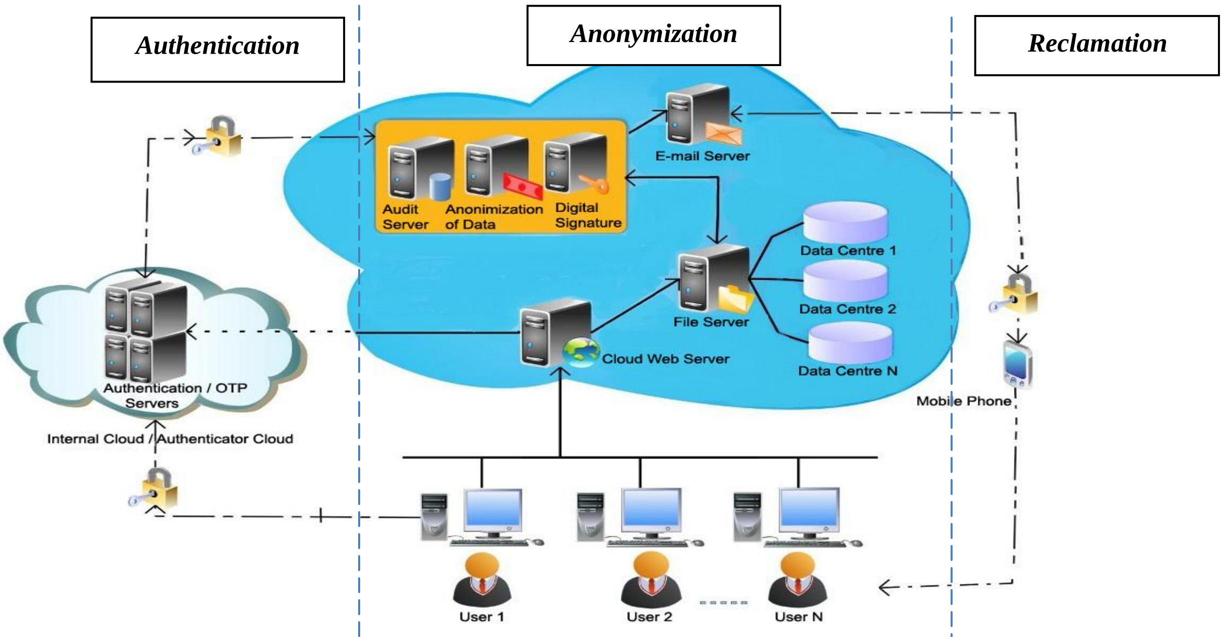

Fig 2: Architecture to Enhance Security of Data in Transit