Blockchain and Machine Learning Applications in Infrastructure Security within Smart Cities

CSE543 Spring 2022 - Group 2

Group Members:

Jacob Jose (Group Leader)

Tristan Blick (Deputy Leader)

Nikhil Bindem

Sasanka Gali

Derek Ratliff

Bishnupriya Pradhan

Tharindu Kumarage

Abstract: The goal of this paper is to highlight possible blockchain and machine learning applications that could be used to improve infrastructure security in smart cities. We have researched ways to apply blockchain to networks of critical infrastructures, Internet of Things (IoT), healthcare, and smart grids. In addition, we have researched ways in which machine learning can be applied to network management, data security and intrusion, malware, spam, and fraud detection.

Table of Contents

1 Introduction

1.1 Motivation and Background

A smart city utilizes information and communication technology in order to share information, and improve efficiency across its industries. One sector that can benefit immensely from the innovations in technology is the infrastructure. Infrastructure consists of all public and private physical structures such as roads, bridges, water supply, sewers, electrical grids, and telecommunication. The latter two are the most critical to a city's day to day functioning and operations. As electrical grids and telecommunications continue to grow with the expansion of smart cities, there is an increasing need for more security measures to protect sensitive information and data.

We chose to write our report on the applications of advancing technologies including blockchain and machine learning within smart cities due to the fact that smart cities are quickly emerging and evolving around the world. By analyzing the power of machine learning and blockchain's potential power within current systems, we may create a clearer understanding for what the future may hold in many cities.

Blockchain was first created by a person (or a group) referred to as "Satoshi Nakamoto." They created blockchain intending to facilitate digital data to be recorded and distributed, which cannot be edited (immutable). These transaction records cannot be altered, deleted, or destroyed, making it a secure way of sending data. Blockchain applications can be used to strengthen and ensure security in industrial networks, healthcare, IoT, and energy smart grids.

Machine learning, simply put, is the simulation of a human brain using a computer. Just as a human can try, fail, and learn from mistakes, a machine learning algorithm can apply these same practices (but much faster) to learn patterns in data in order to make better decisions regarding reasoning about a data set. These algorithms can be applied in a variety of different fields where data analysis is a core task which means that infrastructure security is no exception.

1.2 Goals and Scope of Study

The aim of this study is to find how blockchain can be used in smart cities to ensure security of information shared in infrastructures and across IoT networks. IoT has become widely utilized across a plethora of infrastructure applications such as smart homes, healthcare, transportation, and manufacturing. With this rapid growth comes a substantial increase in the need for security.

The main issue identified today is the lack of security among the different energy entities, efficiency and automated cyberinfrastructure. Cyberattacks cause serious troubles in managing data quality. Further, due to the involvement of third parties, issues like integrity, visibility, and costs of data are increased. Cyberinfrastructure is another major challenge that needs to be controlled. Blockchain is a technology that helps record the value transactions through a cryptographic signature for continuous modification. It is also considered a digital leader. Transactive energy applications' speed, scale and security can be increased due to the application of smart contracts involved in Blockchain. Thus, a systematic path is provided for a decentralized modernization grid in a decentralized manner that helps in connecting the Energy Internet of Things with devices of grid edge. Grid and resilience can be optimized by improving the power grid in the operations and design department. However, cybersecurity helps individuals to protect their systems from cyberattacks. The increase of smart grid connectivity and interaction with buildings will help in creating Smart building automation. It also provides an advanced control system that contains various cybersecurity features. A continuous records list is maintained by a blockchain called blocks. A timestamp is placed in each block and between each block. Without human interaction, the execution of blockchain smart contracts can be possible. It helps provide the resilience to data modification as the block data cannot be altered easily. Smart contracts of Blockchain is a technology that helps in exchanging value without any action of intermediaries as money and information arbiters.

We will be going over the applications of machine learning in network intrusion detection, network fault diagnosis, data fraud and intrusion detection, and malware and spam detection. Within each of these fields, we will review studies using novel frameworks and combinations of machine learning models to evaluate efficient and effective learning methods. Our goal is that this will lead to finding a group of leading solutions to modern problems currently plaguing smart infrastructures today which may have some derivative implemented at a large scale in the future within real smart cities.

2 Overview

2.1 Jacob Overview

As countries move towards digitizing their critical infrastructures, so have their vulnerabilities to external cyberattacks increased. Blockchain technology with its distributed nature is ideal to be used in the protection of Critical Infrastructure Network (CIN) and its components, best in combination with the existing methodologies. CIN is the internal computer network of a country's vital sectors such as defense networks, communication, finance, food supply chain, energy, etcetera. Many strong encryption methods are used to secure the data communicated within networks. But, the centralized nature of such architectures makes them prone to vulnerabilities such as Single Point of Failure (SPOF) especially if the security of the central node is compromised. This increased vulnerability problem can be said to be exacerbated with many Internet of Things (IoT) that are now connected to such networks. Jacob Jose evaluated the applications of blockchains in the infrastructure networks that are critical to a country such as defense, communication, finance, and food supply chain.

2.2 Tristan Overview

Tristan Blick has researched how blockchain applications can be applied to help with data security in the rapidly growing Internet of Things (IoT). The Internet of Things is important to the connection and communication between the infrastructure and other sectors of smart cities. In recent times, a lot of this data has become personal, such as passwords, fingerprints, and face recognition, and there is a need for a way to secure this information. Blockchain has become a go-to application for ensuring data security within the past decade. Blockchain technology can be regarded as a “decentralized logbook with a constantly growing amount of precise data, in which all transactions are carried out digitally”. [28] This could be a possible solution to ensure data security in IoT, however, there is a flaw within the IoT environment.

While IoT provides many benefits, there are also a few limitations that negatively affect possible applications. The most noticeable being IoT’s hardware limitations and in turn its computational abilities. Unfortunately, these computational limitations do not allow for a standard blockchain implementation. Blockchain algorithms are often very resource-intensive and have high energy consumption. This would not work for IoT devices and as a result, researchers have looked to other blockchain applications, such as the lightweight one used in bitcoin, which is simplified and does not require many resources. [28] The proposed solution is to use lightweight blockchain that does not require heavy computation so that the devices in the IoT environment can run safely and effectively.

2.3 Nikhil Overview

This research will take into account a deductive approach for making an interpretivist philosophy in determining the existing usefulness and functionality for blockchain in medical security. Data protection, Data Sharing, and interoperability in population health management are the most urgent concerns in health care security. The solution to this problem is to use Blockchain. When properly implemented, this technology improves security, data interchange, interoperability, integrity, and real-time updating and access. This research will be suitable in analyzing the blockchain technologies capabilities within the medical domain. For getting a solid grip over the knowledge of blockchain applications, there will be usage of secondary data collected from the different authentic journals and websites that will help in highlighting the security aspects developed through blockchain.

Therefore, there will be a proper usage of qualitative research that will be done by collecting information from different journals and articles. In addition, this research will be accomplished in the future through the usage of systematic literature research strategies for determining the blockchain applications within the healthcare domain. Based on the previous research, there will be a proper representation of the different privacy risks for telemedicine facilities and the mitigation measures developed through the deployment of blockchain in a well prospective manner.

2.4 Sasanka Overview

Sasanka is focused on researching applications of blockchain on smart grid to improve its speed, efficiency and security. Like the internet, the smart grid has multiple components like controls, computers, automation, and new technologies working together to improve efficiency. In the case of the smart grid, these technologies work together with the electric grid and respond digitally to the quick change of electric demand. Blockchain, Big Data Analytics, and IoT networks are essential technologies used in smart grids to improve their requirements like security and efficiency. Several factors, including the deregulation of the energy market, increasing demand for electricity, evolution in metering, decentralization, increasing vulnerabilities, led to the transformation of traditional power grids into smart grids.

