SECURING AGGREGATE QUERIES FOR DNA DATABASES

B. S Murthy [1], k. Sai Rama Raju,

[1]**Assistant professor(HOD) , PG DEPT,** Dantuluri Narayana Raju College**, Bhim avaram, Andharapradesh**
Email:- Suryanarayanamurthy.b@gmail.com
[2]**PG Student of MSc(cs) PG, Dantuluri** Narayana Raju College**, Bhim avaram, Andharapradesh**
Email:- Saikalidindi44@gmail.com

ABSTRACT

This project is that our scheme is deterministic, with zero probability of a wrong answer (as opposed to a low probability). We also provide a new operating point in the space-time trade off, by offering a scheme that is twice as fast as theirs but uses twice the storage space. 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) 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 (e.g., “not a G”);

1 INTRODUCTION

DNA or Deoxyribonucleic Acid is the medium of longtime storage and transmission of genetic information for all modern living organisms. 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).

Literature Survey

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 deidentified genomic data can be ldquoreidentifiedrdquo to named individuals using simple automated methods. In this paper, we 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. We 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.

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

Several works focus on protecting biometric computations over genomic sequence records in the context of secure multi-party computations (SMC). Secure outsourcing is a particular case of SMC where a client with low resources (energy, memory, CPU) requests the service of one or more outsourcing agents with abundant resources. Secure outsourcing finds a real projection in the current business models thanks to the proliferation of cloud-based services.

Disadvantages of existing system

- 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.

Proposed System & Alogirtham

In this Project, we consider the framework proposed in where the DNA records coming from several hospitals are encrypted and stored at a data storage site, and biomedical researchers are able to submit aggregate counting queries to this site. Counting queries are particularly interesting for statistical analysis.

Advantages:

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.

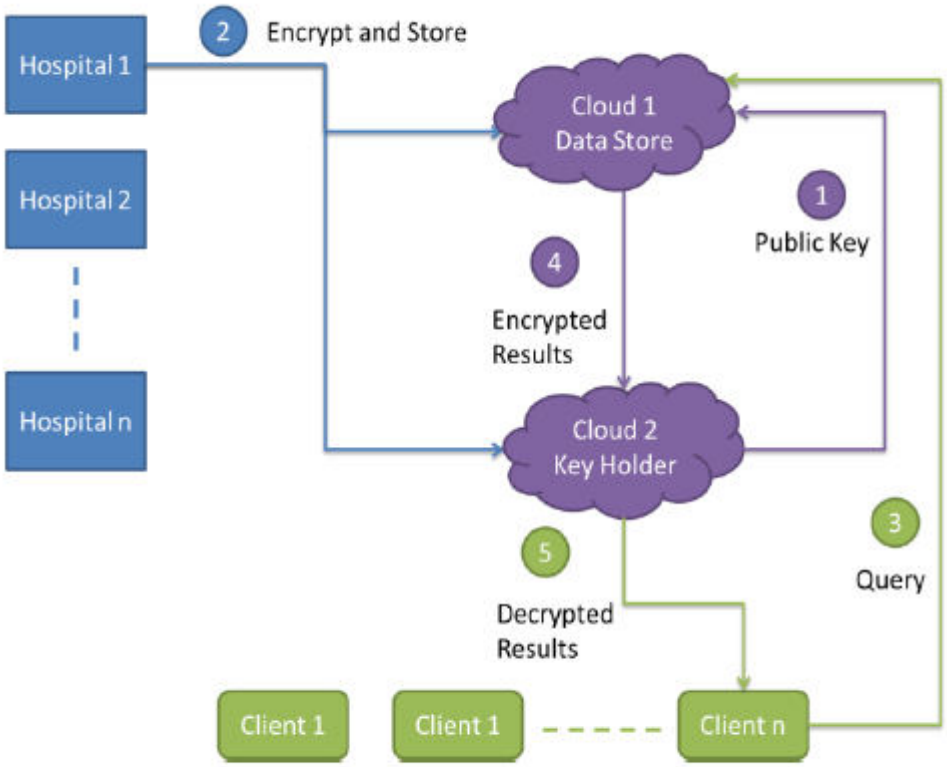

Fig. 1. Framework of secure aggregate queries over encrypted DNA database

Fig 1 System Architecture Diagram

IMPLEMENTATION

Introduction of technologies used

About Java:

Initially the language was called as "oak" but it was renamed as "java" in 1995.The primary motivation of this language was the need for a platform-independent (i.e. architecture neutral) language that could be used to create software to be embedded in various consumer electronic devices.

Applications and applets

An application is a program that runs on our Computer under the operating system of that computer. It is more or less like one creating using C or C++ .Java's ability to create Applets makes it important. An Applet I san application, designed to be transmitted over the Internet and executed by a Java-compatible web browser. An applet I actually a tiny Java program, dynamically downloaded across the network, just like an image. But the difference is, it is an intelligent program, not just a media file. It can be react to the user input and dynamically change.

