iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.55967

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: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Integrating Machine Learning with Cryptography to Ensure Dynamic Data Security and Integrity

Aryyama Kumar Jana[1], Srija Saha[2]
[1]*Department of Computer Science, Arizona State University, Tempe, AZ, United States*

Abstract: ***Keeping user data secure is very important in the modern era, as there are growing concerns on data privacy across the world. With very strict laws in place such as General data Protection Regulation (GDPR) in Europe and California Consumer Privacy Act (CCPA) in California, it has become a necessity for businesses and government agencies across the world to secure data in use, irrespective of whether it is at rest or in transit. Also, secure data transfer is a mandatory requirement for organizations which handle confidential data. Data privacy breach and data leak can impact the organizational value and reduce customer loyalty. In this paper, various methods of securing data transfer using a combination of Cryptography and advanced Machine Learning techniques have been discussed. The high-level idea is that ML should complement the existing cryptography methods used in practice, to build a robust and impeccable secure platform making it almost impossible for intruders to get into the system.***

Keywords: Encryption, Machine Learning, Anomaly Detection, Machine Learning, Cybersecurity, Cryptography

I. INTRODUCTION

With the democratization of data in today's world, sharing data has become ubiquitous across multiple domains and applications. In this scenario, maintaining robust and secure data pipelines and channels is very important to protect user privacy and safeguard sensitive data. The rapid technological advancements aim to create a more connected world than ever before, but this comes at the cost of making personal and sensitive data more vulnerable and prone to attacks and thefts [1]. This has affected the day-to-day activities of people. Right from ordering grocery or household items from an ecommerce website to sharing healthcare records for better diagnosis or transmission of confidential data across businesses or government agencies, the data being shared through various digital channels comes under risk of security breach. The objective of this paper is to dive deep into the methodologies where advanced machine learning algorithms and existing cryptography techniques work in synergy to make the systems resilient to any kind of security issues, ensuring safe and secure data transfer across multiple channels.

There are many use cases where people share their personal information. For example, imagine someone ordered a gift for his/her friend on an ecommerce website, while placing the order, the credit card information must be entered. Now, this information is transferred across multiple locations, in some cases across continents in a matter of a few seconds or less. It is the responsibility of the business to keep the customer data safe during this whole transaction. Similarly, in hospitals or healthcare clinics, patients provide their insurance information and doctors add further personal health related information to it. It is very important to secure this private information while sending or retrieving a patient's medical history or insurance related information. In a manufacturing facility or in a power grid, there is a lot of secure and confidential data transmission in and out of the facility. In all the above use cases, ensuring that the data cannot be accessed by unauthorised people, is of critical importance. In simple terms, an authentication checking mechanism should be in place, which would act as a gatekeeper to identify who the user is, and if he/she has the correct access rights to the data being transferred.

Machine learning is the art of identifying patterns in existing data to predict features in new data. The model tries to learn the patterns within the training data and grows the intelligence to classify unseen data as secure or non-secure [2]. But again, the important thing to consider is the accuracy and loss while training the model. Also, it should be taken care that the model is neither over-fitted nor under-fitted before training, the data quality and quantity is ensured using data pre-processing techniques [3]. Once the model is ready and has all weights assigned as per the training data with minimum validation loss and maximum validation accuracy, the ML model can be used to detect anomalies in the data transfer patterns.

Cryptography, on the other hand, has existed in this world since very old times. It is the art of converting the message to a format that is not understandable by anyone other than the recipient. A similar analogy that can be drawn to this is a letter is locked inside a box and no-one other than the recipient has the key to open the box. In basic terms, the message is encrypted and sent to the recipient. In this case, if the data somehow is in the hands of a middleman, he/she can't understand the data [4].

The decryption key is shared with only the authorized recipients of the data so that they can decrypt the data as required. The most used cryptography algorithms are RSA, ECDSA and SHA [5].

In this paper, both machine learning and cryptography techniques are combined to secure the data. The cryptography techniques with advanced machine learning algorithms can generate unique and unpredictable keys which enables them to encrypt and decrypt the data in a highly secured manner. ML models can create a new direction in the field of cryptography by enhancing and creating new encryption and decryption algorithms. Overall, these two powerful technologies can make the world a more secure place.

