iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38175

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Cryptography: Data Encryption, Decryption and Compression, Database Privacy and Security, Zero Knowledge of Proof

Aksa Isac John[1], Rishabh V. Singh[2], Jane Catherine Olagunju[3]
Sam Higginbottom University of Agriculture, Technology and Sciences
Vaugh Institute Agriculture, Engineering and Technology
B.Tech Computer Science and Engineering

Abstract: ***Evolution and modernization have brought about progress in technology and this has led to the reduction in privacy & internet security due to an increase in cybercrime and threats. As a result of this turn of events, Cryptography is now being used as a means of keeping information of any kind safe from third party individual(s). Research has shown that with the Encryption of information, third party individual(s) have no chance or less chance of getting past this security measure. Hence, Cryptographers keep improving algorithms to make it impossible for a third party to decrypt this information without the key which is where database Privacy and Security come in. The database contains all the information which is a major asset, there are encryptions which can be used at different levels to provide security.***
Lastly, for encryption algorithms which are breached by unknown third-party individual(s), the zero knowledge of proof helps to figure out the identity of this individual. They are an extremely interesting and useful construct. They are fascinating because of their definition, which is mutually opposed, their applicability is very vast in cryptography; they are used to restrict the malevolent users to work according to the protocol. Zero-knowledge serve as a good medium to understand the problems regarding cryptographic protocols.
Keywords: Cipher, Encryption, Decryption, Key, Security, Database, Zero-Knowledge

1. INTRODUCTION

Cryptography is dated back to as far as 1900 BC, 4000 years ago, discovered in the main chamber of the tomb of the nobleman Khnumhotep II, in Egypt. The Egyptians communicated using Hieroglyph which is the use of symbols and pictographs to convey secret messages. These messages were passed between empires and could not be deciphered by a third party.
Later, Scholars began using mono-alphabets which involved replacing alphabets with other alphabets with a secret rule which became the key to unscramble the meaningless message to the real message. This became a substitute for ciphers in 500 – 600 BC.
Cryptography was used between 800 – 1100 in England, throughout Europe, outside Europe; Japan, in World War II by the Germans which by then there were cipher machines.

Data Encryption, Decryption and Compression

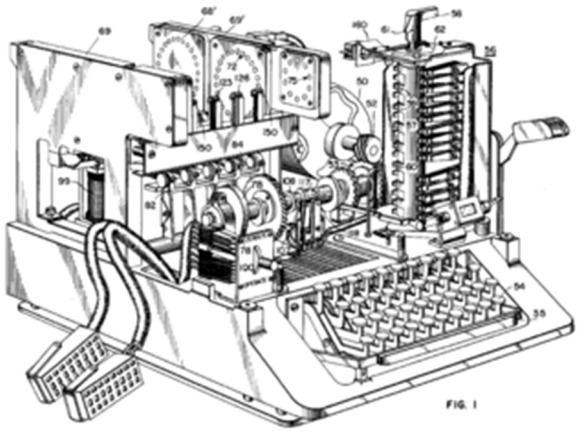
A. Cryptanalysis

This is the study of encryption and decryption of messages, information system analyzation of hidden aspects of a system. Cryptanalysis is used to gain access to a cryptographic security system without the use of a cryptographic key.
Cryptography is used to hide messages with the use of codes and ciphers as a security protocol, while cryptanalysis is used to breach past the security, decipher those codes and retrieve the hidden messages.
Al-Kindi also known as Alkindus first discovered Cryptanalysis in Risalah fi Istikhraj al-Mu'amma (A Manuscript on Deciphering Cryptographic Messages) between 801 – 873. Hence, Alkindus is known as the first codebreaker in history.

B. Cipher Machines

In the 1930s, the U.S. Army cryptologist William Friedman and his assistant Frank Rowlett drew on this simple precept to conceive a cipher machine that was easy to use, simple to rekey, and ostensibly impossible to break. Then in collaboration with the Navy, the two services went to field their device. The Army knew it as SIGABA and the Navy as ECM (Electric Cipher Machine) II. SIGABA was not only the most secure cipher machine of World War II, but it provided yeoman service for decades thereafter.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com



SIGABA, filed in 1944 but not issued until 2001.

