Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 2437–2444

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

An Efficient Hybrid Approach of Attribute Based Encryption For Privacy Preserving Through Horizontally Partitioned Data

Devendrasinh Vashi[a], H B Bhadka[b], Kuntal Patel[c], Sanjay Garg[d]

[a]*Nirma University, Ahmedabad-382481, India*
[b]*C U Shah University, Wadhwan-363030, India,*
[c]*Ahmedabad University, Ahmedabad-380009, India*
[d]*Nirma University, Ahmedabad-382481, India*

Abstract

In the current scenario, the majority of the applications are having the database and it is increasing day by day. The major problem is how to use those databases by sharing it for the societal benefit. So information security is the real concern while sharing the data with the third party. This article aims to implement cryptography on the horizontally partitioned database. There are number of PPDM techniques. Cryptography techniques [9] are one of the solutions for information securities. But each technique is having its own disadvantages. Means if we use individual encryption technique then hackers may misuse the data by decrypting it. So in this research article hybrid approach is used to secure the data. With the help of horizontally partitioned database table. symmetric and asymmetric techniques are used for the partitioned table. In the experimental result symmetric, an asymmetric and hybrid approach of encryption is discussed. This is an approach of implementing encryption on different attribute and different rows in one data set. The tabular and graphical analysis shown that this approach is comparatively good in PPDM.

Keywords: PPDM, Encryption, Cryptography, DES, RSA, Privacy preserving, Security, SMC;

* Corresponding author. Tel.: +91-942-677-4300.
E-mail address: devendra.vashi@gmail.com

10.1016/j.procs.2020.03.296

1. Introduction

Data mining is a very important tool for analysis of any kind of data. Data mining is useful to find out the pattern from the large dataset like banking, online shopping and personal data of an individual from the medical dataset [14]. The medical organization used to give medical diagnosis data [8] to researchers to extract useful data. Sometimes insurance companies also ask the medical data for analysis of the specific age group data for the survey. Even banking sectors are asking the medical report of a person before sanctioning the loan or credit card. So in privacy-preserving data mining, there is a need for protecting personal and sensitive data during data mining. Data should be provided for extracting the useful pattern but by the same time, sensitive and personal data should not be revealed or it should be protected in PPDM. PPDM techniques are mainly to secure multiparty computation (SMC) as data of multiple organizations is used in a data mining process. There are many techniques to preserve privacy like anonymization and perturbation. These two techniques are manly for the large dataset and the basic drawback of both systems is information loss. cryptography is costly but more accurate and reliable [11].

Most of the medical information systems are networked-based having lots of resources like data and software applications that are always vulnerable to attacks. An attack may also take place during data transmissions. To provide security to such IT systems, many cryptographic algorithms are available. In this paper, the experiment is performed using two of the most popular algorithms: RSA and DES. The DES is one of the symmetric key algorithms whereas RSA is one of the asymmetric key-based algorithms. Many security professionals think that developing a fully protected system is almost an impossible assignment [1]. But, according to [2] implementation of a hybrid 128-bit key AES-DES algorithm can enhance the security of the system is a great way. Many researchers worked on performance comparisons of various cryptographic algorithms [3]. This paper describes how DES, RSA, and hybrid (DES+RSA) approach can be used to enhance the overall performance of the system along with privacy [6] preserving data mining through horizontally partitioned data.

The structure of the remaining article is as follows: section 2 describes the proposed hybrid [12] approach of encryption technique and it's an algorithm. Section 3 describes the way partitioned the data horizontally for the given attributes of personal data which may be part of any medical database. Section 4 explains the concepts of symmetric and asymmetric encryption. Section 5 discussed the result analysis with tabular data and graph. Section 6 gives a brief overview of the experiment environment. Section 7 gives an overview of challenges and future work. Section 8 concludes the work of the article.

2. Proposed approach

In this approach privacy preserving is done through hybrid approach. Initially database is collected and converted into excel dataset. Then that excel dataset is converted into the sql server dataset. After that sql dataset is partitioned horizontally into two tables based on the gender. In the proposed system option is given to choose the attributes [5] or [10] to be encrypted in both partitioned table. Also option is given to choose the encryption technique which is to be implemented on the selected attribute in both the partitioned table [13]. After execution of the selected encryption technique, execution time will be shown for that techniques. Each partitioned table can be downloaded and can be merged also. Downloaded table can be shared with the third party for the data mining purpose.

