IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Overcoming credibility and liability issues of genomic data through cryptographic solutions

[1]Mohammed Tanzeel A J, [2] Dr. Srividya B V

[1] Dayananda Sagar College of Engineering, Student, Bangalore, India

[2] Dayananda Sagar College of Engineering, Associate Professor, Bangalore, India

***Abstract*— An individual (data owner) sharing his genetic information (or genomic information findings) with a scientist is considered as a situation in this study (service provider). In this case, (I) the professional organization must ensure that the obtained genetic information (or findings) is truthfully associated with the related individual (and processed effectively) ii) Along with his information, the individual must provide an advanced assent indicating whether the service provider is permitted to access and view the genomic information, and (iii) if his information is attempted to be viewed without his assent, the individual must determine which specialist organization is responsible. As a result, two approaches based on AES encryption and one-time password procedures are developed, which tie data about the validity of information to permission and the phenotype of the person. As a result, in order to verify the data, each gathering must also rely on the right consent and phenotype of the person claiming the data.**

Keywords*— *Genomic data, legitimacy, AES Encryption

I. INTRODUCTION

The fast reduction in the expense of entire genome sequencing and genotyping, today, genomic information is broadly utilized in medical services, research, and surprisingly in sporting genomics. Nonetheless, benefits because of this wide utilization of genomic information show up with possible dangers against people's protection. Genomic information of an individual incorporates security delicate information about him like his actual qualities, inclination to sicknesses, and relatives. Hence, it is urgent to ensure security of a person's genomic information while permitting him to use his information to get certain medical care or sporting administrations. Accordingly, there has been huge measure of exploration endeavors on protection saving handling and secure stockpiling of genomic information. Nonetheless, the credibility and liability issues on genomic information have not been broadly viewed as in the writing. Heaps of individuals share their (anonymized) genomic data for research purposes. Such blessings are fundamental for the assessment neighborhood experts need a great deal of genomic data tests to extend the quantifiable power of their examinations. Additionally, some specialist organizations make calculations on genomic information of people and they are just keen on the aftereffects of such calculations (as opposed to the crude genomic information). In any case, service provider (or specialist organizations) need to ensure that possibly (i) a donated genome indeed belongs to a particular individual, (ii) the results of a genetic test is indeed computed from the correct data of the particular individual.. In this work, we study this believability issue and propose cryptographic procedures that would empower a researcher (or a service provider) to confirm the validity of a denoted genome (or a registered hereditary test). Moreover, as an individual gives his genomic information for research (to a specific substance) or goes through a hereditary test from a specialist organization, he might want to ensure that neither his genomic information nor his hereditary test outcomes will be seen by others. Protection spillage happens when genomic information of the individual or his hereditary test outcomes is openly shared by the specialist co-ops that gather such information at the primary spot. In such occurrences, it is critical to comprehend whom to keep responsible because of such a spillage. Subsequently, (i) the individual needs to furnish a computerized assent alongside his information indicating whether the specialist organization is

permitted to access and see and the genomic information and (ii) if his information is attempted to see without the assent, the individual needs to figure out which service provider is liable for this.

II. RELATED WORK

There have been a few deals with security and protection of genomic information. We momentarily sum up the current endeavors on security/protection of genomic information in the accompanying.

Baldi et al. introduced three well known protection delicate genomic applications: (i) paternity tests, (ii) customized medication and (iii) hereditary similarity testing. In contrast to most past work, this paper zeroed in on completely sequenced genomes. This situation presents new difficulties, both as far as protection and computational expense. For every application, this paper proposed a proficient development, in view of notable cryptographic apparatuses: Private Set Convergence (PSI), Private Set Crossing point Cardinality (PSI-CA), and Approved Private Set Crossing point (APSI). Examinations show that these conventions cause online overhead adequately low to be down to earth today. Specifically, these conventions for security protecting paternity testing are essentially more affordable in both calculation and correspondence than earlier work [1].