Other cipher machines include:

1) *Germany*: Discret, Enigma, T-52, SZ-40/42, SG-41, GK-III, STG-61, H-54, Reverse, RS, FS-5000, ELCRO, Siemens, ANT, R&S, TST, Mi544
2) *UK*: Scrambler, Slidex, Typex, Rockex, 5-UCO, Portex, Singlet, Noreen, BID/610, BID/250, /980, /2510, /2200 KG-40AR, Racal, MEL
3) *USA*: Iron-key, Jefferson, Giddings, SIGABA, SIGTOT, SIGSALY, Telekrypton, Harris, Datotek etc.

Other countries where cipher machines were built include; Switzerland, USSR, DDR Yugoslavia etc.
All these machines have Rotors, Key fill, EMU (Electronic Message Unit), Mixers, Vernam, Codebook, OTP (One-Time Pad), Hotline, Cyklometr etc.

C. Cryptology

This is the study of mathematics like: Number theory, Formulas and Algorithms application that helps in Cryptography and Cryptanalysis.
For the security of data storage and transmission, it must be altered in such a way that any unauthorized individual would not be able to access it. To accomplish this, certain mathematical equations are used.
Over the years, Cryptography has evolved and now, Cryptologists use algorithms to encrypt and decrypt messages.

II. ENCRYPTION

The conversion of a plain text to a cipher text by scrambling the plain text to an unreadable one by using a key is known as Encryption.
A plain text is a readable message while a Cipher text is an unreadable one. A key in cryptography is a character string in an encryption algorithm used to alter the message to a cipher text and back to a plain text. It locks (encrypts) the data so that only someone with the right key can unlock (decrypt) it.

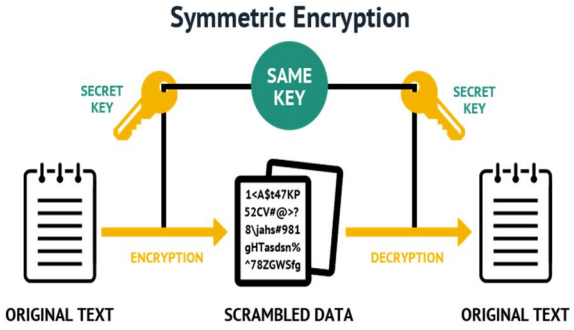
Types of Encryption

A. Symmetric Encryption

In this type of Encryption, only one key is used for both the encryption and decryption of a message.
For Symmetric Encryption (*Secret Key Encryption*) to work, the two parties must communicate secretly to know which key to use and avoid a third-party individual from discovering the key to keep the message safe. This is much faster than Public key but key management can be a huge problem.

Types of Symmetric encryption algorithms include:

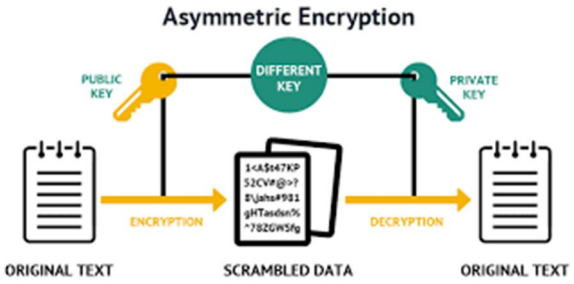
1) *Block Algorithm*: A set length of bits of data is encrypted in blocks and the data is held in the memory as it waits for the other blocks of data to be encrypted. Example: Advanced Encryption Standard (AES), Data Encryption Standard (DES), International Data Encryption Algorithm (IDEA), Blowfish (Drop-in replacement for DES and IDEA), Rivest Cipher (RC) 5, Rivest Cipher (RC) 6
2) *Stream Algorithm*: Data is not retained in the memory; it is encrypted as it streams. Example: Rivest Cipher (RC) 4



B. Asymmetric Encryption

In this type of Encryption, the key used to encrypt and decrypt are different, hence two keys are needed; one to encrypt and the other to decrypt.

In Asymmetric Encryption (*Public Key Encryption*) a pair of public key-private key is used; private key for data encryption and public key for data decryption and vice-versa.



III. DECRYPTION

The conversion of encrypted data or ciphered text back to its original form or plain text is known as Decryption.

Decryption can be done with the use of a secret key used for the encryption of the data and without the use of a key.

The process of decryption is required when a data has been encrypted for safe keeping. The decrypting of the data with the key is the safe type of decryption whereas, the decrypting of a data without the key is the kind of decryption that encryption algorithms of nowadays achieve to put an end to.