Algorithm:

```
Step-1 : Start
Step-2 : create the data set in excel sheet
Step-3 : Upload the data set in the system to convert it into sql server data set
Step-4 : partitioned the sql sever data set horizontally into two data set based on the gender or age
Step-5 : select the attribute from each data set to encrypt
Step-6 : Implement the symmetric encryption for the selected attribute on table-1
Step-7 : Implement the asymmetric encryption for the selected attribute on table-2
Step-8 : download the encrypted data set into excel sheet separately
```

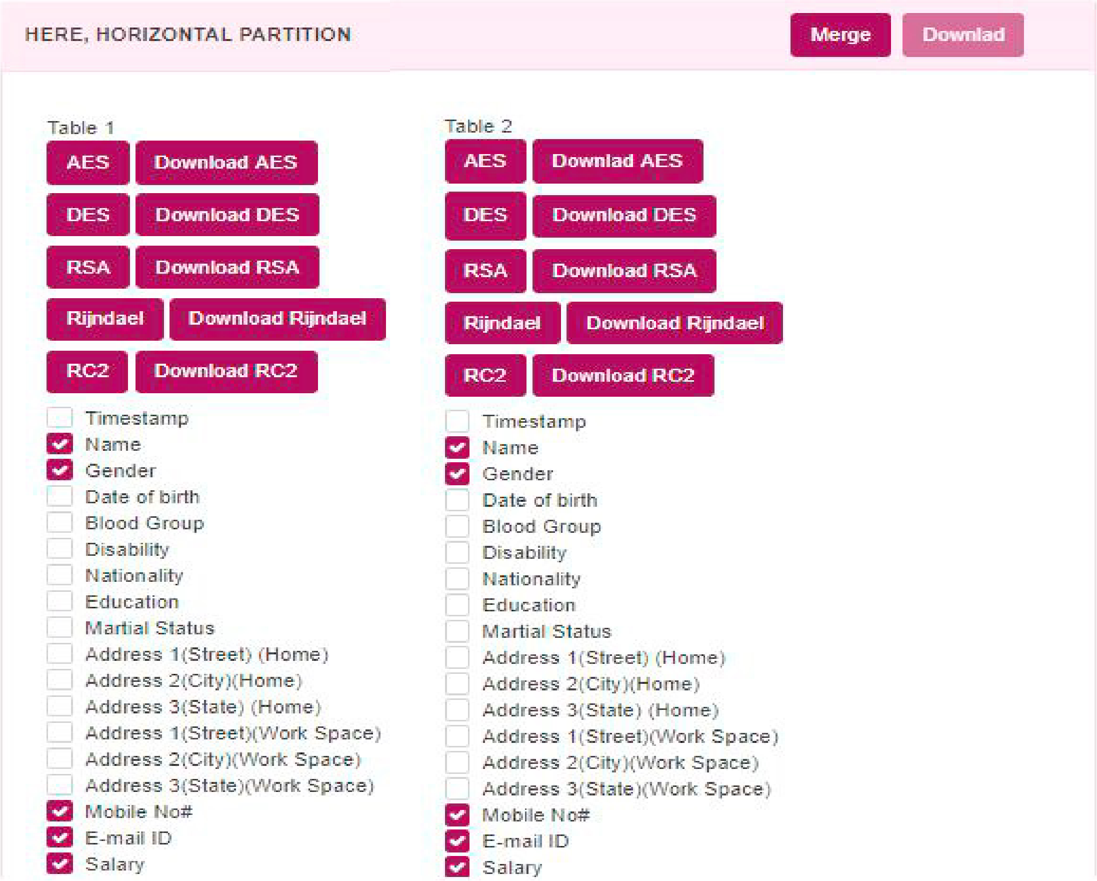
Step-9 : combined both the encrypted excel sheet.
Step-10 : Download the encrypted excel sheet for PPDM purpose.
Step-11 : Stop

3. Horizontally partitioned database

This kind of database partition is a concern with no of partitioning the rows in the multiple tables. If the medical database contains thousands of records and analyst want to survey based on the gender-related data then the whole table will be partitioned into two tables, Male related data rows stores in Male-table and female related data rows stored in Female-table. Both tables can be combined to get the full data set at any time.

