SindhUniv. Res. Jour. (Sci. Ser.) Vol. 52 (03) 279-284 (2020)
http://doi.org/10.26692/sujo/2020.09.41
SINDHUNIVERSITYRESEARCHJOURNAL(SCIENCE SERIES)

Security issues in data at rest in a non-relational Document Database

M. R. JAMALI[++], A. G. MEMON, M. R. MAREE

Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Pakistan

Received 07th March 2020 and Revised 20th July 2020

Abstract: Modern applications required global scope and continued availability. Distributed non-relational databases are used as backend of Big Data to store operational data. Non-relational document databases are designed for high performance, scalability and availability. As increasingly sensitive, confidential, private and personal identifiable information (PII) are being stored in databases server, security issues become growing concerns. Non-relational database document databases are lack of security mechanism and data is stored as plaintext, therefore, it is possible that private, confidential and personally identifiable information data are to be disclosed to insider and outsider including intruder and malicious administrators and it is also possible exposing PII on dark web as commit identity theft and may be victim at risk.

In this paper, a secure Middleware is proposed to perform the required transformation of information using modern asymmetric cryptography to provide the level of security for secure storage and access Big Data operational semi-structured data on non-relational document database in public domain. The proposed system effectively protects private, sensitive and PII data by enforcing privacy, confidentiality, and integrity of data in data at rest in the public domain.

Keywords: Big Data, Non-relational Document Database, Security, Asymmetric Cryptography

1. INTRODUCTION

Big Data is a large volume of data in a variety of formats. It is an important growth area of the Cloud Computing and massive computation run effectively, cost-efficiently and workload to be managed faster as well as cheaper. The cloud environment allows Big Data variety of format storage and analyzing including structured, semi-structured and unstructured data (Ramzan *et al*, 2019). As structured data includes fixed format data such as table. While unstructured data is in unknown form i.e., text files, search engine results, videos and images. Beside it, semi-structured data includes both forms of data. Big Data Technologies are analytical and operational. Hadoop and Map Reduced (Colombo and Ferrari, 2019) have used complex analysis of data for decision making and operational capabilities include capturing and storing data in non-relational databases also known NoSQL (Not Only SQL or Not SQL databases).

Mr. Carlo Strozzi first time used the term NoSQL in 1998 and it is a broad class of database management systems and they do not use SQL (Structured Query Language) as query language. The NoSQL or non-relational follow the CAP theorem (Consistency, Availability, Partition Tolerance). Data consistency after an update or write user sees an update immediately. The system should be available and continue to operate some hardware out of order or software crash and not working. Partition tolerance dynamic addition and removal of the system in a network (Rao *et al*, 2018).Non-relational databases not have transactions and do not follow ACID instead they follow the BASE properties (Basically Available, Soft State and Eventually Consistent). Applications work all the time and need not be consistent and should be in some known state eventually. Document database provides consistency at document level but at the database level it is eventually consistent. Non-relational databases include Key-Value, Document, Column and Graph.

Document Databases stores semi-structured data. It provides high performance, availability, high volume data storage, very simple data model and scalable use in Big Data (Kumar *et al*, 2017).Traditional database scheme is declared before insert tuples and attributes. Collection in document database is not an enforce document structure which makes it more powerful while each document can be retrieved by a unique ID.

In document database data at rest is not secure, and data is stored as clear text. Data are kept unencrypted, there fore confidentiality, privacy and integrity of data are not achieved. These databases have lack of proper data protection mechanism as well as there is lack of research. The data needs to be protected from unauthorized access and transmitted to the intended receiver with confidentiality and integrity.

[++] Corresponding author: Mujeeb-ur-RehmanJamali email: mujeebjamali@usindh.edu.pk

Asymmetric algorithms are superior in security to provide various security services including confidentiality, data integrity, authentication and non-repudiation.These algorithms are classified based as factorization, logarithm and elliptic curve. Integer factorization most prominent algorithm is RSA (Ron Rivest, Adi Shamir and Leonard Adleman). Discrete Logarithm includes Diffie-Hellman (DH) key exchange, ElGamal, DSA (Digital Signature Algorithm). ECC (Elliptic Curve Cryptography) and ECDSA.