Decrypting a file of data without a key is what third party individuals do when they want to gain access into a particular file for information that isn't meant for them. Hence, this era's encryption algorithms are designed in such a way that decryption without a key proves difficult or impossible for privacy and security of the data and information.

IV. COMPRESSION

Data compression is reducing the number of bits in a data to save storage capacity and speed up the transfer of file. In other words, it is a technique used to store data in another format or shrink data in another size to save space and transmission time.

Text Compression is using a single repeat character in place of all the characters not needed indicating a string of repeated characters and replacing a smaller bit string for a frequently occurring bit string. Data compression reduces a text file to 50% or a significantly higher percentage of its original size.

In Data Transmission, the data content or the transmission unit including the header data can be compression. When sending or receiving information through the internet, sizable files may be transmitted in a ZIP, ARC or other compressed format.

Data Compression has two methods:

A. Lossless Compression

This uses a data compression algorithm that reconstructs the original data from the compressed data. Here, data can be compressed and uncompressed without losing data details. This is used in ZIP file format and in UNIX tool GZIP.

Lossless Compression is used when the decompressed file needs to be identical to the original file. Sound or Audio data cannot be compressed well with conventional algorithms

Statistical modeling algorithm for text or text-like binary data include:

1) Burrows-Wheeler transform
2) LZ77
3) LZW
4) PPM

To produce bit sequence, encoding algorithm algorithms include:

a) Huffman coding
b) Arithmetic coding

ZIP archives use a combination of Huffman coding and LZ77 to give fast compression and decompression times and reasonably good compression ratios.

LZ77 is pretty much a generalized form of RLE and it will often yield much better results.

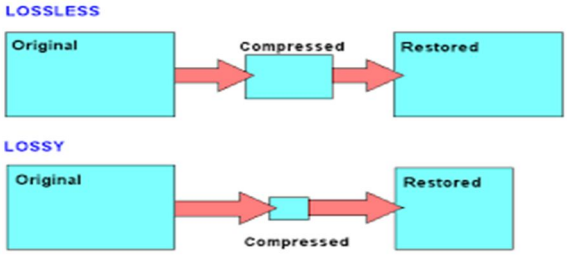
Huffman allows the most repeating bytes to represent the least number of bits.

B. Lossy Compression

This does not allow the exact original data to be restored from the compressed data. Video and audio compression techniques are better suited to this method of compression.

This method of compression allows the decompressed data to be "close enough to the original data" but not identical so it is useful in some way. But repeated compression and decompression of data causes generation loss where the quality of data is reduced with every compression and decompression. This method of compression is used frequently on the Internet and especially in streaming media and telephony applications.

The advantage of Lossy compression over Lossless compression is that in some cases, a lossy method can produce a much smaller compressed file than any known Lossless compression.



V. ENCRYPTION BEFORE/AFTER COMPRESSION

It has been appreciated that there are advantages to eliminating the regularities in the plaintext before encrypting for reasons like the following:

1) Encryption changes the plaintext and makes it random, but Compression removes the redundancies which doesn't work well on a random text.
2) Cryptanalysis method is frequency analysis which relies in finding repeated data and compression reduces the effectiveness which decreases the effectiveness of some attacks.
3) Compressing a data first, an attacker must first decrypt it then decompress it before seeing if the output makes sense. Hence, brute force attack takes longer if the attacker doesn't know to decompress the data first or even to decompress it at all.

Encryption can be done before compression in some situations when it is desired to transmit redundant data over an insecure and bandwidth-constrained channel, it is customary to first compress the data and encrypt it.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

VI. DATA ENCRYPTION IN RDBMS

A RDBMS (Relational Database Management System) can use encryption to protect information in certain situations where the normal security mechanisms of the DBMS are not adequate. For example, an intruder may steal tapes containing some data or tap a communication line. By storing and transmitting data in an encrypted form, the RDBMS ensures that such stolen data is not intelligible to the intruder. Thus, encryption is a technique to provide privacy of data.

A. Benefits of Encryption in Database Security

1) *Peace of Mind:* Knowing your data is completely secure from attack, privacy breaches, or theft allows you to focus on your business, not worrying about prying eyes or potential attacks.
2) *Unauthorized Access:* Prevention Keeping your confidential data confidential is more important than ever in today's business. Encrypting data keeps it safe whether it is lost, stolen, or simply misplaced. This includes keeping data safe against deliberate attacks such as from a disgruntled employee or hacker. It also accounts for the accidental misplacement of data, for example, the safe decommissioning of data on a disk drive that can often times be recovered even after formatting when not encrypted.
3) *Data Protection Act and Certification Compliance:* Encryption is often times mandatory to be compliant with some industry certifications or laws designed to protect and ensure privacy of sensitive data. Applications with databases not meeting these minimum encryption security criteria are ineligible for deployment and therefore not a viable product in that industry.

