Applications of Cryptography in Database: A Review

Huanhao Xu
Management and Systems
New York University, SPS
New York, USA
hx798@nyu.edu

Kutub Thakur
Professional Security Studies
New Jersey City University
NJ 07305, USA
kthakur@njcu.edu

Abu S. Kamruzzaman
Seidenberg School of CSIS
Pace University
NY 10570, USA
ak91252p@pace.edu

Md Liakat Ali
Department of Computer Science & Physics
Rider University
NJ 08648, USA
mdali@rider.edu

2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-4067-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEMTRONICS52119.2021.9422663

***Abstract*—Cryptography is the foundation and core of network security. Privacy protection, sensitive information is particularly important, so whether it is system development or app development, as long as there is network communication. As data plays very critical steps for IOT devices it's very crucial that cryptography applied to the databases as well. A lot of information needs to be encrypted to prevent interception and tampering. Many people use cryptography on a daily basis, not everyone is aware of it. This paper investigates the applications of cryptography in the context of databases and offers suggestions to enhance security and privacy.**

Keywords— security, privacy, public key, encryption

I. Introduction

With the popularity of blockchain and cryptocurrencies, more and more people are adopting the foundation slot of blockchain-cryptography. Because cryptography is one of the keys to the safe operation of the blockchain. There exist three categories of cryptographic algorithms known as symmetric cryptography, asymmetric cryptography (public key), and cryptographic hashing. These types are part of most encryption systems, and we'll discuss each of the next and to find where the strengths and weaknesses are. In addition to this, we need to look at what kinds of risks and attacks on cryptography. Then we can talk about the application of cryptography and database encryption techniques.

Cryptography has been a science for many years, and it has been recorded that first cryptography inscriptions were carved around 1900 BC in the tomb of an ancient Egyptian nobleman, Khnumhotep II. The inscription used some strange sacred carving symbols instead of more common symbols. It is not intended to hide information, perhaps to manifest prestige by changing forms.

In 100 B.C., at the height of the Roman Empire, Caesar used encryption to send secret messages to front-line troops. The "Caesar Code" is probably the most mentioned historical password in the literature. In a replacement password, each character in clear text is replaced by a different character to form a redaction. The variant used by Caesar is password conversion in three locations. Each letter is forwarded with three locations so that the letter A is replaced by the letter D, the letter B is E, and so on. The letter X is replaced by A.

In the early 1970s, IBM customers requested for encryption, and IBM set up an "encryption group" led by Horst-Feistel. A password called Lucifer" was created by the encryption group. In 1973, the National Standards Administration, now renamed to National Institute of Standards and Technology (NIST) proposed group passwords, which became national standards. IBM realized that lots of commercial products are purchased with no encryption support at all. Lucifer also known as DES (Data Encryption Standard) was adopted to provide that support.

DES was hacked in 1997 in a desperate search attack. The DES encryption key size was very small. As computing power grows, all the different combinations of the brute force are calculated to obtain possible clear text information. In the 1980s, there was only one option which was DES. But times have changed. Today, we have more choices, stronger algorithms, faster, and better design. Now, the question is how to choose.

In 1997, NIST again proposed a new group password scheme and 50 proposals came. In 2000, Rijndael was adopted and renamed to AES (Advanced Encryption Standard) [1-2]. In order to provide security in the database, this paper investigates the applications of cryptography in the context of databases and then offers suggestions that can provide security and privacy. The organization of the paper is as follows: section II presents cryptography definition and types, section III presents some of the existing cryptographic algorithm that provide security to the database, section IV is the analysis of cryptography, section V shows the method of analysis, section VI offers some suggestion to enhance security and privacy in the database and section VII is the conclusion.

II. The Definition of Cryptography

Encryption is converting data to make it unrecognizable and useless to unapproved persons; the most secure technique is to use mathematical algorithms and variable values called "keys". The keys you choose, in general, are random strings that are entered through encryption and are integrated into the data transformation. To decrypt the data, you must enter the exact same key.

Cryptography begins with plaintext which is not encrypted. The plaintext is converted into ciphertext with encryption as shown in fig. 1. This ciphertext can be converted back to usable plaintext using decryption. The encryption and decryption is based upon the type of cryptography scheme being employed and some form of key. This process is sometimes written as follows:

$$C = E_k(P) \qquad (1)$$

978-1-6654-4067-7/21/$31.00 ©2021 IEEE

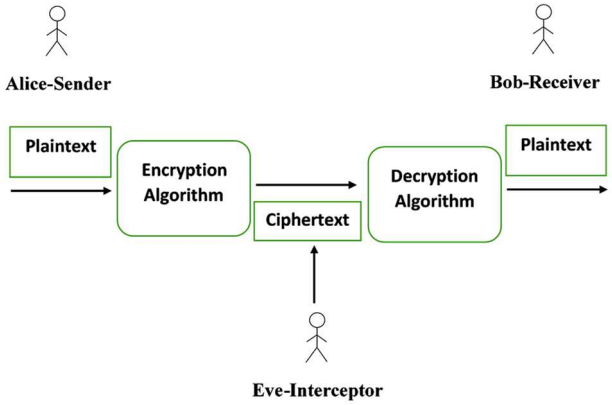

Fig. 1. Cryptography process

$$P = D_k(C) \qquad (2)$$

P = plaintext, raw data, data to be encrypted
C=ciphertext, some camouflage or transformed output of clear text
E=encryption process, the process of turning clear text into a red tape in some way.
D=decryption process, the process of restoring redaction into clear text
k=the key, specialized tools used in encryption or decryption

Cryptography resembles similar mathematical algorithms for encrypting and decrypting messages, while the scientific analysis of cryptography and the destruction of encryption schemes. Cryptography is a secret term that refers to a wide range of research writing, including encryption and cryptographic analysis [3].

Three types of cryptography algorithms are discussed here:

A. Symmetric Cryptography

When making clear or secret edict transformations of information, the password system using the same key is solved and decrypted. Figure 2 shows the symmetric cryptography where both Alice and Bob share the same key.

Security in this scheme depends on the secret key and the encryption algorithm security.

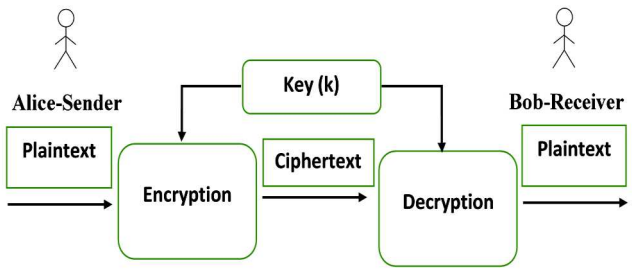

Fig. 2. Symmetric Cryptography

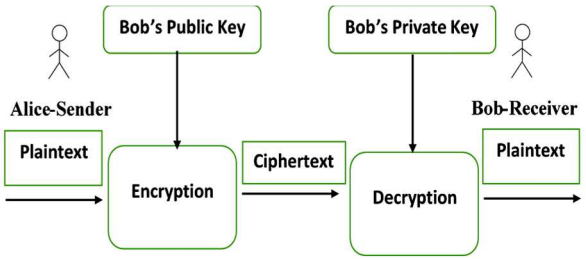

Fig. 3. Asymmetric encryption

Advantages: Algorithm open, fast, high level of secrecy, small footprint.

Disadvantages: Key distribution and management is complex.

Purpose: Encryption with large amount of information.

Representing algorithms: DES algorithm, 3DES algorithm, IDEA algorithm, AES algorithm.

Problem: If the receiver forges a message and falsely is sending it, the sender cannot excuse that it cannot resolve the confirmation of the message and cannot achieve digital signature.

B. Public-Key Cryptography

Encryption and decryption keys are not the same password system when transforming information with clear or secret text. In the asymmetric password system, every user uses one key for encryption and another key for decryption. The key used for encryption is called a public key and the key used for decryption is called private keys which is a secret key for each user. Figure 3 shows the asymmetric encryption where Alice encrypts the plaintext using Bob public key and Bob decrypt the ciphertext using his own private key [4].

Advantages: The sender and the receiver do not require to set up a secure channel to exchange their key, the key space is generally small which reduces the challenges in the key management procedure.

Disadvantages: Slow implementation, not suitable for heavy communication load situation

Purpose: Encrypting critical, core confidential data.

Representing algorithms: RSA algorithm, ElGamal algorithm, elliptic curve encryption algorithm.

Problem: Because the public key is open to the public, if a person uses his own public key encrypted data to send to us, we cannot determine who sent it. If we use the private key to encrypt the data, anyone who knows our public key can decrypt our data.

C. Cryptographic Hashing

The cryptographic hashing is a hash function that is suitable for encryption. Hashing is mainly used to provide integrity property which ensures that the information is correct and no unauthorized person or malicious software has altered the data. This is a mathematical algorithm of any size, a string of map

data fixed size and one-way functions, that is, a function is actually not feasible conversion. Hashing algorithms are useful to protect data from unauthorized modification.

III. The Encryption Algorithms

Day by day people's awareness of the importance of information security is growing with the development of information technology and digital society. In 1997, the National Bureau of Standards announced the implementation of the "American Data Encryption Standard (DES)". Civil forces began to fully intervene in the research and application of cryptography, and the use of encryption algorithms such as DES, RSA, SHA, etc. increased. With the increasing demand for encryption strength, AES, ECC, and so on have recently emerged [4-7].

The following section shows the comparison of the two main types of encryption algorithm used in symmetric systems - DES and AES and two encryption algorithms used in asymmetric encryption algorithms- RSA and DSA.

A. Data Encryption Standard (DES)

DES encryption algorithm is a grouping password, with 64 bits for grouping data encryption and key length is 56 bits. Both encryption-decryption uses a similar algorithm. The DES encryption algorithm is the secret of the key, while the public algorithm, including the encryption and decryption algorithm. In this way, only those who have the same key as the sender can interpret the redaction data encrypted by the DES encryption algorithm. Therefore, the DES deciphering encryption algorithm is the encoding of the search key. For a 56-bit-long key, if searched by the method of exhaustion, the number of operations is 2 of 56 squares [8-9].

B. Advanced Encryption Standard (AES)

In the late 20th century, it is the rapid development of computers, component manufacturing process progress makes the computer processing power much more strong, DES will not provide enough security. On January 2, 1997, the National Institute of Standards and Technology and Technology: NIST announced its desire to recruit advanced encryption standards: AES), to replace DES. The U.S. federal government adopted this block encryption algorithm as a standard as well as has been analyzed by multiple parties and widely used worldwide.

AES encryption is an advanced encryption standard algorithm in cryptography, the encryption algorithm adopts the symmetric packet password system, the key length is supported by at least 128, 192, 256, the group length is 128 bits, and the algorithm should be easy to implement a variety of hardware and software.

The cryptographic algorithm requirements are reversible so that the decryption algorithm can correctly recover the cleartext. Take AES, in the case of key fixing, clear text and redaction are one-to-one correspondence throughout the input space.

Therefore, the various parts of the algorithm are also reversible, and then the sequence of operation of each part is designed to be reversible, the redaction can be correctly decrypted. DES and AES algorithm comparison summary highlighted in Table 1.

TABLE I. Comparison of DES and AES Algorithms

Type	*Key Length (bits)*	*Security*	*Computing Speed*	*Resource Consumption*
DES	56	Low	Medium	Medium
AES	128 192 256	High	High	Low

C. Rivest Shamir Adleman (RSA)

RSA encryption algorithm is the most influential public key cryptography algorithm and is generally regarded as one of the best public key schemes. It was proposed by Ron Rivest, Adi Shamir and Leonard Adleman in 1977. All three were working at the Massachusetts Institute of Technology at the time. RSA is made up of the initial letters of their three surnames. RSA has been recommended by ISO as the first algorithm that can be used simultaneously for both encryption and number signature and is resistant to all known password attacks so far and RSA encryption algorithm used as a public key data encryption standard. RSA uses a very simple numerical fact which is to multiply two large prime numbers, but very difficult to decompose its product type where the product can be exposed as an encryption key.

D. Digital Signature Algorithm (DSA)

DSA is based on integer finite domain discrete pairs, an important feature of DSA is that two prime numbers are exposed, so that when using someone else's p and q, even if you do not know the private key, you can confirm whether they are random, or do it. This is not possible with the RSA algorithm. Compared to RSA, DSA is used only for signatures, while RSA can be used for signature and encryption. Table II shows the comparison between RSA and DSA.

TABLE II. Comparison of RSA and DSA Algorithms

Type	*Maturity*	*Security*	*Computing Speed*	*Resource Consumption*
RSA	High	High	Low	High
DSA	High	High	High	Only for Digital Signature

IV. The Analysis of Cryptography

According to the attacker's knowledge of the clear text, redacted and other information, password analysis are four types: ciphertext-only attack, known-plaintext attack, chosen-plaintext attack, and chosen-ciphertext attack.

A. Ciphertext-only Attack

The attacker had no supporting information in his hands other than the intercepted message. Secret attacks are one of the most common types of password analysis and the most difficult.

This way can be used to attack both the symmetric password system and the asymmetric password system.

B. Known-Plaintext Attack

The attacker uses some ciphertext and also knows the relationship between the part of plaintext and ciphertext. For example, if you are in a conversation that follows a communication protocol, because the protocol uses a fixed keyword, such as "login" and "password", analysis can determine the redaction corresponding to that keyword. If the transmission is a legal document, unit notice, and other types of documents, because most of the documents have a fixed format and some agreed text, in the interception of more documents, you can infer some text, phrases corresponding to the secret.

C. Chosen-Plaintext Attack

The attacker knows the encryption algorithm and can select the clear text and get the redaction corresponding to the corresponding clear text. This is a more common type of Mima analysis. For example, an attacker who intercepts valuable redaction and acquires an encrypted use device and enters any clear text into the device to obtain the corresponding reduction is based on the attacker's attempt to crack the valuable redaction. Select clear text attacks are often used to crack information content encrypted with a public key password system.

D. Chosen-Ciphertext Attack

The attacker knows the encryption algorithm and can select a redaction and get the corresponding clear text. Using this method of selecting a redaction attack, the attacker's target is usually the key used by the encryption process. Digital signatures based on a public key password system are vulnerable to this type of attack.

V. The Method of Analysis

From the analysis of the password, three methods can be used in the process of password analysis, namely, the method of poor attack, the statistical analysis method, and the mathematical analysis method.

A. Poor Attack Method

The idea of cracking the poor attack method is to try all the possibilities to find out the clear text or key. The poor-lifting attack method can be divided into two categories, the poor-lifting key, and the poor-lifting explicit text. The poor key refers to the attacker, in turn, using various possible decryption keys to intercept the secret text, to try to translate, if a decryption key can produce meaningful clear text, then the corresponding key is the correct decryption key. The poor text means that an attacker encrypts all possible clear text while keeping the encryption key unchanged, and if the result of a piece of clear text encryption is consistent with the intercepted message, the corresponding clear text is the message sent by the sender [10-11].

In order to combat the attack, the modern cryptographic system is often designed by expanding the key space or improving the complexity of encryption and decryption algorithms. When the key space is expanded,

- The method of raising the key in the process of cracking need to try more decryption key,
- Improve the complexity of the encryption, decryption algorithm,
- Will enable the attacker whether the use of a poor key or poor method of the password system to crack,
- Each crack attempt requires a higher computational overhead for a perfect modern cryptographic system, and
- The cost of using poor attack methods to crack is likely to exceed the value of a secret edit.

B. Statistical Analysis Method

Statistical analysis is a method to crack by analyzing the statistical laws of clear and text. Some classical cryptographic system encrypted information, secret letters and letter combination of statistical laws and clear text is exactly the same, such a cryptographic system is easy to be cracked by statistical analysis. The statistical analysis method first needs to obtain the statistical law of the dense text, on the basis of which, the statistical law of the dense text is compared with the known clear statistical law, and the correspondence of the clear and dense text is extracted, and then the secret text is cracked.

In order to combat statistical analysis attacks, the password system should be designed to avoid the consistency of the clear text in the statistical law, so that the attacker cannot analyze the statistical law of the paper to infer the clear text content.

C. Mathematical Analysis Method

Most modern cryptographic systems take mathematical problems as the theoretical basis. Mathematical analysis refers to the method by which an attacker solves an unknown amount such as deciphering a key by mathematically using some known quantities, such as the correspondence of some clear and texts, against the mathematical basis and cryptography characteristics of the cryptography. Mathematical analysis is an important way to crack a password system based on mathematical difficulties.

VI. Database And Cryptography

In this part, this paper researches on the database about cryptography. Some knowledge related to cryptography is studied and the combination of this knowledge and database is used [5, 12-13].

The database is a storehouse of data that is prepared, deposited, and managed according to the data structure, and data is the most central property in information systems. A database, like the human brain, is at the heart of all information systems. Once the brain is damaged, it is bound to affect the body function of the whole person. Similarly, if data is lost, destroyed, or leaked in the database, it is bound to cause incalculable damage to the enterprise. Therefore, the encryption protection of data in the database has become an important part of database security. There are three main dangers to the database:

Loss of availability - Loss of availability means that legitimate users cannot use database objects.

Loss of integrity - Integrity loss occurs when a database accidentally or maliciously performs unacceptable operations.

This can happen when you create, insert, update, or delete data. The outcome is corrupted data which results in incorrect decisions.

Loss of confidentiality – It is due to illegal or unintended exposure to confidential information. Loss of confidentiality can lead to unlawful practices, security coercions, and harm to public confidentiality.

Access control – It is a security mechanism of the database management system (DBMS) to prevent unapproved access. After the login process is cleared only through a valid password-protected user account, the user can access the database.

So we will need to take steps to control the risks. The first step is traffic control. It is a distributed system containing a large amount of traffic from one site to another, as well as within a site. The second step is traffic control prevents data from being transmitted as unapproved agents.

The flow policy outlines the channels where information can flow and describes the security classes for data and transactions.

The last step is data encryption refers to encoded sensitive data transmitted on a public channel. Since data is in an incomprehensible format an unauthorized agent can't understand even if he gets access to the data.

How is the database is encrypted? In most cases, the database will use transparent data encryption. For example, the Oracle database used a lot for the encryption as shown in Fig. 4. Database transparent encryption refers to the encryption and decryption of data in the library, the access to the database program is completely unaware. In particular, the application system, which does not need to make any modifications and compilation, can be directly applied to the encryption library.

Two types of data types are unstructured data and structured data. Examples of unstructured are documents and pictures, and the example of structured data is data in a database. Data in both types are important and require encryption protection. Structured data is usually hosted centrally and contains valuable business-sensitive information, so it is especially important to protect it with encryption.

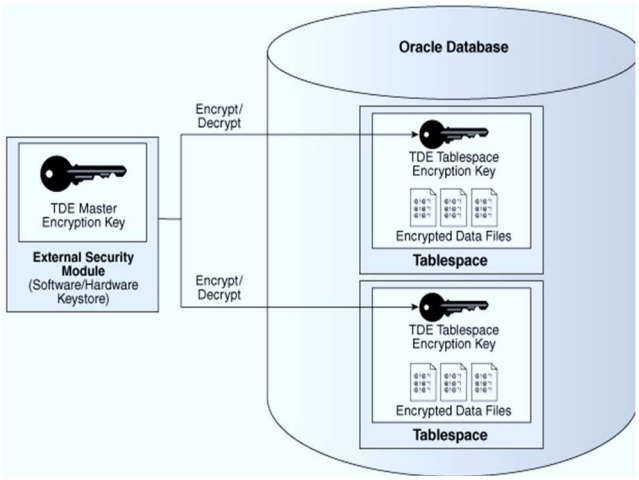

Fig. 4. Cryptography in Oracle Database [14]

Even though the process of decrypting will damage the efficiency of the use of the database the encryption of the database is still necessary to protect the necessary measures to avoid the harsh reality of frequent leakage of sensitive data. Encryption of the Database will increase security significantly. Encryption prevents data from data leakage and malicious destruction and stores data in a confidential manner and presents direct access to the data.

The majority of the encryption solutions application code requires calls to cry forward functions. This is costly because it often requires a deeper knowledge of the application and expertise to program and maintain software. In general, most enterprises spend minimal time on this or don't have relevant expertise to modify existing applications and invoke encryption routines. Oracle transparent data encryption uses nested encryption capabilities deep into the Oracle database to solve the encryption problem.

Application logic executed through SQL does not need to be changed and still works. It can be further explained that the application doesn't need to worry about encryption and decryption. The programmer will write the usual code to insert/update or delete the date in the application table. Oracle database will encrypt data as it writes to the disk and similarly decrypts reading data from disk to maintain the functionality of the application. This is important because current applications typically expect unencrypted application data. Displaying encrypted data can at least confuse application users and even break existing applications [15-16].

VII. Conclusion

Nowadays 5G is taking over its predecessor, the cloud era is coming, research suggests that the future of IT architecture, more is cloud-based design. If more data is stored on the cloud, then if cloud vendors steal and analyze user data, it will seriously violate our privacy. Malicious DBAs and developers, as also mentioned in the threat model, tend to have higher database permissions, even if they don't have permissions, and if they have a way to read the cache, the data will also leak. So how do we protect our data?

Therefore, the current performance of homomorphic encryption cannot meet the normal needs, if commercial to database level, this survey suggests that the need for further study by cryptographers. Encryption is only a small part of database security, more content needs to be the joint efforts of everyone, but also hope to see more people engaged in the field of database security in the future.

References

[1] K. Thakur, M. Qiu, K. Gai, and M. L. Ali, 2015, November. An investigation on cyber security threats and security models. In 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing (pp. 307-311). IEEE.

[2] C. Paar, and J. Pelzl, , 2009. Understanding cryptography: a textbook for students and practitioners. Springer Science & Business Media.

[3] G. C. Kessler. "An overview of cryptography." *the Handbook on Local Area Networks, Auerbach* (1998).

[4] V. Gorbach, M. L. Ali, and K. Thakur, 2020, September. A Review of Data Privacy Techniques for Wireless Body Area Networks in

Telemedicine. In 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (pp. 1-6). IEEE.

[5] U. Maurer. "The role of cryptography in database security." In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pp. 5-10. 2004.

[6] M. A. Saleh, N.M. Tahir, E. Hisham, and H. Hashim, 2015, April. An analysis and comparison for popular video encryption algorithms. In 2015 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 90-94). IEEE.

[7] Kadhem, Hasan, Toshiyuki Amagasa, and Hiroyuki Kitagawa. "A novel framework for database security based on mixed cryptography." In *2009 Fourth International Conference on Internet and Web Applications and Services*, pp. 163-170. IEEE, 2009.

[8] L. Thakur, S, Kopecky, M. Nuseir., M. L. Ali, and M. Qiu, 2016, June. An analysis of information security event managers. In 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud) (pp. 210-215). IEEE..

[9] Ciampa, Mark. "Guide to Network Security." *URL https://books. google. com/books/about/Security+ _Guide_to_ Network_Security_Fund. html*.

[10] K. Thakur, M. L. Ali, N. Jiang, and M. Qiu. "Impact of cyber-attacks on critical infrastructure." In *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, pp. 183-186. IEEE, 2016.

[11] K. Thakur, M. L. Ali, K. Gai, and M. Qiu. "Information security policy for e-commerce in Saudi Arabia." In *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, pp. 187-190. IEEE, 2016.

[12] D. Deshmukh., A. Pasha, and D. Qureshi, 2013. Transparent Data Encryption--Solution for Security of Database Contents. arXiv preprint arXiv:1303.0418.

[13] K. Thakur, M. L. Ali, S. Kopecky, A. Kamruzzaman, and L Tao. "Connectivity, Traffic Flow and Applied Statistics in Cyber Security." In *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, pp. 295-300. IEEE, 2016.

[14] D. Adams, 2014. Oracle Database Online Documentation 12c Release 1 (12.1). Application Development.. Available at https://docs.oracle.com/database/121/ASOAG/introduction-to-transparent-data-encryption.htm#ASOAG10117. Last accessed-Feb 4, 2021

[15] G. J. Simmons., 1979. Symmetric and asymmetric encryption. ACM Computing Surveys (CSUR), 11(4), pp.305-330.

[16] L. Li, K. Thakur, and M. L. Ali, 2020, September. Potential Development on Cyberattack and Prospect Analysis for Cybersecurity. In 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (pp. 1-6). IEEE.