International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

A Privacy Leakage Upper Bound Constraint-Based Approach for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud

A. KANCHANA[1], V. KAVITHA[2], S. CHITRA PRIYAA[3], M. SADHANA[4]

[1]*Assistant Professor, Dept. of CSE, Panimalar Engineering College, Tamil Nadu, India*
[2, 3,4]*UG Student, Dept. of CSE, Panimalar Engineering College, Tamil Nadu, India*

Abstract - *The HEALTH sequences are always encrypted at Cloud1, so Cloud1 cannot access these sequences in clear. The only entity which could decrypt them is Cloud2 which is a trusted entity by the hospital. Cloud1 also does not get any leakage from the queries of clients because he processes the queries in an encrypted Cloud2 is a trusted entity. Cloud2 does not have access to encrypted HEALTH sequences unless he colludes with Cloud1 or a Hospital, whatever the information about the HEALTH is encrypted and stored in the cloud. The algorithm advanced encryption standards is highly practically secured and it is effective in software. It is worth mentioning that our approach is not restricted to a fixed homomorphic encryption technique and therefore, it would be possible to use and inherit the advantages of newly developed ones. In our proposed system, it addresses the problem of sharing person specific genomic sequences without violating the privacy of their data subjects to support large-scale biomedical research projects. The proposed method offers two new operating points in the space-time tradeoff and handles new types of queries that are not supported in earlier work. It may assist the data encryption at the data owners (the hospitals) through pre encrypting a large number of values for the encoding of each letter in the alphabet and transferring them to the data owners. Due to the sensitivity of HEALTH, all these operations have to be performed securely. The goal of securing queries is making both the client and the server ignorant of exactly which sequences match the query but only knowing the aggregated result of the query.*

Keywords: Health Databases, Cloud Security, Secure Outsourcing.

1. INTRODUCTION

There is no universal method to create a protocol for secure multi-party computation and handling aggregate queries on encrypted data is not an exception. Several homomorphic systems only support a subset of mathematical operations, like addition, or exclusive- From a security perspective, only the additive and the multiplicative are classified to be IND-CPA (stands for in distinguish ability under chosen plaintext attack). Partially homomorphic cryptosystems are more desirable from a performance point of view than somewhat homomorphic cryptosystems, which support a limited operation depth. Fully homomorphic systems have a huge cost and cannot be deployed in practice. Sometimes the queries on HEALTH need to take into account various errors such as irrelevant mutations, incomplete specifications and sequencing errors. Therefore, the pattern of the query should be expressed using regular expressions. Many works address practical and privacy-preserving outsourcing of this regular expression type of queries, implemented as oblivious evaluation of finite automata.

- Store vast amount of medical data
- Efficient treatment through the data reference
- Reduce the difficulties to keep the data safe

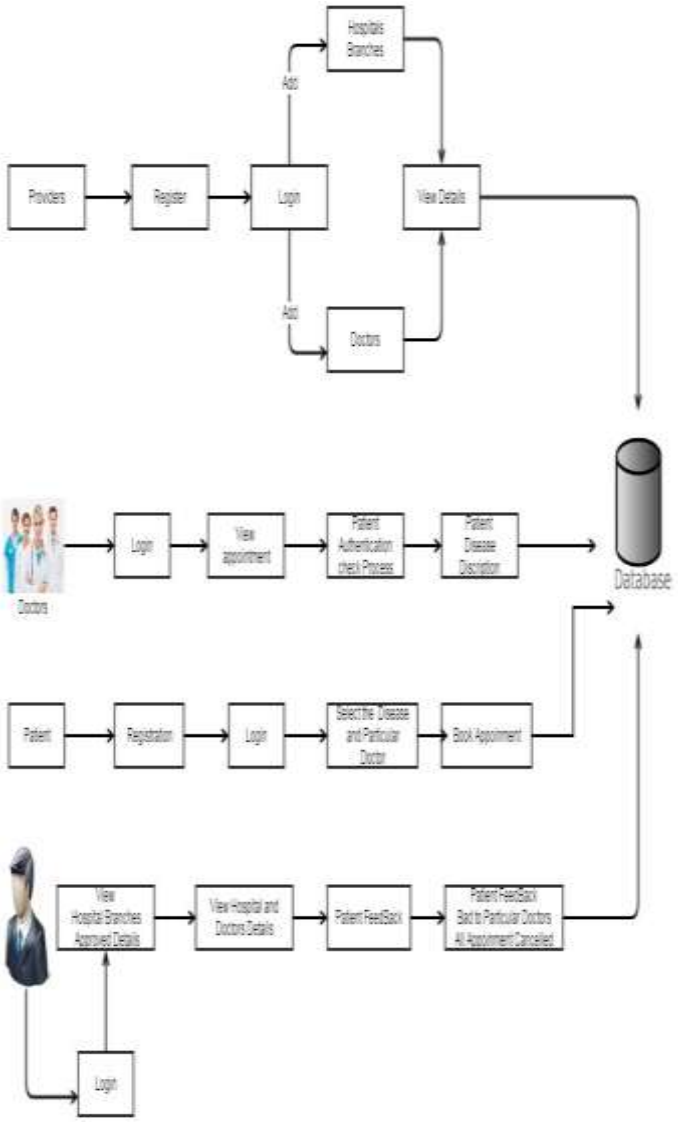

Figure.1 Proposed System Architecture

1.1 Past System Analysis

Human HEALTH data (HEALTH sequences within the 23 chromosome pairs) are private and sensitive personal information. However, such data is critical for conducting biomedical research and studies, for example, diagnosis of pre-disposition to develop a specific disease, drug allergy, or prediction of success rate in response to a specific treatment. Providing a publicly available HEALTH database for fostering research in this field is mainly confronted by privacy concerns.

Today, the abundant computation and storage capacity of cloud services enables practical hosting and sharing of HEALTH databases and efficient processing of genomic sequences, such as performing sequence comparison, exact and approximate sequence search and various tests (diagnosis, identity, ancestry and paternity). What is missing is an efficient security layer that preserves the privacy of individuals‟ records and assigns the burden of query processing to the cloud. Whereas anonymization techniques such as de-identification, data augmentation, or database partitioning solve this problem partially, they are not sufficient because in many cases, re-identification of persons is possible. In past system, there are many disadvantages, they are listed as follows:

(i) In the authors address the longest common subsequence as a private search problem. (ii) In our model, hospitals that have HEALTH sequences do not have the computing and processing capabilities to process researchers‟ requests, so they all store their HEALTH sequences at a server.

(iii) We have presented two new operating points in the space-time tradeoff of the private query problem.

1.2 Proposed System Summary

The proposed system provides a new method that addresses a larger set of problems as well as provides a faster query response time than the technique introduced. Our approach is based on the fact that, given current pricing plans at many cloud services providers, storage is cheaper than computing. Therefore, we favor storage over computing resources to optimize cost. Moreover, from a user experience point of view, response time is the most tangible indicator of performance; hence it is natural to aim at reducing it. Our method enhances the state of the art at both the conceptual level and the implementation level. Moreover, our encoding of the data makes it possible for us to handle a richer set of queries than exact matching between the query and each sequence of the database, including. The proposed approach has lots of advantages, which are summarized as follows:

1. Counting the number of matches between the query symbols and a sequence.

2. Logical OR matches where a query symbol is allowed to match a subset of the alphabet thereby making it possible to handle (as a special case) a “not equal to” requirement for a query symbol.

1. Support for the extended alphabet of nucleotide base codes that encompasses ambiguities in HEALTH sequences.

2. Queries that specify the number of occurrences of each kind of symbol in the specified sequence positions.

1. A threshold query whose answer is „yes‟ if the number of matches exceeds a query specified threshold.

2. SYSTEM APPLICATIONS

2.1 Molecular Biology

Sequencing is used in molecular biology to study genomes and the proteins they encode. Information obtained using sequencing allows researchers to identify changes in genes, associations with diseases and phenotypes, and identify potential drug targets.

2.2 Evolutionary Biology

Since HEALTH is an informative macromolecule in terms of transmission from one generation to another, HEALTH sequencing is used in evolutionary biology to study how different organisms are related and how they evolved.

2.3 Metagenomics

The field of metagenomics involves identification of organisms present in a body of water, sewage, dirt, debris filtered from the air, or swab samples from organisms. Knowing which organisms are present in a particular environment is critical to research in ecology, epidemiology, microbiology, and other fields. Sequencing enables researchers to determine which types of microbes may be present in a micro biome.

2.4 Medicine

Medical technicians may sequence genes (or, theoretically, full genomes) from patients to determine if there is risk of genetic diseases. This is a form of genetic testing, though some genetic tests may not involve HEALTH sequencing.

2.5 Forensics

The HEALTH patterns in fingerprint, saliva, hair follicles, etc. uniquely separate each living organism from one another. Testing HEALTH is a technique which can detect specific genomes in a HEALTH strand to produce a unique and individualized pattern.

3. LITERATURE SURVEY

In the year of 2012, the authors "D. Szajda, M. Pohl, J. Owen, B. Lawson, and V. Richmond" proposed a paper titled "Toward a practical data privacy scheme for a distributed

implementation of the Smith-Waterman genome sequence comparison algorithm", in that they described such as: Volunteer dispersed calculations use save processor cycles of PCs that are associated with the Internet.

The subsequent stages give computational power already accessible just using costly groups or supercomputers. In any case, disseminated calculations running in deceitful situations raise various security concerns, including calculation uprightness and information protection. This paper presents a system for upgrading information security in some disseminated volunteer calculations, giving an essential initial move toward a general information protection answer for these calculations. The technique is utilized to give upgraded information security to the Smith-Waterman neighborhood nucleotide arrangement examination calculation.

Our changed Smith-Waterman calculation gives sensible execution, recognizing most, and by and large all, arrangement matches that show measurably noteworthy closeness as indicated by the unmodified calculation, with sensible levels of false positives. Besides the altered calculation accomplishes a net lessening in execution time and there is no expansion in memory necessities. In particular, our plan speaks to an essential initial move toward giving information protection to a pragmatic and vital certifiable calculation.

In the year of 2010, the authors "M. Blanton and M. Aliasgari" proposed a paper titled "Secure outsourcing of HEALTH searching via finite automata", in that they described such as: this work treats the issue of blunder versatile HEALTH looking by means of absent assessment of limited automata, where a customer has a HEALTH succession, and a specialist co-op has an example that relates to a hereditary test.

Mistake strong looking is accomplished by speaking to the example as a limited robot and assessing it on the HEALTH arrangement, where protection of both the example and the HEALTH succession must be safeguarded. Intuitive answers for this issue as of now exist however can be a weight on the members. Consequently, we propose procedures for secure outsourcing of limited automata assessment to computational servers, which don't take in any data. Our methods are appropriate to limited automata; however the enhancements are customized to HEALTH seeking.

In the year of 2012, the authors "M. Blanton, M. M. J. Atallah, K. B. K. Frikken, and Q. Malluhi" proposed a paper titled "Secure and Efficient Outsourcing of Sequence Comparisons", in that they described such as: we treat the issue of secure outsourcing of grouping examinations by a customer to remote servers, which given two strings λ and μ of particular lengths n and m, comprises of finding a base cost arrangement of additions, erasures, and substitutions (likewise called an alter) that change λ into μ. In our setting a customer claims λ and μ and outsources the calculation to two servers without uncovering to them data about either the information strings or the yield succession.

Our answer is non-intuitive for the customer (who just sends data about the sources of info and gets the yield) and the customer's work is straight in its information/yield. The servers' execution is O calculation (which is ideal) and correspondence, where σ is the letter set size, and the arrangement is intended to work when the servers have just O (σ (m + n)) memory. By using confused circuit assessment novelly, we totally maintain a strategic distance from open key cryptography, which makes our answer especially productive.

4. CONCLUSION

In this paper, we have revisited the challenges of sharing person-specific genomic sequences without violating the privacy of their data subjects in order to support large-scale biomedical research projects. We have used the framework based on additive homomorphism encryption, and two servers: one holding the keys and one storing the encrypted records. The proposed method offers two new operating points in the space-time tradeoff and handles new types of queries that are not supported in earlier work. Furthermore, the method provides support for extended alphabet of nucleotides which is a practical and critical requirement for biomedical researchers. Big data analytics over genetic data is a good future work direction. There are rapid recent advancements that address performance limitations of homomorphic encryption techniques. We hope that these advancements will lead to more practical solutions in the future that can handle larger-scale genetics data. It is worth mentioning that our approach is not restricted to a fixed homomorphic encryption technique and therefore, it would be possible to use and inherit the advantages of newly developed ones.

5. FUTURE ENHANCEMENTS

The need of an online certificate authority (CA) and one unique key encryption for each symmetric key k for data encryption at the portal of authorized physicians made the overhead of the construction grow linearly with size of the group. Furthermore, the anonymity level depends on the size of the anonymity set making the anonymous authentication impractical in specific surroundings where the patients are sparsely distributed.

REFERENCES

[1]M. Kantarcioglu, W. Jiang, Y. Liu, and B. Malin, “A cryptographic approach to securely share and query genomic sequences,” Inf. Technol. Biomed. IEEE Trans., vol. 12, no. 5, pp. 606–617, 2008.

[2]B. Malin and L. Sweeney, “How (not) to protect genomic data privacy in a distributed network: using trail re identification to evaluate and design anonymity protection

systems," J. Biomed. Inform vol. 37, no. 3, pp. 179– 192, 2004.

[3]Z. Lin, A. B. Owen, and R. B. Altman, "Genomic research and human subject privacy," Science (80-.). vol. 305, no. 5681, p. 183, 2004.

[4] A. E. Nergiz, C. Clifton, and Q. M. Malluhi, "Updating outsourced anatomized private databases," in Proceedings of the 16th International Conference on Extending Database Technology, 2013, pp. 179–190.

[5]L. Sweeney, A. Abu, and J. Winn, "Identifying Participants in the Personal Genome Project by Name," Available SSRN 2257732, 2013.

[6]E. Aguiar, Y. Zhang, and M. Blanton, "An Overview of Issues and Recent Developments in Cloud Computing and Storage Security," in High Performance Cloud Auditing and Applications, 2014, pp. 3–33.

[7]P. Bohannon, M. Jakobsson, and S. Srikwan, "Cryptographic Approaches to Privacy in Forensic HEALTH Databases," in Public Key Cryptography, vol. 1751, H. Imai and Y. Zheng, Eds. Springer Berlin Heidelberg, 2000, pp. 373–390.

[8]F. Esponda, E. S. Ackley, P. Helman, H. Jia, and S. Forrest, "Protecting data privacy through hard-to-reverse negative databases," Int. J. Inf. Secure., vol. 6, no. 6, pp. 403–415, 2007.

[9]F. Bruekers, S. Katzenbeisser, K. Kursawe, and P. Tuyls, "Privacy preserving matching of Health profiles," IACR Cryptol. E Print Arch., vol. 2008, p. 203, 2008.

[10]M. J. Atallah and J. Li, "Secure out sourcing of sequence comparisons," Int. J. Inf. Secure vol. 4, no. 4, pp. 277– 287, Mar.2005.

[11]M. Blanton, M. M. J. Atallah, K. B. K. Frikken, and Q. Malluhi, "Secure and Efficient Outsourcing of Sequence Comparisons," Computer Security. 2012, pp. 505–522, 2012.

[12]M. Franklin, M. Gondree, and P. Mohassel, "Communication-efficient private protocols for longest common subsequence," in Topics in Cryptology-CT-RSA 2009, Springer, 2009, pp. 265– 278.

[13]M. Gondree and P. Mohassel, "Longest common subsequence as private search," in Proceedings of the 8th ACM workshop on Privacy in the electronic society, 2009, pp. 81–90.

[14] D. Szajda, M. Pohl, J. Owen, B. Lawson, and V. Richmond, "Toward a practical data privacy scheme for a distributed implementation of the Smith Waterman genome sequence comparison algorithm," in Proceedings of the 12th Annual Network and Distributed System Security Symposium (NDSS 06), 2006.

[15]M. Blanton and M. Aliasgari, "Secure outsourcing of HEALTH searching via finite automata," in Data and Applications Security and Privacy XXIV, Springer, 2010, pp. 49–64.