II. MACHINE LEARNING AND CRYPTOGRAPHY BASED AUTHENTICATION APPROACH

A. Behaviour based Biometric Authentication

Behaviour based biometrics can be used to track unique user behaviour patterns, for example, typing speed, pattern of mouse clicks, etc [6]. This kind of profile can be made using machine learning techniques, by continuous monitoring of the user behaviour patterns. A zero-knowledge proof can be generated from this behavioural biometric user profile, this is a mathematical representation of the user behavioural biometrics. A combination of such biometric behavioural patterns is unique to every user and having duplicates is a rare chance. The behavioural biometric profile of each user in the zero-knowledge proof format is attached to the cryptographic token. Whenever any user tries to decrypt the data, first he/she must provide the behavioural biometric profile, which is then converted to zero knowledge proof and matched with the one stored with the token and then the decryption key is asked with which is then used to decrypt the data. This approach is more secure as compared to a standard cryptography-based security approach as it adds an added layer of security to the data in use.

B. Anomaly Detection for Data Transfer

Underlying anomalies in data transfer processes can be detected using machine learning algorithms [7]. The data is converted into ciphertext using homomorphic encryption. This encrypted data is then transferred through multiple channels to multiple receivers. State-of-the-art machine learning techniques are used on encrypted data for anomaly detection. These algorithms can detect if the data on the receiving side has been corrupted. If the data seems to be compromised, it will be discarded, and source data will be re-encrypted. There is no need to decrypt the data as machine learning models can detect anomalies for both encrypted and decrypted data.

C. Voice Biometrics with Lattice Based Cryptography

Authorised users' voices are recorded and are converted into a mathematical template. The converted mathematical template is then encrypted using lattice-based cryptography and sent through the channel along with the encrypted data. Lattice based encryption is very secure and it cannot be decrypted even by using quantum computing. The data is encrypted by conventional cryptographic algorithms like RSA or ECDSA. The encrypted data along with the encrypted voice template is sent to the receiving end. At the receiver side, authorised users' voices are recorded, and the recorded voices are converted into a mathematical template as well. The converted mathematical template is then encrypted using a lattice-based cryptography algorithm. If encrypted voice templates in both receiver and sender side are matched, users are meant to be authorized and they can access the data.

D. Quantum Key Distribution

Quantum Key Distribution (QKD) is an innovative technique used to generate a highly secure encryption key shared between the sender and the receiver [8]. QKD uses the principles of quantum physics and utilizes photons, the smallest particle of light, to create the key. These photons usually include quantum data including polarization states, in their encoding. While the encrypted data is being transmitted through a channel, if anyone tries to intercept the data, it will disrupt these, alerting both the sender and the receiver about a potential security breach attempt. Fully integrated data transfer security is achieved by using a shared key for encryption that serves to encode data at sender's end and decode it at receiver's end after it has been authenticated and validated. This quantum-safe method offers strong defence against future quantum-based as well as classical threats, making it a novel approach for secure data transport.

E. NTRUEncrypt

The robustness of NTRUEncrypt is predicated on the difficulty associated with lattice problems, specifically the Ring-LWE problem [9]. It is resistant against both conventional and quantum cyberattacks. The data to be transferred is transformed into a vector and multiplied with the public key which is in the form of a lattice.

The private keys are vectors within this lattice. Also, during transmission, an error factor is added as well to introduce some randomness. The original data can be retrieved by using the private key to find the shortest vector within the lattice which is as close as possible to the encrypted vector. This is a quantum-safe approach as it is a np-hard problem as finding the shortest vector within a randomized lattice is a very difficult mathematical problem to solve.

F. Code Based Cryptography

This cryptography is based on error correcting codes. The data to be transferred is corrupted with some error. The private and public key is constructed from error correcting code [10]. The private key is generated from irreducible Goppa code, and the public key is constructed from a random generator matrix of the variation in Goppa Code. The data when reached at the receiver can be only decrypted by users who own the private key. The goal of this technique is to reduce the sizes of encryption and decryption keys. This cryptography technique is very secure and is resistant to quantum attackers. Quantum computers can solve many hard mathematical problems which is the basis of many cryptography techniques. Code based cryptography is based on various mathematical concepts which is inevitable by quantum attacks.

