A Cryptography-Based Secure Privacy Preservation of the Sensitive Data in Healthcare Domain

S. Joseph Gabriel[1*], P. Sengottuvelan[2]

[1]Research Scholar, Department of Computer Science, Periyar University, Salem,
Email: talk2anette67@gmail.com
[2]Professor, Department of Computer Science, Periyar University centre for PG and Research Studies, Dharmapuri, Tamilnadu, India, Email: sengottuvelan@gmail.com
*Corresponding Author

Received: 12.04.2024 Revised: 18.05.2024 Accepted: 29.05.2024

ABSTRACT
With the advent of technical advancements in the field of digitalization there is a lot of data available in the web. But there are also many sensitive data which requires protection from adversaries. Privacy preservation is very crucial to preserve the sensitive data that is stored in the cloud. Cryptographic techniques provide a good mechanism to protect the sensitive data. In this work we discuss about how we use the cryptographic techniques to preserve the privacy of data and we develop a better framework for protecting the sensitive data. In this work we utilized a Diffie Hellman key exchange algorithm for securely communicating between the client and server environment and how this algorithm provides a secure measure is being discussed and it is compared with the RSA algorithm and also a frame work for effective management of the details of the patients and providing proper support to the patients and how there is a tremendous change brought by the ubiquitous computing and how it can be used to enhance the healthcare is also being discussed. There is a comparison of the computation time of the algorithms is being considered and which algorithm is optimal is being suggested.

Keywords: Cryptography, Diffie Hellman key exchange, RSA, Ubiquitous computing

1. INTRODUCTION

Data mining can be viewed as a method for obtaining information from a huge volume of data. Information mining manages the sort of examples that can be mined. Based on the sort of information to be mined, there are two classifications of capacities engaged with Data Mining namely Classification and Prediction, Distinct Function. Cryptography is a procedure of making sure about data and correspondences through utilization of codes with the goal that lone those individuals for whom the data is processed. Accordingly forestalling unapproved admittance to data. The prefix "crypt" signifies "covered up" and postfix graph signifies "composing". In Cryptography the strategies which are used to shield data are gotten from numerical ideas and a bunch of rule-based counts known as calculations to change over messages in manners that make it difficult to interpret it. These calculations are utilized for cryptographic key production, advanced marking, confirmation to ensure information security, navigation on web and to secure secret exchanges, for example we can consider the online transactions. This work discusses the framework of the datamining which is applied to the healthcare domain and how the data is securely shared using the cryptographic algorithms and its performance are being monitored. This research paper is organized as follows the section 2 describes about the recent developments in the research, section 3 describes about the methodology being developed and section 4 describes about the algorithms used and section 5 describes about the performance and section 6 concludes the work

2. RELATED WORKS

Privacy preserving data mining is gaining significant importance in recent period because of the increase in the use of digital means. There were various works which uses cryptographic techniques and some works which doesn't use these techniques and the direction of the research is obtained out of this survey.
Consider a circumstance where various clinical foundations wish to lead a joint investigation for some mutual preferences without revealing inconsequential information. In this circumstance, research concerning indications, finding and medication reliant on various limits is to be coordinated and all the while security of individuals is to be guaranteed. Cryptographic procedures are undeniably inferred for such circumstances where various social occasions collaborate to deal with results or offer non sensitive

