TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 5574-5579

Secure Encryption Mechanism for Mental Health Data

Ashutosh Shankhdhar
Vinay Agrawal
Arushi Mangala Sankalp Chaturvedi
Prateeksha Chaturvedi

Article Info
Volume 83
Page Number: 5574 - 5579
Publication Issue:
May - June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *17 May 2020*

Abstract

The modernization of data analysis in Mental Health research let to a variety of Analysis on this. Today, wellbeing examination and social insurance produce a consistently expanding measure of information. Making these accessible for auxiliary use cases is basic for proficiency gains in wellbeing research, for example by lessening time-and costs-concentrated obtaining of information. A step towards digitization of mental health research made this data easily available to number of users that includes doctors, patients, researchers, statisticians etc. However, allowing this data to be easily accessible to number of users led to problem of security lacking that is to be handled in a very sophisticated manner. In this whole approach, main motive will revolve center around the finding of security concerns and then giving a suitable privacy preserving data analysis to this, along with security additives. In this study we will try to find out the loopholes of existing solutions to the concern. The researcher will be provided with some definite and accurate solution to the present concern. This paper will include the practical implications as well to give a satisfying solution to this. The main motive of this whole study is to first analyze the data present to the data receiving site and then transmitting that data thorough proper secured channel so as to provide the upgradation to the present data problem.

Keywords: *Data Security, Block Cipher, Data Encryption, Symmetric Algorithm, Mental Health Ontology*

I. Introduction

During the last decade the value of data has risen exponentially in every aspect of life due to explosive surge in the amount of data available. It has given birth to a whole new perspective of analyzing and manipulating this data. Today we have countless machine learning and deep learning based models that use those massive chunks of data to give astounding results in almost every problem, whether it is classification, generating trends or predicting the outcome. All these advancements and developments have made data one of the most important commodity. No matter what kind of field database learning algorithms find their usage. There it won't be an exaggeration to call data '*the currency of present time*'.

Then again, to an ever increasing extent scientists accept that breaking down this information past direct clinical consideration of patients (e.g., psychological well-being research, open wellbeing, and other optional uses) can assist them with producing new information and encourage development that helps logical research and improves human services quality and persistent results. This new perspective on human wellbeing and conduct is starting to empower extraordinary walks in social insurance what more, general wellbeing. Data Analysis incorporates various kinds of sub-forms, for example, ascertaining the mean with standard deviation and the recurrence all things considered, including clinic usage, utilization of fiasco or crisis administrations, and the quantity of long periods of first-time medicine visits after admission to emergency clinic Survival was assessed in sedate administration and control gatherings [7].

The same is true for the field of mental health. The data when processed can be used for personal health maintenance, it can be used by caretakers and nurses to maintain the condition of patients as well as by the researcher to generate new algorithms, classification parameter to improve health facility. To elaborate the first case of personal maintenance, the individuals can monitor their own health records in case they find any abnormality in the trend. They can act accordingly and efficiently. Moreover, this also saves the hustle of hiring personal caretakers or nurses. As for researchers the data trends and characteristics helps to modify the problem with much more accuracy, so that they can device counter strategy much more efficiently. With the data analysis the doctors may predict the likelihood of some problems in certain individuals. The development of new medicines and using new techniques also depends upon the responses given by the patients. So data digitization and its availability plays a major role in the advancement of the technology of mental research in context of privacy preserving as well.

However, with the increase of high usage of this data by so many different parties, the misuse of this data is also in the list. Many time the personal health records can be used in one way or another to cause harm to the individual either psychologically, physical, social or economic. For instance, if the records are revealed at the patient work or society, it can cause discrimination or may be the people start judging the individual that may cause traumatic experience, as well as deterioration of social status of the individual. The medical history can also be used by the people to satisfy personal grudges or can be used as threat against the individual. Thus it becomes especially essential to safeguard this data and ensure that only the concerned authorities or personal can get hold of these records. It becomes necessary to check the credibility of the person requesting the data so that data cannot be accessed by unauthorized personals. There has to be mandatory security checks for accessing, manipulating and retrieving the data records. Even while transferring the data proper encryption/decryption is required. There has been a lot of research in the field of protecting the privacy of patients. There has been a lot of research in the field of protecting the privacy of patients. Nevertheless, there is still a long way to go. In our study we diagnose the existing way of achieving this goal as well as devise our own way to tackle this problem. We try to bring the contrast of the prevailing strategy and our own approach. In our approach, we will be having a site naming data fetching site and our data storing site which will be in contact of data service provider. The presence of analysis at the storage site then transferring the data to *DSP* with encryption/decryption algorithm and then when needed giving this data to the authorized identity will be the whole process of working. In this study more efficient algorithms with security checks of data authentication will be used.