Following the list of attributes taken for an experiment for the medical database [10]:

Timestamp
Name
Gender
Date of birth
Blood Group
Disability
Nationality
Education
Marital Status
Address 1(Street) [Home]
Address 2(City)[Home]
Address 3(State) [Home]
Address 1(Street)[Work Space]
Address 2(City)[Work Space]
Address 3(State)[Work Space]
Mobile No.
E-mail ID
Salary
Pan card Number
Diseases
Medication and allergies
Doctor Name
Date & Time of visit



HERE, HORIZONTAL PARTITION

Table 1

□ Timestamp
☑ Name
☑ Gender
□ Date of birth
□ Blood Group
□ Disability
□ Nationality
□ Education
□ Martial Status
□ Address 1(Street) (Home)
□ Address 2(City)(Home)
□ Address 3(State) (Home)
□ Address 1(Street)(Work Space)
□ Address 2(City)(Work Space)
□ Address 3(State)(Work Space)
☑ Mobile No#
☑ E-mail ID
☑ Salary

Table 2

□ Timestamp
☑ Name
☑ Gender
□ Date of birth
□ Blood Group
□ Disability
□ Nationality
□ Education
□ Martial Status
□ Address 1(Street) (Home)
□ Address 2(City)(Home)
□ Address 3(State) (Home)
□ Address 1(Street)(Work Space)
□ Address 2(City)(Work Space)
□ Address 3(State)(Work Space)
☑ Mobile No#
☑ E-mail ID
☑ Salary

Fig. 1. design of implemented system of proposed approach

4. Symmetric and asymmetric cryptography techniques

Major categories of cryptography techniques are classical techniques, modern techniques, and public-key encryption. Based on a number of keys used, cryptography algorithms are categorized into Symmetric key and Asymmetric key cryptography algorithms. The asymmetric key is also known as Two-key or Public-key cryptography [4]. The symmetric key is also known as a single key or private key cryptography.

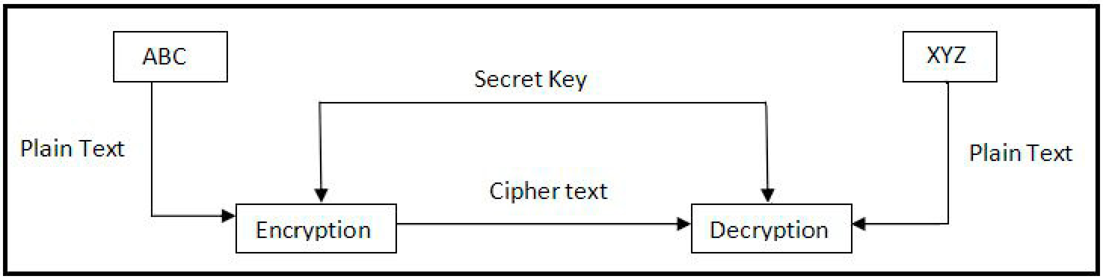


Fig. 2. Symmetric cryptography

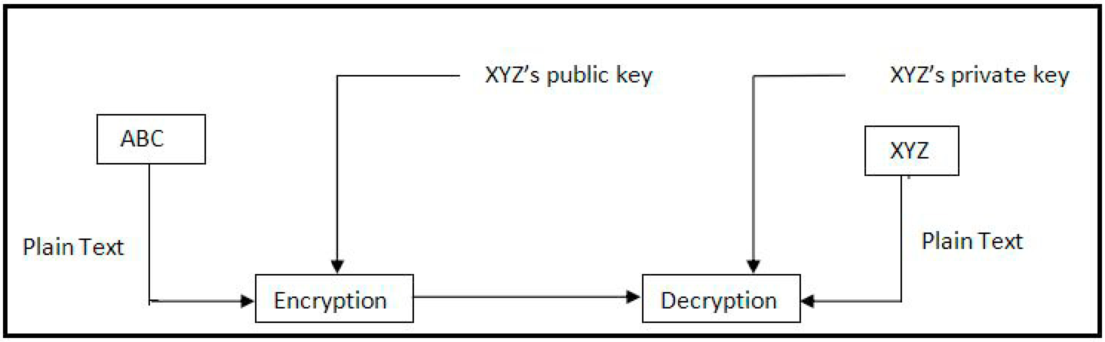


Fig. 3. Asymmetric cryptography

5. Experiment and result analysis

A. Implement symmetric encryption on both horizontally partitioned table:

Table 1. Execution time of different size dataset [7]with DES technique

File name with data size (kb).	Execution time to table1 (milliseconds)	Execution time to table2 (milliseconds)	Total execution time (milliseconds)
	DES	DES	
Sensitive data 1.xlsx (100)	5246.4913	2255.5537	7502.045
Sensitive data 2.xlsx (200)	12320.4785	6711.0791	19031.5576
Sensitive data 3.xlsx (300)	15386.1425	7866.623	23252.7655
Sensitive data 4.xlsx (400)	44923.698	17394.7461	62318.4441
Sensitive data 5.xlsx (500)	51603.7297	23008.4188	74612.1485
Sensitive data 6.xlsx (600)	95470.0559	26598.2158	122068.2717
Sensitive data 7.xlsx (700)	113638.1739	47018.4498	160656.6237
Sensitive data 8.xlsx (800)	130006.0133	49567.0857	179573.099
Sensitive data 9.xlsx (900)	168102.6198	61102.8887	229205.5085
Sensitive data 10.xlsx (1000)	190421.6182	72445.9348	262867.553

B. Implement asymmetric encryption on both horizontally partitioned table:

Table 2. Execution time of different size dataset with RSA technique

File name with data size (kb).	Execution time to table1 (milliseconds)	Execution time to table2 (milliseconds)	Total execution time (milliseconds)
	RSA	RSA	
Sensitive data 1.xlsx (100)	91847.8054	47873.327	139721.1324
Sensitive data 2.xlsx (200)	198769.2087	102330.6142	301099.8229
Sensitive data 3.xlsx (300)	223154.9558	124409.0637	347564.0195
Sensitive data 4.xlsx (400)	425103.3053	227482.6113	652585.9166
Sensitive data 5.xlsx (500)	540896.2435	293430.2817	834326.5252
Sensitive data 6.xlsx (600)	725757.2718	354841.1077	1080598.38
Sensitive data 7.xlsx (700)	863345.5006	426427.1407	1289772.641
Sensitive data 8.xlsx (800)	936562.9757	463212.5778	1399775.554
Sensitive data 9.xlsx (900)	1119045.995	574512.6269	1693558.622
Sensitive data 10.xlsx (1000)	1225847.1096	585248.6902	1811095.8

C. Implement asymmetric and asymmetric encryption on both horizontally partitioned table:

Table 3. Execution time of different size dataset with DES and RSA technique

File name with data size (kb).	Execution time to table1 (milliseconds)	Execution time to table2 (milliseconds)	Total execution time (milliseconds)
	DES	RSA	
Sensitive data 1.xlsx (100)	5246.4913	47873.327	53119.8183
Sensitive data 2.xlsx (200)	12320.4785	102330.6142	114651.0927
Sensitive data 3.xlsx (300)	15386.1425	124409.0637	139795.2062
Sensitive data 4.xlsx (400)	44923.698	227482.6113	272406.3093
Sensitive data 5.xlsx (500)	51603.7297	293430.2817	345034.0114
Sensitive data 6.xlsx (600)	95470.0559	354841.1077	450311.1636
Sensitive data 7.xlsx (700)	113638.1739	426427.1407	540065.3146
Sensitive data 8.xlsx (800)	130006.0133	463212.5778	593218.5911
Sensitive data 9.xlsx (900)	168102.6198	574512.6269	742615.2467
Sensitive data 10.xlsx (1000)	190421.6182	585248.6902	775670.3084