mining results and hence avoiding revelation of fragile information. Cryptographic systems find its utility in such circumstances considering two reasons: First, it offers an inside and out described model for security that consolidates procedures for exhibiting and estimating it. Second an immense game plan of cryptographic counts and works to complete security protecting data mining figuring are available in this space. The data may be passed on among different partners vertically or equally. In vertically allocated data among different colleagues, the individual components may have different credits of same course of action of records and in case of on a level plane isolated data, solitary records are spread out over various components, all of which has comparative course of action of attributes. By far most of the security sparing dispersed data mining counts reveal nothing other than the decisive result. Kantarcioglu and Clifton melded cryptographic methodologies to ensure security in association rule mining over equitably distributed data to restrict information shared and at the same time adding close to no overheads to the mining task. Lindell and Pinkas have analyzed how to deliver ID3 decision trees on a level plane allocated. Yang et al. in their research work have analyzed a response for equally separated data where each customer has a private access just to their own record. Vaidya and Clifton were the fundamental who thought how secure connection rule burrowing should be feasible for vertically distributed. Du and Zhan presented a response for creating ID3 on vertically divided examining two events for mining. Vaidya and Clifton developed a Naive Bayes classifier for sparing insurance on vertically allocated. Vaidya and Clifton in proposed a procedure for gathering over vertically allocated data. All of these procedures are almost established on an extraordinary encryption known as Secure Multiparty Computation (SMC) advancement. cryptographic techniques ensure that the changed data is exact and secure anyway this strategy fails to pass on when more than two or three social events are incorporated. Likewise, the data mining results may break the security of individual records. There exist a nice number of courses of action in case of semi-genuine models yet on the off chance that there should be an event of malignant models less assessments have been made. Dwivedi, A. D., Srivastava, G., Dhar, S., and Singh, R. in their work proposed the utilization of blockchain for safely overseeing the medical care information. In any case, the use of blockchain is exorbitant and requires more data transmission and significant expense for calculation Erlingsson, Ú., Pihur, V., and Korolova, in their work built up a framework which has Randomized aggregable security protection is utilized for getting the insights of publicly supporting from customer programming. This philosophy permits a lot of customer information to be concentrated yet without permitting to investigate the subtleties of the individual trees. It gives high utility investigation of the information which is gathered by the client. This strategy gives a decent measure of protection conservation and it is applied to genuine just as engineered information for testing its validity. Lin, J. L., and Liu, J. Y. C in their examination work comprehends about the issue of privacy preserving Association rule mining. They developed a randomization strategy for the exchange for ensuring the protection of the information. They proposed a calculation for by and by acquiring the itemset which happens every now and again from both genuine exchanges. This work likewise gives some improvement in the mined outcomes.Lohiya, S., and Ragha, L in their work randomized the first information andafterward the speculation approach is applied over the information which is randomized. They found that their strategy ensures the information and combines the protection with a decent exactness and there is no much misfortune in the data and makes the information usable. Sharma, S., Chen, K., and Sheth, A. in their work built up a customized medical services data framework for illness observing and they examined their work and checked it how much security is obtained. Zhang, C., Zhu, L., Xu, C., and Lu, R in their exploration work proposed an forecast framework where there is a protection of security where the clinical information will be scrambled and the information is put away in the cloud worker and it is put away for additional handling like the preparation cycle by utilizing the single layer perceptron learning model. Zhou, J., Cao, Z., Dong, X., and Lin, X. in their work proposed a privacy preserving homomorphic and they built up a protection saving capacity relationship coordinating from the clinical content mining and they found that their model accomplished high security and great execution. Zhu, Y., and Liu, L. (2004, August) in their methodology built up a plan for randomization for protection saving in the assessment and they proposed a system for randomization utilizing the blend models. The impact of randomization on information mining is estimated by the measure of data which is lost and the corruption of the exhibition and they measure by re-enactments and showed how the protection is achieved.

3. Privacy preserving datamining in healthcare

various areas viably use information mining. It empowers the retail areas to show client reaction and encourages the financial area to foresee client benefit. It serves numerous comparable areas, for example, fabricating, telecom, medical care, car industry, training, and some more. Information mining holds amazing potential for medical care benefits because of the remarkable development in the quantity of

electronic wellbeing records. Already Doctors hold tolerant data in the paper where the information was very hard to hold. Digitalization and development of new strategies lessen human endeavors and make information effectively assessable. For instance, the PC keeps a gigantic measure of patient information with precision, and it improves the nature of the entire information the executive's framework. This is where information mining has demonstrated to be amazingly helpful. Researchers are using various methodologieslikegroups, characterization, choice trees, neural organizations, and time arrangement to distribute research. Be that as it may, Healthcare has reliably been delayed to join the most recent examination into regular practice. Information mining has been utilized seriously and generally by various businesses. In medical care, information mining is turning out to be more mainstream these days. Information mining applications can inconceivably profit all gatherings who are associated with the medical care industry. For instance, information mining can help the medical care industry in extortion recognition and misuse, client relationship the board, compelling patient consideration, and best practices, reasonable medical services administrations. There is a lot of information created by medical care exchanges are excessively mind boggling and enormous to be prepared and dissected by traditional strategies.
The framework is as below

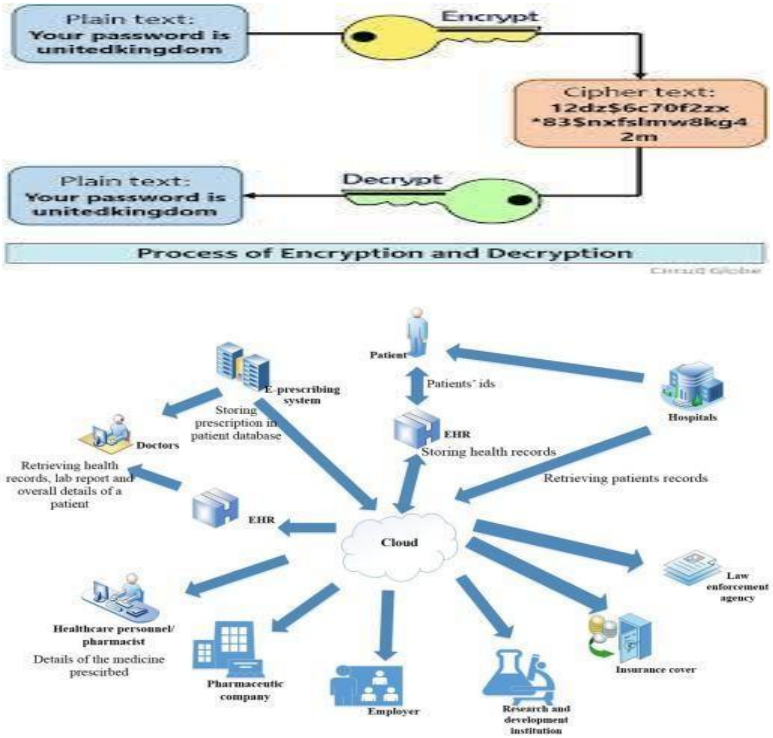

Figure 1. Framework for Cryptography based privacy preservation in health care domain

Cryptographic algorithms for secure communication
Cryptography is concerned with process of conversion of a plain text into a text which is unintelligible to humans. It is a phenomenon of storing and transmission of data and only the persons intended to obtain the data alone can obtain the data and others cannot access the data.

Key benefits arrived out Of Cryptography are as per the following
a. Confidentiality: Data must be obtained to by the individual for whom it is proposed and no other individual aside from them can get to it.
b. Integrity: Data can't be changed away or progress among sender and expected collector with no expansion to data being distinguished.
c. Non-renouncement: The maker/sender of data can't deny their expectation to send data at later stage.
d. Authentication: The characters of sender and collector are affirmed. Just as objective/root of data is affirmed.

Kinds of Cryptography
There are mainly three cryptography types

1. Symmetric Key Cryptography: It is an encryption framework where the sender and collector of message utilize a solitary regular key to encode and decode messages. Symmetric Key Systems are quicker and more straightforward however the issue is that sender and recipient need to by one way or another trade key in a protected way. The most well-known symmetric key cryptography framework is Data Encryption System (DES).
2. Hash Functions: There is no utilization of any key in this calculation.
3. Asymmetric Key Cryptography: Under this framework a couple of keys is utilized to scramble and decode data. A public key is utilized for encryption and a private key is utilized for decoding.

Diffie Hellman Algorithm
Diffie Hellman calculation is basically a convention that is utilized for Exchange of keys. Utilizing this intuitive convention two gatherings will determine a typical secret key by imparting one another. The security of Diffie-Hellman calculation is predominantly founded on the trouble of registering the discrete logarithms.

Algorithm
The algorithm of Diffie Hellman key exchange is as follows and it is explained with general steps and example as follows Begin
Bob and Alice get public numbers P = 23, G = 9
Bob chose a private key a = 4 and Alice chosen a private key b = 3
Alice and Bob process public keys
Bob: x = (9^4 mod 23) = (6561 mod 23) = 6
Alice: y = (9^3 mod 23) = (729 mod 23) = 16
Alice and Bob exchange their public keys Bob gets public key y =16 and Alice gets public key x = 6
Alice and Bob process symmetric keys Bob: ka = y^a mod p = 65536 mod 23 = 9 Alice: kb = x^b mod p = 216 mod 23 = 9
9 is the secret key which is shared between them.
Stop
This algorithm is compared with the RSA algorithm and the level of security and the time taken to access the records by the patient and the medical team were analyzed

5. RESULTS AND DISCUSSION

The entire work was implemented in java and the datasets for the research work was obtained from MIMIC and the time taken to obtain the electronic health record by the patients and the medical team was calculated and the accuracy of the classification algorithms used were calculated and the values are presented in a table1

Table 1: Performance of the Cryptographic algorithms

Sno	Dataset used	Algorithm used	Key size (in Bits)	Computational Time(m)
1	MIMIC	Diffie Hellman	768	4.6225
2	MIMIC	Diffie Hellman	1024	6.3250
3	MIMIC	Diffie Hellman	2048	10.3876
4	MIMIC	RSA	1024	8.356
5	MIMIC	RSA	2048	14.5276
6	MIMIC	RSA	4096	18.8253

From the above table we can observe that the Diffie Hellman key exchange algorithm can perform well when it is compared with the RSA algorithm and it is inferred form the experimentation study that increase in time is proportional to the key size increase between the client and the server and there are various variations being performed in the size of the keys which are used for the encryption and the same process is applied for the RSA algorithm and it found from the experimental analysis that the Diffie Hellman key exchange algorithm performs better in terms of less computational time and it proves to be

an effective cryptographic algorithm. The usage of cryptographic algorithm is very much essential for ensuring a safe communication between the client and server. Information mining can be considered as a calculation which devours information as the information and produces designs like item sets, rules for grouping, rules for Association. Be that as it may, due to datamining there is a danger to the protection of the information in light of the assortment of information and how safely the information is put away and recovered without losing security. Since a large portion of the associations are relied upon and the Diffie Hellman key exchange algorithm proves effective for protecting the privacy of the data when it is exchanged.

6. CONCLUSION

Thus, a privacy preserving framework is being developed which considers the usage of the Diffie Hellman key exchange algorithm for securely transmitting the data the cloud-based administrations for the capacity of information the undertaking of safeguarding the protection is exceptionally pivotal. Hence this work develops a framework for securing exchanging the electronic health record of patients in a secured fashion and also the computational time is also very less.

REFERENCES

[1] Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (pp. 439-450).

[2] Aggarwal, C. C., & Philip, S. Y. (2004, March). A condensation approach to privacy preserving data mining. In International Conference on Extending Database Technology (pp. 183-199). Springer, Berlin, Heidelberg.

[3] Dwivedi, A. D., Srivastava, G., Dhar, S., & Singh, R. (2019).Adecentralized privacy-preserving healthcareblockchain for IoT. Sensors, 19(2), 326.

[4] Erlingsson, Ú., Pihur, V., &Korolova, A. (2014, November). Rappor: Randomized aggregatable privacy- preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067).

[5] Evfimievski, A., & Grandison, T. (2009). Privacy-preserving data mining. In Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends (pp. 527-536). IGI Global.

[6] Kantarcioglu, M., & Clifton, C. (2004). Privacy-preserving distributed mining of association rules on horizontally partitioned data. IEEE transactions on knowledge and data engineering, 16(9), 1026-1037.

[7] Lin, J. L., & Liu, J. Y. C. (2007, March). Privacy preserving itemset mining through fake transactions. In Proceedings of the 2007 ACM symposium on Applied computing (pp. 375-379).

[8] Liu, K., Giannella, C., &Kargupta, H. (2006, September). An attacker's view of distance preserving maps for privacy preserving data mining. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 297-308). Springer, Berlin, Heidelberg.

[9] Lohiya, S., & Ragha, L. (2012, November). Privacy preserving in data mining using hybrid approach. In 2012 Fourth International Conference on Computational Intelligence and Communication Networks (pp. 743-746). IEEE.

[10] Pika, A., Wynn, M. T., Budiono, S., ter Hofstede, A. H., van der Aalst, W. M., &Reijers, H. A. (2019, September). Towards privacy-preserving process mining in healthcare. In International Conference on Business Process Management (pp. 483-495). Springer, Cham.

[11] Pika, A., Wynn, M. T., Budiono, S., Ter Hofstede, A. H., van der Aalst, W. M., &Reijers, H. A. (2020). Privacy-preserving process mining in healthcare. International journal of environmental research and public health, 17(5), 1612.

[12] Qi, X., Mei, G., Cuomo, S., & Xiao, L. (2020). A network-based method with privacy-preserving for identifying influential providers in large healthcare service systems. Future Generation Computer Systems.

[13] Sharma, S., Chen, K., & Sheth, A. (2018). Toward practical privacy- preserving analytics for IoT and cloud- based healthcare systems. IEEE Internet Computing, 22(2), 42-51.

[14] Vaidya, J., & Clifton, C. (2003, August). Privacy-preserving k-means clustering over vertically partitioned data. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 206-215).

[15] Vaidya, J., & Clifton, C. (2004, April). Privacy preserving naive bayes classifier for vertically partitioned data. In Proceedings of the 2004 SIAM international conference on data mining (pp. 522-526). Society for Industrial and Applied Mathematics.

[16] Verykios, V. S., Bertino, E., Fovino, I. N., Provenza, L. P., Saygin, Y., & Theodoridis, Y. (2004). State-of-

the-art in privacy preserving data mining. ACM Sigmod Record, 33(1), 50-57.
[17] Yu, 16. F., & Ji, Z. (2014). Scalable privacy-preserving data sharing methodology for genome-wide association studies: an application to iDASH healthcare privacy protection challenge. BMC medical informatics and decision making, 14(1).
[18] Zhang, C., Zhu, L., Xu, C., & Lu, R. (2018). PPDP: An efficient and privacy- preserving disease prediction scheme in cloud-based e-Healthcare system. Future Generation Computer Systems, 79, 16-25.
[19] Zhou, J., Cao, Z., Dong, X., & Lin, X. (2015). PPDM: A privacy-preserving protocol for cloud-assisted e-healthcare systems. IEEE Journal of Selected Topics in Signal Processing, 9(7), 1332-1344.