High Security Machine Learning Algorithm for Industrial IoT

Harshita Dubey

Department of Electronics and Telecommunication, Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed University), (SIU), Lavale, Pune, Maharashtra, India
Email: harshita.dubey.mtech2021[at]sitpune.edu.in

Abstract: Many leading sectors like agriculture, logistics, healthcare, technology etc. are using Industrial Internet of Things (IIoT) which acts as the major factor of Industry 4.0 as well. The latter used Machine Learning (ML) and Deep Learning (DL) majorly these days. Both machine learning and deep learning technologies have been explored for potential applications in several areas. This combination, when applied to federated learning, is the secret to successful application to create evidence-based medicine as rapidly as possible. In this research, RSA approach is meant to increase the delay-time for the encryption inside a privacy-preserving-blockchain system for the identification of numbers. This paper introduces PriModChain, a framework that combines differential privacy, federated ML, Ethereum blockchain, and smart contracts to guarantee privacy and trustworthiness on IIoT data. Using simulations created in Python with socket programming on a general-purpose computer, the viability of PriModChain was assessed in terms of privacy, security, dependability, safety, and resilience. For the local studies, we used the Ganache v2.0.1 local test network, and for the public blockchain testing, we used the Kovan test network. With the help of the protocol verifier Scyther v1.1.3, we examined the suggested security protocol.

Keywords: IIoT, privacy-preserving, IIoT trustworthiness, blockchain, PriModChain, machine learning, deep learning, security, Industry 4.0, Python

1. Introduction

Industrial Internet of Things (IIoT) is a system that combines the capabilities of computers and communication networks with sensing and acting components like actuators and sensors. The combination of these two factors changes the method in which data is acquired, shared, evaluated, and turned into choices. Applications of Industry 4.0, sometimes referred to as the Industrial Internet, are becoming more prevalent, which has indeed led to an increase in the frequency with which they are used. The production of energy, transportation, healthcare etc. are just some of the important industries that stand to benefit from these applications’ efforts to boost productivity and efficiency. Because it paves the way for predictive analytics and reveals crucial insights that have the potential to transform businesses, machine learning (ML) is an essential component of the Industry 4.0 initiative. This is because ML reveals crucial insights that have the potential to transform businesses. As a result of technological advancements in computing and communication, machine learning (ML) is now able to analyse large amount of data such as that generated by a system that is based on the Internet of Things (IoT) and use the extracted knowledge (such as trained models) in order to improve the real-time decision making in complex situations. ML can do this because it is now able to use trained models. The continual development of computer and communication technology has made this outcome conceivable. In order to increase their commercial value in contrast to that of their competitors, vendors have confined their discernments on product development and upgrades so that they fall within the boundaries of their organisations. This is done in order to compete with other businesses.

Despite this, some companies, such as open banking and smart healthcare, present a substantial problem because of the sensitive nature of human data. IIoT, which is a part of the larger framework of the fourth industrial revolution, also known as Industry, is essential to the way cyber-physical systems and production processes are going to evolve with the assistance of big data and analytics. This framework is commonly referred to as Industry. With the support of real-time data from sensors and other sources, industrial equipment and infrastructures are able to "make choices." This provides them with the ability to generate insights and carry out activities that have been predetermined by humans. In addition, machines are now capable of doing and automating jobs that earlier versions of the industrial revolution were unable to achieve. This opens up a whole new world of possibilities. IIoT is required for use cases that integrate networked ecosystems or habitats. This is true on a bigger scale. One illustration of this is the possibility that cities and factories could one day evolve into what are referred to as smart cities and smart factories. Businesses now have the capability of collecting and analysing a higher amount of data at a faster pace thanks to the widespread use of connected and intelligent devices.

This will help to eliminate the communication gap between the production offices and the general offices, in addition to improving the scalability and performance of the system. The potential of IIoT to offer industrial organisations with a more accurate perspective of how their operations are operating will allow these organisations to make more educated business decisions. During the installation of IIoT, the network's availability, scalability, and security should be given the utmost importance. Availability and scalability may come naturally to industrial activities since they may have been established for a large length of time or have been
in operation for a long period of time. Alternatively, availability and scalability may have been designed into industrial operations. However, many people run into problems with the data security of their operations when it comes to integrating the IIoT into their operations. This is due, in part, to the fact that a large number of companies are still relying on outdated procedures and infrastructure. It is made more difficult to install new technology due to the fact that many of them have been in use for decades and have not been changed during that period.

