High-Speed Encryption using RSA for Privacy-Preserving Machine Learning Algorithm

Reena Devi

Research Scholar, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana

renupardeep52@gmail.com

Abstract- There has been some investigation into whether or not machine learning and blockchain technology may be used to improve healthcare, but the results have been inconsistent. Integration with current encryption techniques that protect user privacy has been a primary focus of research into these technologies. The privacy-preserving algorithm utilizes machine learning to support high data encryption for blockchain server storage using hash codes. This aids in preserving data keywords in form of numbers using high speed encryption and decryption processes for blockchain. As part of this research, propose using the RSA encryption technique to speed up the privacy-preserving identity verification process on a blockchain network.

Keywords- encryption, IOT, machine learning, security.

1. INTRODUCTION

In today's world, every activity, whether done by a human or a computer, is recorded, and all of this information is collected with no clear purpose in mind. Data analysis will be repeated in the future if and when it is deemed necessary [1]. Due to the data's multi-step journey before being processed by various parties, there is a problem of trust to consider. Data may include personal or confidential information that might be misused by analysing organisations [2]. As a result, it is essential that focus all of the attention right now on issues related to the security of private information. Data privacy refers to the ways through which an individual may limit the dissemination of personally identifiable information. Even after a confident introduction, a person could be wary of giving out personal information to a total stranger unless that built up some trust. In the modern age of digital technology, it is customary to speak about "sensitive data" [3] in an effort to protect one's privacy. When seen from a business's perspective, data privacy encompasses a far wider range of activities and considerations than just safeguarding employees' and customers' private information. It has been shown that privacy concerns are a significant barrier to the widespread adoption of AI and ML. Because machine learning needs large datasets for both training and testing, this result was expected. [4] It was also stated that huge data sets are necessary for training and testing in ML.

The popularity of smart home systems that use the Internet of Things has skyrocketed over the last few years, and this trend is only anticipated to continue. The increased popularity of smart home systems might likely be attributed to their common purpose of improving homeowners' quality of life. According to [5], the worldwide market for smart homes is expected to reach $53.3 billion by 2022. Internet of Things (IoT) technologies, such as smartphones and other modern wireless communications, cloud and edge computing, big data analytics, and artificial intelligence, are the primary facilitators of the concept of a "smart home" [6]. In particular, these cutting-edge methods enable manufacturers to maintain a constant link between all of their networked smart home equipment. Massive volumes of data are being created as a direct consequence of the surge in the popularity of smart home devices [7].

When used for centralised machine learning, however, the method's transparency causes problems [8, 9]. For machine learning to really take off, however, need to move toward decentralised learning architectures. Therefore, several large corporations, including Microsoft, are providing financing for research into distributed approaches for enhancing machine learning [10-13]. There is a compelling need to construct as efficient an AI system as humanly feasible owing to the advent of big data technologies. For optimum performance, data is essential for all machine learning algorithms [14]. Information in its raw, unprocessed form, such as a fact, value, string of text, audio file, or visual depiction, may be referred to as data. Information may take many forms, including text, audio, and visuals. This information is not the work of a single group but rather represents the collaborative efforts of several [15-18]. Geographical factors will be used to decide how the data will be shared across the various businesses. This means that research into distributed machine-learning techniques is more important than ever. At now, just one server is responsible for running the machine-learning algorithm and producing the associated results.

The unencrypted version, or "plain text," is contrasted with the encrypted version, or "cypher text," which is explained. Encryption, the process of encoding data or messages so that only the intended recipients can read them, is increasingly common. The onus is on the sender to implement this method, which calls for encoding the plain text, a secret key, and an encryption technique to alter the message. The two processes are inverses: encryption conceals the original message in an encoded one, while decryption reveals it. It is impossible to recover the original plaintext without the encrypted message, the secret key, and some means of decoding. A symmetric key, also known as a secret key, or an asymmetric key may be used to encrypt and decode communications (sometimes called a public key). Cryptography with an asymmetric key, sometimes known as secret-key cryptography, is identical to cryptography with a secret key. These two classes of keys are categorised as symmetric and asymmetric, respectively, in the key hierarchy. Files of any sort (binary, text, etc.) may be encrypted and decrypted using the keys created using the novel symmetric key cryptography approach described in [1], which is based on a randomization process. This method use a random number generator to create the keys. A file of any sort may be encrypted or decrypted using this approach. Protecting the confidentiality of the collected data is crucial in sensor networks, making this method a strong match for such systems.