5 RESULTS AND DISCUSSION

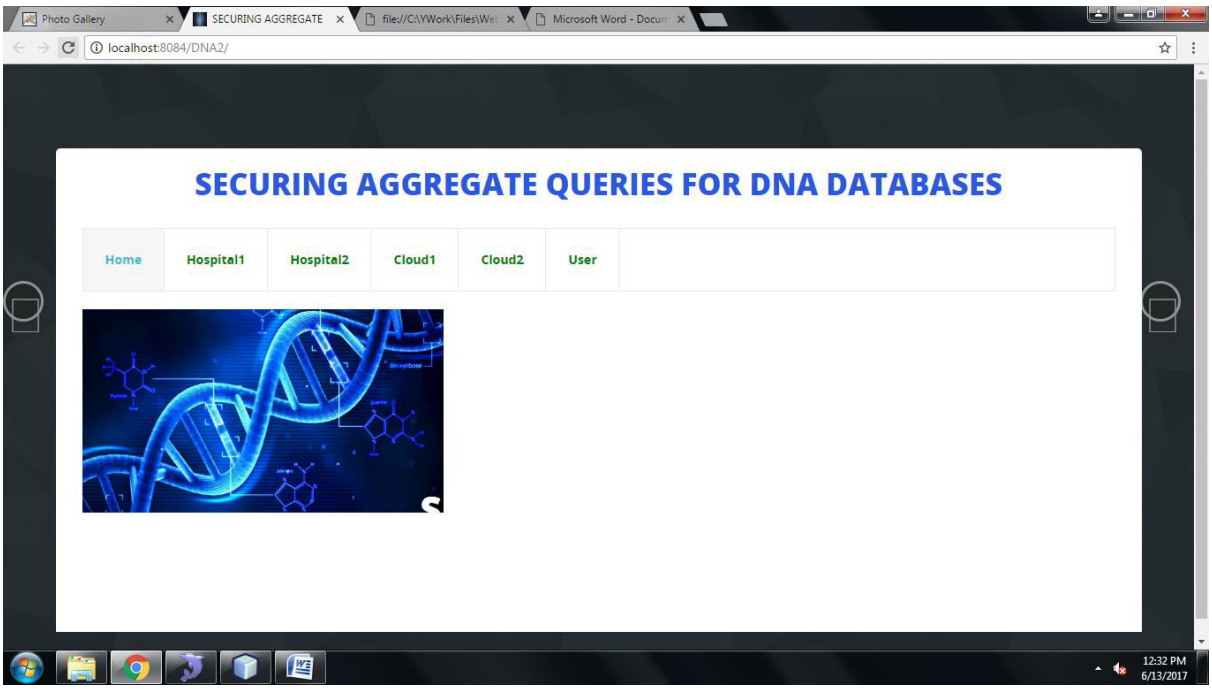

FIG 9.1: Home Page

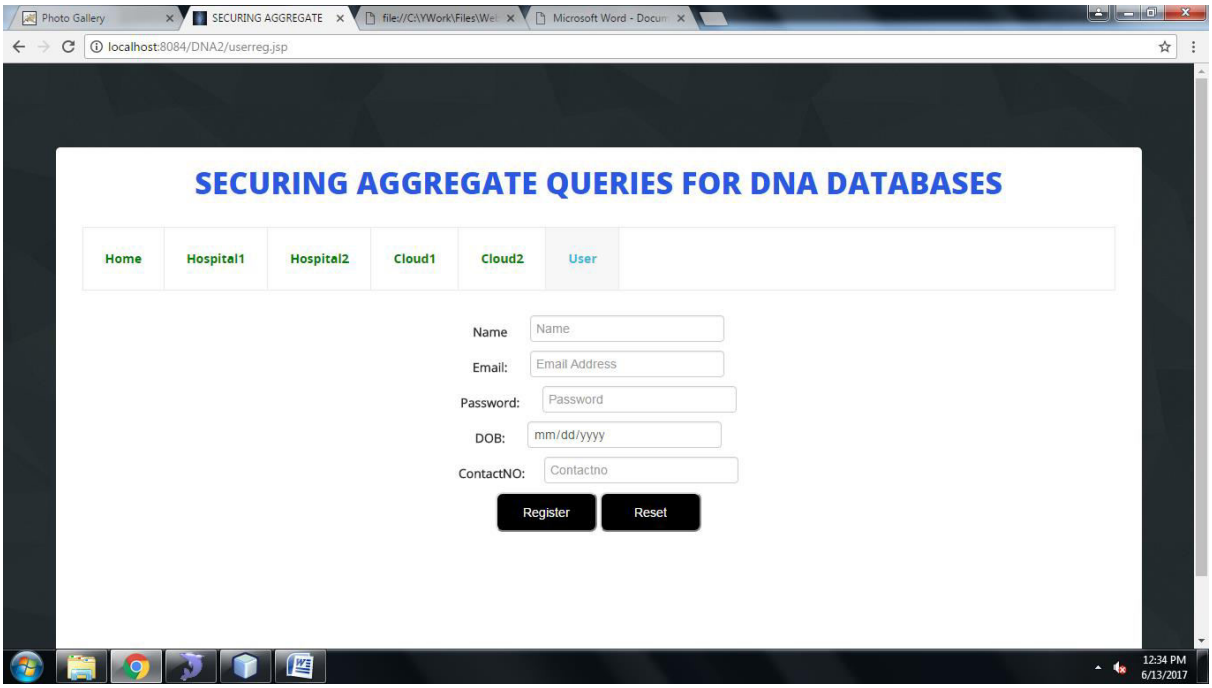

FIG 5.1: User Registration Page

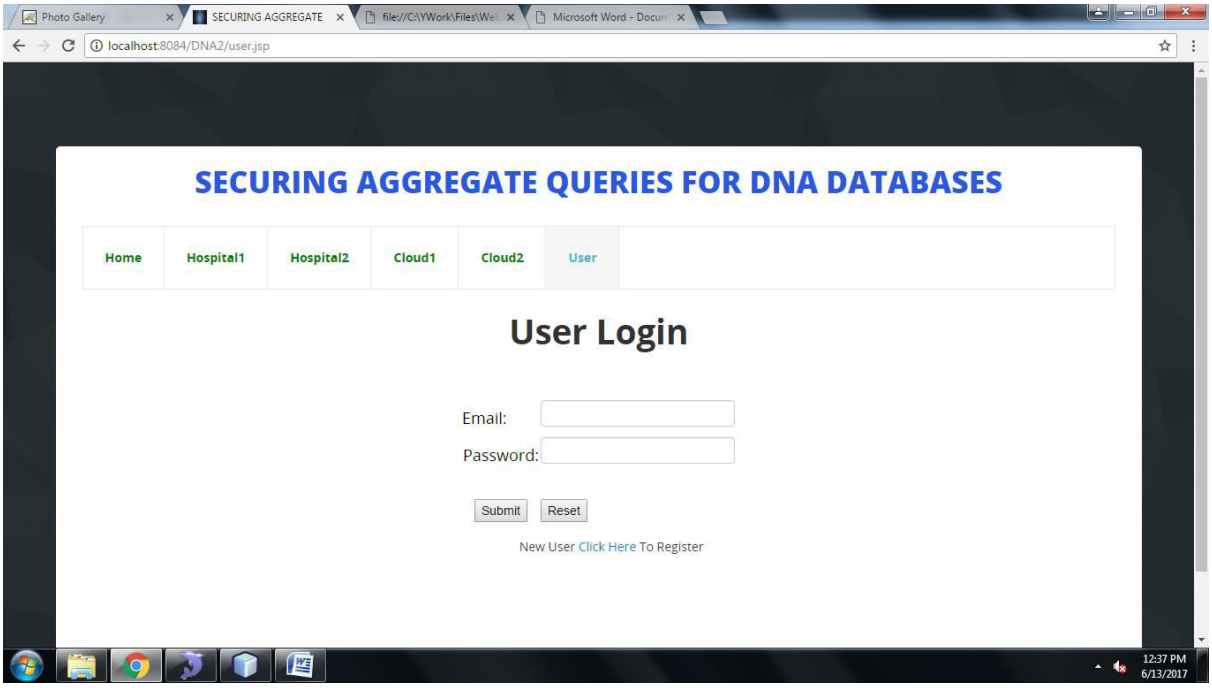

FIG 5.2: User Login Page

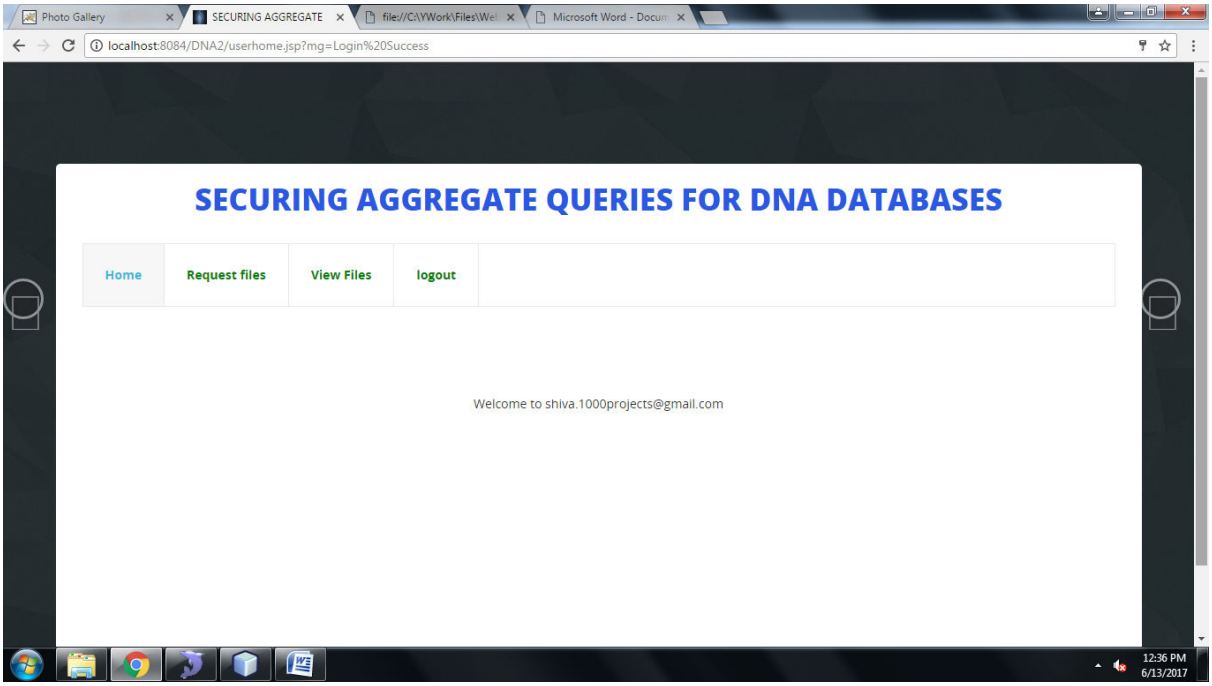

FIG 9.1: User Main Page

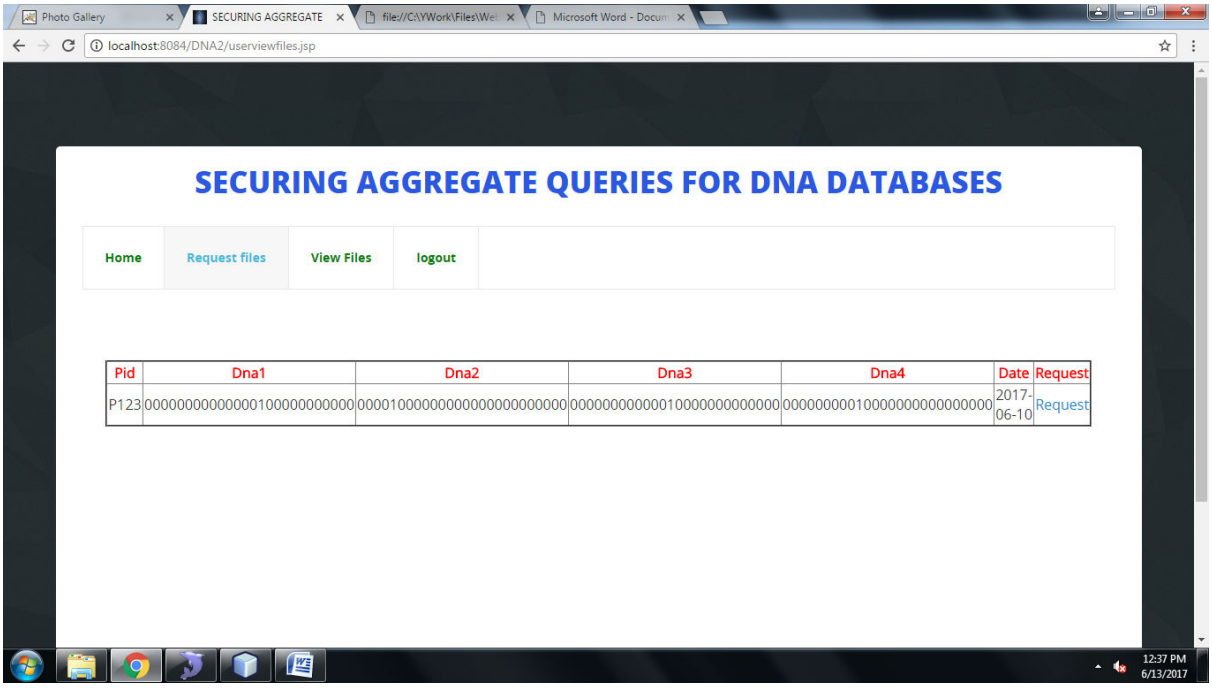

FIG 9.1: User File Request Page

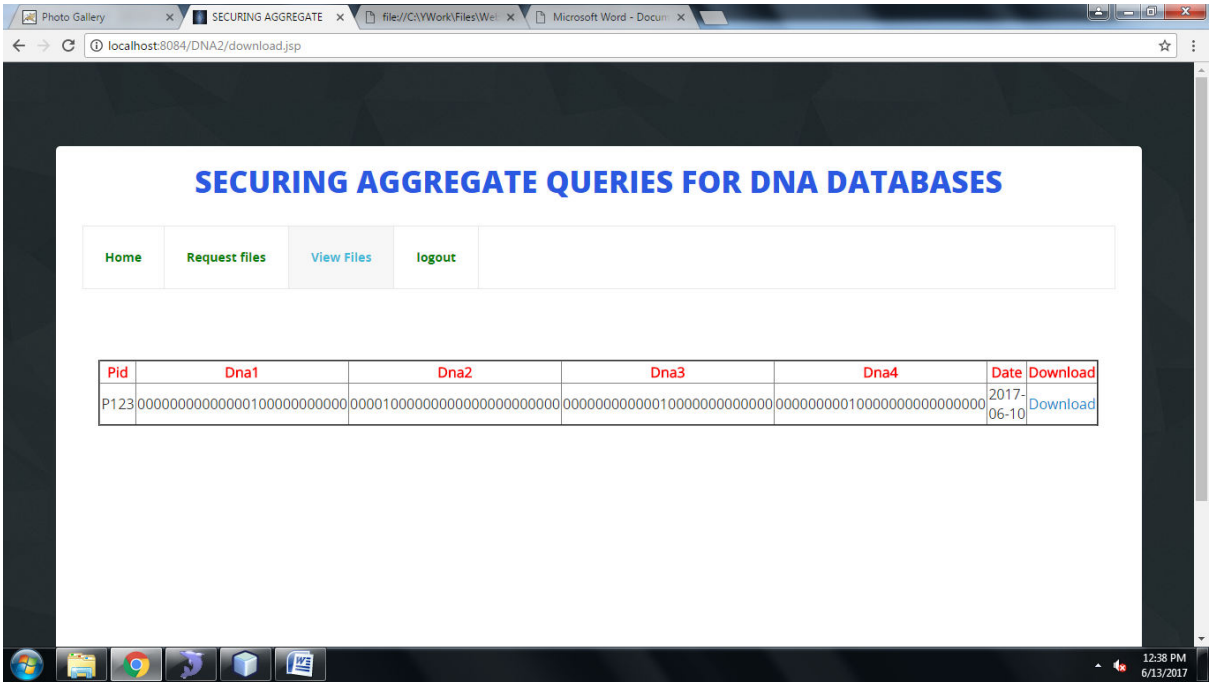

FIG 5.3: User View File Page

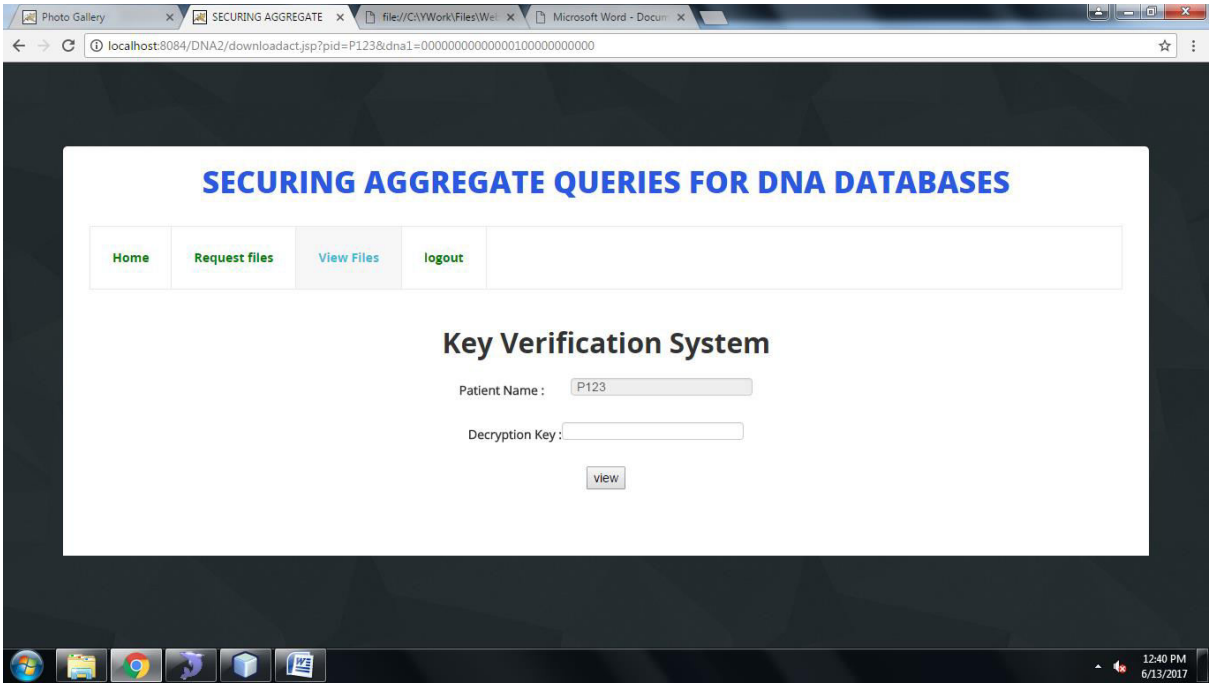

FIG 5.4: Key Verification Page

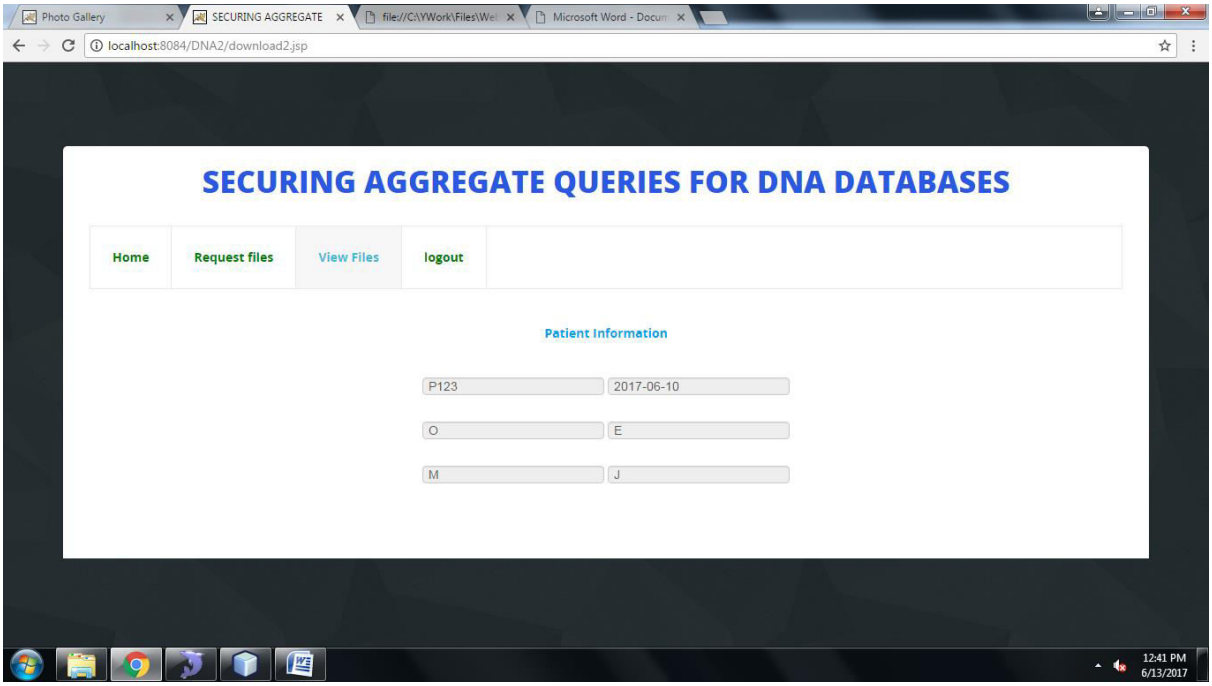

FIG 5.5: Display Patient Information Page

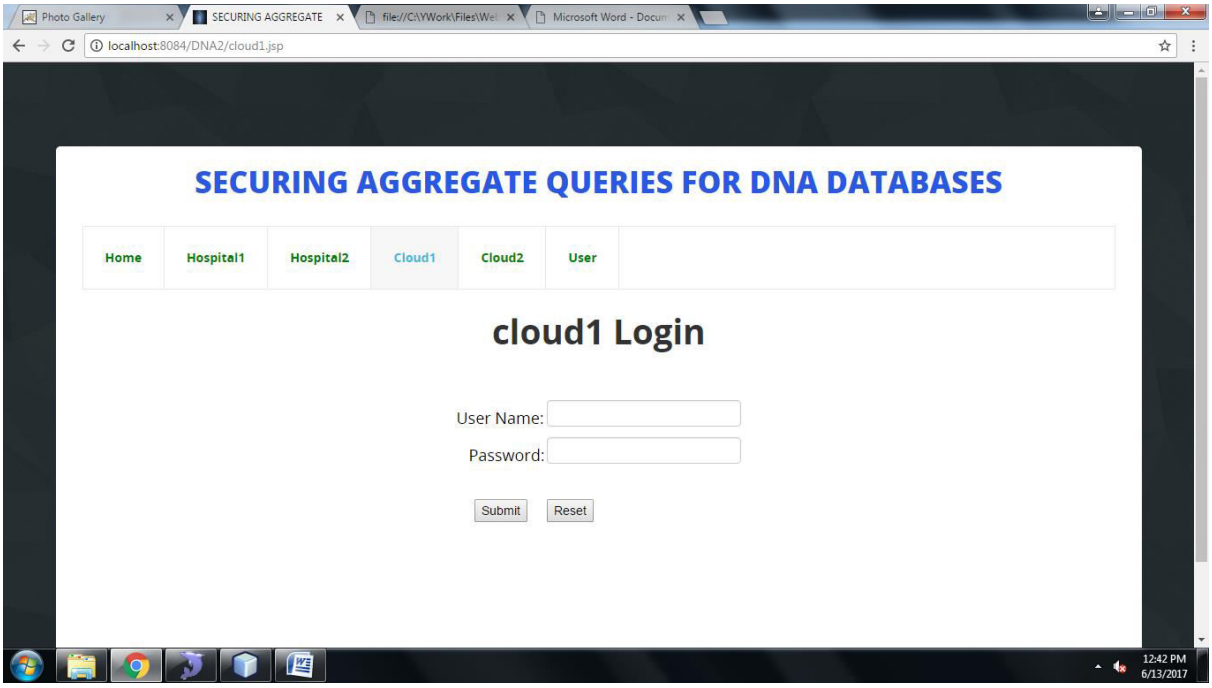

FIG 5.6: Cloud1 Login Page

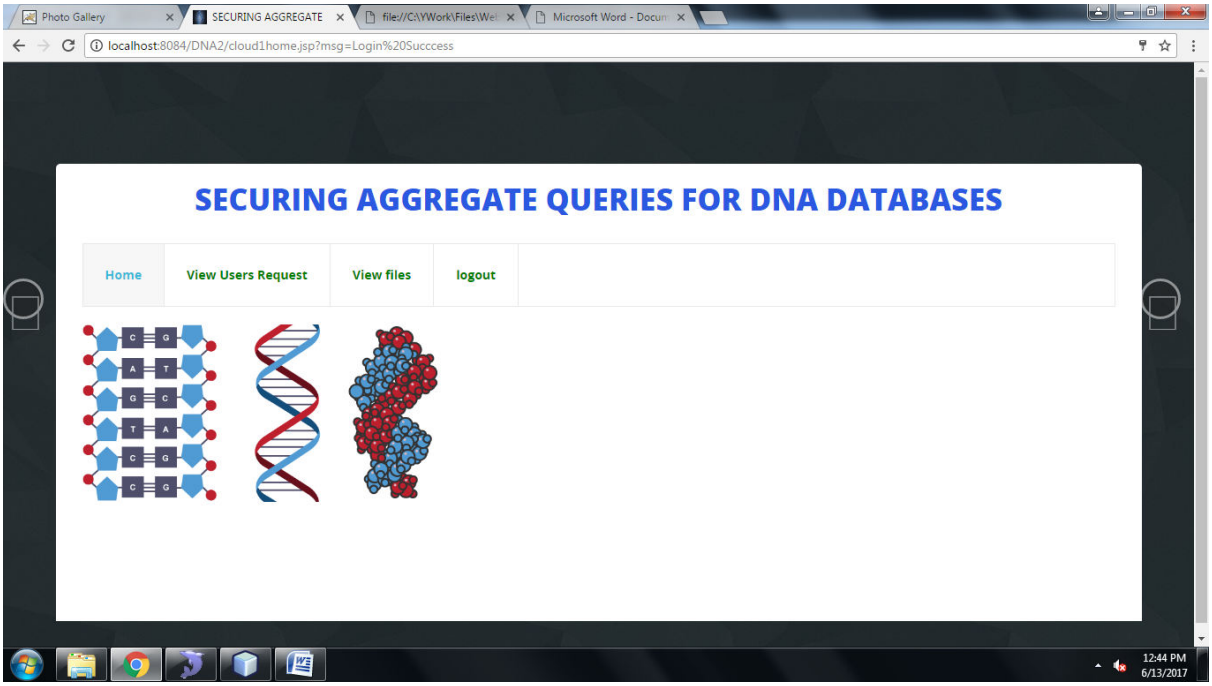

FIG 5.7: Cloud Main Page

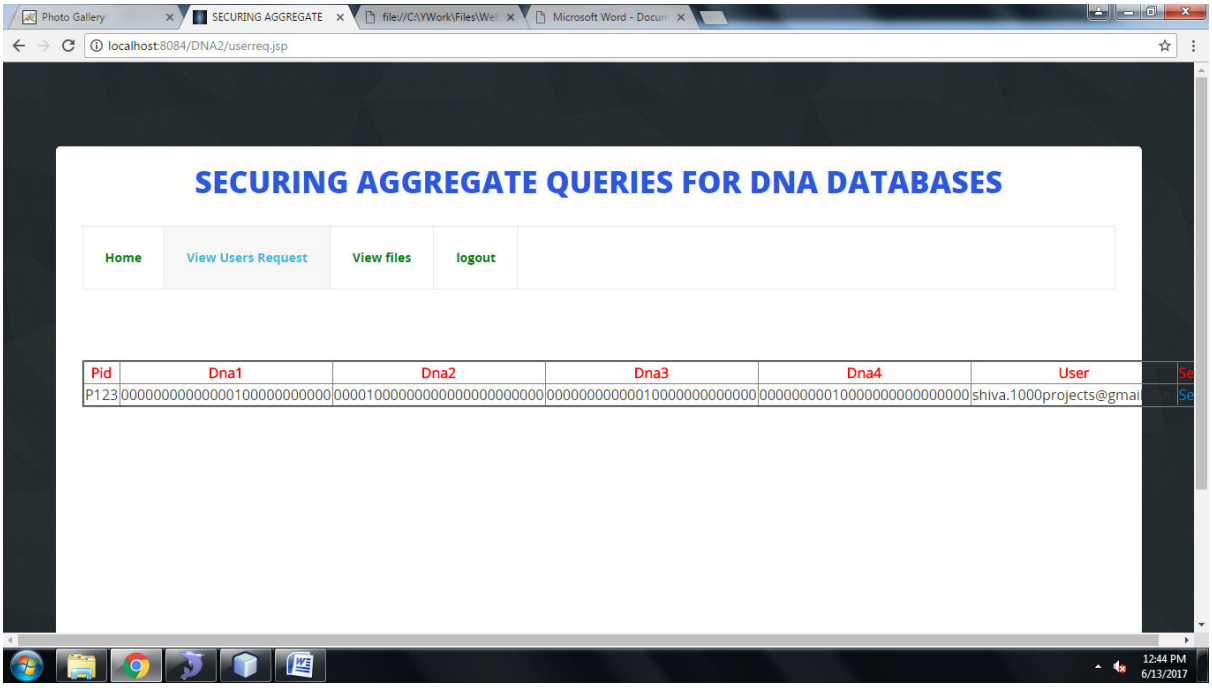

FIG 5.8: View Users Request Page

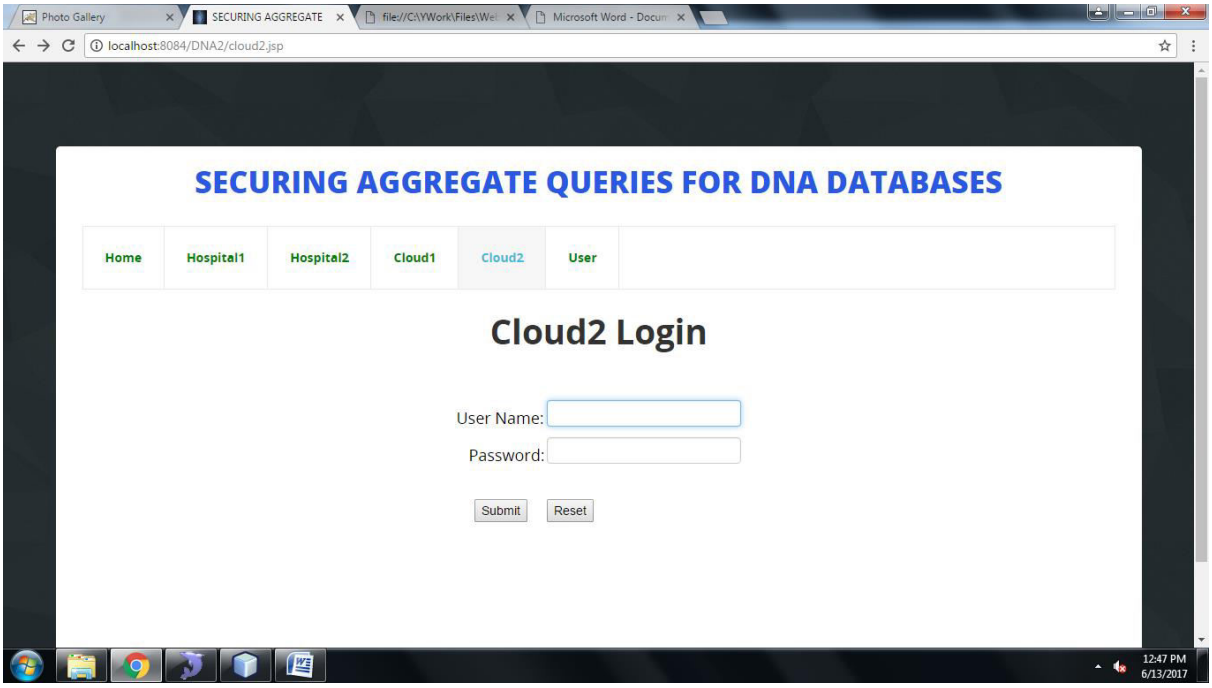

FIG 5.9: Cloud2 Login Page

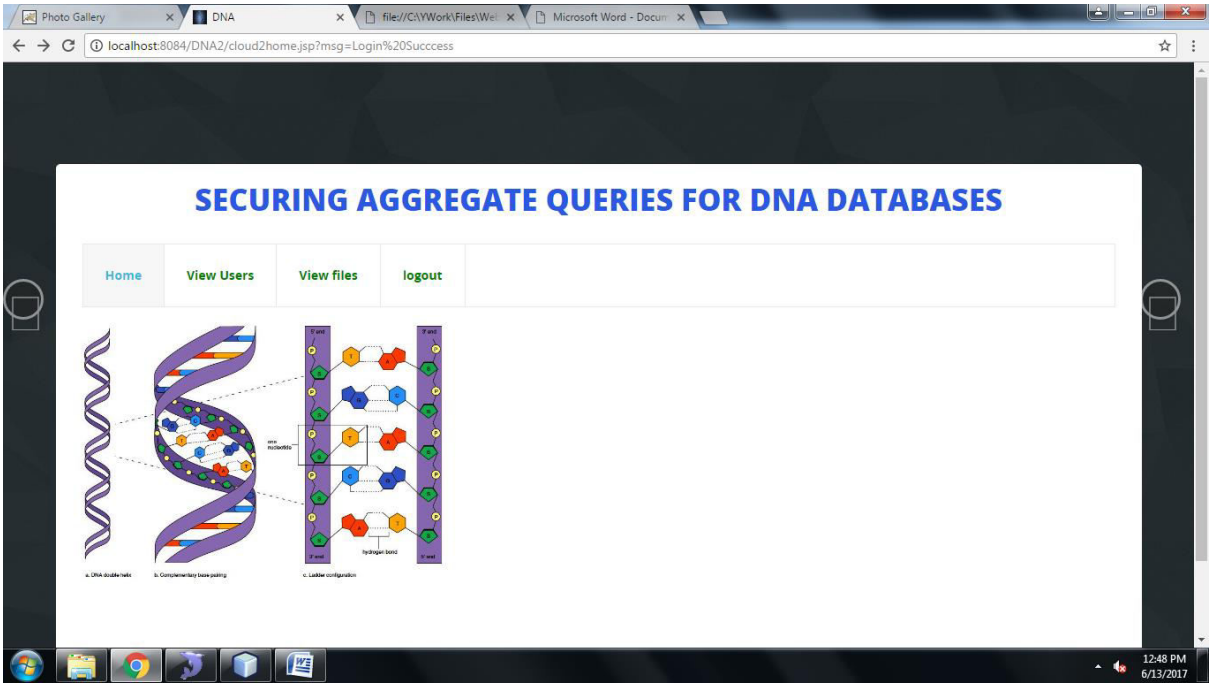

FIG 5.10: Cloud2 Main Page

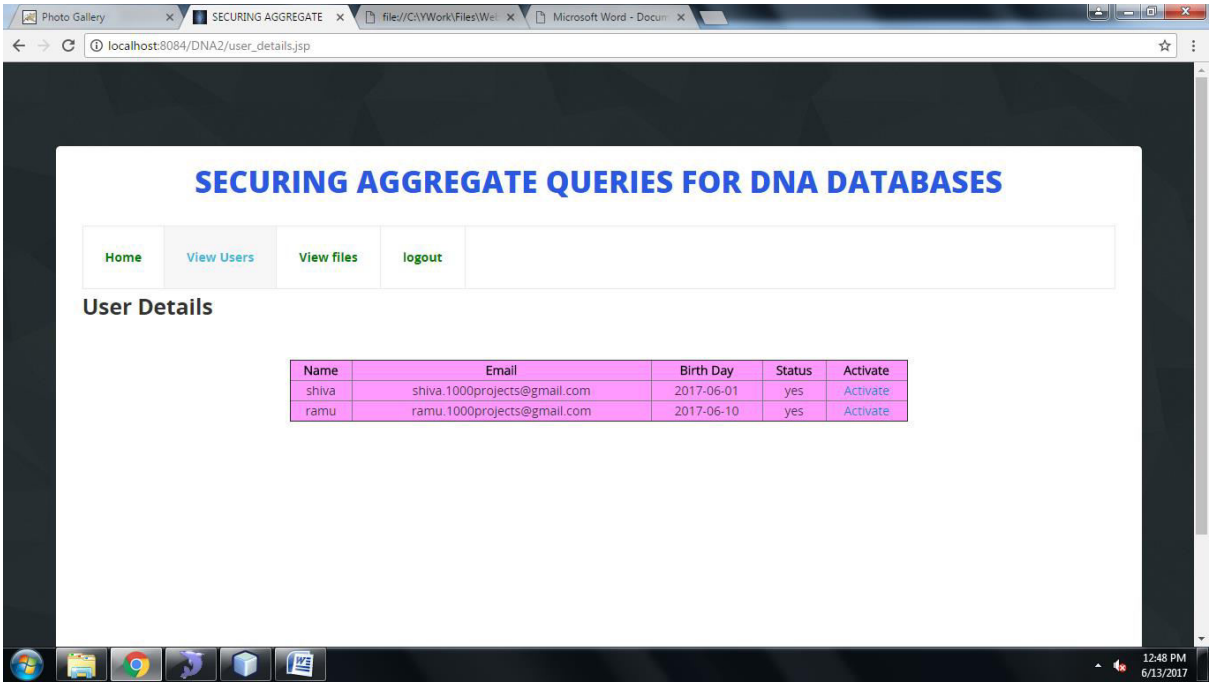

FIG 5.11: Cloud2 View Users Page

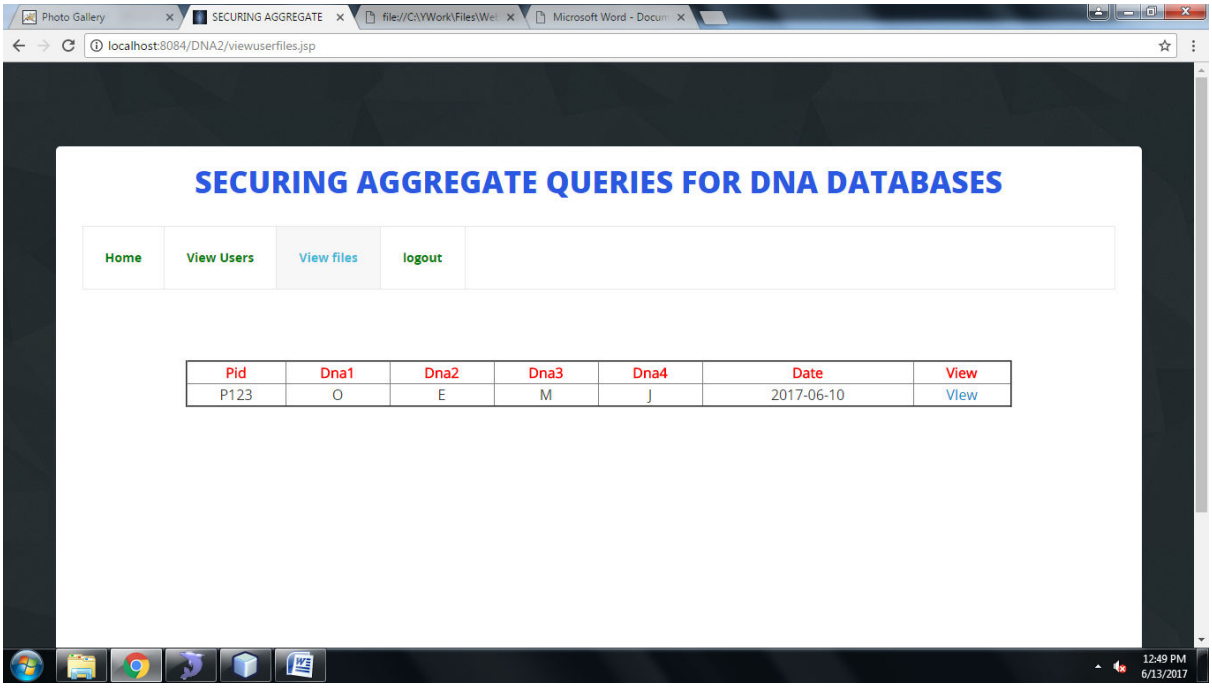

FIG 5.11: Cloud2 View Files Page

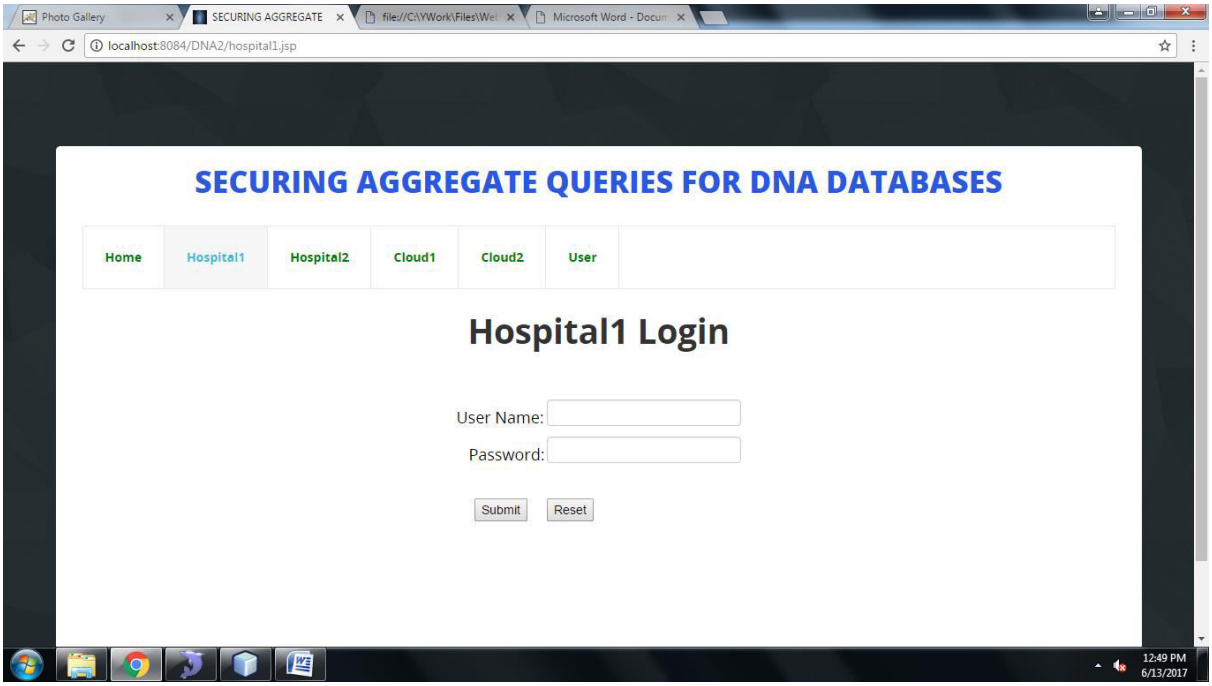

FIG 5.12: Hospital 1 Login Page

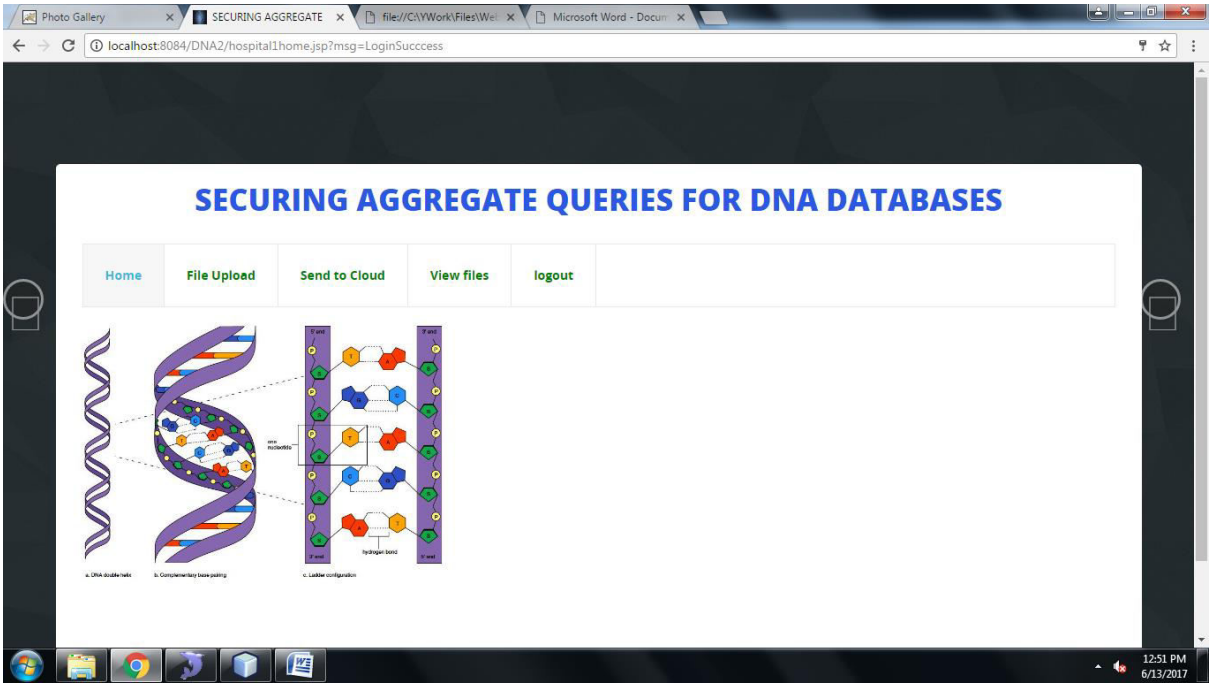

FIG 5.13: Hospital1 Main Page

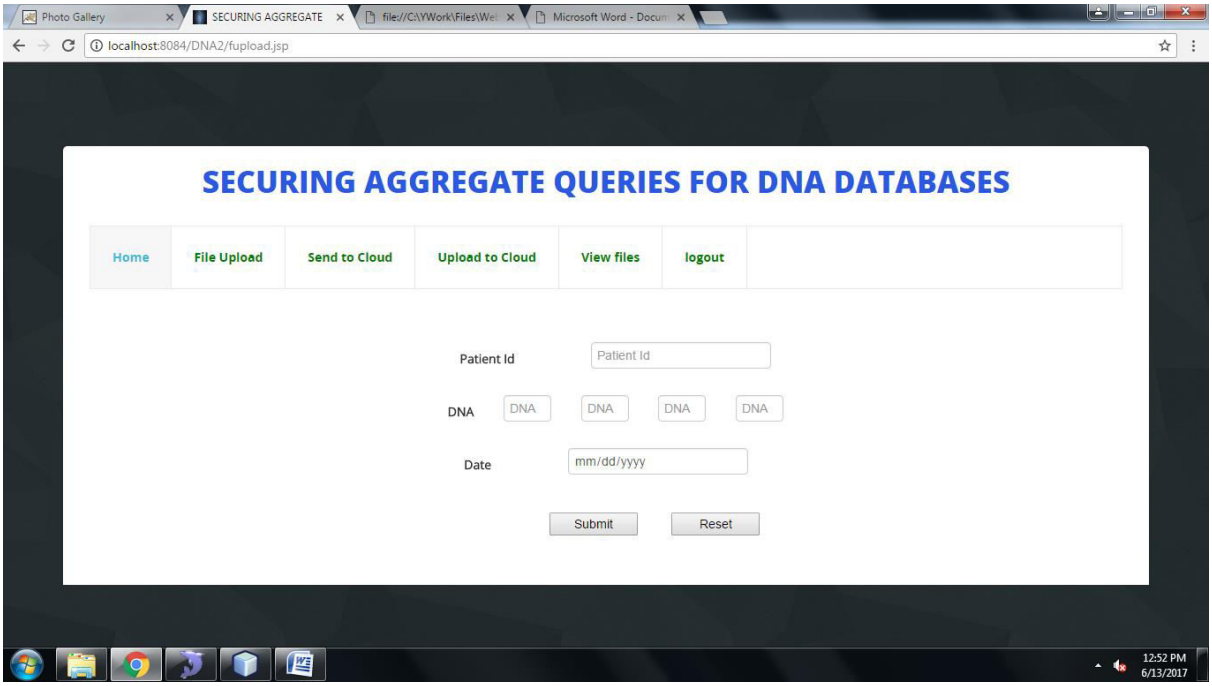

FIG 5.14: Hospital1 File Upload Page

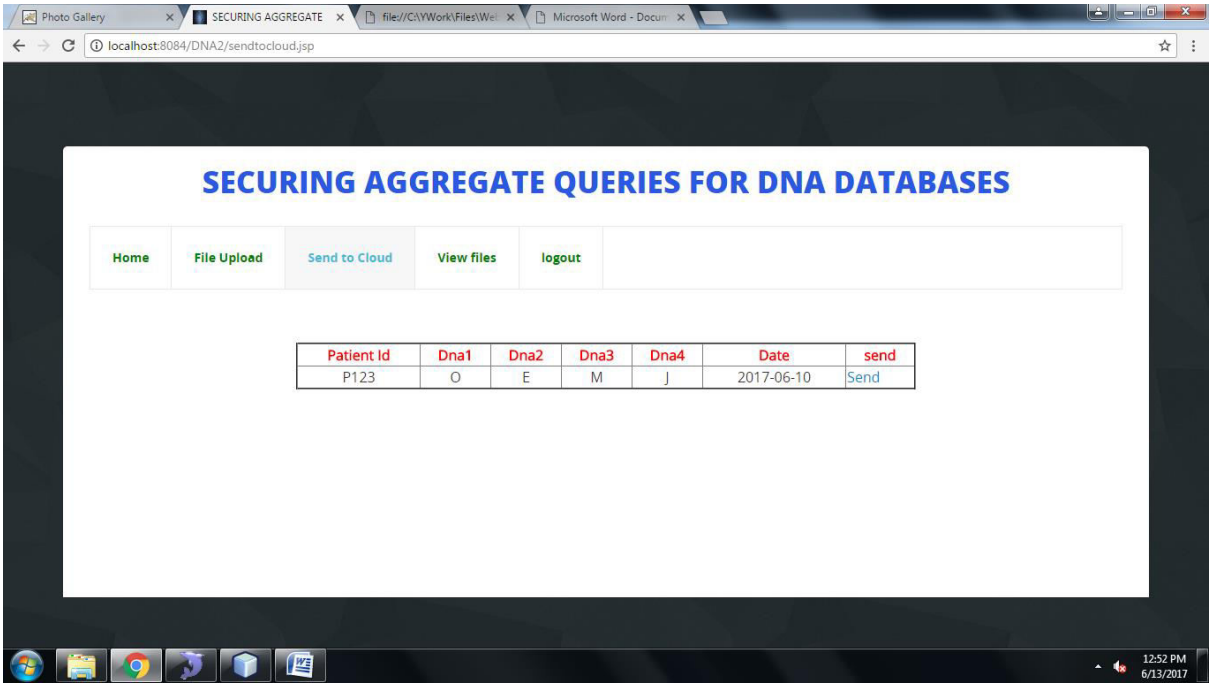

FIG 5.15: Hospital1send To Cloud Page

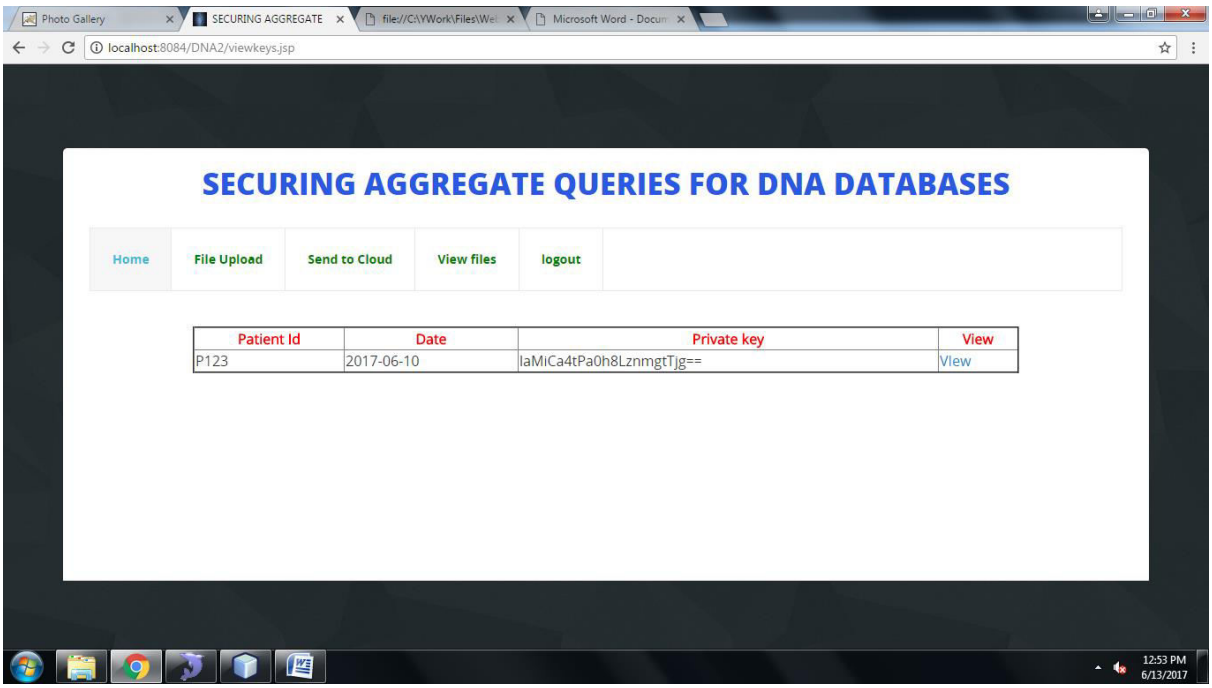

FIG 5.16: Hospital1view Files Page

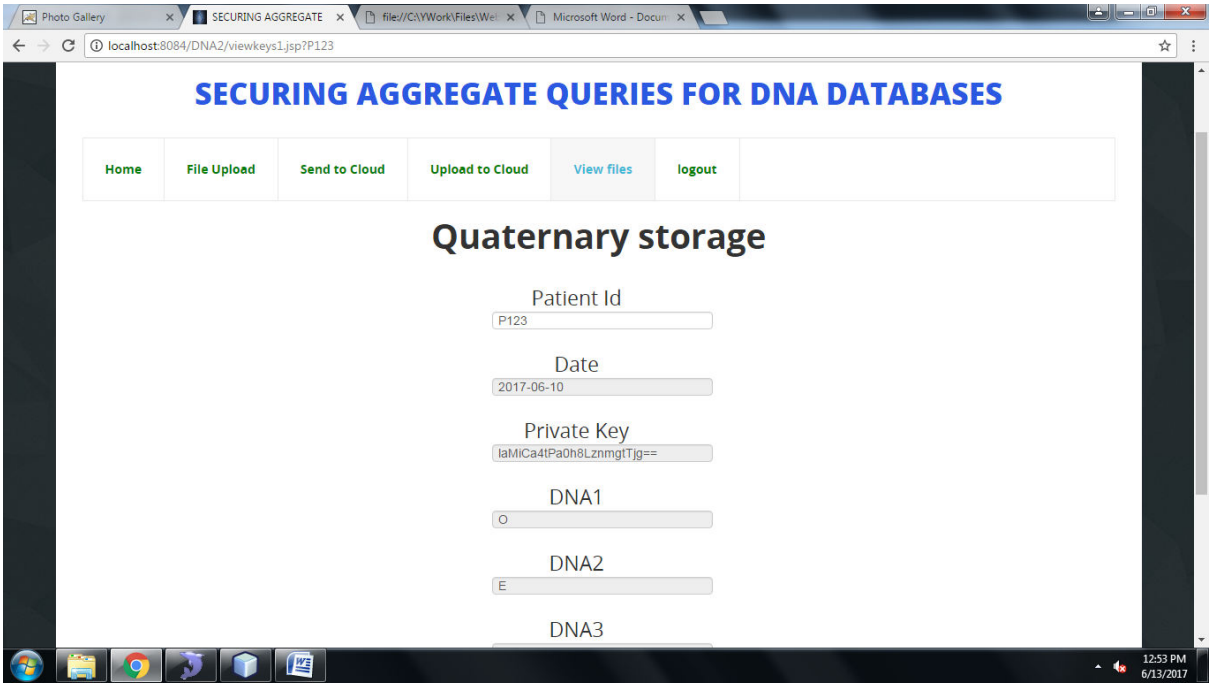

FIG 5.17: View Files Quaternary Storage Page

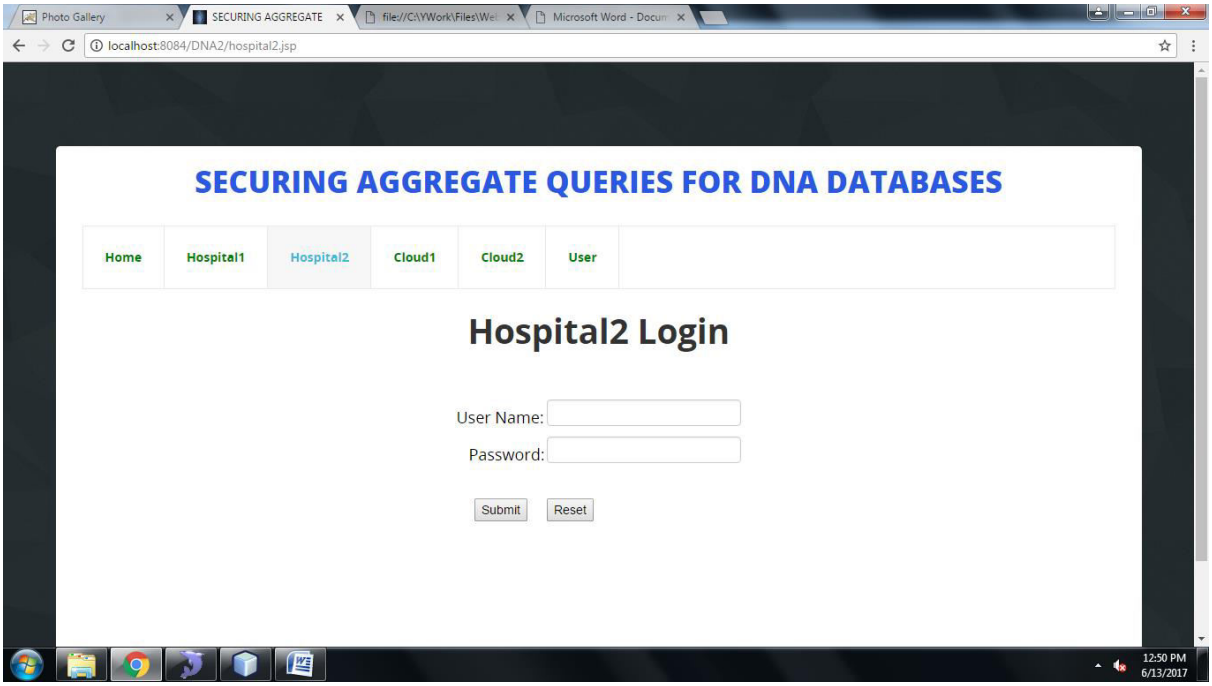

FIG 5.18: Hospital2 Login Page

6. CONCLUSION AND FUTURE WORK

CONCLUSION

In this Project, 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 proposed by Kantarcioglu*et al.* [1] based on additive homophobic 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 tradE off 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.

7. REFRENCES

[1] [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]

[3] [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.

[4]

[5] [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.

[6]

[7]

[8] [4] A. E. Nergiz, C. Clifton, and Q. M. Malluhi, "Updating outsourced anatomized private databases," in *Proceedings of the 16th International Conference* 2168-7161 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

[9] This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2017.2682860, IEEE Transactions on Cloud Computing TCC 11 *on Extending Database Technology*, 2013, pp. 179–190.

[10]

[11] [5] L. Sweeney, A. Abu, and J. Winn, "Identifying Participants in the Personal Genome Project by Name," *Available SSRN 2257732*, 2013.

[12]

[13] [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*