INNOVATIVE SECURITY TECHNOLOGY FOR KYC DOCUMENTS RECORD MAINTENANCE

Abstract

Data security is a practice of physical, administrative and data security. It also includes organizational policy and procedures. Data security is extremely important. Hashing plays a very important role in data security. Hashing is the process of translating a given key into a code. A hash function can be used to generate a new digital fingerprint. Hash algorithms are typically used to show that the file being was not changed by an intruder. Data Hash ensures that you are verified every time you interact with a business. It offers a detailed audit trail, so there’s no uncertainty when it comes to your information. In this paper, we proposed the concept for the secure concept of the KYC verification.

Keywords: Data security, hashing, KYC documents

Authors

Prati Jain
M.Tech. Scholar
Professor,Department of Computer Science &Engineering
Arya College of Engineering and IT, Jaipur, India
jainprati24@gmail.com

Dr. Vishal Shrivastava
M.Tech. Scholar
Professor,Department of Computer Science &Engineering
Arya College of Engineering and IT, Jaipur, India
vishal500371@yahoo.co.in

Dr. Akhil Pandey
M.Tech. Scholar
Professor,Department of Computer Science &Engineering
Arya College of Engineering and IT, Jaipur, India
akhil@aryacollege.in

I. INTRODUCTION

Security measures on your data provide protection from corruption and unauthorized access. Security is broken into different categories including where hardware and software are stored, the devices they're on, any data's shared online. An organizations policies and procedures are located in the data security sector. Companies use tools in this field to mask, encrypt or redact sensitive data from cyber security issues and company regulations. These tools also help you audit and simplify data. [1]

Recently, cyber security experts have been emphasizing the importance of robust data security management. This process helps to protect information against cyber attacks and minimize the risk of human error, which is often the cause of data breaches. While organizations are legally obligated to protect data, there are many regulations that outline these requirements. Organizations in California need to comply with the California Consumer Privacy Act (CCPA), EU organizations must comply with the General Data Protection Regulation (GDPR) and PCI organizations also have to follow the payment card industry data security standard (PCI DSS).[1]

Organizations are legally responsible for 100% data protection and must follow laws such as HIPAA, CCPA, and GDPR in order to do so. Organizations need to prevent their customers from having their data stolen or misused in order to comply with law. These include the European Union's General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA) and the Payment Card Industry Data Security Standard (PCI DSS). [1]

Nomenclature

A radius of

B position of

C further nomenclature continues down the page inside the text box

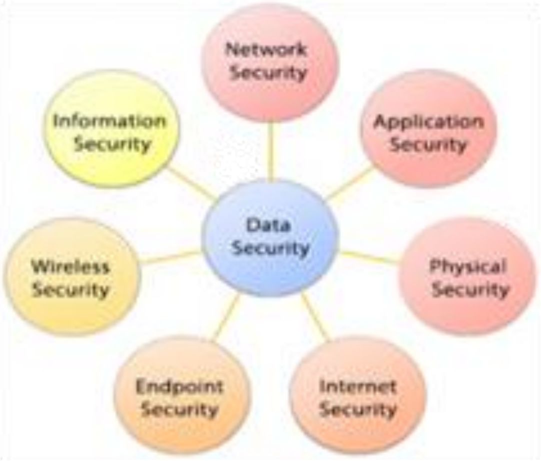


Figure 1: Data Security [Ref. omegasecure.com]

If sensitive data is breached, companies are more likely to experience financial losses and lose customer trust. To avoid the financial repercussions of a data breach and tarnished reputation, it is important for companies to protect their data. In today's society, data security is important. They have all the data from customers and from hospitals like your patient's records. With the right software and hardware in place, hackers can't get it. [2]

You shouldn’t have a data security strategy only because it’s good for business. Data breaches are happening regularly, so you will be ahead of your competition if you are in control of your data. [2]

You need to protect your data from both internal and external corruption because it generates, acquires, saves and exchanges. Companies have to value data as a vital asset. Data can be used for any number of purposes, such as generating, acquiring, saving and exchanging. However, by protecting the data from corruption and illegal access, companies protect themselves from financial loss and brand erosion. [2]

However, because of the government's and industry's regulations, it is necessary for companies to do whatever possible to ensure that they are complying with what is going on. Companies must be in compliance with these regulations and restrict their data to the country they do their business. The government and industry initiated these laws, which makes it critical for companies to comply. [3]

All companies should adopt the CIA triad, which consists of three principles. The confidentiality of the data, integrity of the data (ensuring it’s only use is for what it was intended), and availability of the data for all users.