2. Literature Review

R. Iqbal et al. [1] A quality control system should always include automated defect detection as one of its key components. It has the ability to improve the overall quality of the processes and products that are being evaluated. Simple boundary checking is all that is available at this time for the purpose of problem detection in computer-based industrial assembly lines for automobile instrument cluster systems. Manual analysis of more complicated nonlinear signals is carried out by qualified operators, whose expertise is then used to monitor quality control and the manual identification of flaws in the system. Deep learning is the foundation of a revolutionary technique that offer here for the automated fault detection and isolation (FDI) process. The strategy was validated by applying it to data produced by computer-based manufacturing systems that were accompanied by both local and distant sensing devices. The findings demonstrate that the method successfully represents the various geographical and temporal patterns that may be observed in the data. Under real-time working situations, the method is able to effectively diagnose and identify numerous types of defects.

S. A. Shevchik et al. [2], Additive manufacturing is being heralded as a game-changer in the industrial industry. Nevertheless, despite the great hopes, there are several technological challenges that limit future penetration into larger sectors. The inability to reliably and affordably monitor the process is the primary contributing factor, as is the absence of repeatability in the process itself. This study is a complement to research that have already been conducted in this sector and provides a novel combination of acoustic sensors with a high level of sensitivity and machine learning as a method for process monitoring. A Fibber Bragg grating was utilised in order to collect the acoustic signals that were produced by a genuine powder-bed fusion additive manufacturing process.

M. S. Hossain [3], stated The problem of classifying fruits is an essential one in many different types of industrial applications. A cashier at a supermarket could use a fruit categorization system to assist them to recognise the different types of fruit and their respective pricing. It is also possible to utilise it to assist individuals in determining whether or not certain types of fruit satisfy their nutritional needs. this study, provide a powerful framework for the categorization of fruits that makes use of deep learning. To be more exact, the system is built on two distinct architectures for deep learning.

This study, offer a fruit categorization framework that is based on deep learning. Deep learning is an approach to machine learning. A tiny CNN model and a VGG-16 fine-tuned model were both explored inside the suggested framework as potential candidates for the CNN. For the purpose of evaluating the proposed framework, two datasets of varying sizes and degrees of complexity were chosen.

R. S. Peres et at. [4] It has been said that the manufacturing sector constitutes a data-rich environment, one in which ever-increasing quantities of data are continuously being created by the industry's activities. However, only a very tiny fraction of it is actually used by manufacturers, which makes it a wasted opportunity. As a result, the Intelligent Data Analysis and Real-Time Supervision (IDARTS) architecture that has been suggested presents the rules for the creation of scalable, modular, and pluggable real-time supervision and data analysis systems for industrial settings.

M. Abadi et al. [5], It is used to quantitative methods for studying how information might be leaked from machine learning models about the particular data records that were trained. concentrate on the most fundamental kind of membership inference attack, which is to establish, given a data record and unrestricted access to a model, whether or not the record was part of the model's training dataset.

R. Shokri et al. [6], Investigate in a quantifiable manner how machine learning models disclose information about the individual data records were trained on. concentrate on the most fundamental kind of membership inference attack, which is to establish, given a data record and unrestricted access to a model, whether or not the record was part of the model's training dataset.

Congzheng Song et al. [7], The term "machine learning" (ML) is rapidly becoming more common. Data holders who are not professionals in machine learning but who wish to train predictive models on their data have access to a wide variety of machine learning frameworks and services. It is essential that machine learning models trained on sensitive inputs (such as private photographs or papers, for example) should not reveal an excessive amount of information about their training data.

Matt Fredrikson et al. [8], Applications that are sensitive to privacy are increasingly relying on machine learning, or ML, algorithms. Some examples of these applications include forecasting lifestyle choices, making medical diagnoses, and face recognition. It is possible to discover sensitive genetic information about people by abusing adversarial access to a machine learning model. However, it is uncertain if model inversion assaults may be used to settings other than their own. Introduce a new category of model inversion attacks, which takes use of the confidence levels that are disclosed along with the predictions.