II. Related Work

We are at the stage where we have plenty of solutions for a particular problem. So this problem is also having plenty of solutions. We believe that there is always a ray of improvement in any solution. In this study context to the already present solution related to the problem, we will try to explain loopholes of those solutions and how our solution is somewhat efficient in terms of security as compared to theirs. Existing solution may include several encryption/decryption algorithms, some privacy rules and much more. First existing solution comes under the traditional approaches pointing towards some privacy rules including identification. Large clinical databases were having millions of records in them that can be fetched by malicious hitters having anonymous identity and false motives.

Obviously more examinations are expected to gather a lot of computerized information from patients so as to recognize and approve the most helpful sorts of information and to give proof to clinical utilization of advanced innovation in psychiatry. Until now, most examinations which gather advanced information from patients have utilized solid frameworks planned explicitly for that review. Thusly, most of the writing is centered on the outcomes and less consideration has been paid to highlights of the information assortment stages. Definitely new computerized information stages will develop sooner rather than later that address the requirement for enormous scope information assortment for mental research. Thusly, it is significant that exploration conventions and highlights of these stages be shared among specialists in this field, notwithstanding the outcomes that these stages produce [6].

De-identification [2] was approach to preserve the data where certain no of identity specifier is excluded from the database so that malicious hitter cannot form the data. These specifiers were encrypted and decrypted at the other end. Other approach was to have statistician to check whether the specifiers are properly encrypted or not. But in this approach also there are some loopholes, like if we need to encrypt the very important specifier that are to be transmitted, then we have inefficiency in the encryption data. Another loophole arises when data merging is to be performed for that we need to have whole data but that data was encrypted. Another traditional approach is of notice and consent where notice is being given to the patient and asking for his consent to access the data. Here also the loophole takes place as the cost of contacting lakhs of patients and taking positive responses from them, it is a very troublesome and not at all cost effective. From here we can conclude that we need to have some efficient approach leaving behind these traditional ones but taking their loophole as consideration.

Data anonymization [2] is a process of encrypting the personal information so that the data remains anonymous, hence providing the security to them. Hence, HIPAA principle was used [4][5], so that limited information and data is fetched to reduce the hitters. Another approach was forming the committees having IRB (Institutional Review Boards) for taking the necessary steps for accessing the data of a particular patient having private information. In this approach it is up to that committee that which information has to be revealed or not. As per the time is passing, the hackers are also becoming more efficient. So hence comes another approach of Secure Socket Layer (SSL) [3], use of firewall and information security. This process will include the proper monitoring of thee channel, risk assessment, user information etc. to check the information of each and every detail. But the main loophole in this is that IRB cannot always be sufficiently accountable for the decision taken by them. Data anonymization is also prone to breach if not accurately anonymized the data. This includes encryption of accurate number of specifiers. SSL could be hacked by the hacker easily applying several approach. So there is a large space for improvement in this field.

Another approach to deal with this problem is applying efficient algorithm to secure the data of the patient. AES(Advanced Encryption Standards) [1] is a proper approach towards this problem. Before applying AES, problem of data duplication was also there. AES is a symmetric block cipher and it is used because it resists all unknown attacks but again loopholes were there. It is very hard to implement with software and along with this in every algorithm block is encrypted in the same way so security concerns are there. From these we all can conclude that there is a gap where we can work upon.

Another approach that has been applied is associative rule mining using commutative encryption. In this the machine learning approach is applied for the encryption. The approach is that researcher will just apply the information in form of query and if it abides by the rule of law then they will be allowed otherwise not. The main problem with this is at a time only two persons can work so time consumption along with man power is much more.

So, working on several approaches and dealing with their problem in their approaches, in the coming context this study will propose a solution where above loopholes are rectified and concerns are taken into consideration.

Some online demonstrative programs can be utilized to analyze the accompanying classes of mental issues:

- Adjustment issue
- Alcohol-related clutters
- Anxiety issue
- Drug-related clutters
- Eating issue
- Infancy, youth, and pre-adulthood issue
- Mood issue
- Personality issue
- Psychotic issue [10]

The four different structures are influenced by various security issues relying upon the responsibilities of system owners. In Directed and Acknowledged structures, the mam obligation is from the focal element proprietor, which is capable of organizing the course of exercises. Anyway in Collaborative and Virtual structures the capable isn't so all around characterized. For Collaborative engineering it very well may be expected that the capable is the framework that demand for the joint work to achieve the strategic; the other hand, on account of Virtual every framework is self-mindful, with the hindrance of not having express cooperation from different frameworks side to help a typical security approach [9].

III. Mechanism

The mechanism of whole study is likewise according to the phases mentioned in the previous contexts. If we are taking that data is collected at the hospital management site, then we need to have the data for the analysis part there only. Then we are transmitting that data to DSP for storage, so security of data is to be considered. The whole process of mechanism is divided into two parts-

- Analysis part

- Security part

1) Analysis Part

The data from different patients are collected at the hospital management site. Now millions of records are present at this site. For the proper arrangement of, we need to first do exploratory data analysis on this d data to find duplication anomaly, false record, blanked columns and on the basis of that purpose we are using the logistic regression algorithm of machine learning for the best out of this. Firstly, the data will be processed by different ETL tools and when the prior stage of it is completed then we will head towards the implementation of algorithm.

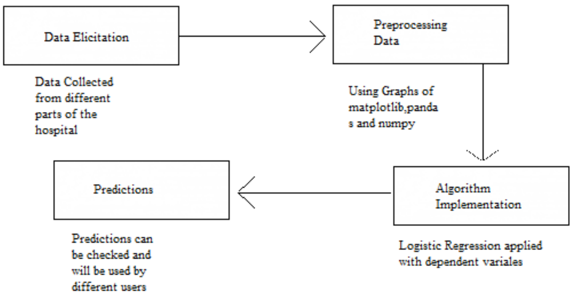

Data Elicitation
Preprocessing Data
Data Collected from different parts of the hospital
Using Graphs of matplotlib,pandas and numpy
Predictions
Algorithm Implementation
Predictions can be checked and will be used by different users
Logistic Regression applied with dependent variales

Fig. 1 – Analytical Processing of Data

Logistic Regression is used because the prediction that has to be given by the hospital management is of binary in nature. So taking different attributes into the consideration the final prediction will be of binary type as YES/NO or 1/0. It is a statistical model which uses dependent variable (binary) and using logistic function.

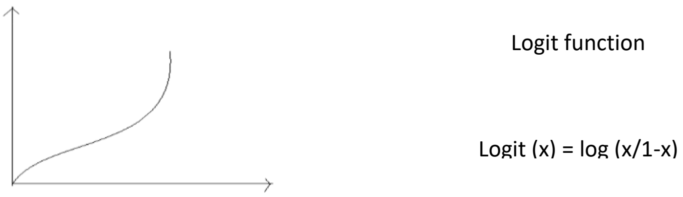


Fig. 2 – Representation of Logit Function

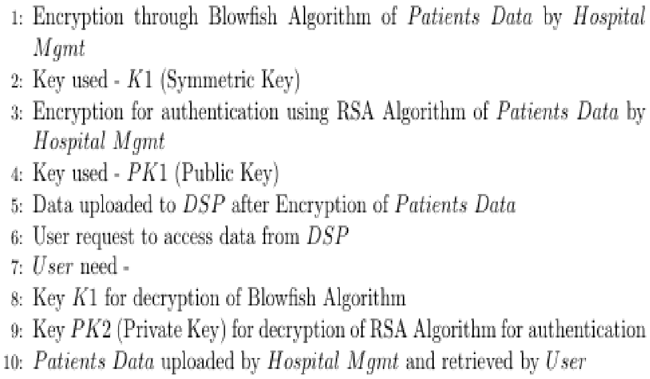
Predictions from this will be used for the analysis of researchers, doctors etc. Here the work of analysis part is done where we are having a sorted data (analyzed) for the future purpose and ready to be uploaded to data service provider site.

2) Security Part

This part is most significant as this is providing the security to the data that is the main motive. For this we are sub –dividing the security process in two parts –

- Encryption/Decryption
- Authentication

3) Algorithm



```
Encryption through Blowfish Algorithm of Patients Data by Hospital Mgmt
Key used - $K1$ (Symmetric Key)
Encryption for authentication using RSA Algorithm of Patients Data by Hospital Mgmt
Key used - $PK1$ (Public Key)
Data uploaded to $DSP$ after Encryption of Patients Data
User request to access data from $DSP$
User need -
Key $K1$ for decryption of Blowfish Algorithm
Key $PK2$ (Private Key) for decryption of RSA Algorithm for authentication
Patients Data uploaded by Hospital Mgmt and retrieved by User
```

For the first part we are using blowfish algorithm which will encrypt the data and then that data will be provided to DSP for storage. When the user wants to fetch the data then user will be having the key that will decrypt the data if and only if owner allows and give that key to user. For providing the security, we have used blowfish because fast, execute in less memory and it much secure then other symmetric algorithm as it is using variable length key. It is a block cipher and along with this it uses 16 rounds so it is secure also. For more securing accessing of the data in this study we are using the authentication process also that will be carried out by RSA Algorithm. In this the algorithm will encrypt the data with a public key and when user needs this data he will have to have a private key for decrypting that. So as a whole, if we combine the whole process we came out to be as –

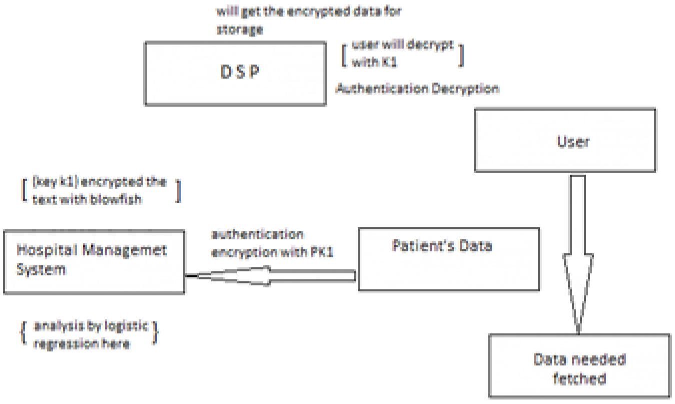

will get the encrypted data for storage
DSP
[user will decrypt with K1]
Authentication Decryption
User
[(key k1) encrypted the text with blowfish]
Hospital Managemet System
authentication encryption with PK1
Patient's Data
{ analysis by logistic regression here }
Data needed fetched

Fig. 3 – Process Description of Proposed Model

So summarizing this whole process, patient will provide the data to hospital management system. Data collected will be analyzed by logistic regression. Then that analyzed data will be encrypted by blowfish algorithm with key K1 and authentication encryption with RSA algorithm with public key PK1. Here double security is provided to the data. Now this encrypted data will be stored by DSP in the form of tables. The data if needed by any user be it researchers or doctor, he will be needing two keys that are private key for authentication (RSA is an asymmetric algorithm) and a blowfish decryption algorithm for decrypting the message. In this way the data will be fetched by any user. This study provided for double security along with proper analysis of data. The algorithm applied are much efficient in taking up the security of the data to another level.

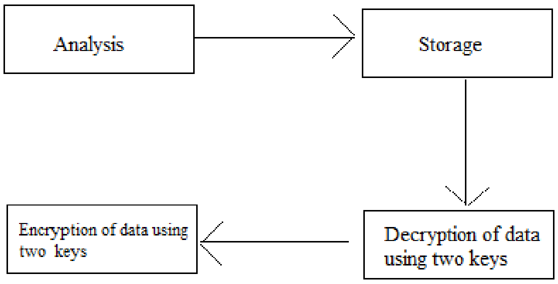

Analysis
Storage
Decryption of data using two keys
Encryption of data using two keys

Fig. 4 – Mechanism Flow Out of Model

The proposed mechanism will not only help the users and clients to maintain confidentiality and integrity of their data but also the level of security which may already be provided will also be enhanced and the cloud realm will become more viable for the data possessors too. A secure system will be ensured in the end and that is the end goal today for any person who requests and uses services of the web [8].

Data with respect to psychological instability is scattered over different assets and it is hard to interface this data, to share it and discover explicit data when required. A Mental Health Ontology can be intended to give a model of emotional wellness ideas and connections that can be utilized to structure a semantic system for some information stockpiling and recovery. Such a semantic system could be utilized for orderly comment of psychological wellness data accessible through different data assets and backing questioning of heterogeneous data sources. A collection of agents can utilize a shared domain ontology as their common knowledge base. This will encourage correspondence and coordination among specialists and backing some significant procedures inside a multi-operator framework such as: issue disintegration and assignment sharing among various operators, results sharing and examination, data recovery, determination and mix and so on [12].

Advantages -

1. Viable way to deal with inside attack by safely circulating the patient information in numerous information servers.
2. Utilizing the Pail lier cryptosystems to perform measurable examination on the patient information without trading off the patients' security.
3. In Proposed framework, due to made sure about circulated database architecture we can accomplish information stockpiling and information examination security.
4. Proposed information recovery procedure permit to recover the information traded off server(s) [11].

IV. Conclusion

Health Information is one of the most delicate sorts of individual data. Mental scatters and medications for mental scatters are commonly more delicate than are other wellbeing conditions. While unseemly use and divulgence of any clinical data may prompt separation in medical coverage or work, mental clutters frequently causes an extra weight of bias and separation. Individuals with mental disarranges may surrender treatment in light of the fact that of disgrace and misjudging. Worry about security may likewise speedy some to pay "from cash on hand" for treatment even secured by protection premiums. Security is an essential prerequisite for emotional wellness explore. In this manner, mental wellbeing specialists must be doubly cautious that records-based look into ensures understanding protection [9].

In this paper we proposed the approach of data security along with the analysis part taken into consideration because the quality of data is the priority in this modern era. We applied logistic regression for the analysis of the data then encryption through blowfish algorithm with RSA algorithm for authentication encryption. Meanwhile with the proper implementation of this approach can reduce the anomaly and crimes that are being carried forward because of lack of security of mental health patients. They obviously have data that is to be preserved and should be given to whom patient wants to give. This approach is a step forward for this nobel cause. Definitely every solution is having a gap of improvement but from our side we are not left with loopholes in this approach. With this approach, efficiency accessibility and security is increased and thus study provides with a better system for this concern.

The three ontology 'measurements' (ailment type, factors and medicines) contain totally different data and are symmetrical to one another. The 'Disease Types' sub-cosmology is progressively a grouping cosmology and is emphatically progressively upheld. The 'Components sub-cosmology is emphatically founded on logical research and uncover distinctive sort of components that may influence our psychological well-being, both emphatically and adversely. The 'Treatment' sub-cosmology is a mix of arranging also, explore metaphysics. Planning new medications is explore work be that as it may, for instance, all the found medications can be progressively ordered. Each of the three 'measurements' are extraordinary from one another and each 'measurement' is special. In any case, together they give a general picture and a decent diagram of mental wellbeing information [12].

Paper concludes up by proposing down to earth suggestions for emotional wellness scientists and future research in the field of protection saving information investigation.

V. References

[1] Dr. B. AnniPrincy, G.Tamilselvi, S.Shruthakeerthi, B.Sowmya, “Privacy Preserving Data Analysis in Mental Health Research Using Cloud Computation”, International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 2, Mar – Apr 2017.

[2] VaibhavLawand, PrathikSargar, AnandBhalerao, PradipJadhav, “Analytical Approach for Privacy Preserving of Medical Data”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181.

[3] Jingquan Li, Xueying Li, “Privacy Preserving Data Analysis in Mental Health Research”, 2015 IEEE International Congress on Big Data, DOI-10.1109/BigDataCongress.2015.23

[4] Nabil Adam, Tom White, BasitShafiq,Jaideep Vaidya and Xiaoyun He, “Privacy Preserving Integration of Health Care Data”, AMIA AnnuSymp Proc. 2007; 2007: 1–5.Published online 2007.

[5] https://privacyruleandresearch.nih.gov/healthservicesprivacy.asp

[6] http://projectideas.co.in/privacy-preserving-data-analysis-in-mental-health-research/

[7] TalayehAledavood, Ana Maria TrianaHoyos,TuomasAlakörkkö, Kimmo Kaski,JariSaramäki, ErkkiIsometsä, and Richard K Darst, “Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype”,JMIR Res Protoc. 2017 Jun; 6(6): e110. Published online 2017 Jun 9. doi: 10.2196/resprot.6919.

[8] AshutoshShankhdhar, Arushi Mangla, PrateekshaChaturvedi, “Encoded Data Methodologies for Cloud Computing Realm’s Security Enhancement,” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1, May 2019.

[9] Miguel Angel Olivero, Antonia Bertolino, Francisco Jose Dominguez-Mayo, Maria Jose Escalona, HariaMatteucci, “Security Assessment of Systems of Systems”, 2019 IEEE/ACM 7th International Workshop on Software Engineering for Systems-of-Systems (SESoS) and 13th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (WDES), DOI - 10.1109/SESoS/WDES.2019.00017.

[10] Hajar Mat Jani, “Benefiting from Online Mental Status Examination System and Mental Health Diagnostic System”, The 3rd International Conference on Information Sciences and Interaction Sciences, DOI - 10.1109/ICICIS.2010.5534712.

[11] Matthieu-P. Schapranow,MatthiasUflacker, Murat Sariyar,SebastianSemler,JohannesFichte,DietmarSchielke, Kismet Ekinciand Thomas Zahn, “Towards An Integrated Health Research Process: A Cloud-based Approach”, 2016 IEEE International Conference on Big Data (Big Data), DOI - 10.1109/BigData.2016.7840929.

[12] Maja Hadzic, Meifania Chen, Tharam S. Dillon, “Maja Hadzic, Meifania Chen, Tharam S. Dillon”, IEEE International Conference on Bioinformatics and Biomedicine, DOI - 10.1109/BIBM.2008.59