ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

An Implementation on Securing Aggregate Queries for DNA Databases

Akshay Bharam, Ashutosh Teknur, Prof. Shailaja Jadhav, Shubham Shinde, Sanket Wankhede

Department of Computer Engineering, MMCOE, Pune, India

Department of Computer Engineering, MMCOE, Pune, India

Assistant professor, Department of Computer Engineering, MMCOE, Pune, India

Department of Computer Engineering, MMCOE, Pune, India

Department of Computer Engineering, MMCOE, Pune, India

ABSTRACT: The issue of sharing individual person genomic sequence arrangements without providing the security of their information subjects to support huge scale biomedical research projects. In any case, expands the outcomes in various ways. One change is that our plan is deterministic, with zero likelihood of a wrong answer. This approach is proven effective in maintaining the privacy constraint against an adversarial server. We introduce cryptographic privacy for queries that allow to performing the most common DNA based identity. The capacity is less expensive than calculation in current distributed computing evaluating plans.This point is motivated by the fact that storage is cheaper than computation in current cloud computing pricing plans. 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: (i) calculate the number of matches between query symbols and a sequence; (ii) 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 (e.g., "not a G"); (iii) support for the extended alphabet of nucleotide base codes that encompasses ambiguities in DNA sequences (this happens on the DNA sequence side instead of the query side); (iv) queries that specify the number of occurrences of each kind of symbol in the specified sequence positions (e.g., two 'A' and four 'C' and one 'G' and three 'T', occurring in any order in the query-specified sequence positions); (v) a start query whose answer is 'yes' if the number of matches exceeds a query-specified threshold. (vi) For all query types we can hide the answers from the decrypting server, so that only the client learns the answer. (vii) In all cases, the client deterministically learns only the query's answer, except for query type (v) where we quantify the (very small) statistical leakage to the client of the actual count.

KEYWORDS: DNA Databases, Cloud Security, Secure Outsourcing.

I. INTRODUCTION

Human DNA data (DNA sequences within the 23 chromosome pairs) are private and sensitive personal information [1]. 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 DNA 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 DNA 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. It follows that the DNA data must be protected, not just unlinked from the corresponding persons [2]. Most privacy laws

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

apply to data "relating to an identified or identifiable natural person", data that cannot be directly or indirectly linked to an individual is not restricted.

Background:-

DNA (or RNA) can be harvested directly from subject samples soon after their receipt, or it can be prepared from stored tissues, blood, serum, saliva, cytological preparations, and pathology specimens [3]. Also, because DNA is an informational molecule, the sequence of DNA stored in a computer is subject to the same considerations as DNA itself. (Note that these guidelines may apply to protein sample analysis as well as DNA and RNA research, if such research falls under the purview of the CPHS.).

Motivation:-

Performing different computational tasks on large biological databases is becoming a more common practice in both public and private institutions. The genomic data stored in these databases may be extremely sensitive: an individual's DNA sequence reveals a great deal of information regarding that individual's health, background, and physical appearance. It has been shown that a sequence can be linked to the corresponding individual simply by recognizing the presence of certain markers.

Objective And Scope:-

- We provides a faster query response time than the technique introduced.
- DNA testing can take into consideration helpful researcher to research the specialists' information.
- The main objectives are providing security to the client queries.
- This work treats the issue of secure outsourcing of succession correlations by a computationally restricted customer C to two servers S1 and S2.

Goal:-

We take advantage of GMP modular arithmetic routines to achieve a much faster implementation of the approach in, as well as for the new approaches proposed in this project.

II. REVIEW OF LITERATURE

1) 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.

The current fast development in the accessibility and prevalence of cloud administrations takes into account helpful on request remote stockpiling and calculation. Security and protection concerns, in any case, are among the top hindrances hindering more extensive reception of cloud advancements. That is, notwithstanding the new security dangers that rise with the reception of new cloud innovation, an absence of direct control over one's information or calculation requests new methods for specialist co-op's straightforwardness and responsibility. The objective of this part is to give an expansive diagram of late writing covering different parts of cloud security. We portray as of late found assaults on cloud suppliers and their countermeasures, and also insurance instruments that go for enhancing protection and trustworthiness of customer's information and calculations. The subjects shrouded in this review incorporate verification, virtualization, accessibility, responsibility, and security and uprightness of remote stockpiling and calculation.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

2) **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.**

Human hereditary or genomic inquire about (now and again called DNA examine) includes the investigation of acquired human qualities. (Note: Although frequently utilized conversely, "hereditary" and "genomic" have fairly unique implications: put just, "hereditary" testing inspects particular DNA material that has a known capacity, while "genomic" testing searches for varieties inside extensive fragments of hereditary material, regardless of whether its capacity is known or not.The utilization of human research subjects in both hereditary and genomic inquire about has encouraged huge logical revelations and accomplishments, by empowering researchers to consider human hereditary variety, to distinguish the hereditary underpinnings of sickness, and to look into how genomic data all the more comprehensively can be connected clinically.

3) **M. Blanton, M. M. J. Atallah, K. B. K. Frikken, and Q. Malluhi, "Secure and Efficient Outsourcing of Sequence Comparisons," *Compute. Secure. 2012*, pp. 505–522, 2012.**

In this venture we treat the issue of secure outsourcing of grouping examinations by a customer to remote servers. The arrangement examination issue, given two strings λ and μ of individual lengths n and m, comprises of finding a base cost succession of inclusions, cancellations, and substitutions (likewise called an alter script) that change λ into μ . In our system a customer claims strings λ and μ and outsources the calculation to two remote servers without uncovering to them data about either the information strings or the yield grouping. Our answer is non-intelligent for the customer (who just sends data about the information sources and gets the yield) and the customer's work is straight in its info/yield. The servers' execution is O(σmn) 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 distorted circuit assessment strategies novelly, we totally keep away from the utilization of open key cryptography.

4) **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.**

We outline correspondence proficient two-party and multi-party conventions for the longest basic subsequence (LCS) and related issues. Our conventions accomplish security as for uninvolved foes, under sensible cryptographic suppositions. We advantage from the to some degree astounding transaction of a productive square recovery PIR (Gentry-Ramzan, ICALP 2005) with the great "four Russians" algorithmic outline. This outcome is the primary change to the correspondence unpredictability for this application over non-specific outcomes, (for example, Yao's confused circuit convention) and, in that capacity, is fascinating as a commitment to the hypothesis of correspondence effectiveness for secure two-party and multiparty applications.

5) **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.**

At STOC 2006 and CRYPTO 2007, Beimel et al. presented an arrangement of security necessities for calculations that take care of inquiry issues. In this paper, we consider the longest basic subsequence (LCS) issue as a private hunt issue, where the errand is to discover a string of (or installing relating to) a LCS. We demonstrate that deterministic choice methodologies don't meet the protection ensures considered for private hunt issues and, truth be told, may "release" a measure of data relative to the whole information. We at that point set forth and research a few security structures for the LCS issue and plan new and effective yield examining and identicalness ensuring calculations that provably meet the relating protection ideas. En route, we additionally give yield inspecting and comparability securing calculations for limited customary dialects, which might be of autonomous intrigue.

6) **K. B. Frikken, "Practical private DNA string searching and matching through efficient oblivious automata evaluation," in *Data and Applications Security XXIII*, Springer, 2009, pp. 81–94.**

In it was demonstrated that the capacity to perform neglectful automata assessment was helpful for performing DNA seeking and coordinating. By negligent automata assessment we imply that one member has a limited state machine and the other member has an arrangement, and toward the finish of the convention the succession proprietor learns whether the machine acknowledges the grouping. A convention was given in, yet it required O(n)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

rounds (where n is the quantity of characters in the grouping) and O(m n) measured exponentiations (where m is the quantity of states in the automata). Both of these variables confine the materialness of this approach. In this paper we propose another convention that requires just O(1) adjusts and lessens the quantity of particular exponentiations to O(n) without uncovering any extra data. We have actualized both plans and have indicated tentatively that our plan is a few requests of size quicker than the past plan.

7) **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.**
Numerous fundamental errands in computational science include operations on singular DNA and genomic groupings. These successions, notwithstanding when anonymized, are helpless against re-ID assaults and may uncover exceedingly delicate data about people. To bolster expansive scale biomedical research ventures, associations need to share individual particular genomic groupings without disregarding the protection of their information subjects. We introduce a generally productive, security saving execution of major genomic calculation without uncovering the crude genomic groupings. Associations contribute scrambled genomic succession records into an incorporated vault, where the head can perform inquiries, without unscrambling the information.

8) **Z. Lin, A. B. Owen, and R. B. Altman, "Genomic research and human subject privacy," *Science (80-.).*, vol. 305, no. 5681, p. 183, 2004.**
Human hereditary or genomic inquire about (now and again called DNA examine) includes the investigation of acquired human qualities. (Note: Although frequently utilized conversely, "hereditary" and "genomic" have fairly unique implications: put just, "hereditary" testing inspects particular DNA material that has a known capacity, while "genomic" testing searches for varieties inside extensive fragments of hereditary material, regardless of whether its capacity is known or not.The utilization of human research subjects in both hereditary and genomic inquire about has encouraged huge logical revelations and accomplishments, by empowering researchers to consider human hereditary variety, to distinguish the hereditary underpinnings of sickness, and to look into how genomic data all the more comprehensively can be connected clinically.

9) **P. Bohannon, M. Jakobsson, and S. Srikwan, "Cryptographic Approaches to Privacy in Forensic DNA Databases," in *Public Key Cryptography*, vol. 1751, H. Imai and Y. Zheng, Eds. Springer Berlin Heidelberg, 2000, pp. 373–390.**
The outcome is a database where the perspective of the server fulfils norms, for example, k-namelessness or l-assorted qualities; however the customer can question and adjust the first information. By uncovering information where conceivable, the server can perform esteem included administrations, for example, information investigation impractical with completely encoded information, while as yet being not able disregard security limitations. Refresh is a key test with this model; naıve utilization of inclusion and erasure operations uncovers the genuine information to the server. This paper indicates how information can be securely embedded, erased, and refreshed. The key thoughts are that information is embedded or refreshed into a scrambled impermanent table until enough information is accessible to securely unscramble, and that touchy data of erased tuples is abandoned to guarantee protection of both erased and undeleted people. This approach is demonstrated powerful in keeping up the protection requirement against an antagonistic server.

10) **P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proceedings of the 17th international conference on Theory and application of cryptographic techniques (EUROCRYPT'99)* , 1999, pp. 223–238.**
We propose a useful plan in view of calculating and semantically secure (IND-CPA) in the standard model. The plan is acquired from a change of the supposed RSA-Paillier plot. This alteration is reminiscent of the ones connected by Rabin and Williams to the notable RSA cryptosystem. Because of the unique properties of such plans, we acquire productivity like that of RSA cryptosystem, provably secure encryption (since recouping plaintext from figure content is as hard as considering) and in-noticeability against plaintext assaults. We likewise build another trapdoor change in

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

light of considering, which has enthusiasm all alone. Semantic security of the plan depends on a proper decisional presumption, named as Decisional Small 2e-Residues suspicion. The strength of this supposition is additionally talked about. Contrasted with Okamoto-Uchiyama's plan, the past IND-CPA cryptosystem in the standard model with one-wayness in view of calculating, our plan is radically more productive in encryption, and presents higher transfer speed, accomplishing a similar development figure as Paillier or El Gamal plans. We trust the new plan could be an intriguing beginning stage to create effective IND-CCA conspires in the standard model with one-wayness in light of considering.

III. EXISTING SYSTEM APPROACH

Existing system Disadvantage

- There is no any universal method for handling aggregate queries on encrypted data is not an exception.
- Partially homomorphic cryptosystems are more desirable from a performance point of view than somewhat homomorphic cryptosystems, which support a limited operation depth.
- 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.
- In the context of DNA data protection, related works can be divided into five groups depending on the function or the query being addressed.
- Secure outsourcing finds a real projection in the current business models thanks to the proliferation of cloud-based services.

IV. PROPOSED SYSTEM APPROACH

Proposed system Advantage

- At the conceptual level, we provide a deterministic scheme, with zero probability of a wrong answer (as opposed to a low probability). This gives confidence to the users that they get exact results to all their queries, without impacting security.
- We also provide a new operating point in the space-time tradeoff, by giving a scheme that is twice as fast as theirs but uses twice the storage space. A variant of this scheme uses only 1.5 their storage space at the expense of additional latency.
    - Our method enhances the state of the art at both the conceptual level and the implementation level.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Proposed system architecture

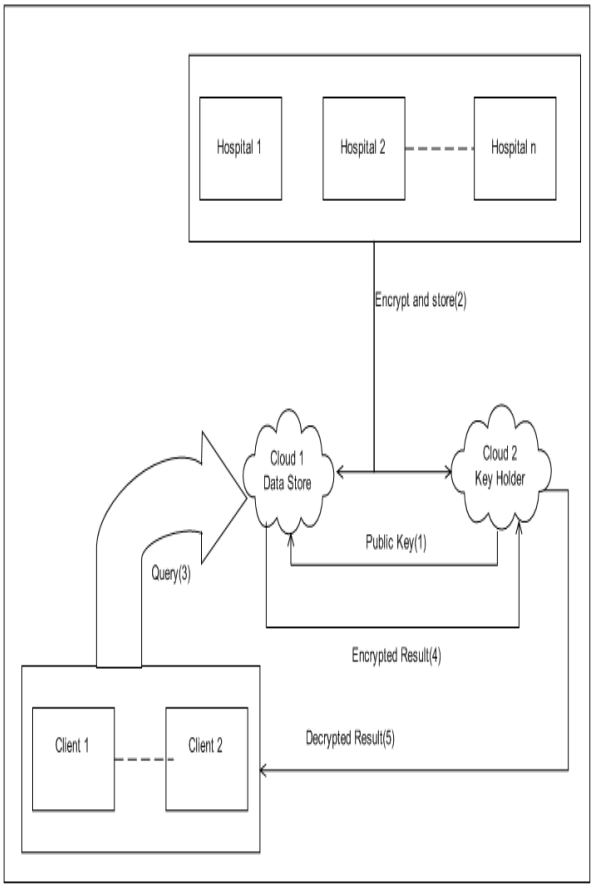

Fig 1: System Architecture

Above system architecture of secure queries of DNA Database which adopts a client-server mode where eachclient is a computer by a client and the servers are data centres or clouds.
Hospitals have requested to the cloud 2 for public key storage the data. Cloud 1 is store the data and cloud 2 is maintaining the key. Hospital store the record on cloud1 with provide the security for encrypt the data. Client have requested to the cloud 1 for DNA report. Cloud 1 check the query and send related record to the cloud 2 then cloud 2 decrypt this record and send to the client.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

V. EXPRIMENTAL SETUP

1. Result Table

2. Let us consider the table 1 for the Encryption and decryption ratio.

	Encryption Ratio	Decryption Ratio
0	0	0
1	0.5	0.3
2	0.73	0.7
3	0.83	0.8
4	0.84	0.81
5	0.86	0.83
6	0.93	0.9
7	1	1

Fig 2. Result Table

3. Result Graph

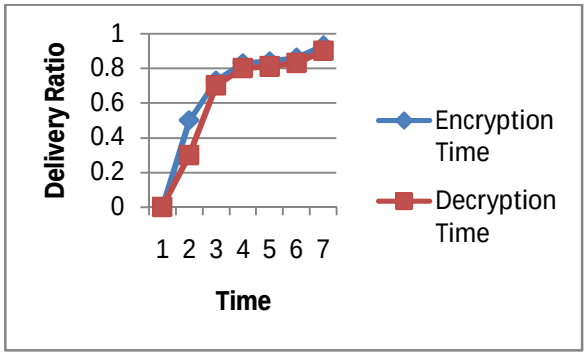

Fig 2. Result Graph

Explanation:
Given graph show the Encryption time and Decryption time.

VI. CONCLUSION

In this project, we have returned to the test of sharing individual particular genomic groupings without damaging the protection of their information subjects keeping in mind the end goal to bolster huge scale biomedical research ventures. We have utilized the system proposed by Kantarcioglu et al. In view of added substance homomorphic encryption, and two servers: one holding the keys and one puts away the encoded records. The proposed technique offers two new working focuses in the space-time tradeoff and handles new sorts of inquiries that are not upheld in prior work. Moreover, the strategy offers help for amplified letter set of nucleotides which is a down to earth and basic necessity for biomedical specialists. Enormous information examination over hereditary information is a decent future work course. There are quick late headways that address execution constraints of homomorphic encryption strategies. We trust that these headways will prompt more commonsense arrangements later on that can deal with bigger scale

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

hereditary qualities information. It merits saying that our approach is not confined to a settled homomorphic encryption procedure and in this manner, it is conceivable to utilize and acquire the benefits of recently created ones.

REFERENCES

[1] 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.
[2] 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.
[3]M. Blanton, M. M. J. Atallah, K. B. K. Frikken, and Q. Malluhi, "Secure and Efficient Outsourcing of Sequence Comparisons," Compute. Secure. 2012, pp. 505–522, 2012.
[4] 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.
[5] 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.
[6] K. B. Frikken, "Practical private DNA string searching and matching through efficient oblivious automata evaluation," in Data and Applications Security XXIII, Springer, 2009, pp. 81–94.
[7]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.
[8] Z. Lin, A. B. Owen, and R. B. Altman, "Genomic research and human subject privacy," Science (80-.)., vol. 305, no. 5681, p. 183, 2004.
[9] P. Bohannon, M. Jakobsson, and S. Srikwan, "Cryptographic Approaches to Privacy in Forensic DNA Databases," in Public Key Cryptography, vol. 1751, H. Imai and Y. Zheng, Eds. Springer Berlin Heidelberg, 2000, pp. 373–390.
[10] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Proceedings of the 17th international conference on Theory and application of cryptographic techniques (EUROCRYPT'99) , 1999, pp. 223–238.