II. LITERATURE SURVEY

H. Adamu, et al. 2022 [4] specified that Text passwords are easily hackable. They can be simple, which in turn leads to users using passwords they had used before. The most secure passwords usually come with the difficulty of memorizing it. That's why most users opt for alphanumerical passwords, which are easy to remember but are also often vulnerable to brute-force and keylogger attacks because they are based on text. A possible new substitution is graphics passwords that take advantage of a person's ability to remember pictures better than text. Existing graphical passwords are vulnerable to shoulder surfing assault. To solve these security flaws, this paper proposes a new authentication method for online applications that uses combinations of one-time passwords, textual, and graphical passwords. The efficacy of the solution was confirmed by usability testing and security analysis.30 people participated in the system evaluation, and all of them found the system to be easy to use, friendly and secure. The security standards were being met. The conclusion from this study is that the system was more usable, with enhanced security when compared to the traditional authentication solutions.

W. Han, et. al 2020 [5] state that Data-Driven password guessing techniques, such as PCFGs (Probabilistic Context-free Grammars), are not efficient in tackling long passwords. The reason is that if the system lacks large-scale training data it will be ineffective for guessing long passwords. For example, without having large-scale data on a specific password type, the system will be unable to accurately guess all of the possible combinations and even if there is adequate large-scale data available, the accuracy of the system depends on how well it has been trained on this data and consequentially obtaining accurate guessing performance for long passwords becomes challenging. Short passwords have a limited vocabulary, so we propose that the grammars of PCFGs be transferred to facilitate guesses on long passwords. TransPCFG is superior to PCFG_v4.1 in producing passwords with a 23.30% guess rate and 56.10% accuracy. This compared to PCFG_v4.1 with an average of 18% and 46%. The more segments you have in a password, the harder it will be to guess. In order to create a long and safe password, you should use four or more segments instead of two-character classes like symbols and numbers that have been popularized.

F. Z. Glory, et al. 2019 [6] People use online services to store their personal information. There has been a number of hacking attempts which make many people worry about the security and privacy of their data. The password authentication system is one of the many ways to protect individual data. With increased internet usage, password security and authenticity have become more necessary and important. But authors recommend using a complicated password to make it more difficult to crack. Our new algorithm makes strong passwords without the need of complicated patterns. This is a password-based system. The inputted information, like words and numbers, will make up the password. We've tested this system by using inputs to the program. Their passwords are strong, we have checked them with 4 popular online password checkers. They have examined that generated passwords can defend against two password cracking attacks. With the Python programming language, they have implemented the system. In the near future, an online interface is planned to be developed so it is free and accessible to everyone. Their first-strategy of generating strong passwords is user-friendly and optimized for being strong as well as being harder to crack than a normal password.

III. PROBLEM STATEMENT

Each of the organization and department required KYC documents for the verification of customer and clients. So, the redundant copies will create the mess as well as the unsecure photocopies and documents can be misused. In order to cater the requirement for the safe and secure KYC document system for the centralized document sharing, which aims for the verification of the user accessing the documents and secure sharing of the documents. So, we are taking the basis of the hash pattern generation for the user validations and also the document sharing.

IV. PROPOSED ALGORITHMS

The proposed algorithm is divided into the two segments first one is authentication and second document sharing.

1. Algorithm for User Enrolment using Encrypted Image

Stage 1: Input User Name

Stage 2: Select the Image which to be encrypted to be used as password

Step 3: Select the Hash algorithm SHA or MD5 which is to be used for the generation of the hash code on the basis of the select image.

Step 4: Set PASSCODE=HASH(1:20) , where HASH is the hash generated on the basis of the SHA or MD5 algorithm selected by user.

Stage 5: Encrypt the Image using the Key generated on the basis of Dimensions of Image and using XOR encryption encrypt the image.

Step 6: Store all the details like user name , encrypted image path , PASSCODE and other details in the table for the registered user.

Step 7: STOP

2. Algorithm for User Validation using Decrypted Image

Step 1: Read Username and PASSCODE.

Step 2: Fetch the Encrypted Image from Database.

Step 3: Decrypt the image.

Step 3: If Details verified in Database then

Grant Access

Else:

Print “Invalid Login Details”

[End of if Structure]

Step 4: End

3. Uploading KYC Data

Step 1: Read the File.

Step 2: Read credentials.

Step 3: Generate SHA-512 hash for File and store in SHAORGFILE.

Step 4: Generate SHA-512 hash for User Data Info and store in SHADATAF.

Step 5: Extract first 20 characters of SHAORGFILE and first 20 characters of SHADATAF and store them in COMBKEY.

Step 6: Perform data steganography using XOR of data and original file.

Step 7: Save the Encoded file.

Step 8: Store the details regarding the data transmission of stegno file in the table corresponding to it in the database,

4. Receiving KYC Data

Step 1: Input the Encoded File.

Step 2: Read Message Sequence Number

Step 3: Read the COMBKEY.

Step 4: If Details Verified in Database then:

- Fetch the Original File from Database.
- Convert to the binary Data.
- XOR operations in the Original and Encoded File.
- Show the Decoded File to the user.

Else:

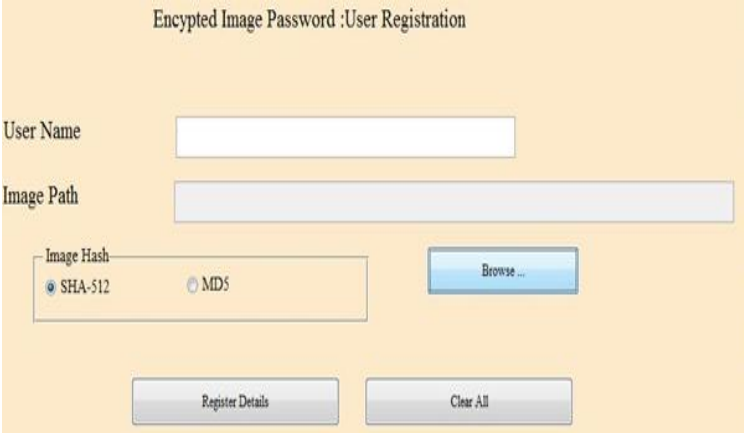
Print “Invalid Details”

[End of If structure]

Step 5: End.

V. IMPLEMENTATIONS

The implementation is performed in the Visual Studio and SQL Server as the database and the some of the forms of the implementation are shown below, in which we have the user registration form.

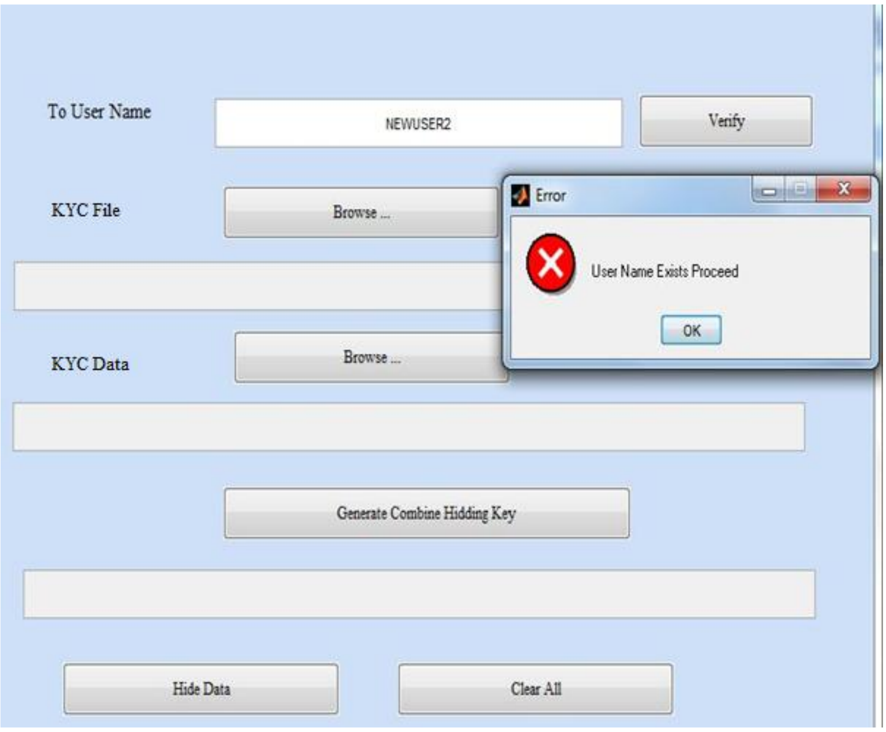


Encypted Image Password :User Registration

Image Hash: ● SHA-512 ○ MD5

Figure 2: Registration of Users

In the fig 2, The hash key (a.k.a. "PASSCODE" or "passcode") is generated based on the username, image and selected hash algorithm. The first 20 characters from the hash code are stored in the database to validate user login process by encrypting and storing details of the encryption in a database for future use.



To User Name: NEWUSER2

Error

User Name Exists Proceed

Figure 3: Upload KYC

The fig 3 shows the Upload KYC documents for the centralized sharing of the data. This concept is about the assembly of COMBKEY. The first step in generating COMBKEY is a hash calculation with an algorithm of SHA-512. With the original image file, a hashed key SHA-512 is generated. The same calculation occurs with the data file and another key (COMBKEY) is calculated by extracting the 20 first characters from both hashes.

VI. RESULT ANALYSIS

The result analysis is performed on comparison with the base papers,

Table 1: An example of a table. (FarhanaZaman Glory et. Al 2019)

OTP	Website/Tool	Result
{urBn17iRfan- 1	Cygnius Password Strength Test	67 bits Entropy
{urBn17iRfan- 1	Password Blue zxcvbn	49 bits
{urBn17iRfan- 1	Cryptool2	Bit Strength 92

Table 2: Base Paper Result (W. Han, et. al 2020)

OTP	Website/Tool	Result
12zxcvbnword1997	Cygnius Password Strength Test	27 bits Entropy
12zxcvbnword1997	Password Blue zxcvbn	39 bits
12zxcvbnword1997	Cryptool2	Bit Strength 68

Table 3: Proposed Work Results

OTP	Website/Tool	Result
12940972653082271991-@-11247074755874897958-#->	Cygnius Password Strength Test	162 bits Entropy
12940972653082271991-@-11247074755874897958-#->	Password Blue zxcvbn	144 bits
12940972653082271991-@-11247074755874897958-#->	Cryptool2	Bit Strength 110

VII. CONCLUSION

One important task for all data analysts is protecting company data from internal and external sources. Data corruption can lead to significant financial loss, loss of consumer trust, and a loss of brand emotion. Data security is the practice of locking down and securing any digital data, preventing potential manipulation or unauthorized access. This includes encrypting customer data at rest and in transit and not having security policies in place to protect your sensitive information from potential breaches. The concept of KYC document sharing found the more secure when compared with the existing approaches of the user validation and verification. This secure concept not only add on the security in the data exchange but also reduce much of the burden of document management.

REFERENCES

[1] Erdem, E., & Sandikkaya, M. T. (2019). OTPaaS—one time password as a service. IEEE Transactions on Information Forensics and Security, 14(3), 743–756.

[2] Golla, M., & Dürmuth, M. (2018). On the accuracy of password strength meters. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security.

[3] Grassi, P. A., Garcia, M. E., & Fenton, J. L. (2017). DRAFT NIST special publication 800-63-3 digital identity guidelines.

[4] Adamu, H., Mohammed, A. D., Adepoju, S. A., & Aderiike, A. O. (2022). A three-step one-time password, textual and recall-based graphical password for an online authentication. 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON), 1–5.

[5] Han, W., Xu, M., Zhang, J., Wang, C., Zhang, K., & Wang, X. S. (2021). TransPCFG: Transferring the grammars from short passwords to guess long passwords effectively. IEEE Transactions on Information Forensics and Security, 16, 451–465.

[6] Glory, F. Z., Ul Aftab, A., Tremblay-Savard, O., & Mohammed, N. (2019). Strong password generation based on user inputs. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 0416–0423.

[7] Li, Y., Wang, H., & Sun, K. (2016). A study of personal information in human-chosen passwords and its security implications. IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, 1–9.

[8] Li, Z., Han, W., & Xu, W. (2014.). A large-scale empirical analysis of Chinese web passwords. Usenix.org.

[9] Ma, J., Yang, W., Luo, M., & Li, N. (2014). A study of probabilistic password models. 2014 IEEE Symposium on Security and Privacy, 689–704.

[10] Melicher, W., Ur, B., Segreti, S. M., Komanduri, S., Bauer, L., Christin, N., & Cranor, L. F. (2016). Fast, lean, and accurate: Modeling password guessability using neural networks. Usenix.org.

[11] Vaddeti, A., Vidiyala, D., Puritipati, V., Ponnuru, R. B., Shin, J. S., & Alavalapati, G. R. (2020). Graphical passwords: Behind the attainment of goals. Security and Privacy, 3(6)..