International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 740-746 www.ijaem.net ISSN: 2395-5252

An Efficient and Scalable Model for Data Security in a Distributed Computing Environment

[1]Nwakalor, chikoadi Favour, [2]Bennett, E. O.
Department of Computer Science,Rivers State University, Port Harcourt, Nigeria

Submitted: 05-05-2022 Revised: 10-05-2022 Accepted: 13-05-2022

ABSTRACT
Ensuring that data transmission between users in computing systems is kept completely undisclosed to the unauthorized entities is always a challenge. Maintaining the accuracy and consistency of data over its entire life-cycle is another challenge in a distributed computing environment, as this instigates hackers to temper and modify data in an undetected manner without permission of its authentic user thereby leading to data breach. Hence, the need to design a more efficient and scalable model for data security in distributed database system that proffers solution in terms of confidentiality and data integrity in a distributed computing environment. The new model uses Hadoop in connection with database to create large space memory location that can handle all the bulk of unstructured information in the distributed computing environment. The evaluation parameter was taken based on the efficiency of the processing speed, risks and free fraud, confidentiality and data integrity. The efficiency was recorded in percentages (%). The proposed system was built using and Object-Oriented Design Approach (OODA), JavaScript and Hypertext Pre-processor (PHP) programming language was employed as the frontend while MYSQL was used as backend for relational database. it was observed that the proposed model performed well when tested and can be used to achieve security principle in a distributed computing environment. The efficiencyof the proposed Model yielded a 70 %.

Keywords: Distributed systems,Scalable model, Efficiency, encryption, decryption.

I. INTRODUCTION

A secure and trusted data allotment in a distributed environment is a very critical research design approach. At present, there are new threats damaging the information systems and data resources thus, security is a goal in every organization. ˝In today’s computer driven world, every business has some sensitive, crucial data and processes that need to be secured. Combining the various definitions of many researchers, A distributed system can be defined as an application that communicates with multiple dispersed hardware and software in order to coordinate the actions of multiple processes running on different autonomous computers over a communication network, so that all hardware and software components work together to accomplish a set of related tasks [1]. Some of the advantages of applications of distributed systemsincludeScalability and Modular Growth,Fault Tolerance and Redundancy, Low Latency, Cost Effectiveness and Efficiency[2].

Different issues are anticipated to emerge as the spectrum of distributed systems and applications grows. Authentication, authorization, and encryption are the difficulties connected with distributed systems in terms of data security. Some of the solutions to the problems of Distributed System proffered by researchers includeAccess Control Based Security, Cryptography Based Approaches, Policy Based Approaches, Pattern Based Security and Quorum Based Security Systems [3].

Nonetheless, there are still some rising limitations in the earlier solutions in data security, some of which areScalability limit, Design challenge,Concurrency control, Openness and Extensibility and Lack of efficient storage [4]. Cluster computing, grid computing, distributed storage systems and distributed databases are all examples of distributed systems in use today [5].

For certain business, having a database that isn't hacked is a must. The computer database security industry is now beginning to catch up with its requirement for improved protection of hardware, software, data and networks. Keeping data secure in a large database system may be a tough and costly undertaking. Almost all large database computer systems are dispersed.

Large enterprise systems must be able to accommodate multiple users running common applications from numerous locations. Hence, ensuring that system and important data transmission between parties is kept completely undisclosed to the unauthorized entities is always a challenge. For the newer distributed architectures, such as wireless and peer-to-peer, this challenge becomes more interesting due to factors such as lack of structure and lack of trust in the network. This has increased the risk of operational disruption, financial losses, legal issues, compliance penalties and reputational damages.

Also, maintaining the accuracy and consistency of data over its entire life-cycle is very paramount. However, up to date, hackers still temper and modify data in an undetected manner without permission of its legitimate user thereby leading to data breach. Thus, the need for a more scalable model for data security in distributed system. Therefore, in this proposed system, knowledge has been derived from the limitations found in the previous technique in solving problems of data security in a distributed system. The idea in the proposed system is ensure and design a database system that can afford the amount of big data in the distributed system, to design several interfaces for each user with a security platform and encryption. Also, design a system for confidentiality limitation in the existing methods.

II. RELATED LITERATURE

Integration of several distributed components leads to new security concerns. As a result, one of the primary challenges in creating distributed systems is security. “The backbone of distributed systems is made up of the web, clusters, grids, and clouds. Security for distributed systems presents a comprehensive view of current security concerns, processes, and solutions. In the context of today's distributed systems, security dictates future paths [6]

[7] presented an architecture for the intelligent database management software system as a distributed active stored centre developed by stepwise refinement system, in which he discussed how existing protocols are sufficient for use in the architectural design to support both data intake and data fusion and visualization principles. The architecture described for the data management software system is scalable, splits or shares the data management software system through gateway access servers, and includes internally duplicated processing components. Control has been proved to be distinct or different from data streams both logically and topologically.

Some researchers introduce a new protection method for the design for the performing of intelligent system for standard manipulation taking place in database system. Existing database systems uses the intelligent layer designed for use and built in into the database [8].

SQL query is one of the flexible user queries obtainable by the system that can accept and converts them into a standard querying. Expressing the relation and mapping system to the stop words removal and semantic matching to distance measure techniques which fully satisfy by the intelligent system in the production and control of the SQL query. Demonstrating the experimental results is aim at the usefulness and meaning of the presented system which has been use.

Also, some authors present a thorough examination of systems that make considerable use of metadata in order to automate knowledge discovery. In terms of metadata classification, the designs of the systems are thoroughly explained [9].

The Meta usage model (CWM,2003) of the common warehouse system also assists in providing the necessary abstractions to model generic representations and the implementation of data mining processes; nonetheless, the contention of data is sufficient for the entire period included in the current system. It primarily focused on the metadata operation during the data mining stage [10].

Distributed system was discussed in terms of different objectives of database models based on classification, access control, attacks, and system failures. Distributed system is multiple redundant within multiple devices and data transferring between devices with different channels [11].

To overcome the challenge of data security [12] suggested a framework for security in a distributed system that focuses on device level system control. This technique considers public key cryptography, software agents, and XML binding technologies. Various technologies, such as Public Key Infrastructure (PKI) and Role Based Access Control (RBAC), are used in the building of secure distributed systems. [13] offered the RBAC approach for developing authentication based on Public Key Certificates (PKC).

III. ANALYSIS OF PROPOSED SYSTEM

System analysis is the process of gathering and interpreting data, identifying problems and finally breaking down a system into its constituent parts. When it comes to using a scalable mechanism for data security in a distributed database computing environment, system analysis is carried out with the goal of examining unstructured data to discover the objectives.

The phrase analysis refers to the process of breaking down a whole into its most fundamental components in order to better understand the nature of the components, their behavior and functionality, as well as how they interact with one another. Analysis defines what the system is capable of doing or its activities. The proposed system provides security in the distribution of various files or information uploaded by different client into a computing environment. The model contained a single name node which comprises cluster of different size blocks that are stored in data nodes, the data nodes are sorted and shuffled. Each of the cluster nodes are encrypted with a security authentication code.

These security authentications are provided and are available to the client to have access to the distributed computing portal. Individual documents are uploaded to it specific size block on the portal and can only be manipulated and control by the authorized user. The new model uses Hadoop which connect to the database to create large space memory location that can handle all the bulk unstructured information.

IV. SYSTEM DESIGN

System design is the activities that direct the setting, stipulation and system optimization for the configuration for the accomplishment of particular objectives. Therefore, the Object-Oriented Design (OOD) Approach is used in the proposed system.The architecture is based on dividing the tasks of a system into individual re-implementable and self-sufficient objects, in which each contains the data and the behavior relevant to the object.

Description of the System Components

Internet Browser: The internet browser provides access for user to login to the Efficient and Scalable model for data Security in a Distributed Computing Environment.

User Interface: The user interface is where individual users that can access the application will type in their details before registering in the network security portal.

Network Security Interface: The users provide their details to the security interface for registration, the security interface then authenticates and encrypt the details as a key. The user registration page contains user name, e-mail address, and password. This component provides checking of individual authentication code for security reason.

Files or Information: These are individual file or information provided by each user. At this point the user uploads the files or information into his database environment in the heterogeneous communication link as a local server.

Heterogeneous communication link: This component receives the files that each user sends. At this interface there is a communication link between each user. Each user can only access his information, individual information is encrypted. At this interface user cannot download information except receiving an approval from the administrator.

Distributed Database Environment: This is the distributed server database environment, where all individual files are stored. Encryption of transaction is applied to each user file and user can only download and access files when he creates a new security account that can be recognized by the admin. The admin gives an approval to a user that intends to access files for access will be given.

Administrator Approval: De-encryption of Transaction is done in administrator portal. This component issues approval to anyone who wants to access files. At this point checking user security code and verification user code takes place. The user enters his initial code and sends to the administrator for verification, after verification is confirmed successful, the approval will be given to that user for accessing and downloading files.

Algorithm 1: File Upload

```
Step 1: Encrypt_file (F)
Step: 2 Obtain public key of recipient Pu={$n$, $e$} to calculate the cipher: C=$M^e$mod(n), where 0≤M<n.
Step 3: To transform Clair text in file F into Cipher text in file F’. Set $a^{\emptyset(n)}$mod(n) = 1 where gcd (a,n)=1 and calculate n=p.q such that $\emptyset(n)$=(p-1)(q-1) then chose e and d to be inverses mod ø(n).
Step 4: For B←1 to numberOfBlock(F) do
Step 5: Else
Step 6: B’=ENC_AES(B,K)
Step 7: else
Step 8: send_to_cloud(F’)
Step 9:For k←1 to SizeOf(K) do
Step 10: else
Step 11: k’=ENC_RSA(k)
Step 12: decryption of the recipient uses their private key Pr={$n$, $d$} and computes: M=$C^d$mod(n).
Step 13: end for statement
Step 14: Save_in_server(K’)
 Step 15: Exit
```

Algorithm 2: File Download

```
Step 1: Decrypt_file (F’)
Step 2: obtained private key Pr={$n$, $d$} and computes: M=$C^d$mod(n).
Step 3: transform Cipher text in file F’ into Clair text in file F
Step 4: For k’←1 to SizeOf(K’) do
Step 5: Else
Step 6:  k=DEC_RSA(k’)
Step 7: else
Step 8: return(K)
 Step 9: apply Phase 2: Decrypt Cipher text
Step 10: For B’←1 to numberOfBlock(F’) do
Step 11: else
Step 12: B=DEC_AES(B’,K)
Step 13: end of statement
 Step 14:  return(F)
Step 15: Exist
```

V. RESULTS AND DISCUSSION

Hardware Requirements:

The minimum hardware requirements necessary for the deployment of the system have the following configuration:

i. Processor: Pentium processor or higher with minimum of 1GHz speed
ii. Memory: 1GB-RAM and 256MB virtual memory, 120GB HDD
iv. Monitor: 32bit Screen Resolution and above

Apart from this minimum configuration, the operating system used may demand higher capabilities for the execution of applications

Software Requirements:

The software requirements for the implementation of the efficient and scalable mechanism for data security in distributed computing environment are categorized into:

i. **Text Editor:** Sublime text or any other preferred IDE
ii **XAMPP**: version 8.2 or higher
iii. **Database**: SQL Server 2005 or higher
iv. **Web Browser** Software Application
v. **Operating System**: Microsoft windows 7 and above, Linux and Mac OS X are all compatible for the deployment of the system.
vi. **Interpreter**: On the local server, the PHP Interpreter must be at least Version 7.
vii. **Firefox Mozilla**, Internet Explorer, Google Chrome.

Certain tools were used in achieving proposed model. These tolls are:

i. **JavaScript (JS)**: The interface design and coding of the system is achieved with JavaScript.
ii. **PHP Hypertext Pre-processor**: The backend (Connection to the database) of the system was achieved using PHP.
iii. **Structured Query Language**: The database designed is achieved using Structured Query (SQL) Management System.
iv. **Hadoop model**: The Hadoop is connected to the MYSQL database to create large memory location for any kind of unstructured data that will be uploaded or downloaded in the distributed computing environment.

A comparative analysis was carried out to determine the performance of the proposed and existing systems in terms of execution time in upload and in download of files with different sizes. The proposed system uses the Rivest, Shamir, and Adelman of MIT (RSA) algorithm. The algorithm is based on an asymmetric encryption and decryption principle. In this algorithm,Public key is distributed to all through which one can encrypt the message and private key which is used for decryption is giving by the admin and is kept secret and is not shared to everyone.

This algorithm suggests the encryption of the files to be uploaded on the cloud. The integrity and confidentiality of the data uploaded by the user is ensured doubly by not only encrypting it but also providing access to the data only on successful authentication. The authorized user can also download any of the uploaded encrypted files and read it on the system. Difference sizes of files or documents were uploaded to the distributed server, the time for execution of these files was determined to indicate the amount of time taken to upload each file depending on its size in kilobyte. Some of the file sizes used in the proposed system are 128kb, 256kb, 512kb, and 1024kb.

The existing system used Blowfish Algorithm, which has Symmetric-key principle, which is a class of algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys may be identical or there may be a simple transformation to go between the two keys. The keys, in practice, represent a shared secret between two or more parties that can be used to maintain a private information link. This requires that both parties have access to the secret key which is one of the main drawbacks of symmetric key encryption used in the existing system, in comparison to proposed system asymmetric key encryption. In comparing the evaluation of the both systems, in the existing system, difference sizes of files or documents were uploaded to the distributed server, the time for execution of these files was determined to indicate the amount of time taken to upload each file depending on sizes in kilobyte. Some of the file sizes used in the existing system are 100kb, 120kb, 150kb, and 180kb.

A graph of files size against time taken was plotted and is shown in figure 1. From results, it is observed that the existing system was able to handle small volume of files in kilobyte which shows that it is limited to handle large files structure, and the execution time taken for uploading files to the distributed server is higher than the proposed system, which means that uploading time of files is higher. In the proposed system, it is observed that it is able to handle larger volume of files and the execution time for uploading files to the distributed server is lower when compared to the studied system. This also indicates that during testing, the system was able to give precise expectation based on the limitation in the existing system.

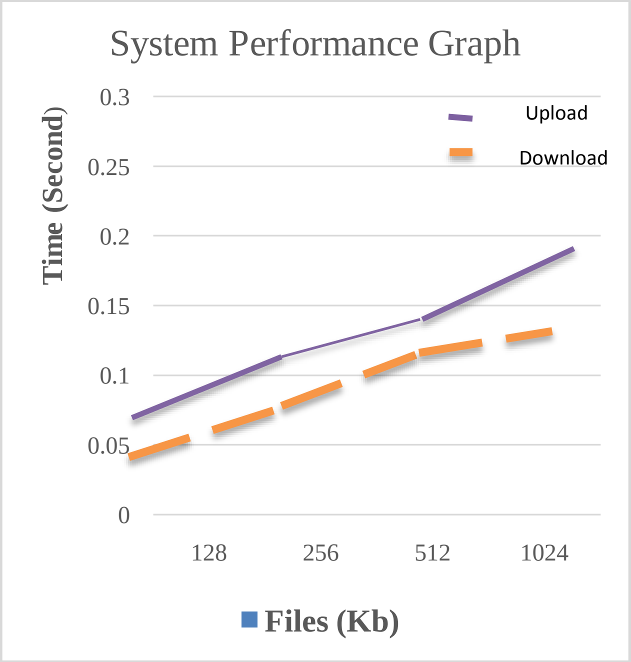

Figure 1: Proposed System Graph Evaluation

VI. CONCLUSION AND RECOMMENDATIONS

The serious challenge associated with distributed computing environment is security principle in terms of confidentiality and data integrity. Ensuring that system and important data transmission between parties or different users in a distributed computing environment is kept completely undisclosed to the unauthorized entities or owner

For the newer distributed architectures, such as wireless and peer-to-peer, this challenge becomes more interesting due to factors such as lack of structure and lack of trust in the network. This has increased the risk operational disruption, financial losses, legal issues, compliance penalties and reputational damagesevolved.

Maintaining the accuracy and consistency of data over its entire life-cycle is very paramount and important. However, in recent times, hackers temper and modify data in an undetected manner without permission of its legitimate user thereby leading to data breach. Hence, the need for a more scalable model for data security in distributed database system is developed. The developed Model in a Distributed Computing Environment has provided solution to the security challenges in a distributed database system. For future improvements and development on this research, concentration on Migration and load balancing and integrating other complex NoSQL databases such as Apache Cassandra and MapReduce to the proposed system is recommended.

REFERENCES

[1]. Grzegorz, W. O. Wojarnik, (2016). Selection of Working Database for the Genetic Algorithm Processing Data of Exchange Quotations, Information Systems in Management. 2, 294-304.

[2]. Watter A. (2015). representation and then learning a transition model in this representation. Multiple studies.International Journal of Database Management Systems (IJDMS), 5(9), 3 – 23.

[3]. Chang-Ji, W. (2013). Using attribute certificate to design role- based access control, 4th International Conference on Parallel and Distributed Computing, 12(31): 216-218.

[4]. Xiao M and Lixonglei, T. (2017). Embedded Database Query Optimization Algorithm Based on Particle swarm. Seventh international conference on measuring Technology 1-14

[5]. Francisca, O. (2018). Information gathering methods and tools: A Comparative Study, Research gate Journals, 1-6, 7 (9), 1-6.

[6]. Belapurkar, A.Chakrabarti, A. Ponnapalli,H.andVaradarajan,N.(2009).Distributed Systems Security: Issues, Processes and Solutions, 2(11):231-332.

[7]. Vijayarani S. (2016). Research in Big Data: An Overview, Informatics Engineering, an International Journal of Database Management Systems (IJDMS), 6(3), 1 – 16.

[8]. 8. Nihalani, k.RajMohan, S. Prabhakar, S. andPrachi, S. (2016). “Integration of Artificial Intelligence and Database Management System: An Inventive Approach for Intelligent Databases”. First International Conference on Computational Intelligence, Communication Systems

[9]. Raymond, B. (2015). Distributed Database Systems. International Journal of Computer Applications (IJCA), 5(3), 1-8.

[10]. Kinsey, R. (2016). Informatica Intelligent Cloud Services automates data migration andintegration, 22(45), 243-261.

[11]. Farid, A. (2014) A Framework of Transaction Management in Distributed Database System Environment. International Journal of Advanced Research in IT and Engineering (IJARIE), 3(2), 35 – 53.

[12]. Touch, D. (1994). Intelligent Support for Exploratory Data Analysis. The Journal of Computational and Graphical Statistics, 14(44): 278-323

[13]. Chang-Ji, W. (2013). Using attribute certificate to design role- based access control, 4th International Conference on Parallel and Distributed Computing, 12(31): 216-218.