III. POTENTIAL USE CASES

Following are a few potential scenarios that highlight the critical importance of secure data transportation in a strongly interconnected information-dependent society:

1) Genomic research: When combined with cryptography approaches, machine learning can safeguard the confidentiality and privacy of genetic data used in collaborative medical research. ML algorithms may be used to discover complicated patterns and detect abnormalities in genetic information, which is thereafter encrypted using cryptography approaches such as homomorphic encryption or fully homomorphic encryption to preserve the confidentiality of the data being transferred [11]. In addition, ML may play an important role in restricting access by developing advanced technologies that verify clients, and cryptographic keys are used for safely authenticating the user access, in accordance with healthcare privacy laws such as the HIPAA.

2) High Frequency trading: High Frequency trading deals with large amounts of data in a few milliseconds [12]. In this kind of trading where instantaneous judgements can result in significant financial rewards, secure data communication is critical. The combination of encryption and ML is essential in protecting confidential data and ensuring stock market integrity. Encryption at every step and safe key handling can be used to protect data during propagation and storage facilities, preventing potential hacking and data theft attempts. ML-powered anomalies and pattern recognition to constantly track market information and trading patterns for abnormalities, unusual actions or possible malicious activity, can help in taking swift actions to limit potential risks [13]. Furthermore, popular cryptographic protocols such as Secure Multi-party Communication (SMPC) enable safe collaborative evaluation of trade techniques and information exchange across multiple organizations while protecting the anonymity of the data and algorithms. The combination of cryptography and machine learning helps to establish a solid security framework in this high-risk, high-reward environment, promoting confidence and maintaining the authenticity of HFT operations while traversing the ever-changing terrain of the stock market.

3) Satellite Communication: Satellite Communication is essential for military operation, government and emergency, disaster relief operations. ML techniques in satellite communications have been proved to expedite and automate the communication [14]. Various ML models like autoencoder, recurrent neural networks are used to analyse network congestion, interference detection and payload configuration. The overall data transfer process is vulnerable to hacking and spoofing. Cryptography techniques are essential to maintain the security and integrity of the data being transferred. Quantum cryptography technology can generate cryptographic keys using the transmission of quantum states of light to enable encrypted communication between ground station and satellite or between two satellites.

4) Remote Surgery: Robotic surgery has largely evolved the healthcare system [15]. Surgery can be performed by doctors remotely via robots. Communication with robots happens using the Interoperable Telesurgery Protocol. As the communication network is publicly accessible by anyone, it is easily vulnerable to cybersecurity attacks like changing the commands sent to robots or hijacking the robot which poses a high security threat to remote surgery. Cryptographic techniques like Advanced Encryption Standard (AES) along with Transmission control Protocol (TCP) and User Datagram Protocol (UDP) enables the encrypted communication between patients and doctors. Different ML models can analyse the operator behaviour and detect any anomaly in the communication.

Also, behaviour-based authentication helps to identify the doctors and patients securely. Together ML and cryptography techniques maintain the data integrity, patient anonymity and secure communication.

5) IoT: Data transmission security on the Internet of Things (IoT) requires a strong mix of encryption [16] and artificial intelligence (ML). Effective encryption protects data while it is being transmitted, and secure messaging standards such as TLS (Transport Layer Security) and DTLS (Datagram Transport Layer Security) ensure information integrity. Verification of devices with encrypted keys, along with machine learning-based behavioural surveillance, improves security. Cryptographic approaches that protect privacy allow for collaborative study without revealing personal information. Recognition of anomalies using machine learning in real time recognizes aberrant trends, boosting preventive security. System integrity, regular upgrades, and behaviour-based biometrics for device identification add layers for further protection. This comprehensive strategy protects IoT data anonymity, reliability, and security while adhering to guidelines, hence strengthening security for the Internet of Things spanning several applications.