E. Ayday, J. L. Raisaro, J. Rougemont, and J.- P. Hubaux proposed protection upgrading advancements for clinical trials and customized medication techniques that utilization patients' genomic information. Zeroing in on hereditary infections usceptibility tests, they fostered another design between the patient and the clinical unit and propose a "protection saving sickness vulnerability test" (PDS) by utilizing homomorphic encryption and intermediary re-encryption. Accepting the entire genome sequencing to be finished by a confirmed organization, they proposed to store patients' genomic information encoded by their public keys at a "capacity and handling unit" (SPU). The proposed arrangement allows the clinical unit to recover the encoded genomic information from the SPU and cycle it for clinical trials and customized medication strategies, while saving the security of patients' genomic information. The framework likewise measure the genomic security of a patient (from the clinical unit's perspective) and show how a patient's genomic protection diminishes with the hereditary tests he goes through because of (i) the idea of the hereditary test, and (ii) the qualities of the genomic data[2].

Mustafa Canim; Murat Kantarcioglu; Bradley Malin introduced a safe system by which individual explicit biomedical information, for example, genomic successions, can be joined and questioned utilizing cryptographic equipment. As opposed to formal security models for biomedical information that "bother" or "sum up" records, our strategies guarantee that information are partaken in its most explicit state. Past a hypothetical premise, they tentatively approved that the engineering is considerably more productive when contrasted with arrangements dependent on costly open key encryption conventions. They additionally diminished the trust prerequisites from numerous outsiders to a solitary outsider, which might be more feasible for certifiable applications. They perceived that this system doesn't expressly address protection infringement that can be straightforwardly separated from the inquiry results, yet are certain that this issue can be settled inside the memory of the safe equipment and plan to address it in future work[3]

N. Karvelas, A. Peter, S. Katzenbeisser, E. Tews, and K. Hamacher introduced an answer which synergistically consolidates ORAM strategies with secure two-party calculation arrangements to over protection safeguarding calculations on rethought, completely sequenced DNA giving full inquiry exibility. The beginning stage was putting away the completely sequenced DNA in little squares and re-appropriates it to a distant worker, covering up simultaneously the entrance designs. The fundamental structure block was another ORAM development that permits the information proprietor to be disconnected in any event, when a reshuf-e is performed. Utilizing secure calculation on the recovered DNA hinders, the calculation can be acted in a neglectful way. Decoupling the information recovery and calculation measure, they got full exibility to adjust to future questions [4]

Khalid Alshafee overviewed various cryptographic strategies applied for the reason on some cloud which primarily secures the information just as the applications on the cloud air. Explicit security method doesn't fundamentally utilize single encryption conspire rather it utilizes distinctive encryption plans for the encryption of the information and convert the information to muddled organization and later on decoded utilizing some extraordinary key. Various encryption procedures at this point are accessible for the assurance of the information in the different applications. The cryptographic plans are seen as fundamental for the data mystery that is saved over the cloud. Disseminated registering shares resources like programming, organizations, stage, and establishment for the clients. So using

the cryptographic techniques inside the cloud will ensure the data security and decency which is by and large required in cloud air. In this exploration, diverse encryption methods utilized in the cloud climate are examined to discover which is generally appropriate in what limit [5].

Zhicong Huang; ErmanAyday; Jacques Fellay; Jean-Pierre Hubaux proposed GenoGuard, a cryptographic framework that offers long haul assurance for genomic information against even computationally unbounded enemies. Decoding endeavors against a GenoGuard figure text under a mistaken key yield a genome grouping that shows up measurably conceivable even to a modern foe. To accomplish this assurance, GenoGuard presents a novel DTE plot that productively encodes a genome succession on a ternary tree with affectability to hereditary recombination and transformation, subsequently catching the profoundly no uniform likelihood dispersion and extraordinary construction of genomic information. GenoGuard also gives protection from foes phenotypic side data (actual qualities of casualties). We give a parallelized programming execution of GenoGuard and exhibit its effectiveness and versatility on a bunch of hubs. GenoGuard hence offers an engaging way to deal with the inexorably significant test of assurance of genomic information [6].

III. PRELIMINARIES

This part gives some significant genomics and cryptography foundation data.

- **GENOMICS:** Genomics is the examination of whole genomes of living creatures, and combines segments from inherited characteristics. Genomics uses a blend of recombinant DNA, DNA sequencing methodologies, and bioinformatics to progression, assemble, and research the plan and limit of genomes. It contrasts from 'old style innate characteristics' in that it considers an animal's full enhancement of acquired material, rather than every quality or one quality thing thus. The human genome is encoded in twofold deserted DNA particles containing two vital polymer chains. Each chain involves clear units called nucleotides (A, C, G, T). The human genome involves around three billion latters. Regardless of the way that more than 90% of these are unclear in any 2 individual, there are contrasts between us due to genetic assortments Single nucleotide polymorphism (SNP) is the most broadly perceived DNA assortment in human people. A SNP is a circumstance in the genome holding a nucleotide, which varies between individuals.

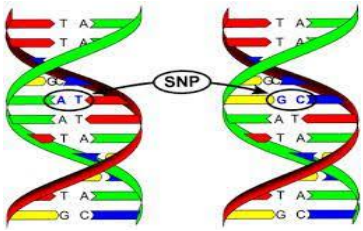

Fig 1: Two sequenced DNA pieces

For instance, the Figure 1 shows the two sequenced DNA sections from two unique people contain a solitary distinctive nucleotide at a specific SNP position.

- **AES ALGORITHM:** Cryptography is the significant interaction, which assumes part in information security. Where clients can store significant and touchy data and afterward move it over unreliable organizations so unapproved individuals can't understand it.

The Advanced encryption standard algorithm (moreover known as the Rijndael calculation) cloud is a symmetrical square cipher calculation that takes plain content in pieces of 128 bits and changes over them to cipher content utilizing keys of 128, 192 and 256 bits. Since the AES algorithm is considered secure, it is within the around the world standard. When it comes to cyber security, AES is one of those acronyms that you simply see popping up all over. That's since it has gotten to be the worldwide standard of encryption and it is utilized to keep a noteworthy sum of our communications secure. The Advanced Encryption Standard (AES) may be a quick and secure frame of encryption keeps prying eyes absent from our information. It is presently days utilized in well-known informing apps like WhatsApp and Flag. AES has an encryption key length of 128, 192 and 256 bits, which can scramble and decode information in squares of 128 bits.

The longest AES encryption key length is otherwise called military-grade encryption. While it is the most secure and the vast majority of the antivirus programming and secret phrase overseeing arrangements utilize 256 pieces, you ought to be completely fine utilizing any of the other two – except if, similar to the US Public safety Organization; you dread future assaults from quantum PCs. It can oppose most if not completely known assaults. AES is additionally quick and minimized on a wide scope of stages. Utilizing ideal execution you can accomplish 1.3 cycles/byte on a solitary center Intel® Core™ i7 Processor Outrageous Release, i7-980X for AES-128 in equal modes.

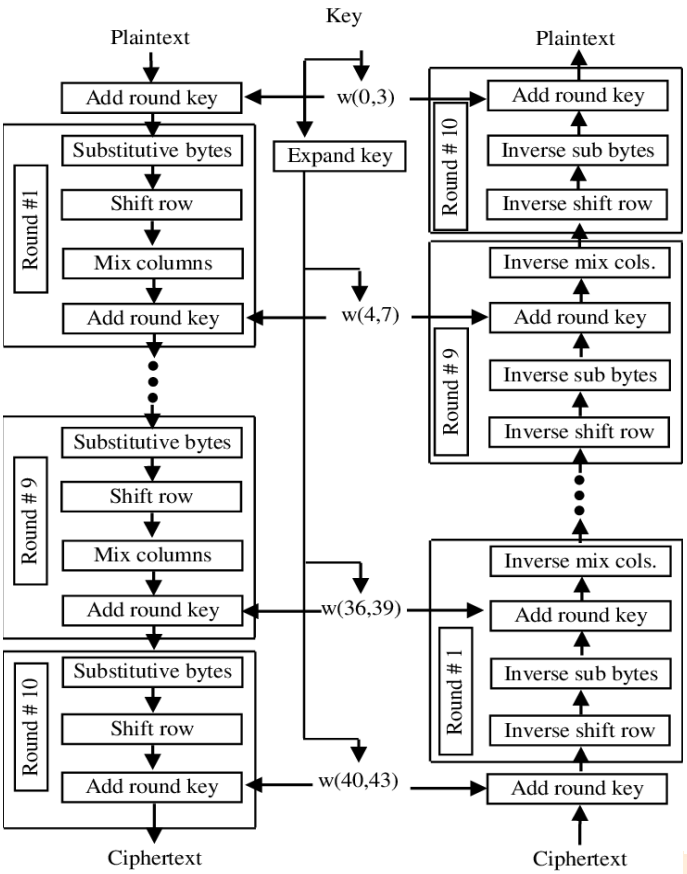

Fig 2: Block diagram for AES encryption and decryption

IV. PRAPOSED SYSTEM

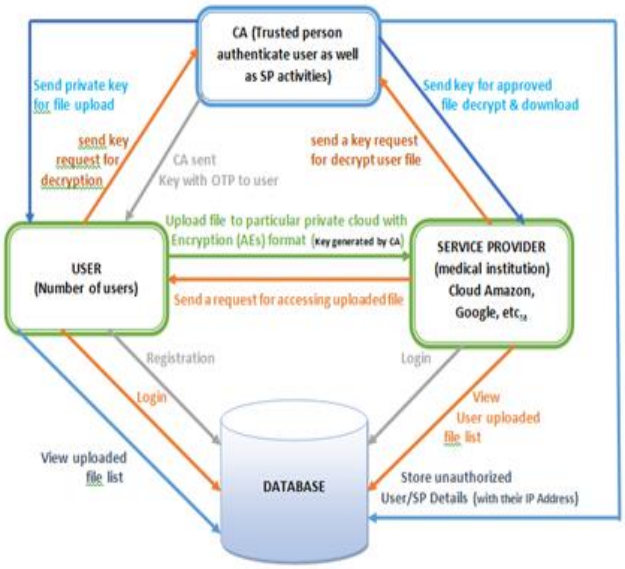

Fig 3: Block diagram

The proposed system mainly consists of the 3 modules certificate authority, user and service provider. Each block has their separate login for this project.

Starting with the CA module, which is a trusted authority which manages both data owner and the service provider. Next is the user module, which allows the user to create login credentials by submitting their information like mail id, phone number, name, etc. After user registration, CA will generate a user identity key for the user which will be stored in the database server. By registering, the Service Provider can gain access to the resources stored in the database server. Now the data user can upload the file to the server. Before uploading the file is encrypted using AES algorithm and the secret key. After the file uploaded, service provider can able to view the user uploaded file in encrypted format. If SP wants to access or view the original format of that file, a request is sent to the user and in return if user wants to share the data with that service provider then the request is accepted and secret key is shared with the Otp via e-mail. And if SP tries to get access of data file wrong secret key and otp more the 3 times, the IP address of the SP is detected and also SP account will get blocked and deactivated.

Similarly if user also wants to access its own uploaded encrypted data from the server, user need to take permission from CA and if the secret key and otp gets wrong, the user account also get blocked and deactivated.

V. RESULTS

- Created a registration and login page with username and password as credentials, using graphical user interface (GUI toolkit) which has huge number of framework available in python. Figure 4 shows the user interface page.

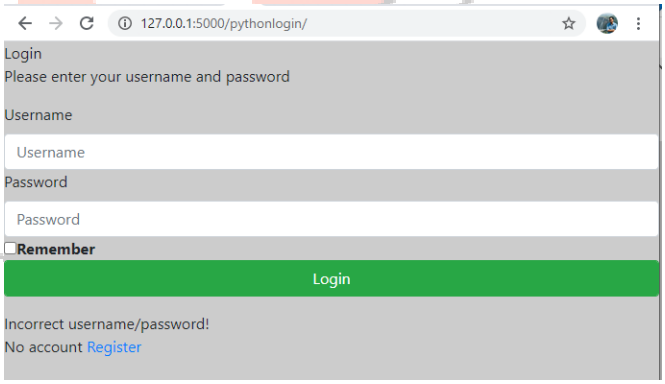

Fig 4: User interface page

- Encrypted and decrypted the database consists of user's genomic data such as Name, SNP-id (Single Nucleotide Polymorphisms), Position of SNP in chromosome, Chromosome number and Ranking using basic python inbuilt libraries. Figure 5 and 6 shows encrypted and decrypted data