2. RELATED LITERATURE

In this section research papers are presented that proposedata confidentiality using cryptography and other technologies. In security issues in Big Data and NoSQL (Ebrahim *et al.,* 2015) stated that document databases did not support authentication, authorization and data encryption. In this research (Altrafi *et al,* 2014) presented that data confidentiality is not achieved due to data is storedas clear / plaintext and further said that many non-relational databases by default did not come with authentication or authorization mechanism, therefore, external method can be used to provide data confidentiality, authentication and authorization.A security plan was proposed (Mohammad Ahmadian *et al,* 2017) proposed a cryptographic module includes data elements, and mappings of cryptographic modules to the data fields and a descriptivelanguage based on JSON notations.In this paper (Hasija and Kumar, 2016)conversion of plain text into different format of compressed BSON type and it is proposed that compression provide security without affecting efficiency. In this study (Kumar and Garg2017) proposed that NoSQL document database hasvarious security issues and it does not support theencryption/decryption by default due to that vulnerable to injection attack and has exposureto DOS attacks. The author suggest that good secure cryptographic system may be used to resolve security issues. The author uses for encrypting and decrypting the data symmetric-key cryptography such as AES, DES, Blowfish for NoSQL document database. It is possible that encryption of data-at-rest of entire drive, encrypts individual files or databases on the disk. It is also possible that encrypt entire documents or individual attributes at the application level.The proposed system is different and a secure Middleware encrypts only fields of sensitive, confidential, private and personal identifiable information are being stored in non-relational document databases.

3. SECURITY ISSUES DATA AT REST

Document Database did not provide security as a priority it is the responsibility of the application to provide security of data at the application level. These databases are week authentication, data are stored as plaintext and no mechanism to verify any alteration of data. Therefore, privacy, confidentiality and data integrity are not achieved. It is the responsibility of developers to provide security at the application level. Insider attackers i.e., includes malicious and trusted employees and administratorcan gain access to useful information. The outside attacker can unauthorized access and disclosure of information. Both insider and outsider may be passive or active attacker.

It is very important for the user that data must be secure when it is in transition and even after it has been stored on the server. The owner of the data needs to be assured that the data stored on the database server is protected against data eavesdropping from outsiders and data needs to be protected even from a malicious insider, if the insider not be trusted. Security mechanisms are required to protect sensitive information on residing public clouds.

4. METHODOLOGY

In this study, data-in-motion is protected by encrypting data in transit using SSL (Secure Socket Layer). Cryptography is used for providing security services i.e.,authentication, data confidentiality anddata integrity (Patel, 2019). In this research,a proof of authentication of user,data confidentiality that is prevention of the disclosure of data to unauthorized user,data integrity that is unauthorized alteration data can be verified.Asymmetric cryptography protect data by converting into unreadable form.

Secure middleware carried out the required transformation and database server not in need to be modified. Only private, sensitive, confidential and PII fields are converted into an unreadable form.

Proposed Methodeach component of system is associated with specific port of the system to provide and receive service as under:

Data Flow **(Fig. 1)**

Step:1
At first, user requests to CA Server for Digital Certificate.

Step:2
User will login after prove the authenticity i.e., Id and password along with Digital Certificate. (Middleware check the user authentication and if valid then authorize).

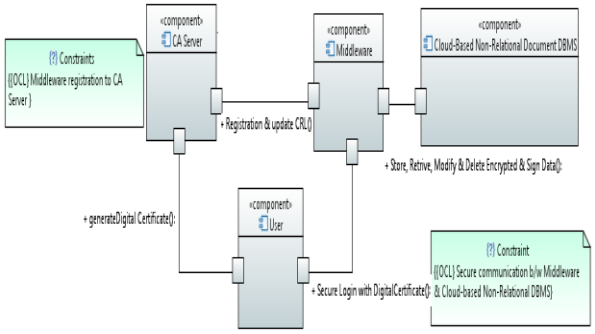
Step:3
After proving the authentication user can access the database and perform operation on data.

Step:4
Middleware encrypts and signs sensitive document's fields including PII using asymmetric cryptography with varying key size as a user security requirement. Middle-ware verifies the integrity of data has not been altered.

Step:5
Middleware performs cryptographic operation on specific fields of a particular document in the collection of the database.

All communication to be done using SSL (provide data encryption at motion/transaction).