6) Blockchain Voting: In the context of web-based election platforms, the combination of blockchain-based technologies [17], encryption and machine learning offers a strong foundation that ensures the security and integrity of the whole election procedure. Blockchain-based technology, having irreversible and distributed ledger, is a vital component for maintaining the electoral system's integrity. It safely captures each vote, which makes it practically immune to changes or malicious manipulations. Cryptography serves a dual purpose here - secures the blockchain and protects each voter's identity. ML adds an extra degree of protection by improving voter authentication procedures, therefore validating the genuineness of every single voter. It may also detect strange voting trends and actions, which may suggest illicit voting. This continuous surveillance technology aids considerably to the protection of the democratic system.

7) Education: The close collaboration of artificial intelligence with encryption acts as an effective safeguard against information theft and privacy violations in the educational environment [18]. By enabling multi-factor authentication for employees and students and strengthening access management to educational apps, machine learning algorithms enhance cybersecurity. To secure academic information and comply with rigorous privacy laws like FERPA, encryption techniques are concurrently used to secure confidential information during motion and at idle state. This potent combination has an impact on online testing, as ML detects and prevents actions towards cheating whereas cryptography encrypts test information for transmission. Secure mediums of communication are guaranteed across the volatile world of distance education by encryption methods, which are complemented by concurrent identification of risks made possible by ML.

8) Supply Chain: Supply Chain is vulnerable to cybersecurity attacks. If any security breach happens in the whole process at any point of time, the whole chain is disrupted [19]. Due to this, businesses can incur huge losses and can lose a huge number of customers as well. ML algorithms can be used to detect behavioural changes in suppliers or manufacturers or customers associated with supply chain platforms. Using these behavioural patterns of users, it is possible to authenticate the user identity in real time. Behavioural biometrics include voice, fingerprint and how users interact with the device. It is possible to detect the anomaly in encryption key using machine learning models and to rotate the key in time to prevent any security issues.

IV. PROCESS FLOWCHART

Fig. 1 shows the detailed process flow chart exhibiting the implementation of a secure data transfer system using the strategic fusion of cryptographic techniques along with Machine Learning (ML) algorithms. The flowchart consists of eight main steps - Data transfer initiation, Data encryption, Data Transfer, Machine learning authentication, Data decryption, Data processing and validation, ML based anomaly detection and Final Verdict. The data transfer is initiated by a sender. Then data is encrypted using various cryptography algorithms like RSA, AES, NTRUEncrypt etc. As part of asymmetric encryption, a key pair is generated - public and private key to encrypt and decrypt the data. The public key can be shared with anyone, but the private key is kept secured by the owner. The sender uses the recipient's public key to encrypt the data and receiver uses sender side private key to decrypt the data. The encrypted data is transferred using secure data transfer protocol through a secure channel to the intended receiver. Before the data is received by the recipient, the user authenticity is checked using machine learning algorithms. The ML models analyze the user behavioral metrics like keystrokes, mouse, facial recognition and voice and help to decide if the user is authorized or not. The authorized user can decrypt the message using some decryption algorithm. Generally, the same decryption and encryption algorithms are used in the process. The receiving user then processes the data and validates the data integrity. After data validity is checked, anomalies in data are detected by ML algorithms.

Anomalies are detected either by analyzing user behavior or by monitoring threats in the system. If the transferred data is detected with anomalies, the system will create an alert and reject the data transfer. If there is no discrepancy found in the data, the data will be utilized for further use.

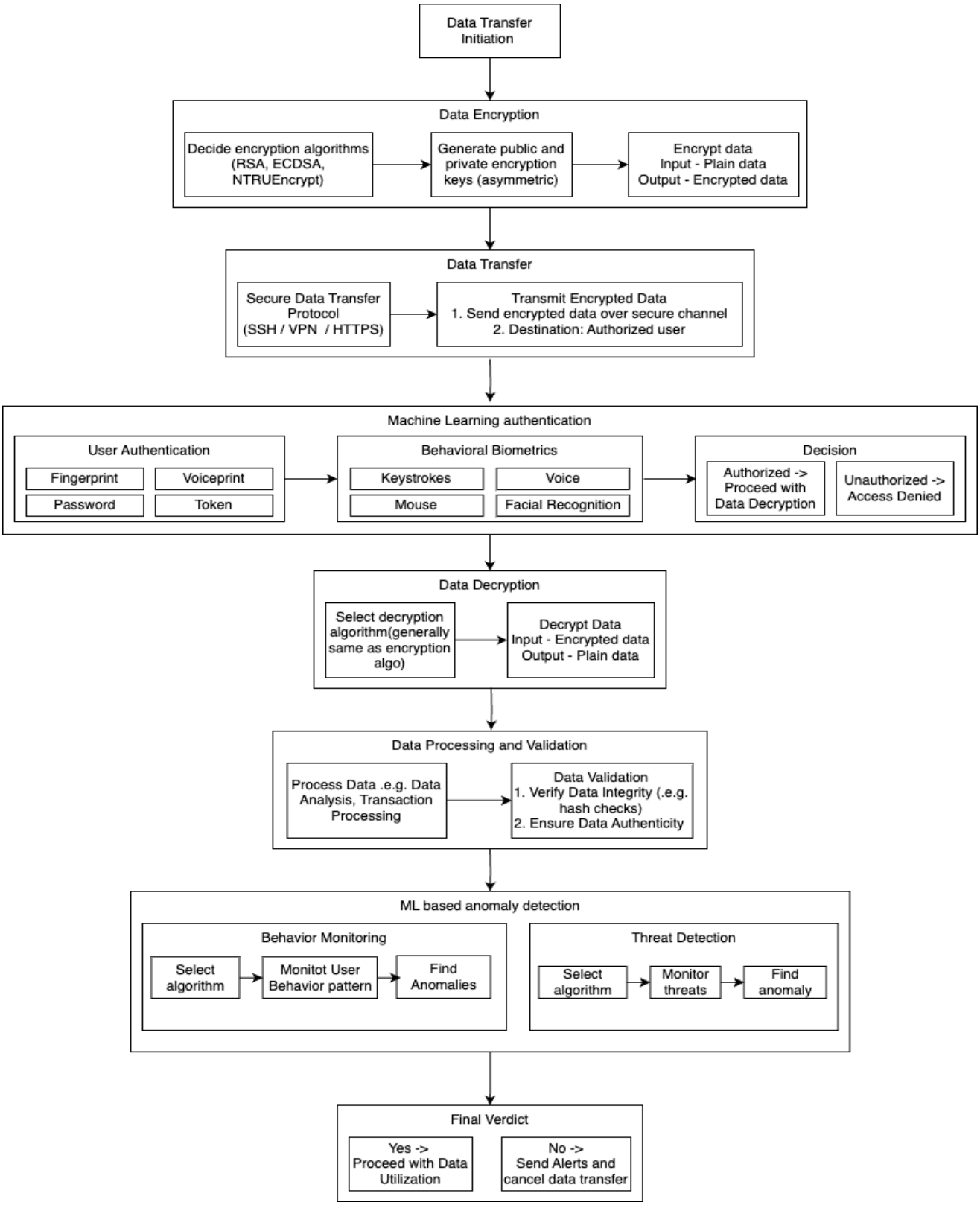


Fig. 1 Flowchart showing the detailed implementation process of a secure data transfer system using ML and cryptography.

V. CONCLUSION AND FUTURE SCOPE

To summarize, the tactical blending of machine learning techniques with cryptographic techniques for facilitating safe transmission of information represents an important change with profound consequences across a wide range of industries, from educational institutions to high-frequency trading and much more. The convergence of these innovations raises the protection of data to new heights, covering an extensive range of precautions. Businesses can assure the confidentiality of private data throughout its complex route via networks and data storage facilities by utilizing the invincible fortress of data encryption, multi-layered user authorization, innovative security techniques, and ML-driven identification of anomalies.

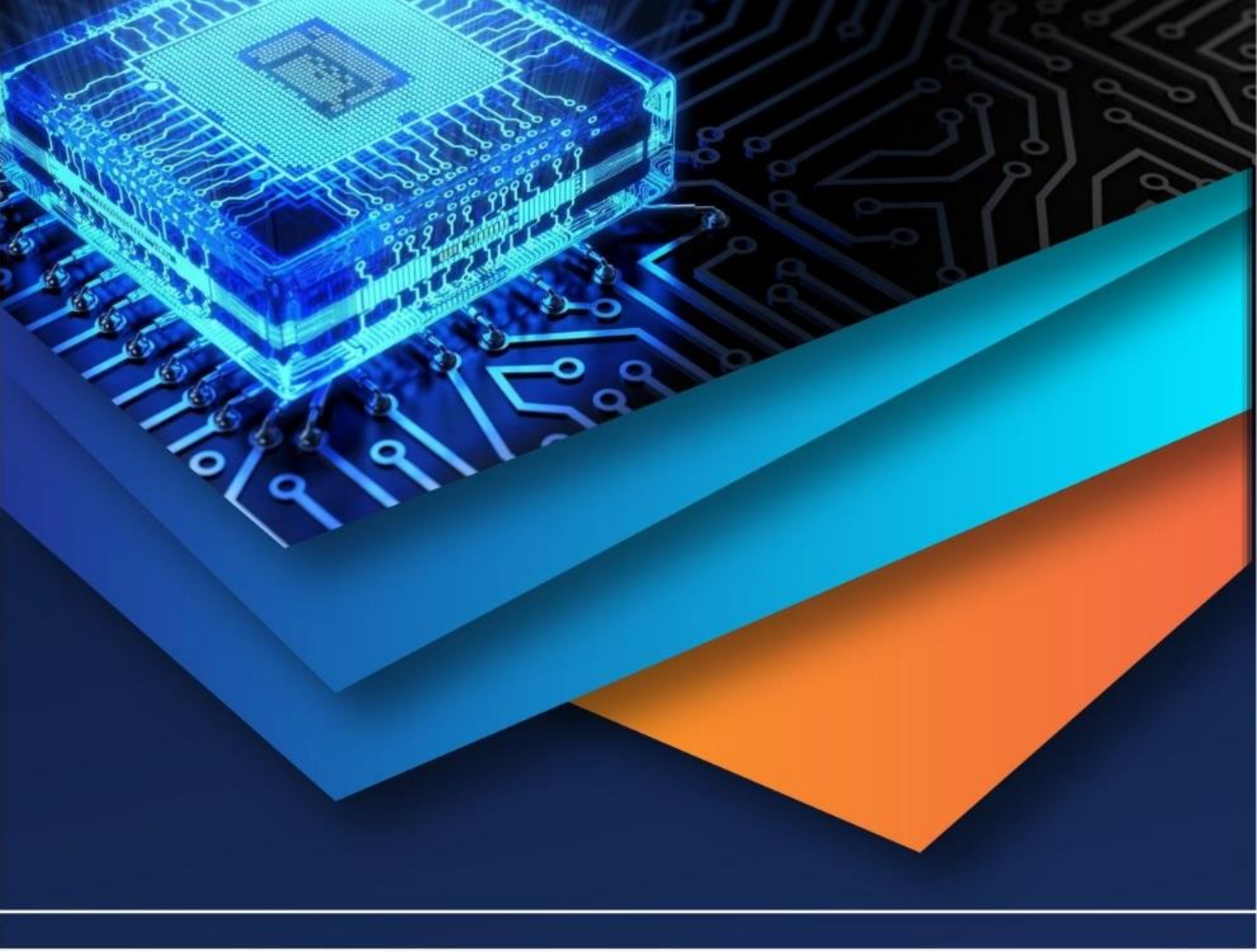
Looking ahead, the array of opportunities for artificial intelligence and cryptography is extensive. The emergence of quantum computer technology necessitates the creation of quantum-resistant techniques for encryption capable of protecting data against imminent quantum risks. Edge computing, given its growing importance, will see the introduction of security theories adapted to the specific limitations of distributed computation of data. In the meantime, the pursuit of compatibility and international norms will establish an intuitive relationship between ML models and encryption protocols, thereby promoting multidisciplinary synergies.

Zero-knowledge proofs will lead the way in securing personal information by enabling coarse-grained oversight of the exchange of information without compromising privacy [20]. Artificial intelligence (AI) will become a beacon in the fight against emerging threats from cyberspace, coordinating quick reaction to threats while thwarting complicated attacks. By creating safe model consolidation and update methods, the ML models' integrity will be strengthened, especially in situations like federated learning.