B. Advantage Encryption

Advantage provides security for your data in a number of important ways including the ability to encrypt stored and transmitted data. For stored data, Advantage can physically encrypt data in tables to protect that data from unauthorized viewing. Advantage supports encryption of table and memo data in both DBF and ADT tables. Support also exists for encryption of ADT database table header information. The Advantage encryption scheme uses key data derived from a case-sensitive password to encrypt data using industry-standard algorithms.

Tables can be encrypted on a table-by-table basis. When a table is encrypted, all record data including memo and BLOB data is also encrypted. In addition:

1) Index data can be encrypted when using data dictionary bound ADT tables.
2) Table header data is encrypted when using data dictionary bound ADT tables.
3) Data dictionaries, which include all database metadata, can be encrypted.

For transmitted data, Advantage supports encrypted communications for client/server communication across the network. By enabling encrypted communication, you ensure that all data is encrypted on your network, including queries, query results, record data, and all client/server interaction.

C. Security Risks to Databases

1) *Excessive Privilege Access:* When users have access rights that allow them to perform other tasks that are not included in their job, which will lead to harmful misuse of such privilege. Let's take an example of our university's administrator. He has given all the access to databases and can change the records of any student. He can misuse the privilege such as changing grades, or marks of a student. As a result, all the different tasks are given a different level of privileges that grants all the access in excess.
2) *Legitimate Privilege Abuse:* Any database users, administrators or a system manager misusing by doing any unethical activity can lead to Legitimate Privilege Abuse.
3) *Privilege Elevation:* Attackers may have the advantage of changing the privileges because excessive exposure may lead to the discovery of flaws. E.g. Ordinary user is given all the access to administrative privileges.
4) *Weak Audit Trail:* A database audit policy makes sure everything is automated, timely and has a proper recording of all the database transactions. Since all the database transactions should have all records and absence of which causes a serious risk to the organization's databases and can cause instability in operations.
5) *Denial of Service:* Pretends or inhibits the normal use of communication facilities. The disruption of an entire network may be the specific target of these attacks either by overloading it with message so as to degrade performance or by disabling
6) *Database Communication Protocol Vulnerabilities:* In the database communication, the large number of security weakness is being identified of all database retailers. Fraudulent activity varies from illegal data access to denial of service to data exploitation.

7) *Weak Authentication:* A weak authentication may give the database more helpless to attackers. Attacker may steal the identity of database users or the login credentials which then helps in the modification of data or sensitive information. If authentication is weak then it helps the attacker to steal data easily.
8) *Backup Data Exposure:* This exposure is an important threat that should be taken care of. Any external media may be exposed to high risks, so they need to be protected from attacks such as theft or destruction.

Fig 3: Databases Security Risks

C. *Database Security Considerations*

1) *Access Control:* Access Control makes sure that there are no interruptions or interference by any attacker neither internally or externally of communications with the databases and other system objects that are according to the policies and control defined. And it also helps in minimizing all the risks that can directly impact the security of the database on the main servers.
2) *Inference:* Inference policy is used to protect the data at a certain level. It happens when the interpretations are from specific data in the form of analysis that are required to be protected at a specific higher security level. It also helps in how to protect the information from being disclosed
3) *User Identification/Authentication:* User Identification/Authentication plays an important role to ensure security. The identification method defines a set of people that are allowed to access the data and it also provides a complete mechanism of accessibility. To make sure, the identity is authenticated and it also keeps the sensitive data safe and also from being modified by any ordinary user.
4) *Accountability and Auditing:* It is required to make sure the physical integrity of all the data that requires defined access to the databases and it is managed through auditing and record keeping. It also helps in the analysis of information that held on servers for authentication, accounting and access of a user.
5) *Encryption:* It is a process that produces a ciphertext for any given plaintext and encryption key. It is a known value to the sender. The sender inputs the encryption key into the encryption algorithm along with the plain text in order to compute the ciphertext.

VII. EMPIRICAL ANALYSIS

This study is done by keen observation of the literature and drawn

1) Frequency: Frequency is the number of occurrences of a repeating commonness. It is calculated in such a way that the paper which has an issue that is not common in some other paper is evaluated as having frequency “I” whereas the papers which are having the common issue have been given the frequency equal to the number of papers having that issue.
2) *Criticality:* It is used to measure the frequency of occurrence of an issue. The criticality factor is divided into four parts i.e. Medium, Moderate, High and Very High.

VIII. ZERO-KNOWLEDGE PROOF

In cryptography, the zero-knowledge proof is a technique in which one party (prover) can covey to another party (verifier) that they know the value, without releasing any information other than they know the value. It is challenging to prove that others have such possession without disclosing the real information. A zero-knowledge proof of knowledge is a special case when a statement consists of the only fact that the prover possesses the secret information.

Zero-knowledge proof must satisfy two properties:

1) *Completeness:* If the convey message is true then the verifier would be satisfied with this information of the prover.
2) *Soundness:* If the convey message is false then the verifier would not be satisfied with the prover except with some probability.

A. *Zero-knowledge Types*

1) Proof of knowledge: this is the knowledge in which some statements sent not whole information to the verifier.
2) Pairing based cryptography: f(x) and f(y) are given, and x and y is not known, it is possible to compute f(x*y).
3) Witness indistinguishable proof: in this verifies should not know or with which witness this proof is made.
4) Multi-party computation: when there is the involvement of different parties, each of them can keep some secrets to themselves. They together form a result.
5) Ring signature: there is an involvement of a key that can only be known to a person it is intended for, not for outsiders.

B. *History*

Zero-knowledge was first conceived in 1989 by Shafi Goldwasser, Silvio Micali, and Charles Rackoff in their paper “the knowledge complexity of the interactive proof system. Then it got upgraded number of times. It turns out that in an internet-like setting where multiple protocols can be applied simultaneously.

C. *Application*

Authentication System: Some users are motivated with an authentication system in which one party wants to prove its identity to another party with some secret information without disclosing their secret intent. This is called zero-knowledge proof of knowledge.

Ethical Behavior: it deals with proper behavior that should be followed during the transaction without any malicious intent while maintaining the privacy of each party.

Blockchains: in this is proposed that the information about the prover and verifier and about the whole transaction is kept hidden and the transaction is valid.

IX. CONCLUSION

With the Encryption of files, information privacy and security are still not guaranteed with the continual aim for humans to breach every encryption algorithm known to man.

The Blowfish algorithm designed to replace the Data Encryption Standard (DES) Algorithm has high speed and claims to have never been defeated.

Twofish Algorithm like Blowfish is free and is used in many encryption programs because it is one of the fastest of its kind.

The Advanced Encryption Standard (AES) Algorithm used by the U.S. Government to protect classified information is extremely efficient and considered impervious to attacks.

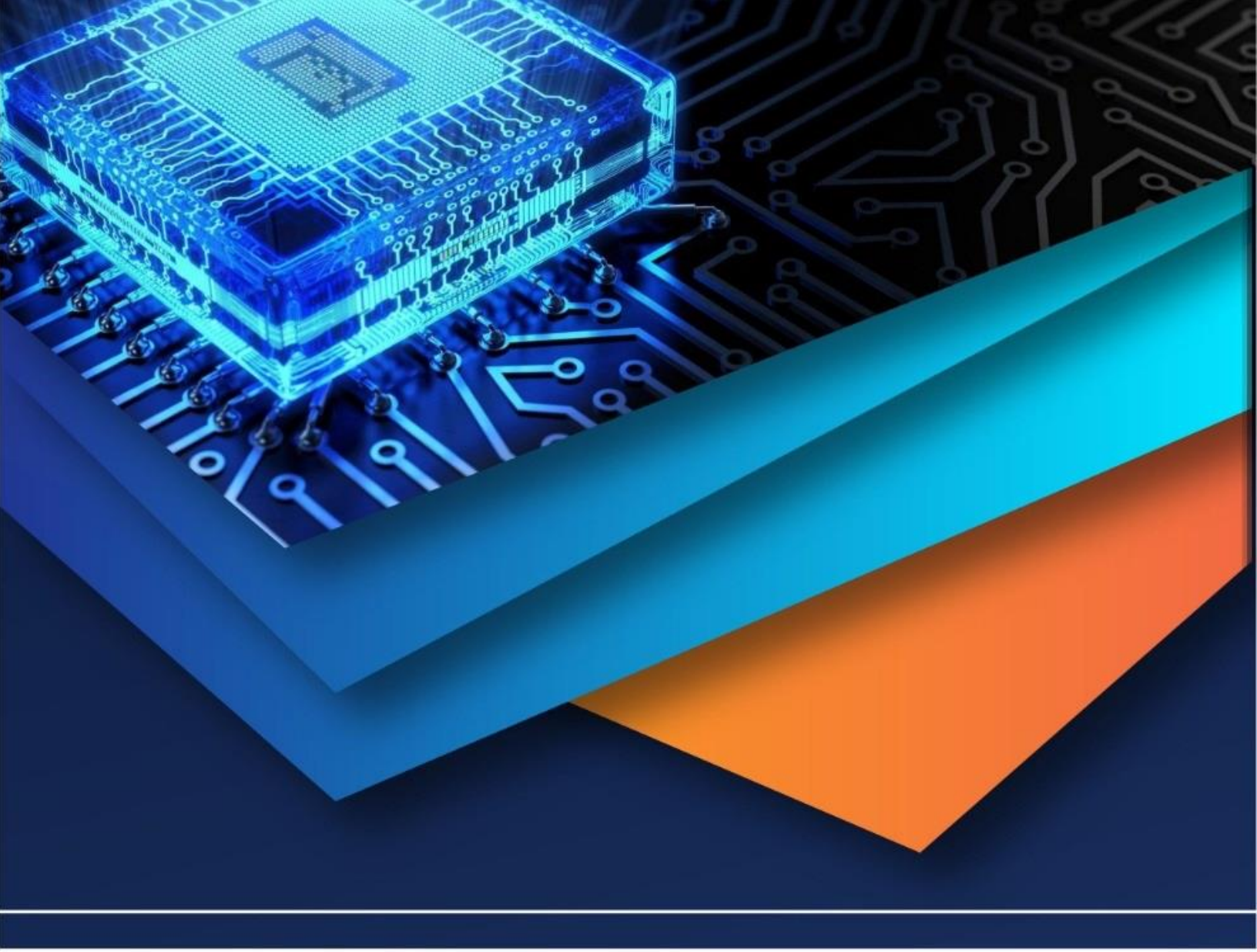
In today’s world of technology, data is helpless to host attacks. Data is the most valuable property for any organization. Security is a big challenge for any organization. It has been concluded that using a strong encryption algorithm can reduce performance but increase security and privacy.

Zero-knowledge proof is an effective technique that enables a trustless transaction between prover and verifier that protect their privacy using mathematics and some probability. It is so helpful that in it some statement is transferred not even the whole information. This encryption scheme is very reliable.

Cryptography is therefore essential in each and every technological era and it keeps advancing with time as technology does. As discussed in the above topics, Cryptography does more than protect data from third party individuals but finds new ways to keep data safe and reveal the identities of third-party individuals that try to hid behind a computer.

REFERENCES

[1] Kumar, Dr Mukesh & Gandhi, Smiley. (2012). Compression and Encryption: An Integrated Approach.
[2] Mucklow, Timothy J. The SIGABA/ECM II Cipher Machine: "A Beautiful Idea." Center for Cryptologic History, National Security Agency, 2015.
[3] https://en.wikipedia.org/wiki/History_of_cryptography
[4] https://cryptomuseum.com/crypto/index.htm
[5] https://searchsecurity.techtarget.com/definition/cryptology
[6] https://www.cloudflare.com/learning/ssl/what-is-encryption/
[7] https://blog.storagecraft.com/5-common-encryption-algorithms/
[8] https://www.educba.com/what-is-decryption/
[9] https://searchstorage.techtarget.com/definition/compression
[10] Basharat, Iqra & Azam, Farooque. (2012). Database Security and Encryption: A Survey Study. International Journal of Computer Applications. 47. 28-34. 10.5120/7242-0218.
[11] Sybase Advantage Database Server Advantage Encryption: https://www.abox.com/PDFM/SYBASE-ENCRYTION.pdf
[12] Uriel Feige, Amos Fiat & Adi Shamir (2015) Zero Knowledge Proofs of Identity, The Weizmann Institute of Science.
[13] https://hackernoon.com/wtf-is-zero-knowledge-proof-be5b49735f27



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY