https://doi.org/10.33472/AFJBS.6.8.2024.2132-2142

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper

Open Access

DATA SANITIZATION AND ENCRYPTION BASED METHOD FOR SECURE DATA STORAGE IN CLOUD COMPUTING

Mr.A.Balthilak[a], Dr.S.Pathur Nisha[b] , Mr.R.Allocious Britto Rajkumar[c] Dr.P.Gomathi[d]

[a] Assistant Professor (SG), Department of Aeronautical Engineering, Nehru Institute of Technology ,Coimbatore, Tamilnadu, India

[b] Professor , Department of Computer Science and Engineering, Nehru Institute of Technology ,Coimbatore, Tamilnadu, India

[c] Assistant Professor (SG), Department of Aeronautical Engineering, Nehru Institute of Technology, Coimbatore, Tamilnadu, India

[d] Professor in Electronics and Communication Engineering, Study World College of Engineering, Coimbatore, Tamilnadu, India

Volume 6, Issue 8, June 2024

Received: 09 March 2024

Accepted: 10 April 2024

Published: 1 June 2024

doi: 10.33472/AFJBS.6.8.2024.2132-2142

ABSTRACT

The Internet of Things (IoT), smart cities, enterprise digital transformation, and the global digital economy are among the latest development trends. Because of the vast amounts of data generated, the continual increase in data storage pressure drives the rapid expansion of the whole storage business. Cloud storage systems become vital in the new era by providing data storage and administration. Governments, businesses, and individual users are all actively transferring data to the cloud. Such vast amounts of data have the potential to generate immense wealth. However, it raises the possibility of unwanted access, data leakage, sensitive information revelation, and privacy breach. Although there are various research on data security and privacy protection, there is still a shortage of systematic. In this work, we provide a complete overview of the literature on data security and privacy issues, data encryption technologies, and appropriate countermeasures in cloud storage systems. Specifically, we begin with a review of cloud storage, including definition, classification, architecture, and applications. Second, we present a detailed study of the issues and requirements for data security and privacy protection in cloud storage systems. Third, data encryption technology and security measures are summarized. Finally, we used Python to develop a seamless encryption and decryption system. We employed the concept of cryptography to hash a given data, upload it to the cloud, and retrieve it as needed.

Keywords— Cloud storage, data security, cryptography, access control, privacy protection

I. INTRODUCTION

Data Sanitization for Cyberbullying, a common problem on social media, poses a major threat to people's mental health and well-being. We use a machine learning-based algorithm to detect instances of cyberbullying on social media sites. Our solution employs natural language processing algorithms to scan textual content and identify cyberbullying situations with high accuracy. Cloud computing is an exciting study topic that is used in industry and academia for a variety of applications, allowing businesses or individuals to outsource their data.

II. CLOUD COMPUTING

Cloud computing is a rapidly growing technology that provides a variety of dynamic, scalable, and pay-per-use services. Cloud computing offers numerous benefits in terms of data storage and access. Data storage refers to storage that is independent of location and requires high-quality data services. Cloud computing differs from traditional computer models in that it is service-oriented, allows for resource sharing, and hosts data in external storage. Enabling resource sharing allows for more efficient use of hardware, which enhances hardware performance in the cloud. Furthermore, resource sharing in the cloud benefits users both economically and practically. In outsourced storage, the phrase data hosting refers to the cloud's ability to give service to consumers quickly and without much waiting time; in other words, data hosting improves data transfer speed. Cloud providers use multiple layers to handle information and provide on-demand services to users. Users can access data in the cloud based on the sort of service they demand, and these services are provided to them via various traditional networks known as cloud storage. The cloud storage contains user profile details, business information, and internet backup details. The cloud storage serves a variety of services to the cloud users’ through a variety of traditional networks and the cloud storage holds the details of the profiles, business logs, and backup of the internet logging details. The problems associated with the cloud includes the Online data backup, data archiving, data compliances, disaster recovery, and compliance regulations. Cloud holds a lot of advantages and the main advantage is that the users are benefitted much mainly, in the business world for the sharing of the data between the service providers as it serves itself as a portability option The users can access the data in the cloud that depends on the type of the service they require and the services are enabled to them using various traditional networks that are simply termed as cloud storage. The cloud storage possesses the details of the user profiles in the cloud, information regarding the business, and the backup details made via internet.