As international information security standards keep evolving, enterprises will commence on a continuous path to integrate their cryptography and artificial intelligence techniques, assuring consistent adherence to local and worldwide privacy regulations. The smooth integration of ML with advanced encryption techniques is still the cornerstone on which the confidentiality and safety of information will eventually be constructed, encouraging confidence, creativity, and a safe atmosphere for the transmission and evaluation of information throughout different industries and fields.

REFERENCES

[1] Dandurand, Luc, and Oscar Serrano Serrano. "Towards improved cyber security information sharing." 2013 5th International Conference on Cyber Conflict (CYCON 2013). IEEE, 2013.
[2] El Naqa, Issam, and Martin J. Murphy. What is machine learning? Springer International Publishing, 2015.
[3] Ghosh, Anirudha, et al. "Fundamental concepts of convolutional neural network." Recent trends and advances in artificial intelligence and Internet of Things (2020): 519-567.
[4] Katz, Jonathan, and Yehuda Lindell. Introduction to modern cryptography: principles and protocols. Chapman and hall/CRC, 2007.
[5] Sury, O. Use of the SHA-256 Algorithm With RSA, Digital Signature Algorithm (DSA), and Elliptic Curve DSA (ECDSA) in SSHFP Resource Records. No. rfc6594. 2012.
[6] Yampolskiy, Roman V., and Venu Govindaraju. "Behavioural biometrics: a survey and classification." International Journal of Biometrics 1.1 (2008): 81-113.
[7] Chalapathy, Raghavendra, and Sanjay Chawla. "Deep learning for anomaly detection: A survey." arXiv preprint arXiv:1901.03407 (2019).
[8] Russell, Joseph. "Application of quantum key distribution." MILCOM 2008-2008 IEEE Military Communications Conference. IEEE, 2008.
[9] Näslund, Mats, Igor E. Shparlinski, and William Whyte. "On the bit security of NTRUEncrypt." International Workshop on Public Key Cryptography. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.
[10] Overbeck, Raphael, and Nicolas Sendrier. "Code-based cryptography." Post-quantum cryptography. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. 95-145.
[11] Aziz, Md Momin Al, et al. "Privacy-preserving techniques of genomic data—a survey." Briefings in bioinformatics 20.3 (2019): 887-895.
[12] Qin, Molei, et al. "EarnHFT: Efficient Hierarchical Reinforcement Learning for High Frequency Trading." arXiv preprint arXiv:2309.12891 (2023).
[13] Kearns, Michael, and Yuriy Nevmyvaka. "Machine learning for market microstructure and high frequency trading." High Frequency Trading: New Realities for Traders, Markets, and Regulators (2013).
[14] Hughes, Richard J., et al. "Quantum cryptography for secure satellite communications." 2000 IEEE Aerospace Conference. Proceedings (Cat. No. 00TH8484). Vol. 1. IEEE, 2000.
[15] Iqbal, Sohail, et al. "SecureSurgiNET: A framework for ensuring security in telesurgery." International Journal of Distributed Sensor Networks 15.9 (2019): 1550147719873811.
[16] Naru, Effy Raja, Hemraj Saini, and Mukesh Sharma. "A recent review on lightweight cryptography in IoT." 2017 international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC). IEEE, 2017.
[17] Fusco, Francesco, et al. "Crypto-voting, a Blockchain based e-Voting System." KMIS. 2018.
[18] Jung, Im Y., and Heon Y. Yeom. "Enhanced security for online exams using group cryptography." IEEE transactions on Education 52.3 (2009): 340-349.
[19] Hamledari, Hesam, and Martin Fischer. "The application of blockchain-based crypto assets for integrating the physical and financial supply chains in the construction & engineering industry." Automation in construction 127 (2021): 103711.
[20] Fiege, Uriel, Amos Fiat, and Adi Shamir. "Zero knowledge proofs of identity." Proceedings of the nineteenth annual ACM symposium on Theory of computing. 1987.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY