Secure Data Mobility in Cloud Computing for e-Governance Application

Ramesh Paudyal[1*], **Subarna Shakya**[2]
[1,2]Department of Electronics and Computer Engineering, Institute of Engineering, Tribhuvan University,
Kathmandu Nepal
[1]er.rameshpaudyal@gmail.com, [2]drss@ioe.edu.np,
*corresponding author: er.rameshpaudyal@gmail.com

Abstract

Due to the rapid technological advancement, traditional e-government systems are getting obsolete because of their inherent limitation of interoperability and accessibility to the highly secured and flexible e-governance services. Migration of such systems into highly secured cloud governance architecture will be a long-term viable solution. However, the adoption of distributed cloud computing has created operational and security challenges. This research work aims to bridge the gap between traditional and cloud-based e-Government systems in terms of data security based on confidentiality, interoperability, and mobility of data among distributed databases of cloud computing environments. In this work, we have created two organization databases by the use of AWS EC2 instances and classified the data based on the Risk Impact Level (RIL) of data by the use of the Metadata Attribute Value (MAV) function. To enhance further security on classified data, we take appropriate security action based on the sensitivity of the data. For the analysis purpose, we implemented different security algorithms, i.e. AES, DES, and RSA in the mobility of data between two distributed cloud databases. We measured the encryption and decryption time along with the file size of data before and after classification. AES performed better while considering the encryption time and file size, but the overall performance of RSA was better for smaller file sizes. Finally, the performance of the data mobility between two distributed clouds databases was analyzed while maintaining the sensitivity level of the data.

Keywords: *Cloud Computing, EC2, e-Governance, MAV, Mobility, Security Algorithms*

Introduction

Every Government seeks to improve its work efficiency by using information and communication technologies (ICT) in its governance system. In this regard, e-Governance came into existence to provide electronic services via internet platforms. E-governance services with the perspective of cloud computing were proposed to restructure the existing e-governance system. On-demand scalability, while considering the resources and universal accessibility via the internet has impacted the reconsideration of e-governance services. This transformation of e-governance services is termed as Government Cloud (G-Cloud). The inherent benefits of G-Cloud like centralization, stability, resource pooling, and so on based on cloud computing along with the enhanced security benefits bolstered the implementation of G-Cloud. As cloud computing is a part of G-Cloud, other significant parameters like convenient infrastructure and service supply are deemed more important while providing significantly lower priority to its security issues. However, G-cloud was developed to enhance e-governance application, thus security is the key issue that needs to be given higher priority (Hada, Singh, & Goyal, 2012).

This work illustrates an approach that supports secured data exchange between e-governance service operations based on a reliable distributed cloud platform. It will help to choose the most appropriate secure data exchange model in G-cloud, based on the data security needs of the e-governance system. Various appropriate data security measures are applied in this approach (Pokharel & Park, 2009).

When we adopt cloud services, security issues are the most common and convoluted challenges which affect the effectiveness of governance in various phases, i.e. application phase, network phase, authentication phase, information phase, and virtualization phase. These security threats are ingrained when we consider cloud computing and significantly impact the successful implementation path of G-

Cloud computing because governmental organizations and their consumer's information are stored at cloud data centers. This critical information of citizen safety is the key to e-governance systems while overall security of the entire datacenter information during the exchange is another aspect of data security that needs to be addressed with ultimate priority for secure G-cloud platform. Thus Confidentiality (C), Integrity (I), and Availability (A) of citizen information need to be secured while mitigating the unauthenticated user's access to the government cloud datacenter. Cryptographic algorithms are required to address the security threat in cloud data center. Also, a single cryptographic algorithm can never be sufficient to ensure the data exchange between multiple datacenters based on the confidentiality of data as the user's access may be compromised. Thus managing access control methods in the cloud environment is the key step in developing a secured G-cloud platform (Sharma, 2017).

This paper focuses on complete infrastructure of the data mobility model to ensure the secure data exchange in G-cloud based on security requirements of data. The data are stored on the distributed G-cloud MAV and cryptographic algorithms to ensure the confidentiality of data. There will be a need to fulfill the security requirements of data based on the security level of data before exchange.

Objectives

The overall objective of this paper is to ensure the secured data mobility between distributed cloud computing environments based on the sensitivity level of the data. The specific objectives are:

- To classify the data based on the Risk Impact Level (RIL) before executing the mobility of data in the cloud
- To implement the different security algorithms based on the sensitivity level of data in a distributed cloud environment

Literature review and related work

Every government is getting involved with the IT industry to improve efficiency and reduce implementation complexity which gave rise to e-Governance services like e-banking, online fund transfer, online registration, and so on via government online portal (Hada, Singh, & Goyal, 2012). Hada, Singh and Goyal (2012), in a paper about security engineering in G-cloud and secure e-governance, analyzed the risks associated with the implementation of G-cloud along with the discussion of security benefits while rebuilding the traditional e-Governance platform to G-cloud. Klymash, Demydov and Baydoun (2019) also illustrated an approach to the creation of a reliable cloud platform, intended to support the secure operation of government electronic services. They have focused substantially on the optimal topological solutions and the appropriate topological design realization of a distributed cloud environments for IaaS and PaaS planes by determining the parameter at the initial phase of the topological network. These approaches bolster the organization's effectiveness to manage multiple distributed data centers in either IaaS or SaaS or PaaS planes. Kaur and Zandu (2016) elaborated a secure data classification model in cloud computing by using a machine learning approach. In the paper, at first security issues were identified and then the development of a framework, which reduces security issues during authentication, was proposed. Data were classified on the basis of security requirements into 'sensitive' and 'non-sensitive' based on confidentiality to address security factor.

Wheeler (2011) also highlighted about maintaining the security and data mobility in the cloud-based e-governance services by maintaining the sensitivity level of data. The sensitivity level of data means risk impact level of data that depends on the organizational requirements. The risk impact level (RIL) of the data or information decides the level of sensitivity based on ISO27005: 2011 (Firoiu, 2015). Metadata extended attribute, MAV was computed and the class of data which need more security were evaluated on the basis of organizational information risk calculation metric. Then, a particular organization can decide the appropriate infrastructure, tool and technology, and platform for securing data mobility on cloud-based e-governance services.

In a study by Hababeh et.al (2018), cloud computing and data mobility faced a malicious threat in personal and confidential data. Traditional security mechanisms were not sufficient to handle the cloud data mobility to maintain the sensitivity level of data. This paper implemented the data classification and security model to reduce the potential threat observed in the data. This model solved the data classification and security issues based on the risk impact level of data. It deals with big data classification and security issues in cloud environments. Hadoop Map Reduce framework was integrated in the cloud environment to solve the security issues aroused during the data mobility in distributed environment. This method can significantly improve data security based on the availability of data. Zardari, Jung and Zakaria (2014) also stressed that to achieve the secured information exchange mechanism in cloud environments, AES techniques are implemented which guarantee the user's information security located on cloud servers. A data classification cloud model was proposed to attenuate the shortcoming of data confidentiality issues based on the distributed cloud environment. To measure classification of data as two mainstreams i.e. Sensitive and Non-sensitive, data was classified using k-Nearest Neighbor (KNN) classifier and, Rivets, Shamir and Adelman (RSA) algorithm were used to successfully encrypt data labeled as sensitive data.

In their research work, Larose and Larose (2014) and Awotunde et.al. (2016) clearly explained the need of security management during the data storage, mobility and exchange for to safeguard the integrity, confidentiality and availability of data as they are located in a wide range of hardware specifications. Security algorithms which employ cryptographic techniques were used to encrypt and decrypt the data. Security algorithms were implemented on the basis of their sensitivity level during storage and transmission. On the one hand, the implementation of cryptographic techniques boosts data security and improves computer security. While, on the other hand, these implementations of security algorithms come at the cost of high resource utilization in terms of time, memory, and CPU usability. Three encryption techniques, viz. AES, DES and RSA were implemented with a comparison based on time of encryption and decryption, thus providing the effectiveness of these algorithms.

From the review of relevant literature as discussed above, we can conclude that data security in the distributed environment is the major issue. A huge amount of data is generated because of the digital and paperless society. These data convey the organization information which plays a key role in the decision making that directly impacts the organizational goal. Rights to information add on some security threats during the mobility of data. E-governance should ensure to protect organization data against potential threats and monitor the data by employing the risk-based security activities throughout the life cycle of data. The effectiveness of e-governance is the foremost problem to enable automatic data-centric security solutions due to the rapid technological advancement. Traditional security and solution were not sufficient to prevent the data threat during the data mobility in the distributed cloud environment (Zardari, Jung & Zakaria, 2014; Kaur & Zandu, 2016; Sharma, 2017). E-governance applications seem to implement their own data governance policy for ensuring the data-centric security solution before executing the data agility. Automatic discovery, detection, and classification of the data with proper security actions are the major problems for any organization in the era of big data. Building a highly confidential data classification and mobility model for data governance will be a viable solution for the given problem.

In the above literature, scholars have used machine learning and multi-criteria decision-making methods. In a study by Hababeh et.al.(2018), the Metadata Attributes Values (MAV) is used to classify data based on the Risk Impact Level (RIL) during data mobility in the cloud computing environment. In this paper, we propose cloud data storage and mobility methods by employing the appropriate security mechanism. This model classifies the data based on the security requirements of the data owner. The appropriate security requirement is calculated by using the MAV function (Hababeh, Gharaibeh, Nofal, & Khalil, 2018) on the basis of the risk impact level of organizational data. To solve the problem we need to design and develop an integrated methodology for sensitive data classification with appropriate security solutions in the cloud computing environment. This methodology aims to fulfill secure cloud data mobility for the e-governance application. This model has been quantitatively and technically analyzed by calculating the data transmission throughput during the mobility of data. This will perform the secure data mobility in the cloud environment by integrating the cloud computing and organizational policy.

Proposed Method

In the proposed method, the cloud based e-Government infrastructure are selected considering the security metrics, classification technique and security management with data mobility. The MAV classifiers are used to classify the data based on the sensitivity level of the data. This classified data is stored in a distributed cloud data center. The next milestone is to automatically define a different security algorithm based on the sensitivity level of data ensuring secure data mobility across cloud databases. Each database has its own security policy as well as requirements that are verified before the mobility of data. Given security algorithms, i.e. RSA, DES and AES are used to encrypt the data that required some form of security verification before the exchange with another database. Figure 1 represents the overall methodology of the proposed research work as illustrated below.

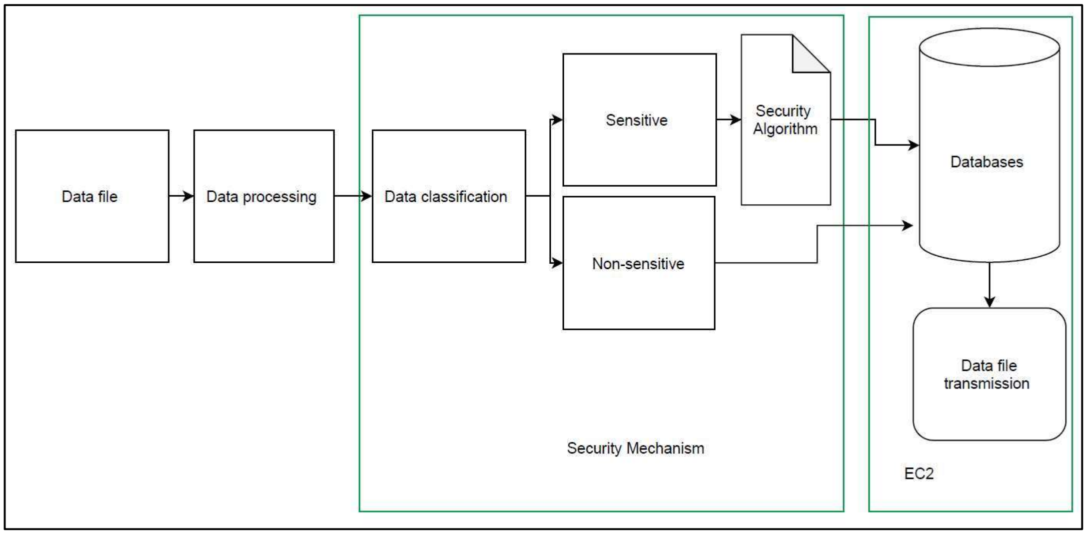

Figure 1: Study Methodology (Source: authors)

Data Collection

Required data sets were locally collected that describes the following identification of the citizen.

The attributes of the sample datasets are described as: Id_number, name, district, Gp_np, ward number, house owner name, gender, age, ctz_number, nationality, phone number, latitude and longitude.

Total Number of attributes = 14

Total number of data = 4661

Initially, we implemented the small dataset; the file size was 732 KB and then the dataset size increased up to 3 GB for the purpose of analyzing the performance of different security algorithms and mobility of data among the distributed cloud databases.

Data Processing

Once the dataset has been imported, our next step was to preprocess the data based on the security requirements i.e. sensitive and non-sensitive. Depending upon the sensitive level of data, different security levels are implemented for integrity protection. The MAV function is used to classify the data based on the risk level of the data. MAV function is a standard function that decides the level of security based on the risk impact level of the data which depends upon the organization structure and function (Wheeler, 2011). The data is classified into sensitive and non-sensitive data, i.e. binary class classification. We have to determine the class of input value based on the risk impact level of the attribute.

MAV Classifiers

The classification process identifies the security needs based on the strategic importance of the data. This importance can be measured using a sensitivity or risk impact of datasets. MAV function is used to identify the security needs of the data as described in equation 3. This classification process maintains the confidentiality level of the data i.e. one of the key security parameters of cloud computing. (Hababeh, Gharaibeh, Nofal, & Khalil, 2018) Classification technique is one approach among many, this function can identify the needs of security based on the Risk Impact Level of data (RIL). In this work, our data set is classified into two classes i.e. sensitive and non-sensitive.

For the risk analysis purpose of data, we consider the given potential risk of cloud computing sabotage, data leakage, data loss, and hacking from an attacker or hacker. These security risk metrics can be defined by using ISO27005:2011 (Firoiu, 2015) standard for minimizing the security threat of data i.e. Assets, vulnerability exposure, threat level and the likelihood of threat. Risk metric value is defined from 0-5 range, i.e. Negligible (0-1), Low (1-2), medium (2-3), high (3-4) and very high (4-5) based on the organization structure and services. Risk impact level (RIL) decides the security control level of data for the protection of data security (Azim, 2017). The security needs of the data can be calculated by using the following steps:

Threat Vulnerability (THRV) = THREAT *VULNERABILITY (1)

RIL = ASSETS*THRV*LTHR (2)

RIL can be measured as, highest risk level (4-5) and lowest risk level (0-1) so, data ware classified on the basis of given function in different classes called sensitive and non-sensitive.

$$MAV = \begin{cases} 0, & 0 \leq RIL \leq 1 \\ 1, & RIL > 1 \end{cases} \qquad (3)$$

This approach uses a metadata of the dataset i.e. provided by the particular organization. Any particular organization has to manage the metadata of their data based on the risk impact level of the attributes of the dataset. These statically managed metadata carry the risk level. In this classification approach predefined metadata values are used for modeling the classification system so, it is able to classify the organization data with high level of accuracy. This makes sure the data loss prevention (DLP).

Security Algorithm

Security algorithms ensure the integrity during the mobility of data in cloud databases. We choose the different asymmetric and symmetric algorithms for securing the data based on the sensitivity level. RSA is considered because of its asymmetric nature while AES and DES are selected based on the symmetric nature of these algorithms.

Rivest–Shamir–Adleman (RSA)

The RSA algorithm is a form of public-key cryptography that uses both public and private keys to protect data in the cloud. The development of public-key cryptography has been the most rapid, and it may represent a radical departure. Because of the use of two different hidden keys along with a secret key, it is also known as the Asymmetric algorithm. The plain text and cipher text in this scheme are integers between 0 and n-1 for some n. The most common size for n is 1024 bits. User data is encrypted before being uploaded to the cloud. If the user requests data from the Cloud provider, the Cloud provider verifies the user's identity and delivers the data. RSA is a block cipher in which each message is mapped to a unique number. RSA is made up of two parts: public and private keys. Public-Key is known by everyone in our Cloud setting, while Private-Key is known only by the user who originally owns the data. As a result, the Cloud service provider does the encryption and the Cloud owner or client does the decryption. The data can only be decrypted with the corresponding Private-Key after it has been encrypted with the Public-Key (Awotunde, Ameen, Oladipo, Tomori, & Abdulraheem, 2016).

Data Encryption Standard (DES)

The aim of the DES algorithm is to provide a standard method for protecting sensitive commercial and unclassified data. The same key is used for both encryption and decryption. For the encryption purpose DES takes a 64-bit long plaintext and a 56-bitkey (8 bits of parity) as input and outputs a 64-bit block (Mahajan & Sachdeva, 2013).

Advanced encryption standard (AES)

The AES algorithm is used for both security and speed. The installation of both hardware and software is much quicker. NIST recommends a new encryption standard to replace DES. It shows how to encrypt 128-bit data blocks in 10, 12, and 14 rounds based on key size. It can be seen on a variety of networks, including mobile devices. It has been extensively tested for a number of security applications. (Sharma, 2017).

Infrastructure and environment setup

This method is implemented in the real cloud environment which is managed by Amazon Web service (AWS) Cloud Service provider (CCP). The configuration of the testing environment is mentioned below.

AWS EC2

It is a web service used for providing a secure and resizable computing capacity in the cloud computing environment. It allows the user to obtain and configure a complete computing resource for a variety of purposes. EC2 offers a variety of choices for storage, networking, processor and operating with the fastest processor in a cloud environment. It provides the most powerful GPU instances for different applications like Machine learning, graphics workload and windows workload. There are a number of instances with a variety of configurations that fulfill every business's needs globally.

T2 micro instance

It is a general-purpose AWS instance type that is able to achieve a baseline level of CPU performance. The baseline performance is governed by CPU credits. It can use a variety of applications such as low-latency, small and medium database, development etc. It provides a full core performance if required. The features of our instances are high frequency up to (3.3 GHz) Intel Xeon processor, 1 vCPU, 6 CPU Credits/hour, 1GB memory, low to moderate network performance and 30 GB SSD Amazon Elastic Block Storage (EBS).

The performance of the AWS depends on vCPU, size of memory, network performance and stored data size or transferred data. The configuration of the cloud depends on their performance based on the user requirements. This model helps to analyze and select the appropriate secured cloud configuration for the e-governance application. This also helps to identify the scalability requirements of cloud computing considering organizational requirements. All data were managed by using the MySQL community server.

Measurement methodology

In this section, the performance of our proposed system is evaluated on the basis of the following parameters. The major performance indicator of our model is the execution time which is transmission time between sender and receiver. This time is calculated by the use of software execution environment and performance of CPU and used memory.

Classification details: It measures the data classification time generated by purpose classifiers. This time is depends on the file size before the classification.

Encryption and Decryption time: It measures the encryption and decryption time taken by the different security algorithms which is AES, DES and RSA.

Execution time: It can be evaluated by used software in the model (execution environment), CPU performance and memory size used in the platform. Data upload to an AWS server.

Delay time: It depends on the specific parameter of the network platform the physical distance between the end-user and location of the cloud storage (AWS region) as well as internal network capabilities and virtual traffic. These factors cannot be adjusted by the end user but only methods can be chosen for data transfer and invocation.

Throughput is the main cloud performance measurements parameter that can be calculated by measuring the time of given two factors i.e. execution time of all components and delay between them that can be calculated by using the above-mentioned parameter.

Mobility of data

Cloud data mobility causes a malicious threat of data that requires the appropriate data protection strategy. Appropriate security management is essential for the reduction of potential security threats in cloud data mobility. In the previous section, we discussed classification and automatic security management of data based on the sensitivity. This entire approach guarantees the secure availability of data during cloud data mobility.

Data Transmission Response Time (DTRT)

After the user authentication between the services provides and the service consumers, data transmission is allowed in a distributed cloud environment. Secure data mobility was achieved among the distributed cloud database on the basis of the consumer's demand. Connection is initiated between two nodes by a secure socket layer (SSL). Data transmission response time (DTRT) is needed for determining the security of the data during transmitting various sizes of files. The transmitting speed of 64Mb/s was provided by our cloud architecture. Hence, to verify the data mobility process, transmission response time, total delay time and throughput were measured for the various data sizes (Zardari, Jung, & Zakaria, 2014).

Data Transmission Delay Time (DTDT)

Data transmission delay time (DTDT) is the difference between the Data transmission response time and standard baseline transmission response time which is given by:

$$DTDT = DTRT - BDTRT \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (4)$$

Data Transmission Throughput (DTT)

Data transmission throughput (DTT) is the amount of data transmission (ADT) and is represented by the file size in bits from the source cloud database to the destination cloud database. It is measured in Mb/s and calculated by using the given below equation.

SDTT = ADT/DTRT…………………………………………………………… (5)

Result and Discussion

In this section, we evaluate and analyze achieved results on the basis of classification time, encryption and decryption time and transmitting time. Performance evaluation and comparison of these algorithms were done in different file size in the above proposed cloud environment.

Classification details

Table 1, provides the classification details of data based on sensitivity level and its file size. In the first phase, 732 KB data is classified into two class sensitive (255 KB) and non-sensitive (481 KB). Then the data size is increase up to 3 GB which is further classified into 1GB of sensitive data and 2 GB non-sensitive data. The time taken for classification by the MAV classification is 14 seconds and 287 seconds respectively.

Table 1: Data Classification Details (source: authors)

Before classification	After classification		Classification time (Sec)
	Sensitive class	Non-sensitive class	
Total file size	File size	File size	
732 KB	255 KB	481KB	14
3 GB	1 GB	2 GB	287

Security measurements

Case I: We applied a security algorithm as per requirements of data for the purpose of storage and mobility over multiple cloud databases. RSA is implemented as an asymmetric algorithm while AES and DES are implemented as symmetric algorithm. RSA generates 1048 bit public and private keys for encryption and decryption purposes while AES generates 64 bit public and private keys and DES generates 128 bit private and public keys. The comparative analysis of these security algorithms is mentioned in Table 2, below.

Table 2: Performance Evaluation of Security Algorithm (Source: authors)

S.NO	Algorithm	file size (KB)	Encryption time (Sec)	Decryption time (Sec)
1	AES	732	7.654901961	4.784313725
	DES		14.35294118	5.262745098
	RSA		34.9254902	23.44313725
2	AES	481	4.171938776	3.435714286
	DES		4.908163265	3.043061224
	RSA		20.85969388	14.47908163
3	AES	255	1.471153846	1.307692308

S.NO	Algorithm	file size (KB)	Encryption time (Sec)	Decryption time (Sec)
	DES		2.451923077	1.0625
	RSA		6.375	4.168269231

As mentioned in Figure 2, the algorithms were compared based on the encryption and decryption time for different data sizes created during the classification phase. The AES algorithm performed the best while considering the encryption and decryption time while the RSA algorithm took the maximum encryption and decryption time. The RSA algorithm took the most time during encryption and decryption due to its need to generate a 1024 bit unique private and public key while the other two algorithms generated the same public and private key for both encryption and decryption purposes. The DES algorithm could be considered if the tradeoff between encryption time and encrypted file size is taken into consideration. Even though the RSA algorithm has the longest encryption/ decryption time and by far the file size was the largest, the RSA algorithm was implemented further due to its highly secured nature even though the file size after encryption was relatively high among these algorithms.

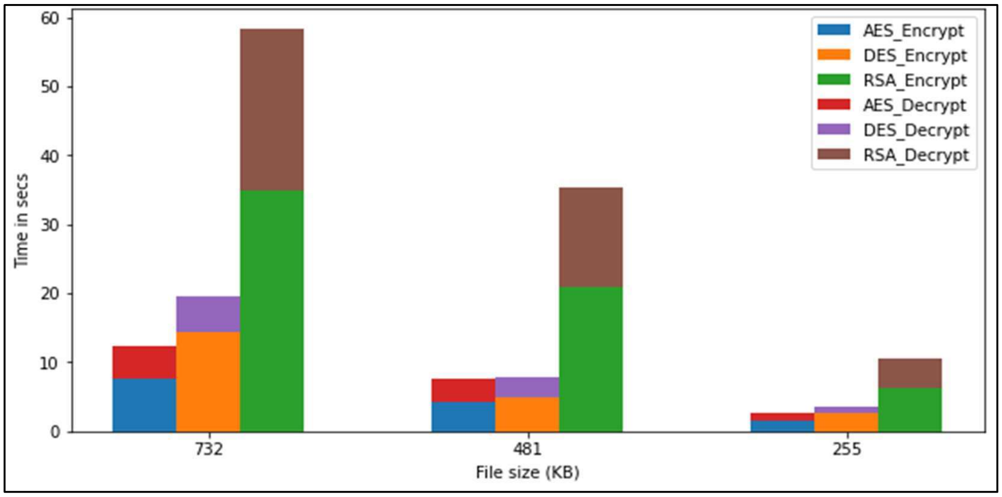

Figure 2: Comparison between AES, DSE and RSA with Encryption and Decryption time (Source: authors)

Case II: RSA algorithm was further used for the encryption of sensitive data. Encryption time and the encrypted file size were compared with original data file size and encryption time. All these tasks were done in the AWS EC2 instance. After encryption, the sensitive file size was 26034 KB from non-encrypted 255 KB, an increase in file size by almost 100 times when compared to the original file size. Even though the file size was humungous, the RSA algorithm was still preferred because of its complex decryption that forbids unauthorized users which significantly increases the data security. The time taken for encryption was measured as mentioned in Table 4, which took 7 seconds for data to be encrypted in the EC2 instance using the RSA algorithm.

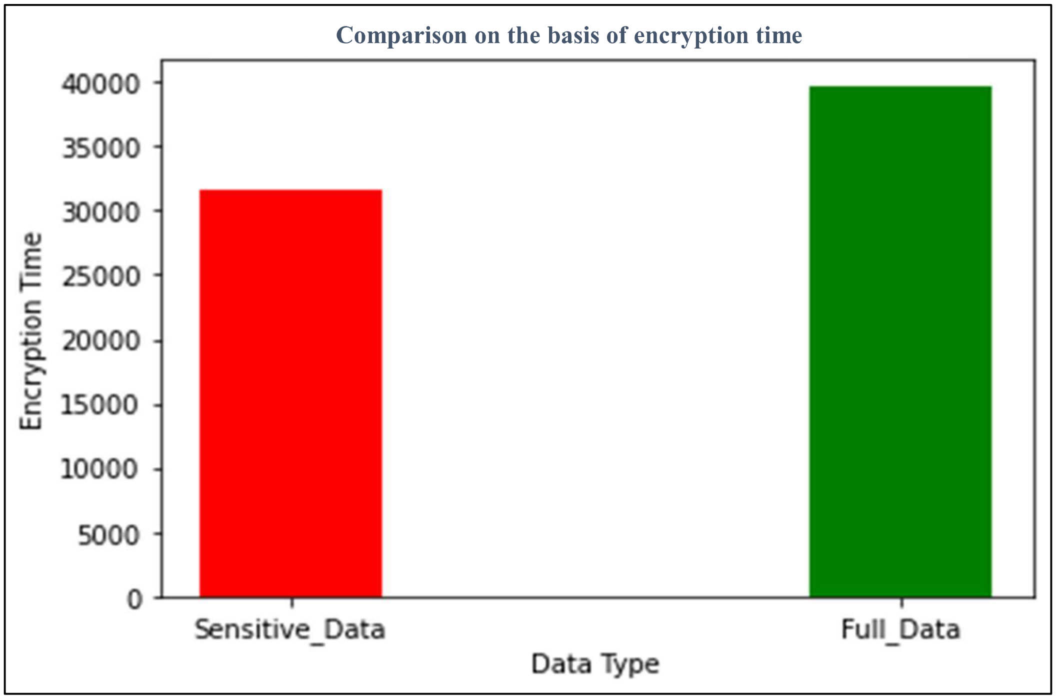

Figure 3: Encryption time sensitive class dataset vs. overall dataset (Source: authors)

Case III: After the encryption of the sensitive data in the cloud database, the comparison of query time was measured. In an EC2 instance, it took 32 seconds to access the sensitive data, 37 seconds to access non-sensitive data and 40 seconds to access the overall data. As in Table 3, even though the file size was increased, the query time for different data was not increased proportionally. With the increase in data size, the access time did not increase because of the advantage of cloud infrastructure in processing huge amounts of data. Thus, a high volume of data won't affect the performance of the cloud-based e-Government services.

Table 3: RSA Calculation file size vs. Execution time (Source: authors)

RSA Calculation		
File size (KB)	After encryption file size (KB)	Database Execution time (sec) (Real time)
255	26034	31.559
481	49107	36.776
732	74733	39.673

Real time is the total time taken to complete the query. From a User's perspective, real time is the total CPU time taken by the user's system for query while from a System's perspective, real time is the total time taken by the System CPU for the execution of a query. Figure 4, clearly shows the comparison of file size before and after implementing the security algorithm that is noticeably analyzed in different cases.

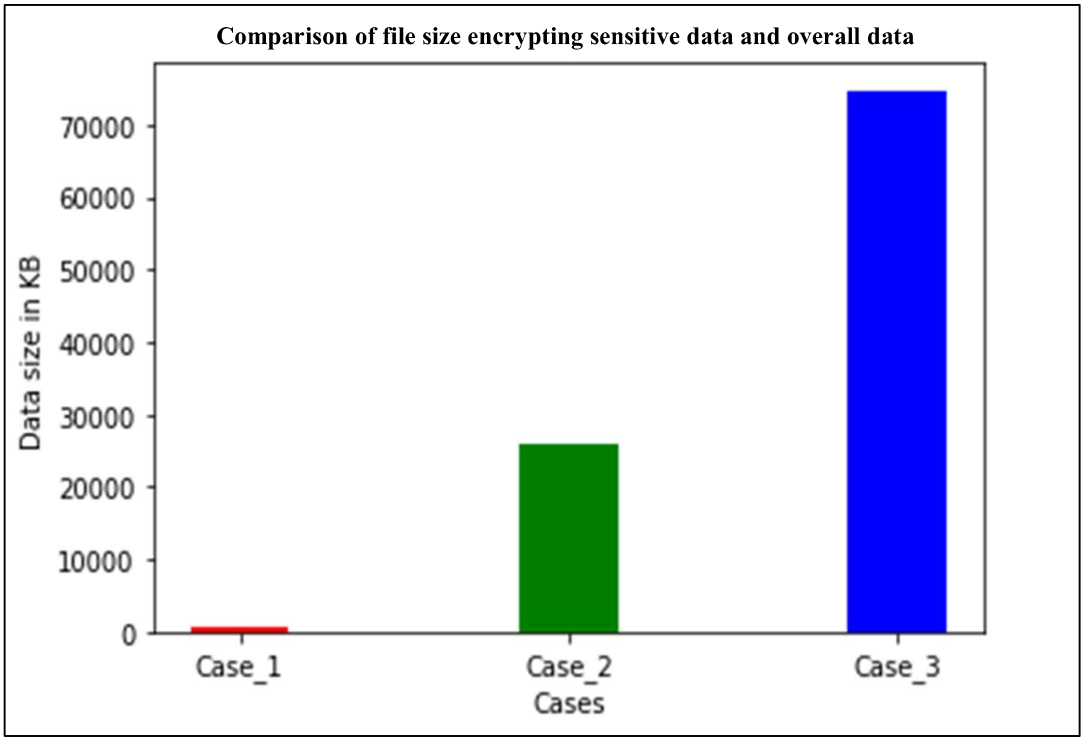

Figure 4: Comparison of File Size with Different Cases (Source: authors)

Case IV: Different security algorithms were implemented in different file sizes for encryption and decryption purpose. Firstly, we implemented a small data size (KB) and evaluated the encryption and decryption time in the cloud environment. After then for the comparisons purpose data size increases up to GB. The performance of the large file size is mentioned in Figure 5.

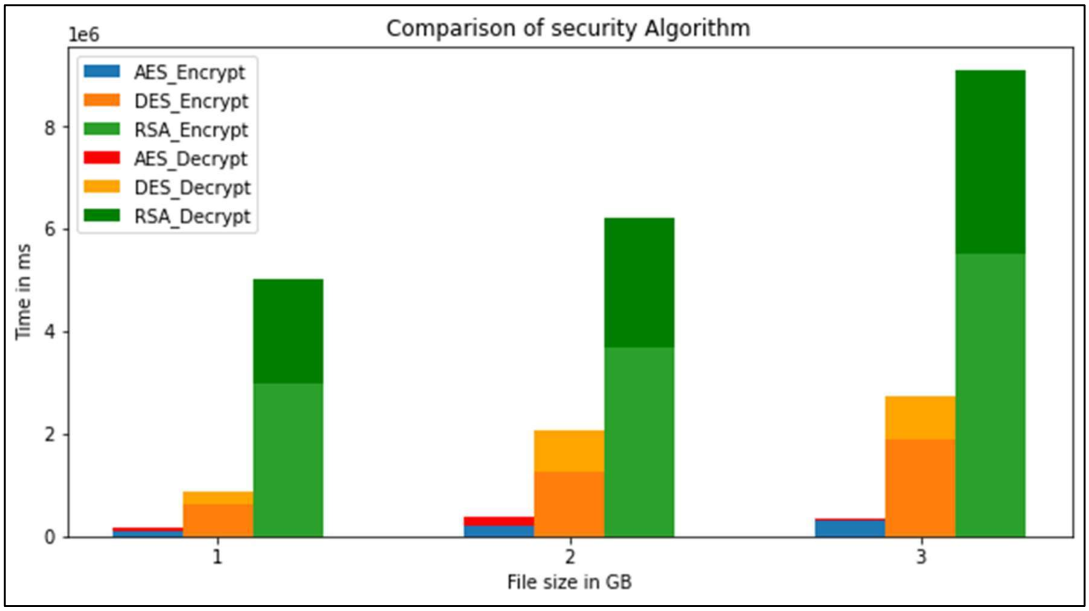

Figure 5: Comparison of Security Algorithm in Large Data Size (Source: authors)

For the encryption and decryption of the larger data, the table below shows the time taken to encrypt and decrypt our data. For the 1 GB sensitive data, the encryption time for AES, DES and RSA algorithms were 100.5 sec, 635 sec and 3000 sec respectively. The decryption time taken by those algorithms were 62.8 sec, 232.8 sec and 2013.6 sec respectively. While for the 255KB sensitive data, the encryption time for AES, DES and RSA algorithm were 4.17 sec, 4.90sec and 20.85 sec respectively. Similarly, their decryption time were 3.43, 3.04 and 14.47 seconds respectively. Here, when comparing the encryption and decryption time for different data sizes, it can be acknowledged that with the increase in the size of the input data, the encryption and decryption time did not increase in the same ratio. The larger the file size, the smaller the total time taken for encryption and decryption. This significant decrease in time was achieved leveraging the performance of the cloud environment with a huge volume of data. The larger data size improved the performance of these algorithms in the cloud environment.

Mobility measurements

Case I: During the exchange phase, Role-based access of data is provided on the user level. The user privilege was defined by the central data center, while access is only granted to specific users based on authentication. Also, user privilege is defined clearly while creating the user ID at the central data center. The user's views were limited on the basis of the privilege level of the user, thus, different users can have different views of the database instance. Securing data mobility between two clouds instances on AWS EC2 was based on user demand. Thus, a flexible system can be created where users can be added by the central data center on the basis of the requirements of the organization. Also, User authentication is validated by service providers as well by the consumers and the use of SSL connection between source and destination cloud guarantees further security during the exchange. In order to successfully exchange data, the service provider must validate the authentication details or user privileges. We measured the transmission response time and total delay time for processing the query in order to measure the system's efficiency.

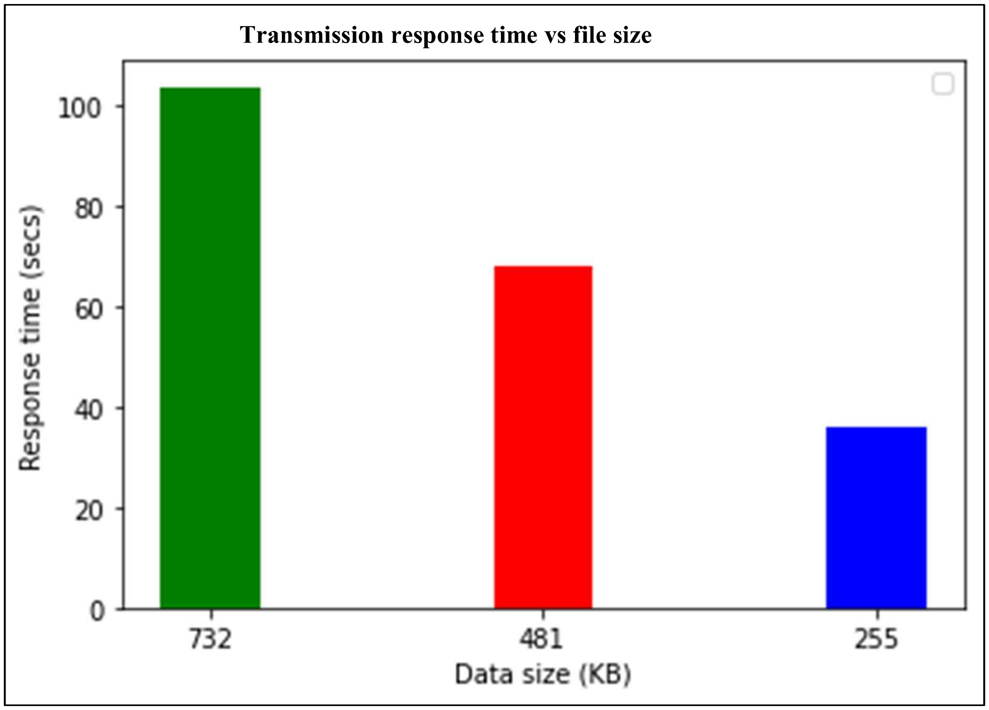

Figure 6: Data Transmission Response vs. File Size (Source: authors)

Case II: As shown in table 4, the transmission response time to send 1 GB data is 141.5 sec and transmission delay time is 16.4 sec while 57.48 MB/sec throughput is achieved. Similarly, for 2 GB data, the transmission response time is 297.8 sec, the transmission delay time is 43.2 sec and a throughput of 55.017 MB/sec is achieved. Finally, for 3GB data, the transmission response time is 627.4, transmission delay time is 78.2 and a throughput of 52.228 MB/sec is achieved. While analyzing the transmission response time, increasing file size did not significantly increase the transmission response time thus, it can be concluded that with an increase in file size improved our transmission response time.

While considering transmission delay time, an increase in file size significantly affected our delay time. With an increase in data size, the transmission delay time was also increased. When we consider the transmission throughput, a larger data size achieved lower throughput for overall file transmission. Thus we can conclude that when we increased the file size from KB to GB the throughput of the system increased significantly up to a certain file size. After that, the increase in file size started to decrease the overall throughput of the system.

Table 4: Data Mobility Details for Larger Size (Source: authors)

Data size	Data Transmission Response time (sec)	Data Transmission Delay time (sec)	Data Transmission Throughput (MB/sec)
1GB	141.5	16.4	57.4848
2GB	297.8	42.2	55.017
3GB	627.4	78.2	52.228

Limitation and future work

In this research, the scope is limited to the implementation of security algorithm on the basis of sensitivity level of data in cloud environment. Further research can be conducted by the implementation of big data classification and security in cloud computing environments for cloud data governance application.

Conclusion

In this research, at first, the data classification was performed using the MAV in cloud computing environments. Then, we studied the comparative analysis of security algorithms based on the encryption and decryption time, file size and total execution time. The RSA algorithm is used for the encryption of highly sensitive data due to its complexity in deciphering by intruders. As security is our primary concern, even before the exchange, the use of RSA algorithm is justifiable even though we have to tradeoff file size and query execution time. Since our system is implemented in a cloud environment there is no limitation of storage and processing capacity. Thus, implementing the RSA algorithm won't have a significant impact on cloud-based e-Government System. The data exchange between two databases is achieved on the basis of user privilege, which further boosts the security management and mobility based on the sensitivity level of data. Hence, to achieve a highly secured cloud government system needs to consider data classification and appropriate security management before storing and mobility of data in distributed environments.

Conflict of Interests: The authors hereby declare that there is no conflict of interest.

References

Awotunde, J. B., Ameen, A. O., Oladipo, I. D., Tomori, A. R., & Abdulraheem, M. (2016). Evaluation of four encryption algorithms for viability, reliability and performance estimation. *Nigerian Journal of Technological Development*, 74-82.

Azim, M. (2017). Security Assessment Model for A Cloud based e-Governance System based on Fuzzy Comprehensive Evaluation Method. *International Journal of Computer Science and Information Security (IJCSIS), 15*(4).

Firoiu, M. (2015). General considerations on risk management and information system security assessment according to ISO/IEC 27005: 2011 and ISO 31000: 2009 standards. *16*(249), 93.

Hababeh, I., Gharaibeh, A., Nofal, S., & Khalil, I. (2018). An Integrated Methodology for Big Data Classification and Security for Improving Cloud Systems Data Mobility. *IEEE Access, 7*, 9153-9163.

Hada, P. S., Singh, R., & Goyal, D. (2012). Security Engineering in G-Cloud: A Trend towards Secure e-Governance. I*nternational Journal of Computer Applications, 46*(13), 33-38.

Kaur, K., & Zandu, V. (2016). A secure data classification model in cloud computing using machine learning approach. *International Journal of Grid and Distributed Computing, 9*(8), 13-22.

Klymash, M., Demydov, I., & Baydoun, N. A. (2019). The "Data Embassies" Concept as a Secure Communication Core for e-Gov Implementing in Emerging States. *In 2019 IEEE 20th International Conference on Computational Problems of Electrical Engineering (CPEE)* (pp. 1-4). IEEE.

Larose, D., & Larose, C. (2014). *Discovering Knowledge in Data: An Introduction to Data mining* (Vol. 4). John Wiley & Sons.

Mahajan, P., & Sachdeva, A. (2013). A study of encryption algorithms AES, DES and RSA for security. *Global Journal of Computer Science and Technology.*

Pokharel, M., & Park, J. S. (2009). Issues of Interoperability in E-Govemance System and its impact in the Developing Countries: A Nepalese Case Study. *In 2009 11th International Conference on Advanced Communication Technology. 3*, pp. 2160-2164. IEEE.

Sharma, S. (2017). Enhance data security in cloud computing using machine learning and hybrid cryptography techniques.*International journal of advanced research in computer science, 8*(9).

Wheeler, E. (2011). *Security risk management: Building an information security risk management program from the Ground Up.* Elsevier.

Zardari, M. A., Jung, L. T., & Zakaria, M. N. (2014). Data Classification Based on Confidentiality in Virtual Cloud Environment. *Research Journal of Applied Sciences, Engineering and Technology, 8*(13), 1498-1509.