III. DATA SANITIZATION

Data sanitization is the process of purposefully and permanently erasing or destroying data from a storage medium such that it cannot be recovered. Normally, when data is wiped from storage media, the media is not truly erased and can be recovered by an attacker with access to the device. This raises severe issues about security and data privacy. Sanitization cleanses storage media, ensuring that no data remains on the device and that no data can be recovered, even with modern forensic techniques. As the functional lifetime and storage capacity of storage systems expand, IT assets frequently retain sensitive company data after being decommissioned. The various data sanitization methods are

- Physical Destruction
- Data Erasure
- Cryptographic Erasure
- Data Masking
- Data Discovery and Sanitization
- Data Masking and Discovery with Imperva

IV. SECURITY AND DATA STOROGE IN CLOUD

4.1. SECURITY IN CLOUD

The growth of Cloud computing is experienced by the users such that they upload the data, either a sensitive data or insensitive data, in the cloud to share the data or for saving the cost. With all these benefits, the Cloud is facing numerous problems in terms of data security and privacy that hinder the growth of cloud computing. In cloud computing, the outsourcing of the data occurs and the users wish to protect their data such that it is accessible to the authorized users. It is highly required that the decisions regarding the access policies of the data depend on the data owners engaged in cloud computing. Cloud platform uses the infrastructure computing that offers the resources, services and information to the requested users. As an example, one can say that the cloud users utilize the services to accomplish their role in a pay-as-you-go way that safeguards them for huge capital investment required for operating their own IT infrastructure. There is no limitation inside the cloud such that the users have a concern towards privacy questioning that if the private data is protected data while referring to the data of the cloud. In case if the users detect that the data is an unprotected privacy data, they stop using Cloud computing that highlights that the privacy protection is a huge concern in cloud computing. Some of the security issues may arise due to the following reasons like: Insecurities due to the usage of the hardware by a huge number of the cloud users. The second reason is due to the physical boundaries brought about by the virtual technology. In other words, the physical boundaries may get affected when a logic server possesses number of virtual servers, but a virtual server belongs to a number of logic servers. The privacy and security issues will affect the customers of the cloud service badly and hinders the smooth development of the services related to the cloud. The main reasons behind the security and data integrity issues are:

1.The capacity of the attackers, such as internal and external attackers to attach the cloud. 2.The security risks of the cloud, mainly in places, where relevant considerations of attacks and Countermeasures are made. 3) Increasing and growing risks of cloud security.

4.2. Privacy Protection in Cloud

During the deployment of the cloud server as a broker, privacy seems to be critical in data publish-subscribe systems and hence, the trust of the cloud server fades. The CSP in cloud computing may share the data among themselves or sometimes need to send the data to the recognized users. While sharing the data, the service providers ensure that the user or the requestor, who received the data, imposes similar privacy policy on the received data and CP may be required to send the data with a sticky policy. The receiving end can implement the sticky or privacy policy only when the sender and the receiver hold the same privacy policy. However, there is no well-defined standard regarding the policy for preserving the privacy of the data. The main reason for having no standard privacy policy is that the rule sets of different policy languages differ and hence, it is impossible to develop a privacy policy in a single language. Thus, it is impossible for a single policy language to develop a privacy policy. The various policy languages include the XACMLv2 (OASIS 2005), XACMLv3 (OASIS 2010), PERMIS, P3P (David Chadwick et al.2008), Keynote (W3C 2002), and so on. For illustration, XACMLv2 never assist delegation of authority, whereas the languages XACMLv3 and PERMIS support it. The policy language XACML assumes that it possesses a stateless Privacy-Preserving Data Publication, but never works for a state- based policy rule. PERMIS supports state-based rules and hence, it possesses the tendency to assist both dynamic and static Separation of Duties (SoD). The main role of P3P is to denote the privacy policies, but the other languages authorize the policy languages. Hence, it is well understood that the cloud authorization service should assist multiple Privacy-preserving Data Publishers (PDPs). The decision based on the authorization should be kept simple despite of the complex policy handling authorization infrastructure. The ubiquitous nature of the internet assists the CSPs to sustain and get closer to the global market and audience. CSPs aim at rendering the required hardware to the users along with the software and resources. Moreover, it possesses the effective approach for creating, updating, and accessing the data by the users. The CSP manages the transactions of the customers, manages other cloud-related issues and so on. The security methods of the communication are adapted to the cloud and the same kind of attacks is demonstrated in the cloud. Moreover, the CSP must look after the privacy of the data of the users such that they are not intruded by the malicious attacks and it is highlighted that the service provider itself can act as an intruder or the attacker of the user data. In general, the privacy requirements are three in number that includes the data privacy, tag privacy, and trapdoor privacy.

4.3.Various Kinds of Privacy Protection Methods

The working challenge in providing the services to the cloud users is to offer the secure and privacy-enhanced authentication of the users. The users, who share the sensitive information including the information regarding the finance, health reports, and so on seek privacy. The privacy is offered by hiding the identity of the user and enables the identification for which a lot of techniques, such as cryptographic tools, anonymous authentication methods, zero-knowledge protocols, and group signatures are used. The access to the data is enabled only for the authenticated users and this process is managed by the CSPs. The protection to the data by the CSPs should ensure that they are not prone to the attacks of the malicious users. Moreover, there are hundreds of users accessing the services of the cloud services simultaneously such that there is a need for a much efficient process that reduces the computational overhead of the cryptographic methods. Privacy protection is the strategy that ensures the protection of the individual information or the sensitive details before they are published and the privacy protection methods are classified into three. One of the privacy protection methods is perturbative method. The perturbative method causes some change in the individual elements of the original data. The second privacy protection method is generalization methods that are engaged at swapping the original data with less precise values. Finally, the synthetic data generators ensure the privacy protection that establishes the synthetic data that seems similar to the original data. The other techniques to enable the privacy of the data are using the data sanitation, data distortion, cryptographic, blocking, and anonymization. Moreover, in cloud, the Public auditability authentication requires better standard to sustain the privacy of the data through provable secure storage. The privacy-preserving techniques of the new generation have been already on an implementation that mainly targets on the economic development and cooperation. The process of storing a highly confidential data on the cloud in the future is highly advanced both technically and effectively through a well-defined framework and this framework is the hope of the government, private and public sectors.

4.4.Encryption-based Methods

An easy and better way to maintain the privacy of the data belonging to the user relies on the concept of data encryption that is performed before it is outsourced to the cloud, which is followed by the Amazon IS service.

Data encryption is the effective method to deal with the privacy of the data and whatever is the advantage of the data encryption-based data privacy protection, the traditional encryption methods fails to encrypt the data or trapdoors that generates a number of the copies of cipher texts corresponding to the data in the Cloud system. The generated number of the copies is equivalent to the number of Cloud users. Even though there is a lot of the privacy securing methods, there is no proper balance in the Privacy and Utility (PU). A basic method to ensure the privacy of the sensitive data is about encrypting the data before it is outsourced to the cloud and fetches back the data by applying the keyword-based search on encrypted data. The encryption-based privacy protection methods offer resistance against the illegal accesses, but increments the computation overhead. The overhead in computation is addressed by the data owner’s while using the mobile devices and data files of large size. Attribute-Based Encryption (ABE) is anticipated as an effective solution for realizing the fine-grained and scalable control systems and in this type of encryption, the rights to access the data is provided to the individual users. Mainly, Ciphertext-Policy Attribute- Based Encryption (CP-ABE) provides an option to the data owners for pointing a policy to access the data from a number of attributes, and grants permission for encrypting the data based on the access policy using the public key components. Moreover, the encryption-based methods contribute much towards protecting the sensitive data.ABE plays a promising role in data encryption in the cloud storage systems. The privacy of the data is well assured through the usage of the attribute-based access control schemes that is developed on the top of ABE schemes. To be clear, the individual users in the system possess a number of attributes that are inherited in their secret key. Prior to the publication of the data, the publisher establishes a policy to access the data of the user and this policy specifies the attributes that an authorized user is strictly required to have. Followed by the definition of the access policy, the encryption of the data is progressed that depends on the access policy offered by the publisher. Thus, it is clear that the authorized users hold the required attributes as per the specification in the access policy and only the authorized users are permitted to decrypt their data. The effective enabling of the authorized access to the data based on the attribute-based access control helps the publishers to finalize the policies to data access without the prior knowledge of the number of users available in the system at the perusal of the publishers. The significance of the attribute- based access control is that the control method develops only a copy of the data, which is encrypted. ABE aims at protecting the privacy of the data, and therefore, ABE can be employed for protecting the privacy of the subscription. A most genuine method for encrypting the subscription trapdoor is performed using the ABE method, but using a different set of parameters. However, ABE method for subscription trapdoor need an authority, who takes charge of managing the attributes and generation of the key such that the tags are generated for individual data published or published trapdoors for the individual data subscriber

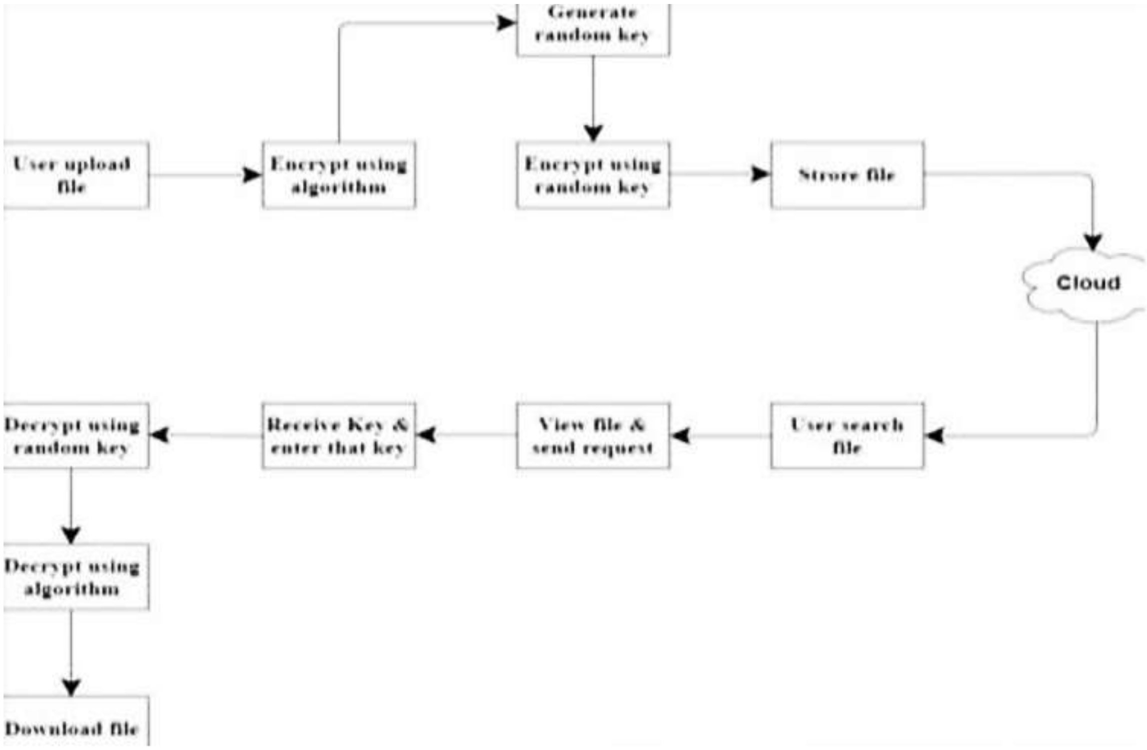
4.5.Decryption

Decryption is activated when the attributes of the user and the respective access policy match each other. In addition to the application of ABE for designing the secure access control, the requirement to enable the protection of the user privacy is essential in case of the access control systems. The challenge above is solved using the anonymous ABE and the extension of anonymous AB99E. In anonymous ciphertext policy attributebased encryption (CP-ABE), the decryption is progressed as: 1. Establishing the secret attribute key at the user perusal. 2. Check if the attribute key meets the requirement of the access policy. 3. If the attribute corresponding to the secret key meets the requirements of the data access policy of the ciphertext, the user can decrypt the data. In other case, the user can only guess the access policy that was used by the data owner. Thus, the Anonymous CP-ABE is employed for the military applications and commercial applications. In these applications, the access policy remains a sensitive information and hence, these access policies require protection.

4.6.Sanitization based Method

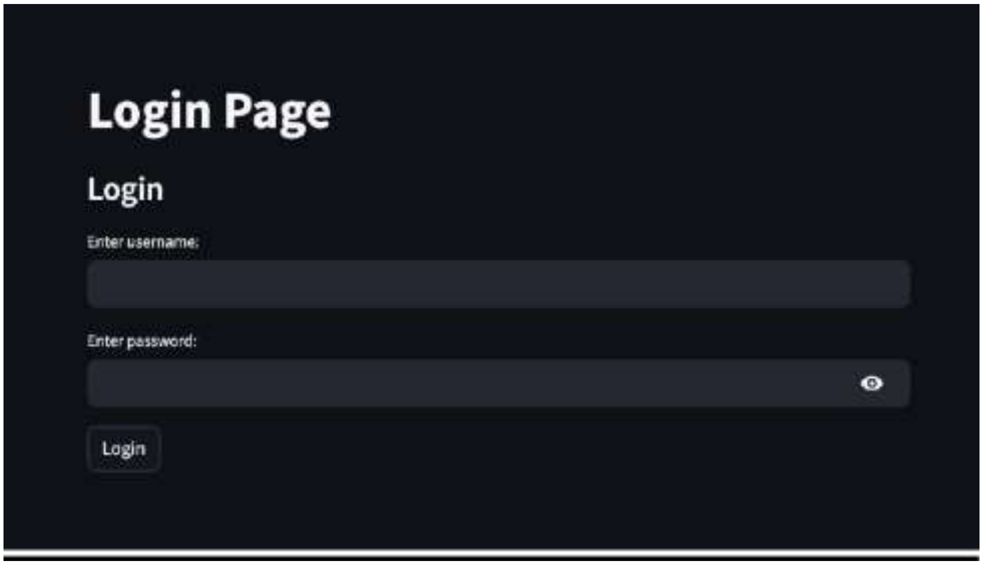
The decoupling approach enables us to share and filter the data effectively in our day-to-day life and mainly, in the Data publish-subscribe. Data publishers are nothing, but they are simply the entities like banks, firms that have regular investments, or some research institutions. These publishers publish their data to their customers or the data users. Data subscribers or the users subscribe to a data based on their interests through the subscription trapdoors. The enormous amount of data in volume and velocity raises the difficulty regarding the storage, management, sharing, analyzing, and visualizing using the existing tools and infrastructures. For the past decades, the growth of the technologies related to the data collection and data analysis seems to be tremendous and specifically, the exponential growth is addressed in the area of Privacy-Preserving Data Publication (PPDP) that has been a boon to the academicians and industries. In other words, the data requires sanitization through the privacy mechanisms that guarantee privacy such that there is no leakage in the privacy of the sensitive information prior to the publication of the data for the public. There are a series of privacy-protecting mechanisms existing in the literature, such as l-diversity, kanonymity differential privacy, differential privacy, and closeness, among which differential privacy emerges as an effective approach for protecting the privacy of the data. Differential privacy allows addition or deletion of the single record that never influence the results of the privacy analysis. To ensure the privacy of the data from plaintext data publishing, there are a lot of contributions from the side of research people and industrial communities providing solutions for data publication that enable the confidentiality of the data. Most of the privacy-preserving techniques utilize the generalization concept for minimizing the term domain of the original data and remove the Quasi- Identifiers (QIs). QIs are simply a set of scanty combination of terms that possess the tendency to locate a record.

V. SYSTEM ARCHITECTURE



VI. PROPOSED WORK AND IMPLEMENTATION

We have used PYTHON to create a seamless encryption and decryption based system. we have used the concept of Cryptography to hash a given data and upload it in the cloud and also retrive it when needed. The code streams a locally hosted page that lets us login using a special id which is logged for 1st layer of security and then details are asked to encrypt. When the user inputs data and clicks encrypt, the details are not stored anywhere in system, rather it is hashed and directly upload to cloud storage as '.enc' file and when needed those are decrypted to view the details. The phone number of the person is declared as their ID. Also the special IDs and password of staffs or admin users is stored in 'cred.enc' which is also a encrypted file. Even when a data breach takes place and files get leaked, when encrypted files are open, the attacker can only view a randomly written hashed stream of characters.



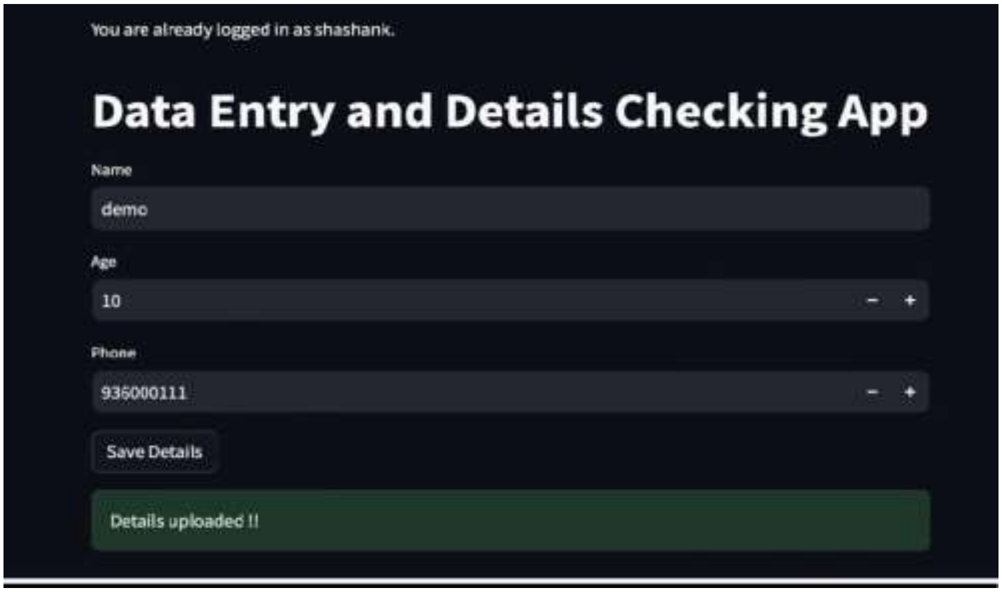
Login Page

Login

Enter username:

Enter password:

Fig 8.1. Login Page



You are already logged in as shashank.

Data Entry and Details Checking App

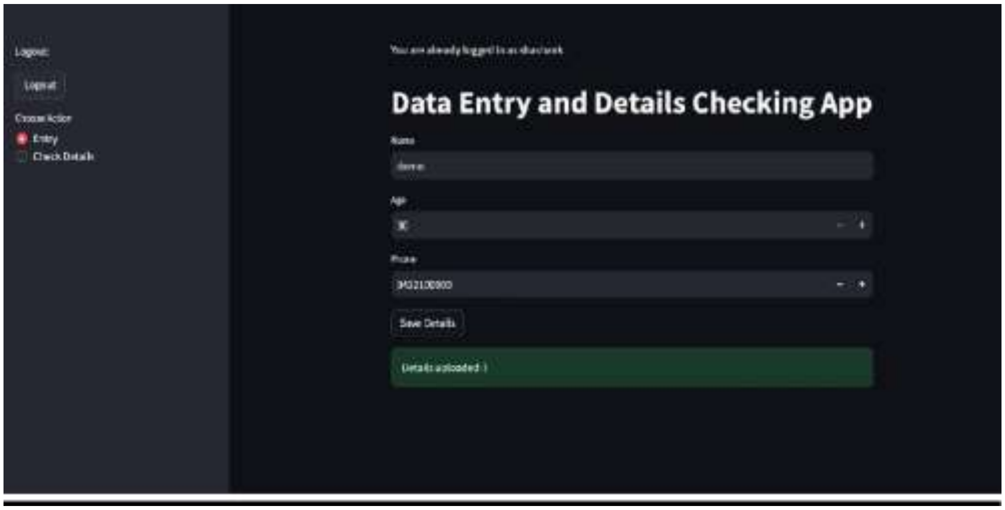
Name: demo

Age: 10

Phone: 936000111

Details uploaded !!

Fig 8.2 Data Entry Basic Profile



You are already logged in as shashank

Data Entry and Details Checking App

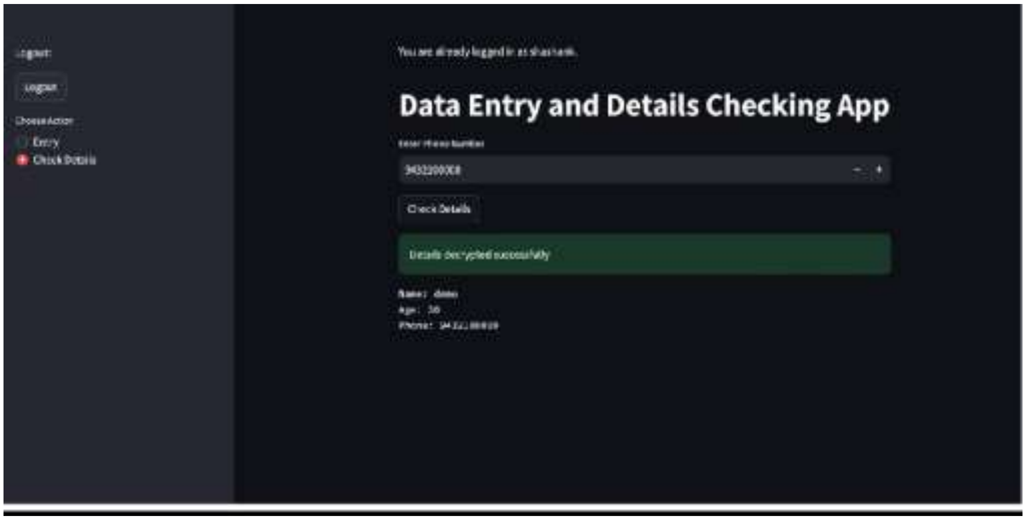
Name: demo

Age: 30

Phone: 9432100000

Details uploaded :)

Fig 8.3. Data entry and details uploding



You are already logged in as shashank.

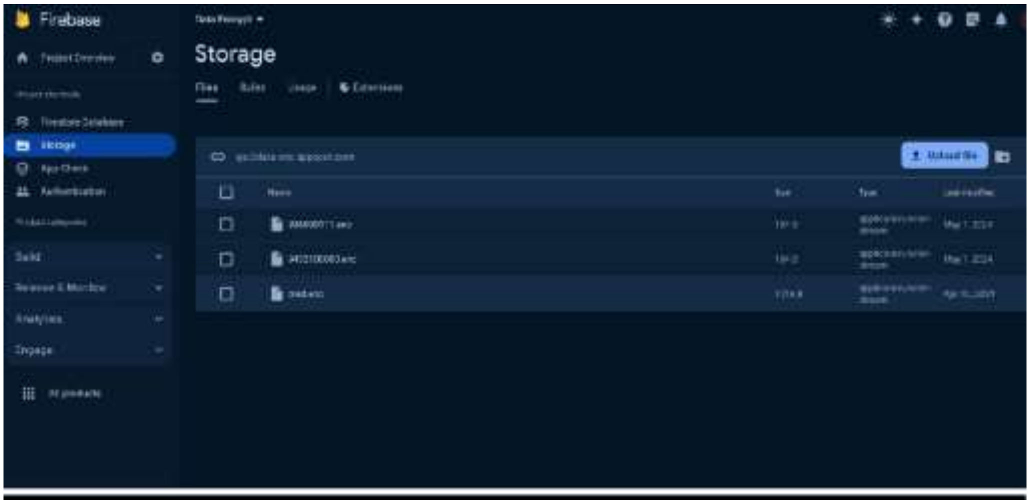
Data Entry and Details Checking App

Enter Phone Number: 9432200008

Details decrypted successfully

Name: demo
Age: 30
Phone: 9432200008

Fig 8.4. Decrypting User Data



Name	Size	Type	Last modified
[illegible].enc	[illegible]	[illegible]	May 1, 2024
9432100000.enc	[illegible]	[illegible]	May 1, 2024
[illegible].enc	[illegible]	[illegible]	[illegible]

Fig 8.5 Storage of Encrypted Data

Fig 8.6. Encrypted File

VI CONCLUSION

In the world of cyber-attacks and data breaches, encryption of data and active monitoring and logging must be done to secure and prevent data from being leaked which leads to many social, economic, and national issues. This project doesn't mean to blame or find mistake in present security used, but this proves to add an additional layer of security. In conclusion, Python as an coding language can also be used as a mean of security, where programming meets security.

REFERENCES

1. Aditya Desai ,Shashank Kalaskar,OmkarKumbhar,and Rashmi Dhumal,(2021) Cyber Bullying Detection on Social Media using Machine Learning, ITM Web of Conferences 40, 03038 ICACC.
2. Berti.F, O. Pereira, T. Peters, and F. X. Standaert,(2017) ''On leakageresilient authenticated encryption with decryption leakages,'' IACR Trans. Symmetric Cryptol., vol. 2017, no. 3, pp. 271– 293.
3. Casemore .B,(2019) ''Network modernization: Essential for digital transformation and multicloud,'' IDC, Framingham, MA, USA, White Paper US45603019.
4. Chillotti .I, N. Gama, M. Georgieva, and M. Izabachène,(2020) ' TFHE: Fast fully homomorphic encryption over the torus,'' J. Cryptol., vol. 33, no. 1, pp. 34–91.
5. Dong .C, K. Yang, J. Qiu, and Y. Chen, (2019)''Outsourced revocable identity– based encryption from lattices,'' Trans. Emerg. Telecommun. Technol., vol. 30, no. 11, p. e3529.
6. Du .L, K. Li, Q. Liu, Z. Wu, and S. Zhang, (2020)''Dynamic multi-client searchable symmetric encryption with support for Boolean queries,'' Inf. Sci., vol. 506, pp. 234–257.
7. Ge .C, W. Susilo, L. Fang, J. Wang, and Y. Shi, (2018)''A CCA-secure key-policy attribute-based proxy re-encryption in the adaptive corruption model for dropbox data sharing system,'' Designs, Codes Cryptogr., vol. 86, no. 11, pp. 2587–2603.
8. Hu .C, R. Yang, P. Liu, T. Li, and F. Kong,(2019) ''A countermeasure against cryptographic key leakage in cloud: Public-key encryption with continuous leakage and tampering resilience,'' J. Supercomput., vol. 75, no. 6, pp. 3099–3122.
9. He .D, N. Kumar, S. Zeadally, and H. Wang,(2018) ''Certificateless provable data possession scheme for cloud-based smart grid data management systems,'' IEEE Trans. Ind. Informat., vol. 14, no. 3, pp.

1232–1241.

10. Lee .K,(2020) ‘‘Comments on ‘Secure data sharing in cloud computing using revocable-storage identity-based encryption,’’ IEEE Trans. Cloud Comput., early access, doi.
11. Wang .F, L. Xu, K.-K.-R. Choo, Y. Zhang, H. Wang, and J. Li, (2020)‘‘Lightweight certificate-based public/private auditing scheme based on bilinear pairing for cloud storage,’’ IEEE Access, vol. 8, pp. 2258–2271.
12. Wang .X, T. Luo, and J. Li,(2018) ‘‘A more efficient fully homomorphic encryption scheme based on GSW and DM schemes,’’ Secur. Commun. Netw., vol. , pp. doi: 10.1155/2018/8706940.
13. Wang .X, X. Cheng, and Y. Xie,(2020) ‘‘Efficient verifiable keyaggregate keyword searchable encryption for data sharing in outsourcing storage,’’ IEEE Access, vol. 8, pp. 11732–11742.
14. Wei .J, W. Liu, and X. Hu,(2018) ‘‘Secure data sharing in cloud computing using revocable-storage identity-based encryption,’’ IEEE Trans. Cloud Comput., vol. 6, no. 4, pp. 1136–1148.
15. Xiong .H, H. Zhang, and J. Sun,(2019) ‘‘Attribute-based privacypreserving data sharing for dynamic groups in cloud computing,’’ IEEE Syst. J., vol. 13, no. 3, pp. 2739–2750.