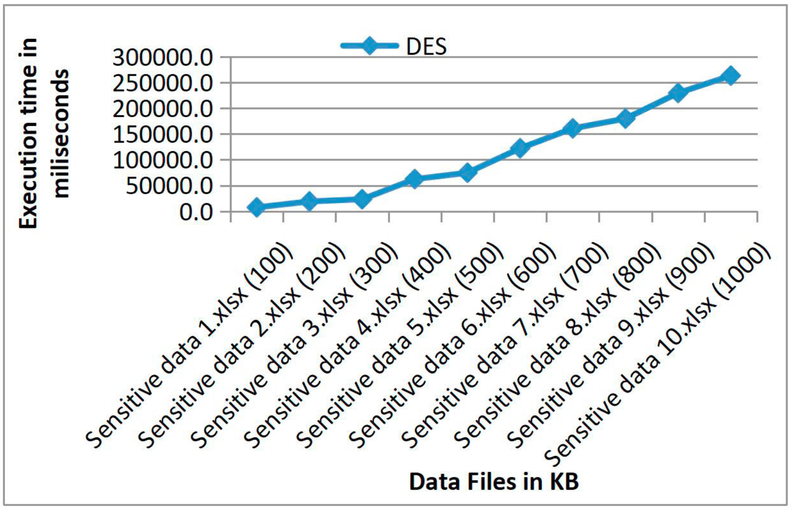


Fig. 4. Performance analysis using DES technique based on execution time

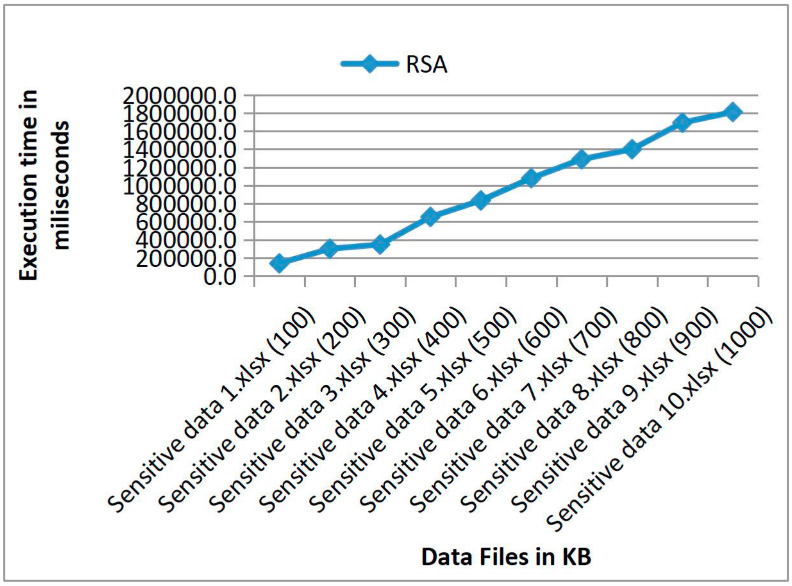


Fig. 5. Performance analysis using RSA technique based on execution time

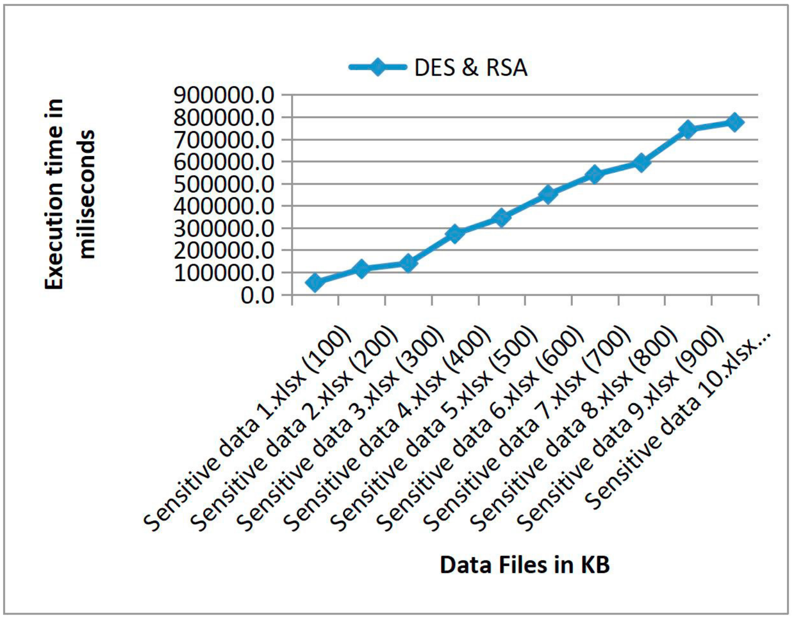


Fig. 6. Performance analysis using DES and RSA technique based on execution time

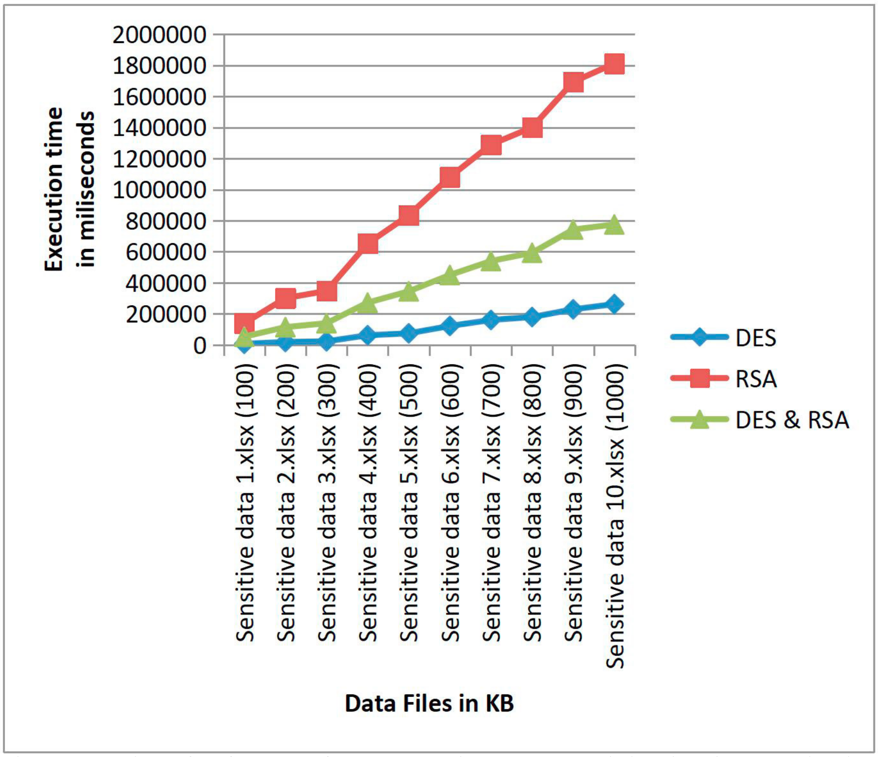


Fig. 7. Comparison of Performance of DES, RSA and DES & RSA technique based on execution time

During the experiment total, 10 database files were used of a different size. Initially, small size files were used for encryption then gradually the size of the database file is increased by 100 kb as per Table 1, Table 2 and Table 3. Initially, DES symmetric encryption technique is implemented on both the horizontally partitioned tables and observed the execution time as per Table 1. Then RSA asymmetric encryption technique is implemented on both the horizontally partitioned table and observed the execution time as per Table 2. Then DES technique is implemented on table 1 and RSA encryption technique is implemented on table 2 and observed the execution time as per Table 3. as per Fig. 1 Name, Gender, Mobile, Email and salary are the common attributes are elected to encrypt for all three separate approach of encryption so that due to common parameter and attributes execution time can be analyzed to justify the proposed hybrid approach of encryption.

Fig. 4 depicts the performance analysis of the DES encryption which shows that execution time increased in milliseconds for the database files increased in kb. Fig. 5 depicts the performance analysis of the RSA encryption which shows that execution time increased in milliseconds for the database files increased in kb. Fig. 6 depicts the performance analysis of the DES and RSA encryption which shows that execution time increased in milliseconds for the database files increased in kb.

6. Experiment environment

Following hardware and software were used for the above experiments:
Processor: Intel® Core(TM) i3-3110M CPU @ 2.40GHz
RAM: 8 GB (7.88 GB Usable)
Operating system: windows 10 pro (64 bit)

7. Challenges and future work

Every institutes and organization are collecting personal data nowadays for without any reason. Even patient visit any hospital for just medical checkup or for any diagnosis, the hospital will ask almost all personal data which might not be needed. It is a very critical issue that how to provide privacy to each and every person nowadays. Security of the data and protecting privacy is the major challenge for every organization.

After implementing this approach and based on the study of different symmetric and asymmetrical encryption technique, a new hybrid approach can be developed with cryptography vertically partitioned database.

8. Conclusion

The experimental result indicates that the performance of the Hybrid of approach (DES and RSA) algorithm is also better than using only DES or only RSA based on execution time on data files. Also, experimental results indicate that the use of DES on both the sub-tables (table1, table2) is the better option than using RSA based on execution time. Hence for time constraint systems DES need less time but less secure, RSA needs more time and more secure. And as per proposed hybrid approach need average time and more secure than DES and RSA so it is advisable to enhance the overall performance of the system by using a hybrid approach. so this attributed based encryption with horizontal partitioned is more suitable in privacy-preserving data mining as in privacy-preserving main concern is to protect privacy for sensitive attributes of an individual.

References

[1] Bilal B (2015), “Framework for Choosing Best Intrusion Detection System”, BIJIT - BVICAM’s International Journal of Information Technology, Vol. 7, No. 1, 821-826.

[2] Jignesh P., Rajesh B., Vikas K. (2012), “Hybrid Security Algorithms for Data Transmission using AES-DES”, International Journal of Applied Information Systems, Vol. 2, No. 2, 15-21.

[3] Patel, K. (2019) “Performance analysis of AES, DES and Blowfish cryptographic algorithms on small and large data files”. International Journal of Information Technology, 1-7. https://doi.org/10.1007/s41870-018-0271-4.

[4] https://support.microsoft.com/en-in/help/246071/descript-tion-of-symmetric-and-asymmetric-encryption, Retrieved on January 20, 2018.

[5] Akash Siddhpura, Prof. Daxa V. Vekariya (2018), "An approach of Privacy Preserving Data mining using Perturbation & Cryptography Technique",ISSN: 2454-4248, Volume: 4 Issue: 4, 255 – 259.

[6] Anand Sharma and Vibha Ojha (2010), "IMPLEMENTATION OF CRYPTOGRAPHY FOR PRIVACY PRESERVING DATA MINING", International Journal of Database Management Systems (IJDMS) Vol.2, No.3, 57-65.

[7] Surbhi Sharma and Deepak Shukla (2016), "Efficient multi-party privacy preserving data mining for vertically partitioned data", 2016 International Conference on Inventive Computation Technologies (ICICT).

[8] Atsuko Miyaji, Kazuhisa Nakasho, Shohei Nishida, (2017), "Privacy-Preserving Integration of Medical Data ", Springer.

[9] Dr. Hitesh Chhinkaniwala, Sanjay Garg, (2011), “Privacy Preserving Data Mining Techniques: Challenges & Issues”, Proceedings of International Conference on Computer Science & Information Technology, CSIT - 2011.

[10] Devendra I Vashi et al.(2017), “Critical Study and Analysis for Deciding Sensitive and Non-Sensitive Attributes of Medical Healthcare dataset Through Survey and Using Association Rule Mining”. Int J Recent Sci Res. 8(5), pp. 17218-17222. DOI: http://dx.doi.org/10.24327/ijrsr.2017.0805.0307.

[11] Devendra I Vashi et al (2017), "Challenges & Opportunities in Privacy Preserving Data Mining for Healthcare Dataset", International Conference on “Research and Innovations in Science, Engineering & Technology”, ICRISET-2017.

[12] Dixit, Pooja, et al. (2018) "Traditional and Hybrid Encryption Techniques: A Survey." Networking Communication and Data Knowledge Engineering. Springer, Singapore, 2018. 239-248.

[13] Lalithambikai, A., & Vanitha, M. (2018). “An Efficient Technique for Cryptography with Enhanced Key Security”. International Journal of Pure and Applied Mathematics, 118(8), 479-484.

[14] Karapiperis, D., Gkoulalas-Divanis, A., & Verykios, V. S. (2018, September). “FEMRL: A Framework for Large-Scale Privacy-Preserving Linkage of Patients”, Electronic Health Records. In 2018 IEEE International Smart Cities Conference (ISC2) (pp. 1-8). IEEE.