PROPOSED SYSTEM (UML COMPONENT DIAGRAM)




```
_id: "1"
Empid: 3110
First_Name: Binary('dKbCDaqqm1bp2A/HX+hJZynX6jYrBEeUlQmAWspjWAtjNblUQJAbtfzA1n0KW5EtIaXSmrlE9+uS0rQab/CNSkf7+mx3KnxXKz+o...')
Surname: "Jamali"
Gender: "M"
Email: Binary('fNbB4QSUvcMqwaeqrq3VEGRBvpbMVmbZYcTkLtVRZJX1w84jLeKKv14WGF3DQX99uoHwBPdRNT9Hr7mxbpDsSZXg4FN3De1gdQV0...')
DOB: Binary('n65JP46YEyB9GSDwt02jwFCEGJsf+Z+S7sivUnb+4kOVECk0e62BM4WvR8zcTmW7pngEM7TjK0VHBxkURnmOMHe+e0tNe+8ckIze...')
Date_of_Joining: "26-10-2009"
Salary: 55000
PhoneNo: "03332607423"
Address: Object
    Present: Binary('EAPuhnLRCg6eCFV36/bhm6Qqx4RnZjlxucOOvdvvL0AkZ892BQ3Cz7toEqR9xhgILr0GxxPzXeKyPIC/UfWwo5K+VYEAxh0XkMQi...')
    Perminent: Binary('EAPuhnLRCg6eCFV36/bhm6Qqx4RnZjlxucOOvdvvL0AkZ892BQ3Cz7toEqR9xhgILr0GxxPzXeKyPIC/UfWwo5K+VYEAxh0XkMQi...')
Bank_Account_Info: Object
    AccountNo: Binary('KVwJsng+0ndSmhhc0DdNVV2zUgxCkmjRMqH2iAzJYX1qXfxu/z82PvCgr0UMRFd3NHluHTe4PF63PUicsqShG9nA1hYVeDUuYSUk...')
    Bank_Name: "HBL"
    Branch: "Sindh University Jamshoro"
    IBAN_No: Binary('V/ueaHnUi6IrKyWpLofnmZkhngzPTmCID1CSdC7oGhy/sr301J3n38DdzoJn+g1QWiuNM3PeGEy+5bg/I/sa0KCclmFfGocKduxZ...')
    Next_of_Kin: Binary('Dao1X2IaoNNcUDahMO/ONTol0a00car2F2rf33JFqOtlykiVje3S9woVPERONlOCvmwql0xWidEmIHytgCM/T4o4jA2+uWkSyJ6k...')
    Card_Type_Code: "DC"
    Card_Type_Name: "Debit Card"
    Card_No: Binary('ogEJ0PBrKteIqywUiF9az4uGY8mjj388mjDUE5z/0h8uAf+SGZAFyNK4zU1Zn3DAh8H7QmzoTVuZ0Q7xoRyv9hixgq0hKk+NSo88...')
    Issue_Date: Binary('n+Hcgt7EzGeU/8HLumLR/xbegVrrbLlGNR9xr5ASXbsc6TeITMI2rW89DNq5zj/vsP/qPYxj+x/n5KEphjQx4Tp08/yjoxttdgT2...')
    Expiry_Date: Binary('Y8O+GHouTBJTdLo5uhSvE2Tf/6QXxA5eBBbXfQRadjVH0yt6CN4YWJQBb6J8X9wUQSpv4qEx8JD7Gz33eUF2q7lvSkHC0RScdoBb...')
    Card_Pin: Binary('oEKp9sZ4NfyKa9D45GEGsPGBYFt+aRooyVoUbWJw/3AQ58kmDPaHZ+gJPhlUH9cR+WaHr7SJqOAn4gAQvOlkESNY+qaeVGJ8fQRK...')
    Credit_Limit: 25000
```


Fig.2

5. HARDWARE AND SOFTWARE REQUIREMENT

The prototype implementation of the system and experiment was conducted on Processor Intel Core i 3, (4 CPUs) 3.3 GHz, RAM 2048 MB, 500 GB Hard disk and Window 8.1 Pro (64-bit). Document Database (MongoDB 3.4), Java 1.9 and asymmetric cryptographic API (JCA and JCE). Two Data Sets, one 1.5 Million documents of 11 fields and 0.5 Million documents with 21 fields CSV file imported.

6. RESULTS AND DISCUSSION

Projection **(Fig.2)** of the document with encrypted fields of the document and not sensitive data in the collection of the database to be stored as plaintext. Secure Middleware transform only selected private and confidential fields into the encrypted and unreadable format and data is stored in public domain.

There are 10 executions of each operation i.e., encryption field, insertion and updating document in the collection of the document database. In **(Fig. 3)** varying key sizes of asymmetric cryptography used for performing encryption operation for private, PII and sensitive fields of 223 Bytes (12 Fields) out of 24 Fields of the document in the collection of the database. RSA 1024 bit key mean time in a millisecond is 310 and standard deviation and dispersion from mean up and down are 40.32. RSA 2048bit key meantime is 323.83 and its standard deviation is 47.75 and RSA 4096 bit key meantime is 339.83 and dispersion from mean is 57.56. As the size of the key is increased result becomes more secure but encryption time is increased provide a level of security.

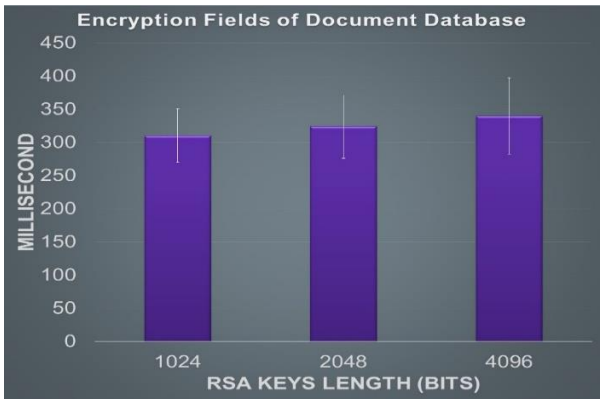


Fig. 3

In creation fields of a document in the collection of the document database operation in (**Fig. 4)** mean time in a millisecond with varying key size new document insertion in the collection of the document database and encrypted fields 223 Bytes (12 Fields) along with 90 Bytes (12 fields) plaintext. The mean time is in millisecond 22.33, 22.5 and 29 respectively of RSA key size 1024, 2048 and 4096 and their standard deviation and dispersion are 3.72, 5.00 and 4.81 in a millisecond. The performance of the document database is very high in insertion and creation of the documents in the collection of the database and result shows mean time in the millisecond and their standard deviation.

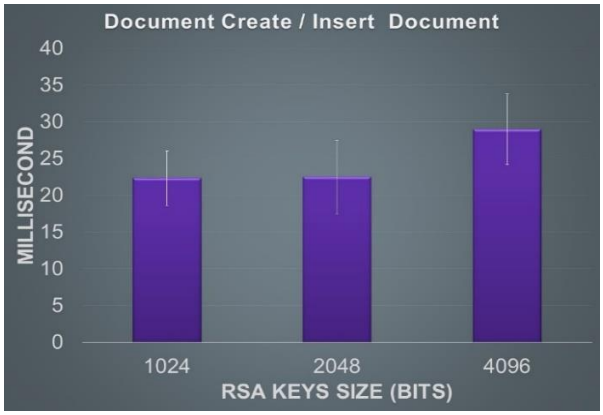


Fig.4

In **(Fig. 5)** updating encrypted 223 Bytes Fields (12 Fields) and (12 Plaintext Fields). The mean time in millisecond 23.83, 23.83 and 25.83 respectively of the RSA key size 1024, 2048 and 4096 are their standard deviation and dispersion is 5.45, 6.88 and 1.60. In the update operation, the performance of the document database is very high and results mentioned mean time of each operation and their standard deviation from mean time in millisecond.

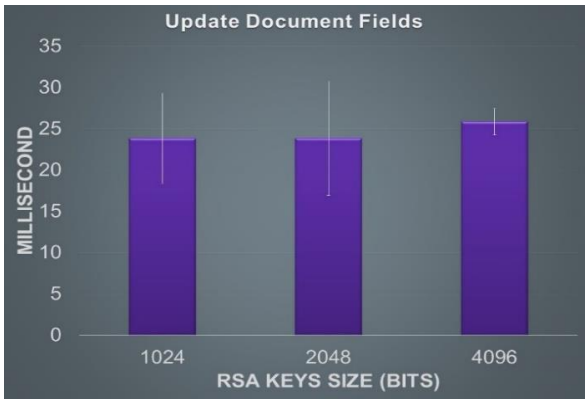


Fig.5

7. CONCLUSION

In this paper, a secure Middleware is proposed to provide transformation of private, confidential and personally identifiable information into encrypted and unreadable form using asymmetric cryptography with varying the key size. It provides a level of security according to user requirements. Experiment and result confirmed that proposed secure Middleware is easy,useful, useable and provides enhances security. All non-relational database to be used without any modification.

REFERENCES:

Asadulla, K. Z. (2014). "NoSQL databases: new millennium database for big data, big users, cloud computing and its security challenges". IJRET: International Journal of Research in Engineering and Technology.

Charmi P. and R. Sheth, (2017). "Encrypting Data of MongoDB at Application Level", Advances in Computational Sciences and Technology.

Cornelia G., R. Gyrodi, G. Pecherle, and A. Olah, (2015), "A Comparative Study: MongoDB vs. MySQL". 13thInternational Conference on Engineering of Modern Electric Systems (EMES).

Duygu T., Ramazan and S. Sagiroghlu, (2015). "A Survey on Security and Privacy Issues in Big Data", IEEE.

Ebrahim S. and M. A. Nematbakhsh, (2015). "A Survey on Security Issues in Big Data and NoSQL". Advances in Computer Science an International Journal.

Hanen A. and F. Gargouri, (2017), "MongoDB-Based Modular Ontology Building for Big DataIntegration". Springer J. Data Semant (2018) 7:1–27. Https://doi.org/10.1007/s13740-017-0081-z

Haina Y., X. Cheng, M. Yuan, L. Xu, J. Gao, and C. Cheng, (2016). "A Survey of Security and Privacy in Big Data", IEEE.

Haralambos M. and P. Katsaros, (2015). "Security-aware elasticity for NoSQL databases", IEEE.

Hitesh H. and D. Kumar, (2016). "Compression & Security in MongoDB without affecting Efficiency" ACM ISBN 978-1-4503-3962-9/16/03.
DOI: http://dx.doi.org/10.1145/2905055.2905155.

Jitender K. and V. Garg, (2017). "Security Analysis of unstructured data in NOSQLMongoDB Database", IEEE International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), 300-305.

Kamlesh K. H. and S. Mary, S. Bhanu, (2014). "Sensitive Data Protection of DBaaS using OPE and FPE". Fourth International Conference of Emerging Applications of Information Technology.

Krunal S. and J. Patel, (2015). "EncryScation: A Novel Framework for Cloud IaaS, DaaS security using Encryption Obfuscation Techniques", IEEE.

Kuntal, P., (2019), "Performance analysis of AES, DES and Blowfish cryptographic algorithms on small and large data files", Springer Int. J. inf. Tecnol. https://doi.org/10.1007/s41870-018-0271-4

Mohammad, A. (2017). "Secure Query Processing in Cloud NoSQL", IEEE International Conference on Consumer Electronics (ICCE).

Mohammad A., F. Plochan, and D. C. Marinescu, (2017) "Secure No SQL: An approach for secure search of encrypted NoSQL databases in the public cloud". International Journal of Information Management.

Natalia M. and A. Makhmudova, (2016). "Survey of Big Data Information Security", IEEE 4thInternational Conference on Future Internet of Things and Cloud Workshops.

Obay G. A., M. A. Mohamed and M. O. Ismail, (2014). "Relational vs. NoSQL Databases: A Survey". International Journal of Computer and Information Technology.

Praveen M. (2014). "Big Data working Group Big Data Taxonomy". Cloud Security Alliance.

Pietro C. and E. Ferrari, (2019). "Access control technologies for Big Datamanagement systems: literature review andfuture trends", Springer Colombo and Ferrari Cybersecurity.
https://doi.org/10.1186/s42400-018-0020-9.

Ramalingeswara T. R., P. Mitra, R. Bhatt, and A. Goswami1, (2018). "The big data system, components, tools, and technologies: asurvey", Springer Knowledge and Information
Systemshttps://doi.org/10.1007/s10115-018-1248-0.

Shabana R., I. S. Bajwa, B. Ramzan and W. Anwar, (2019), "Intelligent Data Engineering for Migrationto NoSQL Based Secure Environments", IEEE Special Section on Advanced Software andData Engineering for Secure Societies.

Sethuraman S. and A. Nair, (2015) "Security Maturity in NoSQL Databases – Are they Secure Enough to Haul the Modern IT Applications?", IEEE.

Sergey B., S. Murzintsev and A. Tskhai, (2016). "Detecting Text Similarity on a Scalable No-SQL Database Platform", IEEE.

Shady H. and Z. Zainol, (2017), "Document-Oriented Data Schema for Relational Database Migration toNoSQL", IEEEInternational Conference on Big Data Innovations and Applications. DOI 10.1109/Innovate-Data.2017.13.

Sharvari T. (2016). "Non-Relational Databases in Big Data", ACM.

Sundhara K., K. B. Srividya, and S. Mohanavalli. (2017), "A Performance Comparison of Document OrientedNoSQL Databases", IEEE International Conference on Computer, Communication, and Signal Processing (ICCCSP-2017).

Xingbang T., B. Huang and MinWu, (2014). "A Transparent Middleware for Encrypting Data in MongoDB", IEEE Workshop on Electronics, Computer and Applications.

Zoran H., D. gimnazija and V. Croatia, (2016). "Comparative Analysis of Cryptographic Algorithms", International Journal of Digital Technology.