2. Role of IIoT and IOT

IIoT as well as Machine Learning (ML) uncover insights that were previously hidden in data, which paves the way for speedier, more automated responses as well as enhanced decision-making. By ingesting data in the form of photographs, videos, and audio recordings, machine learning for the Internet of Things has the potential to improve intelligence, forecast future trends, and spot irregularities. After providing a brief overview of IIoT, author of this piece of research goes on to a discussion of machine learning. In today's society, matters of this kind are taking on an increasingly significant role for a variety of reasons. Secondly, we have a vast variety of ML applications that were successfully implemented in IIoT.

These applications are acknowledged as having a great deal of significance and value within the academic world as well as within the corporate world. The examination of data and the sharing of information are the cornerstones around which IIoT is constructed. Utility of Internet of Things is directly proportional to the amount of data that is made accessible, for example by means of sensors. A decrease in functionality may occur if certain data are missing; for instance, a heating system that is unable to detect the temperature of the room will be unable to control the temperature in an appropriate way if it is unable to determine the temperature of the room. In the lack of data, there is a possibility that one's security and safety might be compromised, as is the case with smoke detectors and window sensors, for example.

People from all over the world have been able to observe the transmission in real time of many different kinds of structured and unstructured data. The latter has been taken from a variety of sources over the course of the past several decades. These sources include social media platforms, transportation and communication systems, gadgets and sensors. On the basis of the information that is now available, the International Data Corporation has made a projection that between the present and the year 2025, about 180 zeta bytes' worth of data will be generated. An whole new sector of business has emerged all over the world as a direct consequence of the massive amount of data that has been generated recently. This sector of industry is known as the data economy. There are many ways in which the oil industry of the past and the digital world of today may be compared. One comparison that can be drawn is between the two. Comparable to the value of oil after it has been refined and refined some more, the digital world of today is very valuable once it has been completely cleansed and treated. Because of this gradual but consistent growth of the data economy, there has been a rising interest in the idea of IOT.

3. Machine Learning and Deep Learning

Both the methodologies of Deep Learning (DL) as well as Machine Learning (ML) have the calibre of resolving various security issues with the help of their techniques of embedded intelligence. These methods may also protect a network from malicious intrusions. In this article, we take a methodical approach to analysing the security needs, possible attack sites, and available solutions for IOT connected networks. After that, we had a conversation about the drawbacks of these security solutions, which brought us to the realisation that it is essential to make use of machine learning and deep learning methodologies. IIoT has never previously made advancements at the breakneck speed at which it is doing so at the present. It is anticipated that, by the year 2025, the number of devices connected to the Internet of Things will reach 20.4 billion and this would result in a contribution of around $267 billion to the economy of the whole globe.

IIoT is causing a revolution in many various elements of our lives, including our health, our living surroundings, our supply chains, our factories, and our farming practices, to mention just a few of these areas of our life that are being affected. In a word, the benefits that are now being offered by the Internet of Things are huge, and it is quite likely that they will continue to increase in the not-too-distant future as more and more cutting-edge technology becomes available. On the other hand, it is anticipated that the Internet of Things (IIoT), with a compound annual growth rate (CAGR) of 14.4 percent between the years 2017 and 2021, would be the most significant user of internet resources. It has become an extremely profitable target for malevolent users as a result of the quick rate at which it is spreading, and it is exposed to a varied variety of various forms of attacks as a result of this vulnerability. These approaches are now being used to handle the following: Additionally, we investigated the limits of using machine learning models to protect IIoT networks. This investigation, which may help pave the way for new research areas in the years to come, can be found here.

The utilization of ML models is popular in the sectors of security research since the last several years. Because IIoT devices routinely produce enormous amounts of data, this strategy for securing latter systems has the capability of becoming a viable choice in the not-too-distant future. This potential is due to the fact that this method of securing IIoT systems has had the potential to become a viable choice. The training of machine learning algorithms might make use of this data in some capacity. The primary objective of this study is to provide a comprehensive review of the research that has been conducted on the application of machine learning to the field of Internet of Things security. This review will focus on research that has been conducted on the use of ML to improve performance of IIoT security. Till now we have touched briefly on concerns about the privacy of sensitive data as well as the potential impact that these
worries may have on the network security provided by the Internet of Things. Two basic categories, edge devices and gateway devices, are workable divisions for IoT. A device with sensors and/or actuators that consumes little power and uses few resources is referred to as an edge device. Cisco is responsible for the creation of the phrase "edge device." Edge devices are often designed to perform a specific purpose, such as collecting temperature data and transmitting it to a central hub. Examples of such functions include: In comparison to devices near the network's edge, those located in the gateway often have access to a greater amount of resources. A gateway device's purpose is to link edge devices to the Internet and to gather data from edge devices before bringing it to a centralised location. Additionally, a gateway device's role is to transport data from edge devices to a centralised location.

Because of the enormous number of connected devices, the immense amounts of data that are exchanged between them, and the potential impact that these gadgets will have on our day-to-day lives, we need to implement security measures. There are a lot of obstacles to overcome when it comes to the process of integrating security into a network that contains IoT-based devices. To begin with, the components that comprise of IoT come with a wide range of forms and configurations. Users have access to a broad range of customizable choices, including but not limited to the devices that can be utilised, the communication protocols that can be used, the data types that can be transferred and shared, the resource levels of devices, and the settings for the system. Because it is made up of such a diverse collection of individual components, the Internet of Things is notoriously difficult to safeguard in its whole. The sheer quantity of different gadgets that are all linked to one another is a second factor that contributes to the complexity of the situation.

Academics now have a new topic to investigate thanks to the billions of connected devices, especially in regard to the nominal functioning, resilience, and security of these systems. We propose that, in order to address the challenges of securing IoT linked devices, capabilities of machine learning should be included in IoT gateways so that they can contribute to the protection of the system. This would allow the gateways to help address the challenges that are presented. Learning from Experience, Examples, and Analogies is an area that falls under the umbrella of Artificial Intelligence (AI), which is the umbrella term for the discipline as a whole. Within this domain, computer programmes are given the power to learn from their own faults. The capabilities that are incorporated within the programme grow more intelligent as a direct result of the learning that takes place, and the programme also has the capacity to make informed judgments as a direct result of the learning that takes place. Artificial neural networks (ANNs), also known as convolutional neural networks, and genetic algorithms are two of the approaches that have seen the most success in the field of machine learning (GAs).

### 4. Problem Statement

In today's culture, data are gathered without any explicit aim in mind; every action, regardless of whether it was conducted by a person or a machine, is logged. In the current world, the collection of data is a process that happens continually. After it has been determined whether or not it is essential to do so, the data will be evaluated at a later point in time. The fact that the data will travel through numerous phases before being processed by a big number of different persons causes a trust issue, which is a problem that occurs as a direct result of the existing condition. There is a risk that the data will include sensitive or personal information that, if received by the entities that are taking part in the analytical stages, may be utilised in a way that is immoral. There is a chance that the data may contain sensitive or personal information that would entail a risk. If something like this were to take happen, it would be recognised that the data had been hacked. As a direct consequence of this, it is extremely vital for us to give serious attention to the difficulties that are related with maintaining the privacy of people's personal data at the current point in time. The phrase "data privacy" refers to the multiple ways in which a person may manage the method in which a given piece of data is employed in connection to the relevance of that data. These approaches include encryption, passwords, and other security precautions. Table 1 below further tells us about the factors that make AES better than RSA method of encryption.

<table>
<thead>
<tr>
<th>Factors</th>
<th>AES</th>
<th>RSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year developed in</td>
<td>2000</td>
<td>1978</td>
</tr>
<tr>
<td>Length of key</td>
<td>128, 192, 256 bits</td>
<td>&gt;1024 bits</td>
</tr>
<tr>
<td>Encryption process</td>
<td>Faster</td>
<td>Slower</td>
</tr>
<tr>
<td>Size of message Block</td>
<td>128 bits</td>
<td>Min 512 bits</td>
</tr>
<tr>
<td>Power consumption</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Algorithm</td>
<td>Symmetric</td>
<td>Asymmetric</td>
</tr>
</tbody>
</table>

Currently in digital technology, the protection of an individual's personal data is considered as the most sensitive data in nature. When viewed from the perspective of an organisation, the concept of data privacy involves a significantly wider range of activities and concerns than simply protecting the personal information of an organization's workers and customers by itself. This is because data privacy involves a number of different aspects of an organization's operations. This is as a result of the fact that one of the fundamental issues with data privacy is the maintenance of the veracity of the data itself. It has been widely concluded that worries about the privacy of data present a barrier to the production of commodities that are driven by artificial intelligence and machine learning. This conclusion was reached after the following research was conducted: This conclusion was arrived at as a direct consequence to the fact that machine learning involves use of massive datasets in both the training and testing stages of the process. This was the primary cause of how this conclusion was made. In addition, it was discovered that in order for machine learning to work well, both the training
and the testing phases need extremely huge data sets. This is the case in order for machine learning to operate correctly.

5. Related Work

The Industrial Internet of Things (IIoT) and related developments like Industry 4.0 are driven to take use of the extensive distribution and heterogeneity of the whole industrial value chain to embrace business advantages in the cutthroat market. Although there are many benefits to this, there are also trust concerns in internal activities and communications due to the substantial integration of diverse technologies and concepts. It is important to consider the value of trustworthiness from a variety of angles, including quantification, measurement, standardisation, certification, and the implementation of cutting-edge cybersecurity frameworks and standards. A theoretical measurement that aims to gauge the level of dependability required of a component, a system of interconnected subsystems, or a system is a trustworthiness level matrix. The National Institute of Standards and Technology (NIST) framework for infrastructure cybersecurity and the European Union Agency for Network and Information Security are two examples of cybersecurity frameworks for IIoT systems.

Chamikara and the other individuals he worked with in the sector of The Fourth Industrial Revolution, also known as Industry 4.0. Industrial Internet of Things (IIoT), is already revolutionising numerous important industries, including the healthcare sector, agriculture, mining, transportation, and energy. One of the most important forces behind Industry 4.0 is the Internet of Things (IoT), which largely relies on Machine-Learning (ML) in order to use the tremendous interconnection as well as massive amount of data made accessible by IIoT. As a consequence of the fact that machine learning models built on sensitive data are vulnerable to adversarial attacks, their full potential in Industry 4.0 and other applications is restricted when they are implemented in this way. In light of the results of this research, it has been suggested that a system known as PriModChain be developed in order to maintain the integrity of IIoT data and guarantee that it can be relied upon. Differential privacy and Ethereum blockchain like applications will be used to guarantee the truthfulness of data collected from IIoT devices.

Chaabouni and his team researched and found that IoT is rapidly becoming widespread throughout the world at a startlingly rapid rate. Cyberattacks which uncovered previously undisclosed defects in intelligent networks, highlighted fundamental weaknesses in these networks. As a consequence of the Mirai virus infiltrating video surveillance equipment, distributed denial of service attacks (DDoS) were launched against the system, which brought the Internet to a grinding halt. Based on a complete analysis of the many defence options that are now available, this article assigns the risks and difficulties associated with maintaining the safety of IoT networks to one of four distinct categories. As part of this research, we investigated pre-existing network intrusion detection system (NIDS) implementation tools/datasets in addition to the free and open-source network sniffer software, with a particular focus on intrusion detection systems (IDS) (NIDS). After this is complete, an evaluation of pre-existing network intrusion detection systems (NIDS) and their potential applications in the Internet of Things is carried out. This evaluation considers the systems' designs, detection methods, validation procedures, risks that are mitigated, and algorithm implementations in order to determine which are the most promising.

6. Requirements & Proposed Framework

A Software Requirements Specification, abbreviated as SRS and sometimes known as a Software Requirements Specification, is an exhaustive description of the behaviour of a system that is going to be built. A requirements specification for a software system is the name given to this particular piece of writing. It is made up of a group of use cases that, when put together, explain every conceivable contact that customers will have with the product. The SRS has more in store for you than just use cases; in addition to those, there are non-functional requirements. When we talk about "nonfunctional requirements, " we're talking about criteria that put boundaries on the way a product may be designed or carried out (such as performance engineering requirements, quality standards, or design constraints).

Detailed description of the requirements for the system: A compilation of information that has been organised and structured in such a manner as to fulfil the requirements of a system. It is the responsibility of a business analyst, who may also be referred to as a system analyst, to analyse the business demands of their customers and stakeholders in order to assist in identifying business issues and putting up potential solutions. This is done in order to ensure that the business meets the requirements of its customers and stakeholders. Within the context of the systems development lifecycle, the BA frequently plays the role of a liaison between the business side of an organisation and the information technology department of the organisation or with external service providers. This can be done either internally or with third-party vendors. There are three types of requirements that are applicable to projects, and they are as follows:

- Product requirements are papers that explain the features of a system or product.
- Business requirements are a statement, in terms of the business, of what must be given or accomplished in order to provide value (which could be one of several ways to accomplish a set of business requirements).

This page provides a more in-depth discussion of this architectural layout. The purpose of this design is to safeguard the personal information of users while preserving the integrity of the data generated by Internet of Things (IoT) devices. It was possible to determine whether or not the PriModChain protocol was feasible in terms of maintaining the privacy and safety of the users, as well as dependability, safety, and resilience, by using a general-purpose computer to run simulations that were written in Python and made use of socket programming. This allowed
for the possibility of determining whether or not the protocol was practicable. Because of this, it was possible to establish whether or not the procedure could really be carried out.

The RSA algorithm is an example of symmetric encryption. In order for this method to work correctly, it has to be fed random numbers from a generator. The idea of "asymmetric" alludes to the fact that in order for it to function, it needs two separate keys, which are respectively known as the "Public Key" and the "Private Key." This need is what gives rise to the term "asymmetric." On the other hand, the Private Key is something that must be guarded from others at all times, in contrast to the Public Key, which, according to its name, is accessible to everyone and every person. We make use of differential privacy, federated machine learning, the Ethereum blockchain, and smart contracts as part of the PriModChain architecture, which is detailed in this article, in order to provide privacy and trustworthiness for IIoT data. This architecture is also described in this paper.

The layered architecture of PriModChain is shown in Fig. 2, where each layer focuses on how various technologies are used to enforce certain trustworthiness properties. In addition, the figure shows the preferred on-chain and off-chain data storage options for each layer, where on-chain refers to EthBC data storage and off-chain refers to IPFS data storage. The ML model parameter datasets are too big to be stored on EthBC, hence PriModChain employs IPFS as the off-chain data storage method.

The criteria or framework for the process provide an overview of the activities that are carried out by the organisation that is expanding. In the PriModChain framework, the smart contract, DISTEN, CENTAUTH, IPFS, and EthBC are organised as shown in Fig. 1. The framework we've developed, dubbed PriModChain (Privacy-preserving trustworthy machine learning model training and sharing framework based on blockchain), tackles the privacy and trust concerns with machine learning in IoT systems. Differential privacy, federated learning (FedML), smart contracts, and the Ethereum blockchain are all combined in PriModChain (EthBC). For off-chain data management, PriModChain makes use of the interplanetary file system (IPFS). To create a global representation of the distributed machine learning knowledge in a distributed IIoT environment, the proposed framework (PriModChain) employs FedML. FedML offers the ability to train an ML model using both streams of data and static data.

7. Architecture

The architecture of this is primarily influenced by the base paper. The author of the paper used a symmetric encryption technique known as AES for the goal of preserving the secrecy of deep learning or machine learning models. The writer of the report that will be carried out in the not too distant future suggests that it would be doable to investigate new techniques that would either lengthen or shorten the amount of time required for encryption so that the RSA approach could be used instead.
test network for the reasons of testing the public blockchain. Both of these networks were used for the purposes of validating the blockchain. These two networks are both regarded as being local test networks in their respective areas. Scyther v1.1.3 protocol verification, which is available as a free download from the internet, was used in order to check the suggested safety protocol in order to get the outcomes that were sought for.

8. Conclusion

In an IIoT environment, we proposed a brand-new framework called PriModChain that may be utilised for reliable machine learning and sharing. To impose privacy and trustworthiness on ML in the IIoT, PriModChain combines the principles of smart contracts, blockchain, federated learning, differential privacy, and interplanetary file system (IPFS). While differential privacy imposes privacy on the ML models, federated learning serves as the overall framework for federating and sharing ML models. The Ethereum blockchain and the integration of smart contracts give the framework traceability, transparency, and immutability. With safe P2P content distribution, IPFS delivers immutability, low latency, and quick decentralised archiving. The viability of the suggested framework was examined in terms of privacy, security, dependability, safety, and resilience.

References