	A	B	C	D	E	F
1	SNP-ID	NAME	POSITION	CHROMOS	RANKING	
2	882834	rs1297986	39248147	19	926	
3	255781	rs9939609	53786615	16	728	
4	445791	rs7903146	1.13E+08	10	681	
5	508598	rs4680	19963748	22	629	
6	882838	rs8099917	39252525	19	618	
7	672239	rs6265	27658369	11	551	
8	816100	rs3	31872705	13	540	
9	463641	rs1800629	31575254	6	393	
10	680382	rs25531	30237328	17	389	
11	426	rs2234693	1.52E+08	6	329	
12						
13						

Fig 5: Database

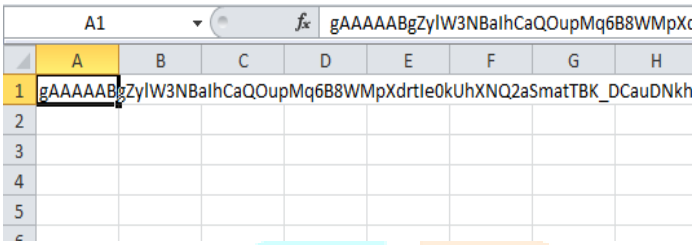

Fig 6: Encrypted database

VI. CONCLUSION

In this project, a cryptographic plan is proposed to share genomic information and genomic test outcomes. The proposed plans are between individual and researcher (service provider). Utilizing the proposed plans, from one viewpoint, a service provider can check the legitimacy (or authenticity) of genomic information it gets from data owner. Then again, the individual, through a computerized assent, can ensure that the service provider won't get access and view his genomic information without his consent. The proposed plans depend on AES encryption and this cryptographic crude empowers us to connect the data about the authenticity of the information to consent and the identity of the individual. Security and reasonableness of the proposed plans is likewise talked about. The proposed plans can be effectively received by existing chips away at protection safeguarding handling of genomic data.

REFERENCES

[1] P. Baldi, R Baranio, E De Cristoaro, P GAsti, and G Tsudik, "Countering GATTACA: Efficient and secure testing of fully sequenced human genomes," Proceedings of ACM CCS '11, pp. 691-702,2011.

[2] E. Ayday, J. L. Raisaro, J. Rougemont, and J.-P. Hubaux, "Protecting and evaluating genomic privacy in medical tests and personalized medicine," in WPES, 2013.

[3] M. Canim, M. Kantarcioglu, and B. Malin, "Secure management of biomedical data with cryptographic hardware," IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 1, 2012.

[4] N. Karvelas, A. Peter, S. Katzenbeisser, E. Tews, and K. Hamacher,"Privacy-preserving whole genome sequence processing through proxy-aided ORAM," in Proceedings of the 13th Workshop on Privacy in the Electronic Society, 2014, pp. 1–10.

[5] Khalid Alshafee," Encryption Techniques in the Cloud" in International Journal Of Scientific & Engineering Research, Volume 7, Issue 7, July-2016

[6] Z. Huang, E. Ayday, J.-P. Hubaux, J. Fellay, and A. Juels, "Genoguard: Protecting genomic data against brute-force attacks," in n Proceedings of IEEE Symposium on Security and Privacy, 2015.

[7] P. S. L. M. Barreto, H. Y. Kim, B. Lynn, and M. Scott, "Advances in cryptology — crypto 2002," M. Yung, Ed. Springer Berlin Heidelberg, 2002, pp. 354–369.

[8] M. Kantarcioglu, W. Jiang, Y. Liu, and B. Malin, "A cryptographic approach to securely share and query genomic sequences," IEEE Transactions on Information Technology in Biomedicine, vol. 12,no. 5, pp. 606–617, 2008.

[9] E. Ayday, J. L. Raisaro, U. Hengartner, A. Molyneaux, and J.-P.Hubaux, "Privacy-preserving processing of raw genomic data," in DPM, 2013.