F. Fraile et al. [9], Specifically in the context of horizontal integration of operational systems in factories as part of information systems in supply chains, the Industrial Internet of Things (IIoT) is having a considerable influence on the manufacturing sector. This technology can be used by manufacturing companies to create data streams along the supply chain that monitor and control the processes of manufacturing and logistics. In the end, the goal is to make these data streams interoperable with other software systems and enable smart interactions among supply chain processes.

Pathum Chamikara Mahawaga Arachchige et al. [10] It has been suggested that the internet of things (IoT) is altering important sectors such as healthcare, agriculture, finance, energy, and transportation, among others. The Internet of Things platforms are constantly being upgraded with new technologies such as the combination of software-defined networks (SDN) and network function virtualization (NFV) in the edge-cloud interaction. The exceptional accuracy that deep learning (DL) achieves when taught with a huge quantity of data, such as that created by the Internet of Things, is one reason why it is gaining in popularity. On the other hand, when trained on extremely sensitive crowd-sourced data, such as medical data, DL algorithms have a tendency to leak private information. The currently available methods for protecting users' privacy while using DL depend on the more conventional server-centric techniques, which need significant computing power.

Yue Wang et al. [11], Within the context of the data collecting scenario, this work investigates how to provide differential privacy by making use of the randomised answer. The randomised algorithm that is being conducted by the client reports a perturbed value to the untrusted server based on the value that has been provided by the client. The use of randomised replies in surveys allows for quick approximations of accurate popular-on statistics while still respecting the privacy of the individual respond- dents.

3. Implementation

The AES technique, sometimes known as the Rijndael method, is a symmetrical block cypher mechanism. It can encrypt and decrypt data using keys of 128, 192, or 256 bits in length. The Advanced Encryption Standard (AES) is frequently utilised because of its security and dependability. The model loading procedures are shown in Fig. 1. As can be seen in Fig.2 and 3, a key is first produced, encrypted, and kept on the server, and only then does its hash code be added to the blockchain.

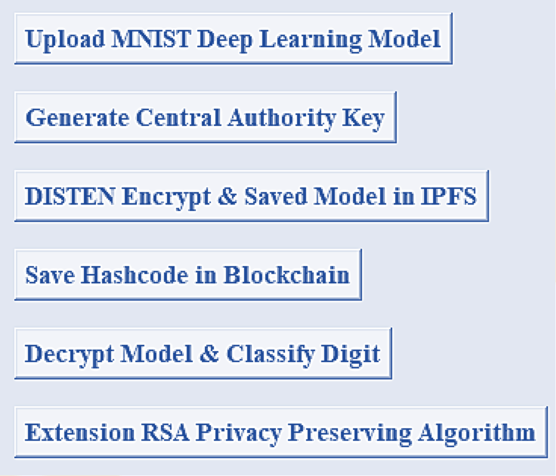

Upload MNIST Deep Learning Model
Generate Central Authority Key
DISTEN Encrypt & Saved Model in IPFS
Save Hashcode in Blockchain
Decrypt Model & Classify Digit
Extension RSA Privacy Preserving Algorithm

Fig. 1 GUI functions

1. Upload MNIST deep learning model

Modified National Institute of Standards and Technology Database deep learning model will be uploaded into h5 model.

```
uploadModel():
    global filename
    text.delete('1.0', END)
    filename = filedialog.askopenfilename(initialdir="model")

    text.insert(END,str(filename)+" loaded")

    pathlabel.config(text=str(filename)+" loaded")
```

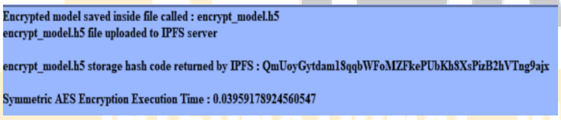
2. Generate central Authority key

Generate the general authority key for encryption.

```
generateKey():

    text.delete('1.0', END)

    global key

    key = getKey();

    text.insert(END,"Key generated by Central Authority = "+str(key)+"\n\n")
```

3. DISTEN Encrypt & Saved Model in IPFS

It has to encrypt and save on server's model. After the saving this, it will be update the encryption time. I.e. how much time it will be taken for uploading and what things are generate here.



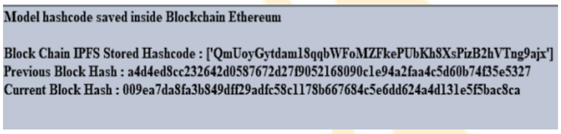
```
Encrypted model saved inside file called : encrypt_model.h5
encrypt_model.h5 file uploaded to IPFS server

encrypt_model.h5 storage hash code returned by IPFS : QmUoyGytdam18qqbWFoMZFkePUbKh8XsPizB2hVTng9ajx

Symmetric AES Encryption Execution Time : 0.03959178924560547
```

Fig. 2 Encryption

An alternative term for public-key encryption is asymmetric algorithms. Both the sender and the receiver utilise their own private key throughout the encryption and decryption processes in an asymmetric method. Here are the sender-specific keys: In order to use a public key, you must also have access to the corresponding private key. The data is encrypted using the public key, and decrypted using the private key. An individual cannot use a public key to decode information. The private key is associated with the public key, but not vice versa. The private key is only known by its owner, in contrast to the public key which is accessible to the public. Because of this, any third party in possession of the user's public key may communicate with them. It is only possible for the user to decrypt his or her own communication using a private key known only to themselves. For the advancement of machine learning and artificial intelligence, it is crucial that the data used in these areas be trustworthy and protected. Given the nature and scale of rich data, traditional methods of breaking into a data centre for practice purposes will be severely constrained. Furthermore, most of the data and resources required for efficient training of machine learning models are held by a very small number of major technological enterprises, which is both problematic for centralization and risky for people's privacy.

4. Save Hasecode in Blockchain

The AES and encryption time will save in Blockchain.



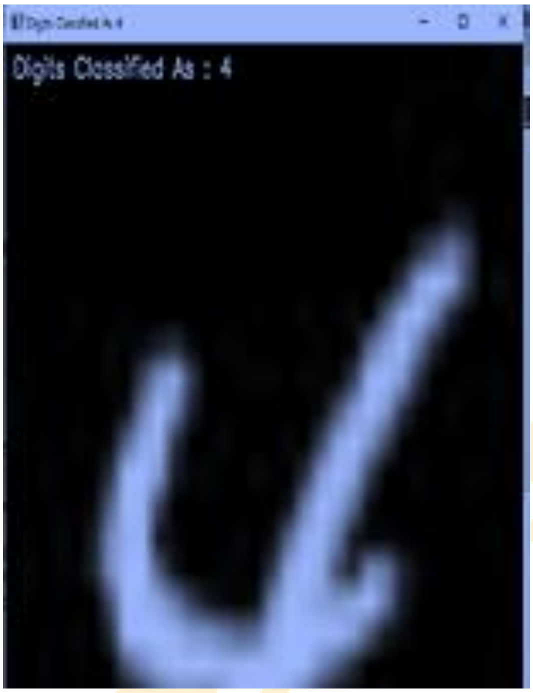
```
Model hashcode saved inside Blockchain Ethereum

Block Chain IPFS Stored Hashcode : ['QmUoyGytdam18qqbWFoMZFkePUbKh8XsPizB2hVTng9ajx']
Previous Block Hash : a4d4ed8cc232642d0587672d27f9052168090c1e94a2faa4c5d60b74f35e5327
Current Block Hash : 009ea7da8fa3b849dff29adfc58c1178b667684c5e6dd624a4d131e5f5bac8ca
```

Fig. 3 Blockchain hash code

The RSA method requires a pair of public and private keys in order to function since it is asymmetric (i.e. two different, mathematically linked keys). When a key pair is formed, one key is made public and the other is kept secret and is never shared with anyone. You can see an example of the sorted numerical output in Fig. 2.

5. Decrypt the Model &Classify Digits

It will be decrypted for classification of any digit number.



Digits Classified As : 4

Fig. 4 Classification of digit

6. Extension RSA Primary Preserving algorithm

It will be use random number for identification (RSA). Which is the things are of propose in extension work. Extension is better than propose.

4. Results

In the context of encryption, a person's "public key" is used, while their "secret key" is required for decryption. This is an illustration of RSA, an asymmetric cypher. The Advanced Encryption Standard (AES) is a symmetric cypher since it employs the same key for encryption and decoding. As will be demonstrated below, RSA helps reduce the amount of time needed to decrypt messages.

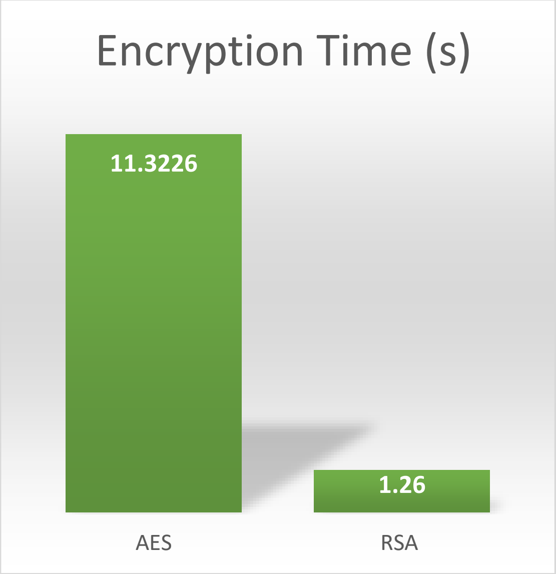


Fig. 5 Encryption Time Result

The result for Encryption time is shown in figure 5 and table 1.

Table 1 Result for Encryption Time

	AES	RSA
Encryption Time (s)	11.3226	1.26

5. Conclusion

The IIoT is causing revolutionary change across all industries, including the energy sector, farming, mining, transportation, healthcare, and more. That change is being ushered in by the Industrial Internet of Things. The Internet of Things is a major force in what is being called the Fourth Industrial Revolution. To make the most of the vast amounts of data and interconnections made possible by IIoT, machine learning (ML) is widely employed. Due to its central role in facilitating the Fourth Industrial Revolution, the Internet of Things (IoT) is a critical enabler of this revolution. On the other hand, access to the Internet of Things network is open to anyone who wants it. Blockchain technology was used in the creation of this network. As a result, in the event of a network security breach, all users will have access to potentially sensitive information, such as transaction records and encryption keys. Since this infrastructure is publicly accessible, it can be used by an adversary to steal sensitive user data without the users' knowledge. In order to solve this issue, the authors of this study propose a quick and secure encryption method that can be implemented in industrial IoT applications.

References

[1] R. Iqbal, T. Maniak, F. Doctor, and C. Karyotis, “Fault detection and isolation in industrial processes using deep learning approaches,” IEEE Transactions on Industrial Informatics, vol. 15, no. 5, pp. 3077–3084, 2019.

[2] S. A. Shevchik, G. G. Marinelli, C. Kennel, C. Leinenbach, and K. Wasmer, “Deep learning for in situ and real-time quality monitoring in additive manufacturing using acoustic emission,” IEEE Transactions on Industrial Informatics, 2019.

[3] M. S. Hossain, M. Al-Hammadi, and G. Muhammad, “Automatic fruit classification using deep learning for industrial applications,” IEEE Transactions on Industrial Informatics, vol. 15, no. 2, pp. 1027–1034, 2018.

[4] R. S. Peres, A. D. Rocha, P. Leitao, and J. Barata, “Idarts–towards intelligent data analysis and real-time supervision for industry 4.0,” Computers in Industry, vol. 101, pp. 138–146, 2018.

[5] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, “Deep learning with differential privacy,” in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016, pp. 308–318.

[6] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, “Membership inference attacks against machine learning models,” in Security and Privacy (SP), 2017 IEEE Symposium on. IEEE, 2017, pp. 3–18.

[7] C. Song, T. Ristenpart, and V. Shmatikov, “Machine learning models that remember too much,” in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017, pp. 587–601.

[8] M. Fredrikson, S. Jha, and T. Ristenpart, “Model inversion attacks that exploit confidence information and basic countermeasures,” in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. ACM, 2015, pp. 1322–1333.

[9] F. Fraile, T. Tagawa, R. Poler, and A. Ortiz, “Trustworthy industrial iot gateways for interoperability platforms and ecosystems,” IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4506–4514, 2018.

[10] Arachchige, Pathum Chamikara Mahawaga; Bertok, Peter; Khalil, Ibrahim; Liu, Dongxi; Compete, Seyit; Atiquzzaman, Mohammed (2019). *Local Differential Privacy for Deep Learning. IEEE Internet of Things Journal, (), 1–1.* doi:10.1109/jiot.2019.2952146

11 Y. Wang, X. Wu, and D. Hu, "Using randomized response for differential privacy preserving data collection." in EDBT/ICDT Workshops, vol. 1558, 2016

.