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 1, January 2018

Reliable Accessing of Massive Queries Using Cryptographic Approach

T.Roger Jees Smith[1], D.Mythili[2], M.Padmavathi[3]

Assistant Professor, Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai, Tamil Nadu, India[1]

B.E, Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai, Tamil Nadu, India[2]

B.E, Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai, Tamil Nadu, India[3]

ABSTRACT: The DNA 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 DNA sequences unless he colludes with Cloud1 or a Hospital, whatever the information about the DNA 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 DNA, 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: DNA Databases, Cloud Security, Secure Outsourcing.

I. 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 homomorphism 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 homomorphism cryptosystems are more desirable from a performance point of view than somewhat homomorphism cryptosystems, which support a limited operation depth. Fully homomorphism systems have a huge cost and cannot be deployed in practice. Sometimes the queries on DNA 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.

DOI: 10.15680/IJIRCCE.2018. 0601077

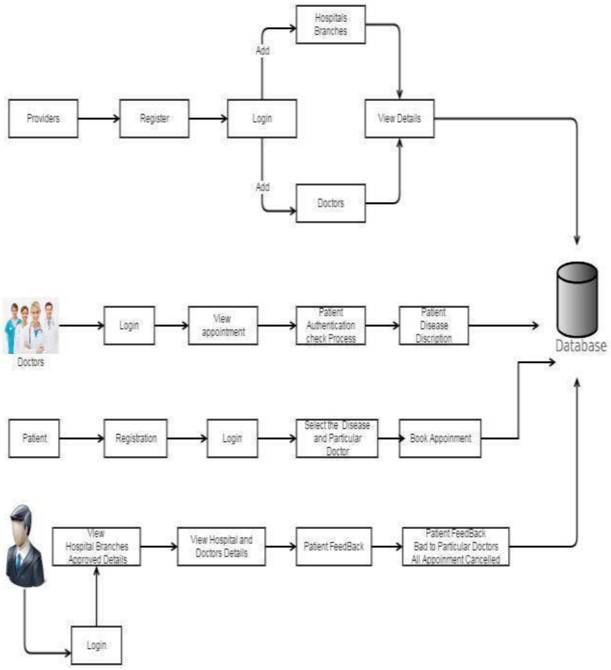

Hospitals Branches
Add
Providers
Register
Login
View Details
Add
Doctors
Doctors
Login
View appointment
Patient Authentication check Process
Patient Disease Discription
Database
Patient
Registration
Login
Select the Disease and Particular Doctor
Book Appoinment
View Hospital Branches Approved Details
View Hospital and Doctors Details
Patient FeedBack
Patient FeedBack Bad to Particular Doctors All Appoinment Cancelled
Login

Fig.1 Proposed System Architecture

II. PAST SYSTEM ANALYSIS

Human DNA data (DNA 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 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. 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 DNA sequences do not have the computing and processing capabilities to process researchers‟ requests, so they all store their DNA sequences at a server.

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

DOI: 10.15680/IJIRCCE.2018. 0601077

III. PROPOSED SYSTEM

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:

(i) Counting the number of matches between the 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.

(iii) Support for the extended alphabet of nucleotide base codes that encompasses ambiguities in DNA sequences.

(iv) Queries that specify the number of occurrences of each kind of symbol in the specified sequence positions.

(v) A threshold query whose answer is „yes‟ if the number of matches exceeds a query-specified threshold.

IV. SYSTEM APPLICATIONS

A. 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.

B. Evolutionary Biology

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

C. 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.

D. 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 DNA sequencing.

E. Forensics

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

V. LITERATURE SURVEY

In the year of 2008, the authors "M. Kantarcioglu, W. Jiang, Y. Liu, and B. Malin" proposed a paper titled "A cryptographic approach to securely share and query genomic sequences", in that they described such as: present a novel cryptographic framework that enables organizations to support genomic data mining without disclosing the raw genomic sequences. Organizations contribute encrypted genomic sequence records into a centralized repository, where the administrator can perform queries, such as frequency counts, without decrypting the data. they evaluate the efficiency of our framework with existing databases of single nucleotide polymorphism (SNP) sequences and demonstrate that the time needed to complete count queries is feasible for real world applications experiments indicate that a count query over 40 SNPs in a database of 5000 records can be completed in approximately 30 min with off-the-shelf technology. To support large-scale biomedical research projects, organizations need to share person-specific genomic sequences without violating the privacy of their data subjects. In the past, organizations protected subjects' identities by removing identifiers, such as name and social security number; however, recent investigations illustrate that identified genomic data can be identified to named individuals using simple automated methods.

In the year of 2004, the authors "B. Malin and L. Sweeney" proposed a paper titled "How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems.", in that they described such as: the erosion of privacy when genomic data, either pseudonymous or data believed to be anonymous, are released into a distributed healthcare environment. The increasing integration of patient-specific genomic data into clinical practice and research raises serious privacy concerns. Various systems have been proposed that protect privacy by removing or encrypting explicitly identifying information, such as name or social security number, into pseudonyms. Though these systems claim to protect identity from being disclosed, they lack formal proofs. Algorithmic proofs of re-identification are developed and we demonstrate, with experiments on real-world data, that susceptibility to re-identification is neither trivial nor the result of bizarre isolated occurrences. We propose that such techniques can be applied as system tests of privacy protection capabilities.

In the year of 2012, the authors "Z. Lin, A. B. Owen, and R. B. Altman" proposed a paper titled "Genomic research and human subject privacy", in that they described such as: Public genetic sequence databases are a critical part of our academic biomedical research infrastructure. However, human genetic data should only be made public if we can adequately protect the privacy of research subjects. Individual genomic sequence data (such as SNPs) are quite "identifiable" using common definitions, while our efforts to understand disease susceptibility or therapeutic opportunity require access to large genomic data sets.

Interest in understanding how genetic variations influence heritable diseases and the response to medical treatments is intense. The academic community relies on the availability of public databases for the distribution of the DNA sequences and their variations. However, like other types of medical information, human genomic data are private, intimate, and sensitive. Genomic data have raised special concerns about discrimination, stigmatization, or loss of insurance or employment for individuals and their relatives. Public dissemination of these data poses nonnutritive privacy challenges. the hurdles may be greater than had been suspected. Suppose that 10% of SNPs are randomly changed in a sequence of DNA, a fairly major obfuscation that would not please many genetics methods.

The following figure illustrates the Home page of the proposed system design.

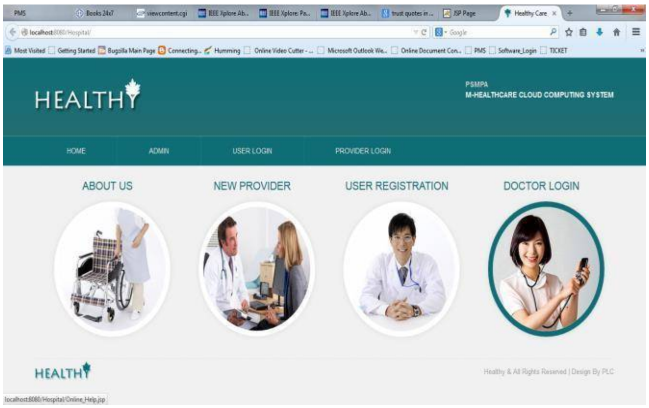
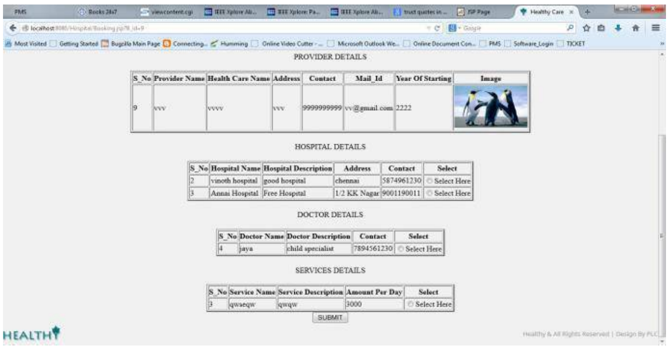


Fig.2 Home Page Design

The following figure illustrates the Provider Registration Details of the proposed system design.



PROVIDER DETAILS

S_No	Provider Name	Health Care Name	Address	Contact	Mail_Id	Year Of Starting	Image
9	vvv	vvvv	vvv	9999999999	vv@gmail.com	2222	

HOSPITAL DETAILS

S_No	Hospital Name	Hospital Description	Address	Contact	Select
2	vinoth hospital	good hospital	chennai	5874961230	○ Select Here
3	Annai Hospital	Free Hospital	1/2 KK Nagar	9001190011	○ Select Here

DOCTOR DETAILS

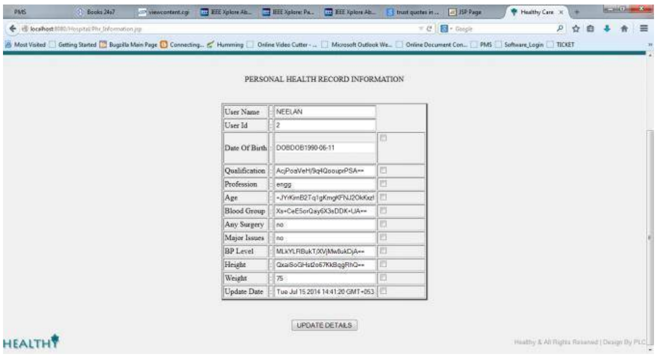
S_No	Doctor Name	Doctor Description	Contact	Select
4	jaya	child specialist	7894561230	○ Select Here

SERVICES DETAILS

S_No	Service Name	Service Description	Amount Per Day	Select
3	qwaeqw	qwqw	3000	○ Select Here

Fig.3 Provider Registration

The following figure illustrates the setting PHR Details Privacy of the proposed system design.

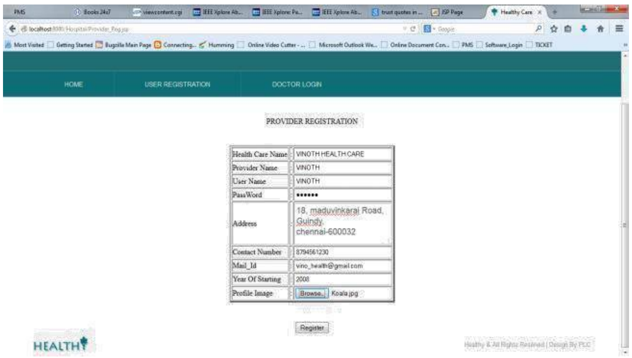


PERSONAL HEALTH RECORD INFORMATION

User Name	:	NEELAN	
User Id	:	2	
Date Of Birth	:	DOBDOB1990-06-11	□
Qualification	:	AcjPoaVeH/9q4QoouprPSA==	□
Profession	:	engg	□
Age	:	+JYrKimB2Tq1gKmgKFNJ2OkKxzl	□
Blood Group	:	Xs+CeE5orQay6X3sDDK+UA==	□
Any Surgery	:	no	□
Major Issues	:	no	□
BP Level	:	MLkYLRBukT/XVjMw8ukDjA==	□
Height	:	QxaiSo/3Hst2o67KkBqgRhQ==	□
Weight	:	75	□
Update Date	:	Tue Jul 15 2014 14:41:20 GMT+053	□

Fig.4 Setting PHR Details Privacy

The following figure illustrates the Services View of the proposed system design.

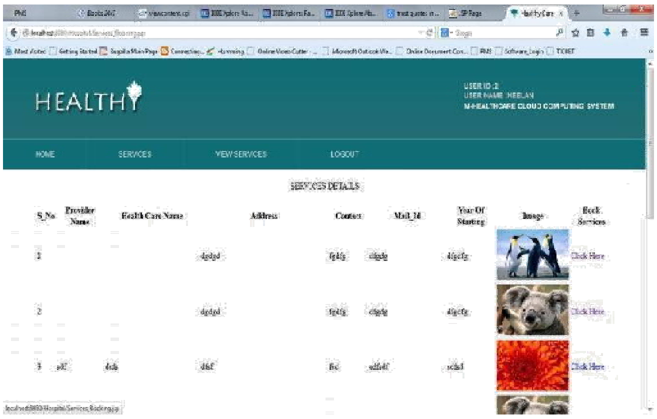


PROVIDER REGISTRATION

Health Care Name	VINOTH HEALTH CARE
Provider Name	VINOTH
User Name	VINOTH
PassWord	••••••
Address	18, maduvinkaraj Road, Guindy, chennai-600032
Contact Number	8794561230
Mail_Id	vino_health@gmail.com
Year Of Starting	2008
Profile Image	Browse... Koala.jpg

Register

Fig.5 Services View



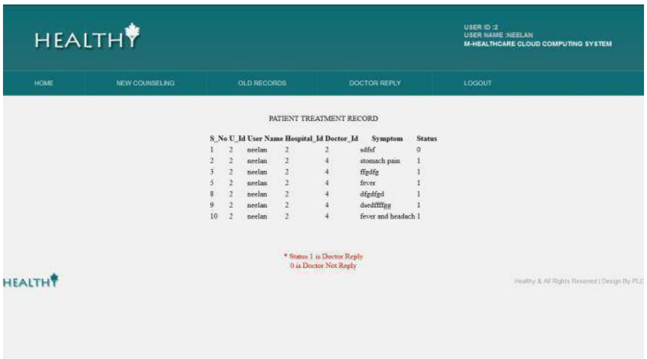
HEALTHY
USER ID :2
USER NAME :NEELAN
M-HEALTHCARE CLOUD COMPUTING SYSTEM
HOME
SERVICES
VIEW SERVICES
LOGOUT
localhost:8080/Hospital/Services_Booking.jsp

SERVICES DETAILS

S_No	Provider Name	Health Care Name	Address	Contact	Mail_Id	Year Of Starting	Image	Book Services
1			dgdgd	fgdfg	dfgdg	dfgdfg		Click Here
2			dgdgd	fgdfg	dfgdg	dfgdfg		Click Here
3	sdf	dsds	dsf	fsd	sdfsdf	sdsd		Click Here

Fig.6 Overall Service Details

The following figure illustrates the Doctor Treatment Details of the proposed system design.



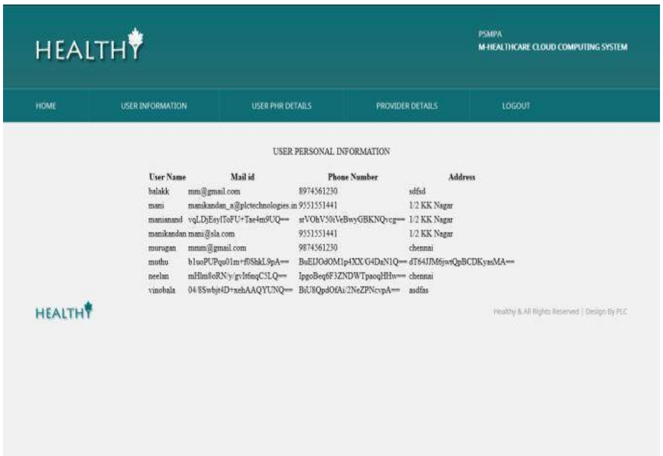
PATIENT TREATMENT RECORD

S_No	U_Id	User Name	Hospital_Id	Doctor_Id	Symptom	Status
1	2	neelan	2	2	sdfsf	0
2	2	neelan	2	4	stomach pain	1
3	2	neelan	2	4	ffgdfg	1
5	2	neelan	2	4	fever	1
8	2	neelan	2	4	dfgdfgd	1
9	2	neelan	2	4	dsedfffffgg	1
10	2	neelan	2	4	fever and headach	1

* Status 1 is Doctor Reply
0 is Doctor Not Reply

Fig.7 Doctor Treatment Details

The following figure illustrates the User Personal Report Details of the proposed system design.



USER PERSONAL INFORMATION

User Name	Mail id	Phone Number	Address
balakk	mm@gmail.com	8974561230	sdfsd
mani	manikandan_a@plctechnologies.in	9551551441	1/2 KK Nagar
manianand	vqLDjEsyIToFU+Tae4m9UQ==	srVOhV50iVeBwyGBKNQvcg==	1/2 KK Nagar
manikandan	mani@sla.com	9551551441	1/2 KK Nagar
murugan	mmm@gmail.com	9874561230	chennai
muthu	b1uoPUPqu01m+f0ShkL9pA==	BuElJOdOM1p4XX/G4DaN1Q==	dT64JJM6jwtQpBCDKyasMA==
neelan	mHlm8oRN/y/gvIt6nqC5LQ==	IpgoBeq6F3ZNDWTpaoqHHw==	chennai
vinobala	04/8Swbjt4D+xehAAQYUNQ==	BiU8QpdOfAi/2NeZPNcvpA==	asdfas

Fig.8 User Personal Report

VI. CONCLUSION

In this paper, we have revisited the challenge 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.

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. 5,606–617, 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 SSRN2257732, 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 DNA 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. Secur., vol. 6, no. 6, pp. 403–415, 2007.

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

[10] M. J. Atallah and J. Li, "Secure outsourcing of sequence comparisons," Int. J.Inf. Secur., 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,"Comput. Secur. 2012, pp. 505–522, 2012.
[12] M. Franklin, M. Gondree, and P.
Mohassel, "Communication-efficient private protocols for longest common subsequence," inTopics in Cryptology-- CT-RSA 2009, Springer, 2009, pp. 265– 278.
[13] M. Gondree and P. Mohassel, "Longest common subsequence as private search," inProceedings 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 DNA searching via finite automata," in Data and Applications Security and Privacy XXIV, Springer, 2010, pp. 49–64.