Smart grids play a crucial role in smart cities for electricity production without any loss for various functions. They help improve the efficiency of production, present opportunities for conservation, and most importantly, enable coordination between the city control Center and other infrastructure domain operators for smooth functioning of the city. A smart city is all about various entities working together as a single entity in various conditions. Hence, even under

extreme conditions, the critical functions of a smart city. The smart grid does play a significant role in smart cities, but it has a few vulnerabilities which should be taken care of to improve its speed security and reduce transactive costs. Blockchain technology can help improve all the functionalities and overcome security-related vulnerabilities in the smart grid.

2.5 Derek Overview

Derek Ratliff focused on researching the applications of machine learning to improve the quality of network security within smart cities. Many different filtering algorithms and tools exist such as Snort, IPTables, Splunk, etc. but these tools are anything but perfect. As the intricacies of cyber-attacks evolve, so too must all critical networks in order to handle these attacks. Since smart city networks will be among the most essential, these will be at the forefront of security improvements. One of the many things that machine learning excels at is pattern recognition, which fits perfectly with anomaly detection. An Intrusion Detection System (IDS) can monitor network traffic and learn what falls into the category of normal traffic and sort out anything that doesn't. Of course, this leads to a non-zero false positive rate when applying machine learning principles but that is why using the proper models is crucial for the success of anomaly detection.

There have been many developments in the uses of machine learning to improve these Intrusion Detection Systems. Since systems following this structure rely on pattern recognition, they can excel in Denial of Service attack detection, deducing what type of IoT device sent a packet, and even finding the exact physical location of a physical network attack or failure. The fast and efficient heuristics developed by machine learning algorithms allow networks to become increasingly more secure as time goes on, independently increasing in quality. Since a smart city is a highly networked area, all of these principles will be exceptionally important. These developments will be a major step in preventing cyber-terrorism, repairing worn network equipment, and even for simple logging.

2.6 Bishnupriya Overview

Bishnupriya Pradhan focused on researching machine learning applications of data security in smart cities. Two threats to data security: malware and spam - particularly targeting mobile devices, are studied in detail.

Interconnectivity, one of the biggest strengths of smart cities, can also be considered its greatest weakness. Users engage with smart city ecosystems in various ways using smartphones and mobile devices. For example, a user's smartphone becomes their mobile

driver’s license and ID card with digital credentials which speeds and simplifies their access to the city and local government agencies. The user must give personal data to some devices which leads to vulnerabilities. While compromising a single smartphone may not seem like a big deal, it can serve as an access point for lateral movement across the entire network of connected devices. For example, at the application level, a device can be attacked through a SMS Trojan and the attacker can get money from it, or when the device is connected to some unsecured network, it can get access to and release other important data within the system.

For efficient malware and spam detection in mobile devices, several traditional and machine learning based approaches have been studied.

2.7 Tharindu Overview

Tharindu Kumarage's research focus resides in machine learning applications for data security in smart cities. In particular, he recognized two important areas of the smart city data threat landscape that have the potential to interrupt the seamless performance of different infrastructures. These threat areas are 1) Fraud detection and 2) Intrusion detection. An important note here is that these threats will be analyzed at a system level rather than a network level. This distinction helps to decouple machine learning applications for data security with machine learning applications for network security.

IBM defines data security as a practice where digital information has been protected from unauthorized access, corruption, or theft throughout its entire lifecycle [32].His sections follows the same definition and builds upon how to employ machine learning applications to prevent unauthorized access, corruption, or theft in the form of fraud and/or intrusions.

3 Detailed Results

3.1 Blockchain Applications in Critical Infrastructure Networks

The subjection of CINs to cyberattacks has increased since they have been modernized [27]. A key factor that motivates adversaries is the high level of disruption a successful attack can cause to a country. The modus operandi often starts with a seemingly innocuous email that masks a script which then gains access to the networks. The most common types of cyber-attacks include Phishing, Zero-day attacks, Brute force Attacks with Password Spraying, Denial of Service Attacks, and Malware. It is estimated that by the year 2025, cyber-attacks can cost more than $10 trillion [26]. The Ukraine power grid hack, the WannaCry Ransomware attack are some examples of cyberattacks that have caused substantial economic losses in the last 10 years. When the critical infrastructure of a nation shuts down, not only that industry has a financial loss, but a domino effect also transpires across all other sectors in the country. This is very dangerous to the national security as well as economic aspects of a country. Hence it is vital that a country's critical infrastructures be secured from unauthorized access.

Defense Networks

Proper authentication is important in defense networks. A lot of military equipment is connected to networks so that they can be controlled from a far distance remotely. These machines by themselves communicate with many smart devices to gather data. Since a large number of connections must be made within a short time, ensuring only the authenticated nodes are connected in the network is crucial. Blockchains with their immutable and distributed nature can greatly facilitate verifying the identity of nodes in the network. For instance, when a command is sent from one entity to another, the system can hold the command until the validity of its sender is confirmed by nodes in the blockchain. The details of participants can be stored in the blockchain across the network after verification. As each block in the blockchain is added after getting approval from the majority of participants, the chances of blocks with illegitimate entries from entering the blockchain are next to impossible. In this way, connections and commands can be verified and executed with minimal time spent for authentication. In addition, blockchains can work with smart contracts.

The distributed nature also helps in avoiding SPOF, as copies of the data are stored across all nodes in the network and so are available almost all the time. This distributed architecture to prevent SPOF is not something new; many organizations had already set up their data centers across geographical regions. In mission-critical operations of the military, a distributed architecture can reduce the time it takes to relay a message to a single data center

and back. A key advantage that blockchains put forward is that, with blockchains, if one of the network components fails, they can easily be precluded from the decision-making process. Furthermore, as blockchains use hashing to link the blocks, blockchains can be said to be robust to most cyberattacks. Trying to decrypt a hash requires high computational resources that are mostly affordable only by state-sponsored groups.

Another application of blockchains is for reducing the Distributed Denial of Service (DDoS) attacks on computer networks by using blockchain (Figure 1). The IP addresses of those machines that cause DDoS attacks on networks can be stored in a Federated or Consortium blockchain, which is a blockchain that is distributed across the networks of organizations. Since altering data from a blockchain requires calculating hashes of the subsequent blocks, it is very unlikely for an entity to alter the IP addresses stored in the blockchain's blocks. This blockchain could help reduce DDoS attacks in a manner that would be easier to use than the prevalent method of using captchas. The notable advantage is that the user is not asked to fill any captchas as his/her IP address can be monitored in the background. Moreover, since the entries in the blockchain are not removed, the adversaries could potentially run short of IP addresses to commence further attacks.

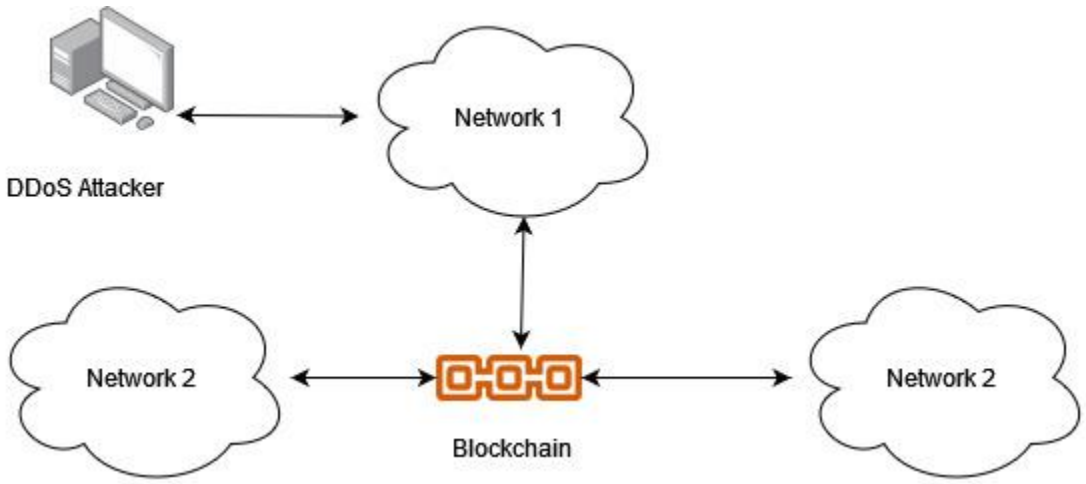


Figure 1: DDoS Attack Prevention using Consortium Blockchain

Communication Networks

Due to advancements in computing power, blockchains can now be used in communication networks. Blockchain emails are email systems that make use of blockchains[11]. In blockchain email systems, the authenticity of the sender is verified by using blockchains – whenever an email is transmitted, its details are recorded onto the blockchain that is often stored on the private network of the organization. Recipients can verify the legitimacy of the email from the blockchain. Email addresses that attempt to impersonate that of an authenticated user will have to take control of more than half of blockchain nodes, this is next to impossible with the technology available now. Hence blockchain email can help filter out malicious emails in communication networks.

Furthermore, emails and their attachments can be archived while ensuring their integrity using off-chain storage mechanisms of blockchains. In off-chain storage, information that is contained in the email such as the email-body, attachments like image files, documents, etcetera can be stored on cloud-based storage [22]. A hash of these files is stored on the blockchains. Whenever this archived data is taken, it can be verified with the hash stored on the blockchain. This can help to identify those files that have been modified without authorization.

However, a consequence of using blockchains for emails is the limitation of storage facilities. As the number of email addresses increases, so will blockchain's storage requirements. A possible aftereffect would be increased storage requirements in the participant nodes. The rising number of emails means that the storage constraints on each node would be very high. But, this could be resolved by using blockchain emails for mission-critical communications.

Finance

Blockchain was initially developed for the cryptocurrency Bitcoin. Although not as fast as cryptocurrency, blockchain itself is also gaining popularity in the financial industry. Smart contracts are now being employed by financial organizations to automate transactions. Given the benefits, banks have now started to use blockchain networks to transfer money across countries. The transparent nature of blockchain makes them very suitable as the transactions are recorded in each of the participating nodes. The traceability property of blockchain ensures that in the event of a discrepancy, the transactions can be verified on the blockchain. Since blockchains operate in a peer-to-peer manner, by making transactions via blockchain, the banks can substantially reduce their operational costs and increase speed and efficiency [5].

Presently to transfer money across countries, banks use secure networks offered by

organizations such as Fedwire, Society for Worldwide Interbank Financial Telecommunication (SWIFT), Clearing House Interbank Payments System (CHIPS) etcetera to send information regarding the transaction, such as information of sender and recipient [10]. Although suitable today, these methods cannot be said to be the best. An international wire transfer could take two business days to be completed. Furthermore, the intermediary systems also charge a fee for the transactions done. Smart contracts with blockchains can help banks avoid these expensive intermediates. As smart contracts execute based on predefined conditions and are documented in blockchain, they can execute faster. With smart contracts, banks can store their transactions to a blockchain that is accessible to both parties. Moreover, as the number of nodes on a blockchain increases, it can be said to be more secure. As then, a large number of participants would be available for consensus.

Food Supply Chain

The networks that manage food supply chains have also implemented blockchains to track their data (Figure 2). The traceability property of blockchain allows the industry to keep track of the food product's journey from production, all the way to the consumers. In the case of traditional systems, data is stored in different servers managed by different vendors/participants. A drawback of this approach is that data stored in discrete servers are not accessible outside; this has the vulnerability for the data to be altered. In the case of blockchain systems, the transactions are sent to each participant node, which allows the data to be always available for inspection. Many notable corporations such as Walmart Inc, Nestlé use blockchain to track their food inventory [13].

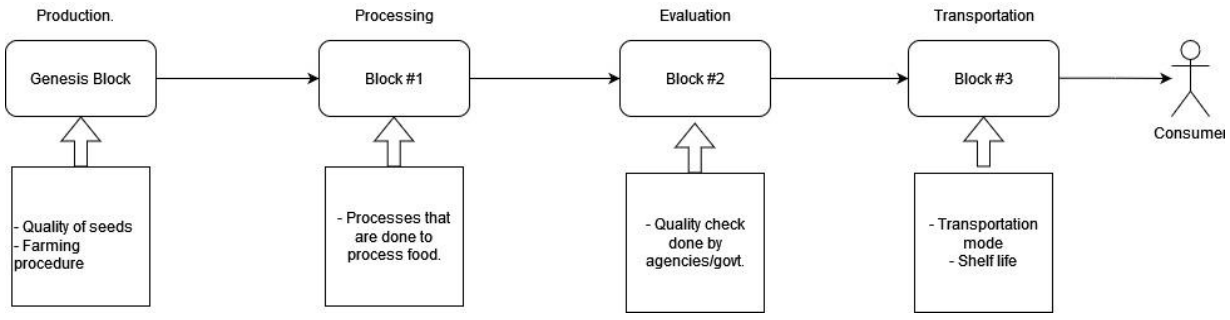


Figure 2: Blockchain in food supply chain

3.2 Lightweight Blockchain Applications for IoT security

Smart cities require a smooth process for communication between peers for each industry to work well with one another. This is especially evident in infrastructure, where it is important that city planners can quickly communicate with designers, engineers, and more; and vice versa. The Internet of Things (IoT) is the network of devices that make this communication fast and seamless. The IoT network plays a key role in smart city communication and offers many ways to improve infrastructure efficiency. IoT consists of all devices that communicate and interact over the internet and transfer data. In today's world, nearly every aspect of human life now utilizes some form of internet-capable device. As a result, a lot of personal information is being collected. This information includes uniquely identifying data such as biometrics, photos, videos, and audio. The IoT environment must be able to handle, store, and transmit this information in a secure and reliable way in order to preserve security and safety.

Most smart cities are attempting to maintain the same level of security and privacy that was in place before the implementation of more advanced technology. However, the IoT network of devices currently has many security risks and vulnerabilities that need to be addressed to ensure the security of infrastructure across smart cities. A possible solution for these security vulnerabilities is blockchain technology, which can provide a more advanced security environment compared to central database security.

A decentralized security model can allow for faster growth of technologies and greater information security, however, it also leaves room for risks to be overlooked as there can be a lack of consistent security measures. Despite this flaw, decentralization is important in order to maintain anonymity and to ensure that no one user controls the system. Blockchain also offers immutability, high fault tolerance, high availability, transparency, and auditability. Immutability prevents users who can access the blockchain from modifying or changing any existing stored results or transactions. High availability and transparency give everyone access to view transactions and results in order to allow for auditability. Anybody with access to the blockchain can audit and verify that all transactions and results are valid.

Blockchain has become very popular over the past two decades, especially in the financial industry with cryptocurrencies. In this field, blockchain has proven to be a reliable method for logging and verifying transactions between users. Figure 3 below shows how each block of information is linked to one another in the chain [29]. Blockchain's decentralized data management system would be an efficient solution to the privacy challenges faced by IoT networks in smart cities.

Figure 3: graphic of how a blockchain works [29]

The reason blockchain applications have not been implemented right away on the Internet of Things is due to energy consumption and scalability. Blockchain applications are often resource-intensive which limits their applications. The IoT network has limited computational resources and capabilities, which has led to vulnerabilities and a greater need for data security. In addition, this lack of resources prevents the usage of the standard blockchain. A standard blockchain implementation requires a large number of resources in order to process the security algorithms. As a result, there have been many studies on how blockchain technology can be integrated in IoT devices.

Most of the proposed solutions from the original studies used a public blockchain platform to secure and manage IoT devices. However, using a public blockchain platform adds performance limitations and an unnecessary extra hardware and network cost. Using these blockchain platforms for time-sensitive IoT applications is not efficient and as a result, researchers have been trying to find a way to utilize blockchain technology without dealing with the large overhead.

One example of a possible solution was using Hyperledger Fabric, a permissioned blockchain platform to manage and secure IoT devices. Hyperledger Fabric uses a lightweight consensus mechanism which has better performance. However, there is still one underlying issue with Hyperledger Fabrics, scalability and overhead [1]. After failures in implementing a standard blockchain technique, researchers started to investigate other applications that utilize blockchain technology, but do not require the resources and overhead.

Most notably, in 2019, Dorris et al. proposed a new type of blockchain architecture, lightweight scalable blockchain, that is “optimized for IoT requirements while also providing end-to-end security” [2]. These applications include simplified algorithms that still ensure data security. The solution was to implement a lightweight blockchain architecture like the one that makes up the Bitcoin infrastructure [29]. A lightweight architecture simplifies the blockchain algorithm without losing data security. The purpose of this type of blockchain is to help secure applications that need data security and reliability but lack the resources. The Internet of Things is one of those applications where lightweight blockchain could be very useful due to IoT’s

computational limitations. Due to IoT's computational limitations, the implementation of a lightweight blockchain may be extremely beneficial in negating these constraints.

To ensure scalability in a Lightweight Scalable Blockchain (LSB), IoT devices, cloud storages, and service providers are organized into clusters. The cluster heads are responsible for managing the blockchain. This involves verification and storage of individual transactions, a line of communication between nodes. In order to reduce memory and packet overhead of the blockchain, the data of IoT devices is stored in the cloud, off-the-chain. This is one way in which the LSB helps to reduce the overhead.

In addition, the flow of data to and from the IoT devices is kept separate from the transaction flow. Transactions are broadcast among the overlay nodes while data packets are routed toward their destination. This separation results in reduced delays and packet overhead for transferring data. Table 1 below summarizes the methods that allow the LSB to meet the most essential security requirements.

Requirement	**Employed method**
Confidentiality	Encryption is used for all transactions.
Integrity	Each transaction includes a hash of all other fields contained in the transaction.
Availability	An overlay block manager sends a transaction to its cluster members only if a key contained in the transaction matches one of the entries in its keylist. This ensures that the cluster members only receive transactions from authorized nodes.

Authentication	Each node should have a stored genesis transaction in the blockchain to be authenticated. As transactions are chained to the genesis transaction, a node is authenticated when it has the private key corresponding to the output public key of a transaction stored in the blockchain.
Non-repudiation	Transactions are signed by the transaction generator to achieve non-repudiation. Additionally, all transactions are stored in the blockchain, so involved parties in the transaction cannot deny their complicity in a transaction.

Table 1: security requirements and resolution Lightweight Scalable Blockchain [2]

LSB uses non resource-intensive algorithms that eliminate the need for computing hashes prior to appending a block to the blockchain, thus using less hardware for computations. This makes the application IoT friendly as it does not need to use up valuable resources. In addition, LSB uses a distributed trust method to decrease the processing time for validating new blocks as the overlay block managers continue to build up trust in one another. This allows the program to scale based on usage over time, leading to more effective and efficient security measures.

Finally, several security analyses have shown that LSB is secure against a broad range of attacks. This was a major concern amongst researchers because the LSB does not require as much overhead and as a result security measures may be compromised. However, after several analyses, studies have found that security is still very strong using LSB and meets all the requirements, such as those given in table 1 above. LSB is a strong contender to improve security within IoT as it continues to bring a high level of security and anonymity for IoT users while offering a much lower overhead than a standard blockchain application [2].

3.3 Blockchain Applications in Healthcare Security

In the medical system, Block chain innovation is often used to store and share patient information amongst facilities, testing centers, pharmaceutical corporations, and clinicians. In the healthcare profession, blockchain technologies can reliably identify the most important and even hazardous errors. The technology assists health organizations in gaining an insight & increasing the assessment of health files. Researchers have investigated Block chain as well as its substantial advantages in healthcare in this research. This research will review and describe 14 important Blockchain medical applications. Blockchain is crucial in dealing with fraud in clinical studies; the potential exists here to enhance the data effectiveness for treatment. This can help to alleviate the worry of manipulating data for healthcare by enabling a one-of-a-kind backup and recovery pattern with the greatest amount of safety.This offers data accessibility wide range, interconnectedness, traceability, and verification. Healthcare data must be retained secure and private for a variety of reasons. Blockchain contributes to the distributed encryption of information in health coverage & helps in avoiding specific threats.

Objectives :

Findings are produced for analyzing block chain applications in the healthcare domain for the purpose of improving security, which includes proper diagnosis of the benefits in both performance and safety in telemedicine and healthcare systems through the use of various block chain facilities.

- To discover and debate Blockchain technology enablers for rejuvenating healthcare services.
- To explore and discuss relevant Blockchain applications in healthcare.
- To discover the Unified Production System of Blockchain application implementation in medical amenities provision.
- To recognize Blockchain technology's ability to help the global healthcare culture.
- To research Blockchain and its applications in healthcare..

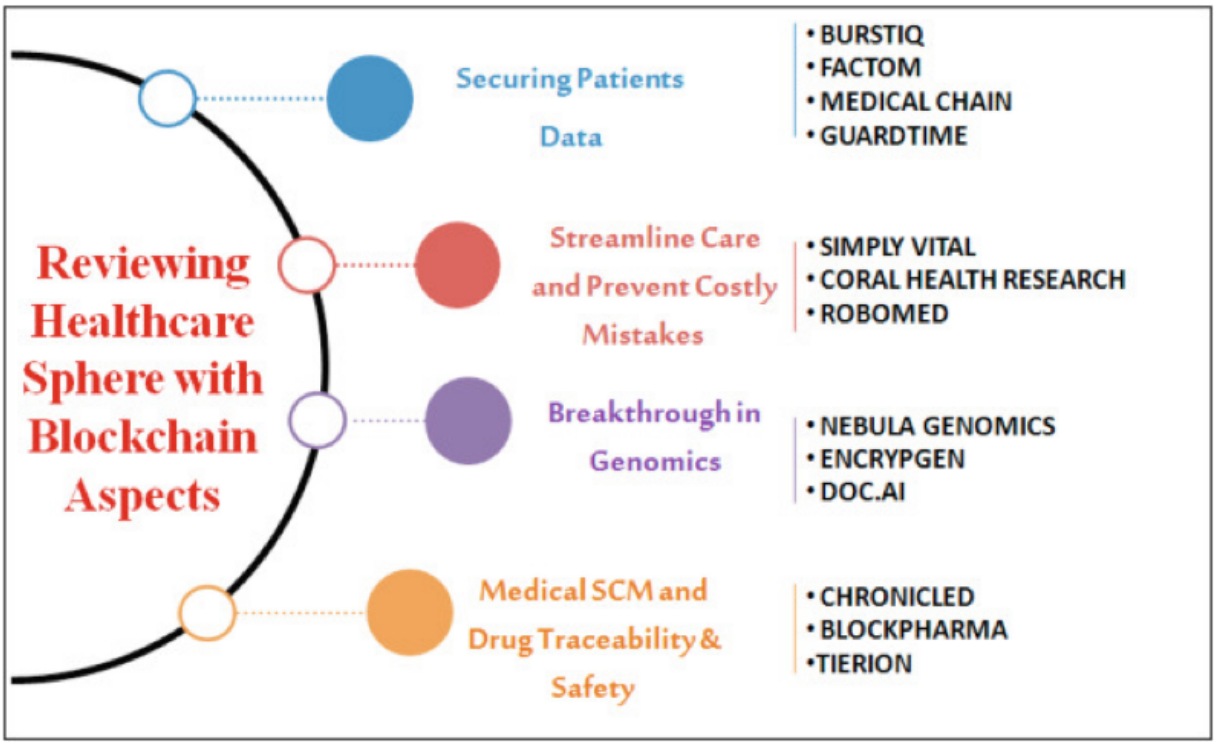


Figure 4: Blockchain enablers in healthcare facilities [6]

Data storage for various patients:

Before or during the various rounds of clinical research, a large amount of individual care and healthcare information is collected [16]. Quality assessments, blood tests, estimates, and wellbeing polls are all provided. This may produce more accurate results, indicating the presence of a paper or other information. Healthcare professionals will look over the data and presume its validity, which they will confirm by comparing it to the real text stored on the Blockchain.

Analyze the consequences of specific techniques :

Scientists can rapidly test a given procedure on a large portion of a patient group using validated access to personal health information [17]. This has a number of notable benefits that increase the quality of care for a variety of patient groups. Medical enterprises will be able to collect data in the real world and supply patients with a wide range of spectral efficiency prescription drugs or solutions thanks to the Blockchain architecture. Hospitals' work is made easier by blockchain, which has the majority of the information in front of it. They will adequately educate individuals on how to consume the medicine based on the results. This would provide

doctors with real-time updates on a person's present condition based on wearable data, as well as alert clinicians.

Transparency and security:

These are two important factors to consider where Blockchain provides superior security and transparency while allowing clinicians to focus more resources on clinical results [17]. This would allow them to fund medical research and treatments for any rare disease. In a medical system, seamless data exchange amongst medical application developers can help to increase diagnostic quality, appropriate therapies, and cost settings. Blockchain helps diverse health system entities to keep in touch and share information for increased security and openness[18]. Individuals can trade and observe their knowledge by utilizing such a network.

Validation:

Analytics are used to verify transactions on a Blockchain before they are added to the chain. The validity of the material is safeguarded until it can be encoded, password-protected, and archived. Healthcare institutions, technical entrepreneurs, and the medical profession are all seeking an opportunity to discover what they'll do now and in the future to make services more inexpensive. Blockchain may bring in a revolution in the help to boost when healthcare administrators can readily validate data.

Keeping track of health records:

For patient records, blockchain has the potential to be a game-changer. The exchange of health information, the storage of eHealth, the management of coverage, and the performance of administrative activities are all examples of its applications. Patients can communicate personal health data to a Blockchain through an app. Virtual Blockchain agreements improve the interaction between detectors and cognitive devices [31]. In most cases, telehealth is spread throughout multiple healthcare systems. All information will be consolidated on the blockchain, and patients will be able to access historical data. The aggregation of all data in one place can provide new information about a clinical condition. As a result, the Blockchain concept would ensure the data's authenticity and legitimacy while also safeguarding the privacy of individuals.

Falsely identifying content :

The blockchain will promote transparency and the discovery of fake content. Individual

and consumer medical studies will continue to be simple to assess. The smart contract is ideal for obtaining approval and keeping protocol documents and results accessible to the public. The innovation allowed the general public to see what happens in medical research for the first time [24]. This idea aims to be user-friendly while also providing patients with secure real-time access to their hospitals.

Patient monitoring :

The trustworthiness of a Blockchain allows health professionals to ensure that they have access to emergency equipment when it is needed. Physicians could even spend hours watching patients and responding to health issues remotely [4]. Using Blockchain and medicine, it is possible to improve temperature tracking inside nursing stations, bed usage, and equipment organization. The blockchain healthcare infrastructure is being developed to provide health care providers and facilities with a permanent online presence.Blockchain and IOT advancements consolidate to advance supply chain flexibility and accountability,bringing better transparency for effective healthcare management.

Enhancement of safety :

Blockchain increases patient safety, resolves concerns with medical validity and pharmaceutical traceability, and allows for secure connectivity. This is the only way to alter the present supply chain strategy and prevent fraudulent drug manufacturers from bringing medications to market. Regardless of the type of healthcare facility or organization, Blockchains will allow all data to be processed in a single location. The Blockchain system's openness will allow clinicians to quickly view substantial health records to aid in diagnosis and the development of such a speedier and more precise therapy [6].

3.4 Blockchain for Smart Grid security and resiliency

In the grid modification, application of Blockchain smart contracts gives a chance to increase the speed, scale, and security of the modern grid [20]. The data's speed, security and control are required to increase real-time transactive energy and distributed energy resources. In the operations and designs, grid resilience improvement is essential for energy delivery systems that help in modernizing and help in exchanging and consuming the energy; at the edge of the grid, the energy delivery systems require high-end security and trust for verifying the data integrity and managing complex Transactive and distributed energy resource (DER)

exchanges. For real-time energy transactions, the required speed, scale, control, and security are not matched with visibility, control, and security of grid edge devices [20]. As a result, to fill these optimization and security gaps, smart energy contracts of Blockchain are used to help improve the grid resilience art. This provides an automated cryptographic signature with distributed ledgers for increasing the trust, integrity and resilience of the energy delivery systems. For verifying time, user data blockchain are used, which also help in protecting the data.

Objectives:

The report's main goal is to provide information related to Pacific Northwest National Lab (PNNL), its unique testbeds, and its integration process for accomplishing the desired goals discussed below.

1. In the energy sector, the transaction costs are reduced and secured distributed energy exchange can be facilitated.
2. On the transactive campus of Pacific Northwest National Lab (PNNL), the different blockchain solutions went under testing. It helps in investigating implementation issues and helps in reducing cyber-related risks.
3. The security applications of blockchain technology are validated and verified for Transactive energy.
4. At different levels, communication is essential for investigation between the blockchain ledger and control system.
5. Different regulatory and standards models are mandatory for facilitating the decentralized security measures adoption and sustainability like Blockchain.
6. Market standards are required to improve interoperability, cost control and complexity mitigation.
7. Cyber security is imperative to investigate for controlling the costs of Blockchain.
8. To understand the different blockchain solutions that help in implementing the protection measures.
9. To comprehend the failure of Blockchain when delivering and developing improvements.

Advantages of Blockchain

Blockchain provides various benefits like high-end security and optimization when applied to electricity infrastructure. From the perspective of security, Blockchain helps enhance

trust and transactive energy data integrity where it supports multi-factor verification with the help of distributed ledgers. Although, for data anomalies and real-time response, Blockchain can provide autonomous detection to unauthorized attempts. It helps in improving the applications, network and sensor infrastructure. The other benefits of Blockchain are discussed below.

1. Trustworthiness is enhanced, and data integrity is preserved.
2. With the help of distributed ledgers, multi-factor verification is supported.
3. Transaction data integrity is secured.
4. Eliminate intermediaries and control the energy exchange costs.
5. Distributed Energy Resource (DER) transaction adoption and monetization are facilitated. In real-time, the transactions are implemented, and settlement is done on actual consumption [20].
6. DERs and Electric Vehicles (EV's) generate the excess exchange at the consumer level, facilitated efficiently by smart contracts of Blockchain.
7. To maintain the stamp data blocks of ordered time, high-end security distributed escrow is enabled that cannot be updated retroactively.
8. Data anomaly detection can enhance the detection and responding time ability of cyberattacks.
9. In the energy sector, transaction costs can be reduced with the help of Blockchain.
10. The operators of the distribution system can leverage blockchain to receive the transaction data of energy requisite for changing costs of the network from consumers.
11. Data requirements and constraints can be reduced through transmission system operators for clearing purposes.

The application of smart contract in the energy area

The following model (figure 5) illustrates how blockchain technology is applied to the power system to help lower costs by eliminating third parties and enhancing arbitrage opportunities. This model depicts how individuals have the chance to create and sell energy to one another. Smart contracts/ programs enable peer-to-peer energy exchanges, among other things by allowing customers and suppliers to trade with one another in other words, rather than interacting through a multi-tiered structure, electricity grid operators, distributors, and telecommunications network operators, the producers and consumers engage in a direct exchange of energy and these transactions are highly secured and much faster. In 2016 the first

case was recorded of the blockchain system where the neighbors in New York exchange the energy through the decentralized blockchain system [20]. With the help of the model in Figure 5, the process of Blockchain was explained, which helped to eliminate the need for intermediaries in the transaction process—the generation of electricity, which consumers and blockchain purchase, is updated by meters of Blockchain.

In the distributed ledger, a unique timestamped block is created for verification. The system operators use the Blockchain at the distribution level for receiving energy transaction data so that network costs can be charged from consumers. For transmission operators, data requirements are reduced, and speed is increased for the clearing transactions because the transactions may be carried out and paid based on real usage. In the load ledger of Blockchain, the smart contracts are built and help record transactions through Blockchain, which helps in enabling the advanced metering infrastructure (AMI), from Distributed Energy Resource (DERs), Electric Vehicles (EVs), etc., smart contracts of blockchain help in facilitating the consumer level exchange. This helps provide the additional storage facility, and from bulk, energy system substation load balancing can be done efficiently.

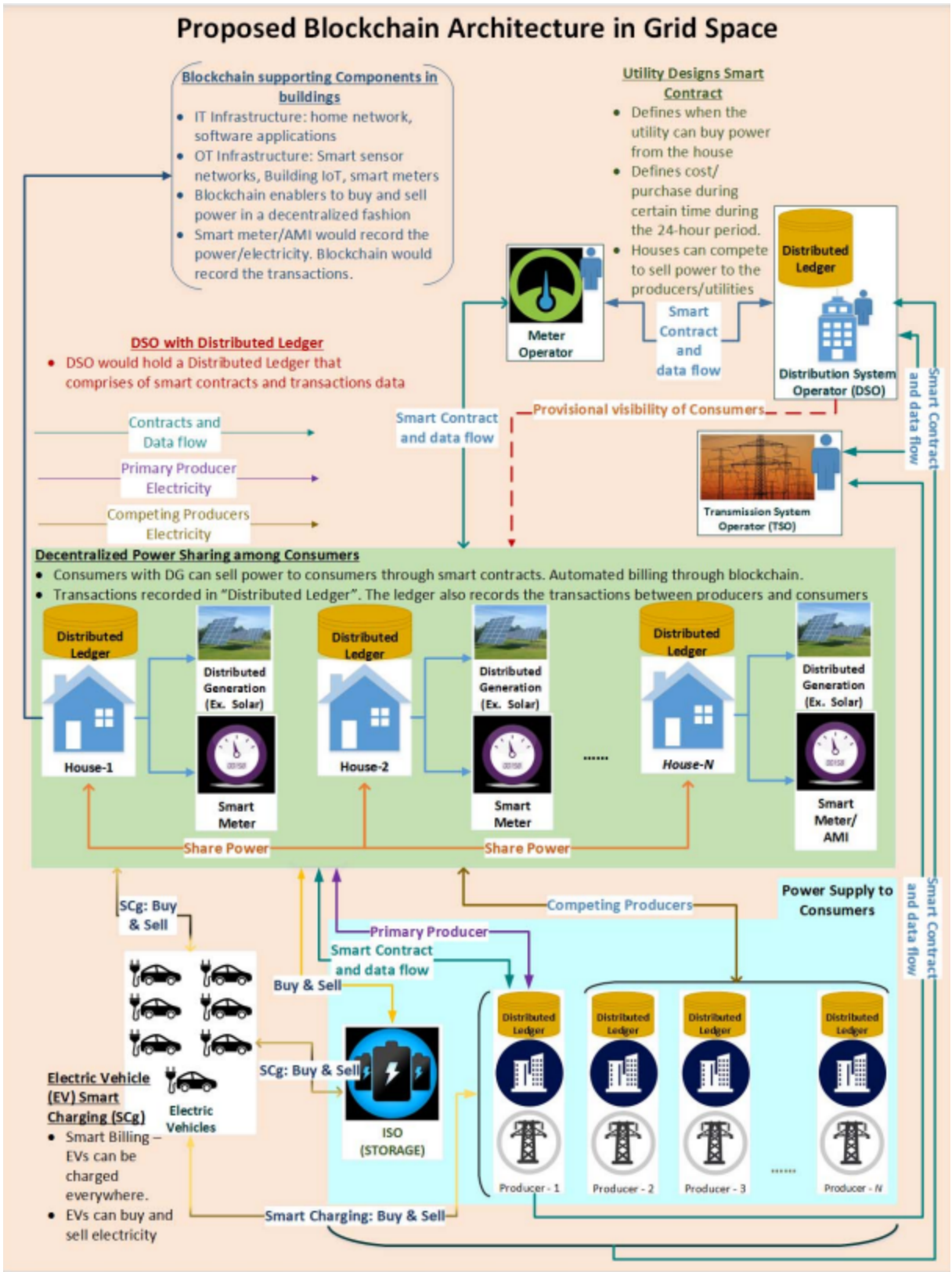


Figure 5: In the Energy Infrastructure the Application of Blockchain [20]

Cybersecurity of Blockchain

Blockchain offers a one-of-a-kind method for disseminating trust that includes the cybersecurity value proposition. Cybersecurity comes with a variety of advantages. A distributed escrow, for example, can be used to keep track of stamped data blocks that are next to impossible to be changed or modified. As a result, data integrity and trustworthiness are improved. Keyless Signature Infrastructure (KSI) and other blockchain implementations, increase integrity measures and assist boost in reliability of authentication and encryption which isn't burdensome and are cost efficient in deploying keys [20]. Blockchain applications with effective features, increases trustworthiness and data provenance. With the use of Blockchain, communications can be made more secure. The communications can be secured with the application of Blockchain from different operational technology and control systems. It also includes an advanced crypto signature that provides data signer, data authenticity and data asset signature timing [20].

Guardtime is the most effective and largest blockchain supplier that provides a keyless signature facility to give a high-end secured global platform that allows customers to employ cryptographic signature technology with ease [20]. With cryptographic signature verification and distributed infrastructure, data integrity, competition, and real-time energy exchange for microgrids, as well as generating & sale of energy from concept to implementation, are all conceivable. Increased data integrity in the Blockchain can assist detect cyber-attackers and improve Distributed Energy Resource (DER) grid connection resiliency. Current approaches are employed in energy distribution, which also assist in connecting buildings-to-grid (B2G) during the susceptibility of cyberattacks. The energy management system, on the other hand, received Distributed Energy Resource (DER) data that may be tampered by attackers who can take advantage of insecure connectivity while compromising Distributed Energy Resource (DER) input and output signals. This sort of assault can be prevented via Blockchain, which includes a ledger where electricity transaction time and consumption can be recorded.

In Grid Space, the Questions Related to Blockchain Cybersecurity

Both grid modernization and state of the art technology can be improved with the help of smart contracts/ programs of Blockchain. In this process, Blockchain increases the electricity infrastructure's trustworthiness and efficiency. Blockchain runs on a cryptographic signature that helps prevent the energy transaction manipulation and energy delivery systems configuration. Due to the Blockchain's improvised version, the resistance of the process is high in the grid power. Smart contracts also do away with the need for third parties to verify and enforce contracts. The data can operate with high speed, total confidentiality, and secure control thanks

to the use of transactive energy in real-time. In this part, we'll look at some of the most important cybersecurity questions relating to Blockchain.

What is the process implemented by Blockchain in preventing adversaries from attacking a transactive retail market and posting their fake signals on the market [20]?

Blockchain does not provide a hundred percent solution for security threats, instead it increases the security measures by providing authentication, encryption, and ability to provide verification of data [20]. Yes, we can prevent attacks in the transactive retail market by using Guardtime's Blockchain which is a permission-based process.

The first step is that the KSI of Guardtime offers a distributed ledger that helps in recording all transactions. Data authenticities are maintained, the signing entity is verified and signing time is recorded. KSI cryptographic technology of Blockchain is the first security layer. It helps in assigning a data signer, which maintains data authenticity and signing time to a data asset in the transactive retail market. In the digital form, the format of data is signed.

The second step is the Blockchain KSI's stack. A high reductant and high-end secured platforms are provided by this layer which helps in leveraging the accessibility, and entities can participate in the singing events of cryptographic.

What techniques does blockchain use to prevent consumers from hackers from entering behind the meter systems [20]?

The behind the meter systems access may not be prevented by Blockchain. However, the necessity of a third party can be reduced or removed by blockchain to clear transactions. Due to this step, the chance of attack can be reduced. But the main purpose is to secure the data integrity after the attacker enters the system. In this case, configuration manipulation can be detected by blockchain and helps in isolating malicious actors. It can also help in providing tampering forensic evidence, which is sufficient for baselines reconfiguration.

How do DER devices and nodes can be unambiguously identified by blockchain and help in preventing spoofing [20]?

Blockchain has two security layers for data validation. For an entity or data block to be a part of the Public Key Infrastructure, it should be witnessed by a trust anchor provided by Keyless Signature Infrastructure (KSI). Then and only then the data block could be validated. And without it, it is not allowed to be a part of the infrastructure. In this way, spoofing is prevented. It also provides data information and signing time without any disturbance.

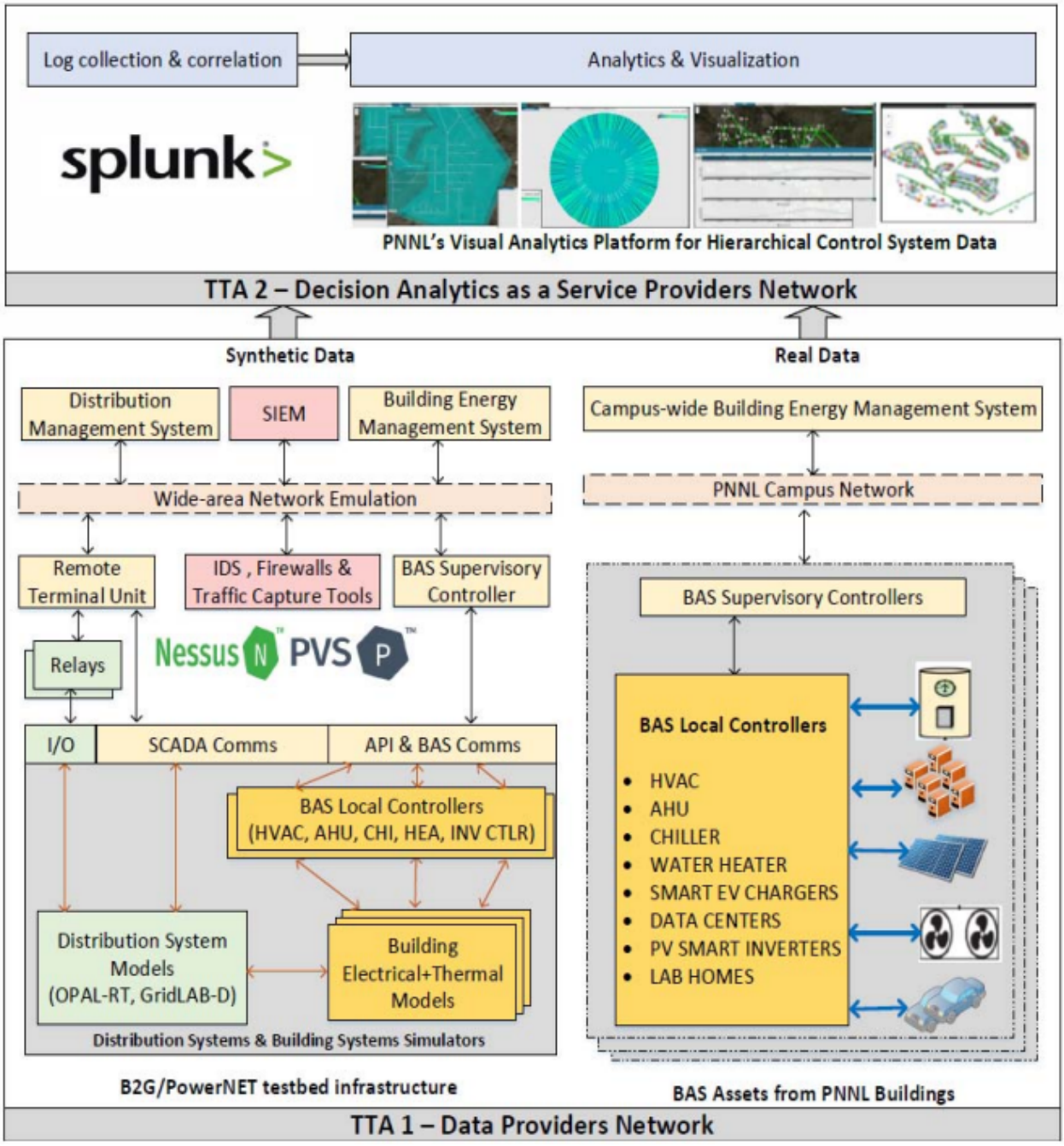


Figure 6: Cybersecurity Testbed of Buildings to Grid [20]

Researching blockchain cyber risks, vulnerabilities, and mitigations in the context of protecting the grid's edge and providing secure transactive energy solutions might be extremely beneficial to grid cybersecurity and resilience research. This project leverages PNNL's buildings-to-grid cybersecurity testbed to establish and validate the application of blockchain technology to securing distributed energy exchange at both speed and scale (Figure 6).

Researchers may model and simulate energy delivery systems from the distribution substation to the end user at PNNL's B2G Cyber Testbed. The model incorporates not just physical system aspects, but also the underlying cyber control system, which leverages industry-grade hardware and communication protocols to mirror real-world control hierarchies properly. Operational Technology (OT) and Information Technology (IT) communication are used at three levels between ledger and control systems: first, local level- controllers of building automation systems (BAS). Second, system-level- at the Building Energy Management System (BEMS) level, where controlling is done. Lastly, a system of systems-level-control at the distribution control center [20].

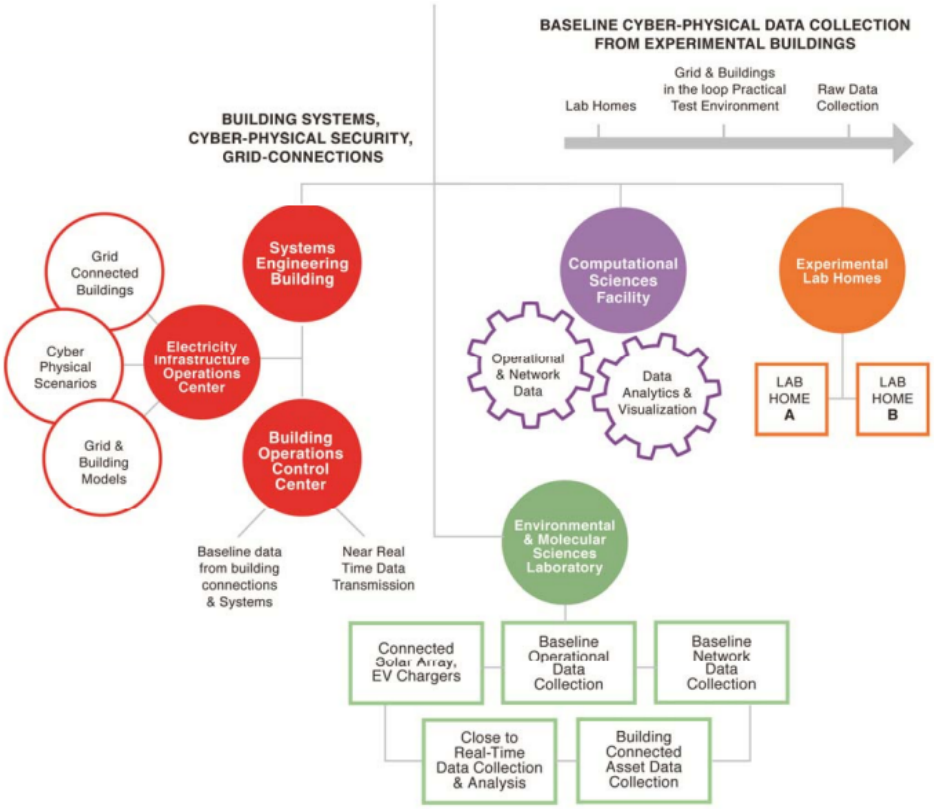


Figure 7: Cybersecurity Integration of B2G in PNNL and Transactive campus [20]

Improving Transactive Energy Solutions of Blockchain, by combining B2G Testbed of PNNL and Connected Campus.

This diagram (Figure 7) depicts the high-level rendering of Pacific Northwest National Lab's (PNNL) buildings-to-grid cybersecurity testbed integration with the infrastructure that supports the transactive campus.

While the Buildings-to-Grid Cybersecurity Testbed at PNNL can simulate a variety of cyber-attack scenarios, threats, vulnerabilities, and blockchain mitigations, the Connected Campus at PNNL can analyze and test blockchain applications at the cyber energy nexus at the required speed and scale [20]. Combining these two distinct setups might help enhance blockchain applications for transactive energy security, speed, and scalability. This was a model developed by the PNNL, Washington State University and the University of Washington [20]. This research is done for the first time where testing is done at this scale using the building load's transactive control. Numerous buildings and devices are involved in it.

Transactive Campus of Pacific Northwest National Lab (PNNL)

Transactive energy is defined as a collection of economic and control systems that enable a dynamic supply and demand balance throughout the whole electrical infrastructure, with value as a key operational metric. A connected campus based on transactive energy, or simply transactive campus, is an environment built on such a transactive energy framework to execute transactive building load management involving several buildings and equipment. The Transactive campus of PNNL included systems of diagnosis and automation in numerous buildings at a time. It leads to identifying the operational area issues for managing the building loads. The increased visibility, controlling power of load transactions, costs, signs, instructions, and other essential data are provided by Blockchain to the operators that connect with a ledger of Blockchain to improve the following areas. First, security with control, second, through smart contracts trustworthiness, and third, through ledgers integrity and visibility of data. PNNL designs the transactive energy for managing devices' data and helps in decision making to connect buildings' environments. With the help of intelligent controls, communication can be possible and help adjust the energy loads. A realistic testbed is offered by the transactive campus for performing the framework testing of blockchain, first, for implementing the Blockchain on various connected buildings as overlaying security architecture in PNNL. Second, smart contracts are built for defining the exchange and consumption of energy, load management, and peer-to-peer transactions. Third, endpoint cyber security can be enabled for optimizing management controls of energy.

The transactive concept combines financial signals with dynamic control techniques to change the timing and quantity of energy consumption in devices, buildings, and campuses, resulting in better efficiency and cheaper energy costs while simultaneously providing more flexibility to the power grid.

3.5 Machine Learning Applied in Intrusion Detection and Network Management

Smart cities, just like any other network, need to have some sort of measure for intrusion detection and intrusion prevention. There are currently two methods used for intrusion detection: signature-based, and anomaly-based. Signature-based methods are implemented by a network engineer providing some patterns for malicious or unallowed traffic so that the incident may be reported and logged. The other method is anomaly-based which is much more difficult to implement since it detects abnormal behavior. Of course, this method requires an abundance of data to signify what exactly qualifies as normal behavior. Because of this anomaly detection lends itself very well to machine learning.

Some of the most important assets to protect from attackers within a smart city is the city infrastructure. When an attack is performed, there must be measures taken to categorize, log, and counter the network assault. One effective solution proposed by [3]. is applying the use of Restricted Boltzmann Models (RBM) to gather a set of high level information of a system from a large amount of data which is sent to a Feed-Forward Neural Network (FFNN) in order to train it for detecting and categorizing Denial of Service (DoS) attacks. Since this is a highly generalized solution proposal, it can likely be extended to many other systems of all sorts so that the networks of the future may be entirely prepared to detect and classify any sort of attack as needed.

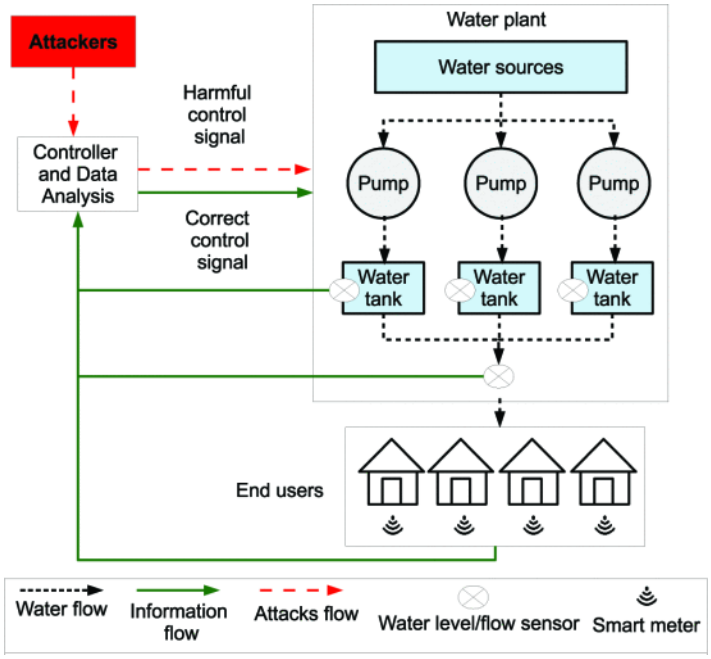


Figure 8: An Attack on a Smart Water Plant [3]

The specific example used in this study by [3] is a network for a smart water plant. There are many different methods for executing a DoS attack against a water plant and therefore, there should be some control analyzer to categorize the different types of attacks present, if there are any at all. A visual representation of a basic DoS attack against a smart water plant is shown in Figure 8. For this system, an anomaly detection system should be used to discover potentially malicious control signals. “Recently, deep learning algorithms have been applied for anomaly-based intrusion detection systems to enhance the detection accuracy for malicious users” [3]. The creators of this intrusion detection system (IDS),

too, agreed that an anomaly-detection approach would be much more effective than a signature based approach.

Within this solution domain, the main actor is the FFNN, which actually does the categorizing of control signals. This may raise a question of the usefulness of the RBM within the solution; however it is actually able to roughly double the performance of the detection and classifications. “The experimental results showed the ability of the proposed framework to detect distributed Denial of Service (DDoS) attacks, where the RBM model was able to enhance the performance of the classification algorithm.” [3]. Looking at figure 9, shows that in order to attain a false positive count with an RBM which is similar to the false positive count of the same algorithm without an RBM requires approximately 2.5 times the number of control signals. This shows that the RBM truly does exceptionally well in giving high level information about the infrastructure to the FFNN instead of only using raw data.

Without Refined Boltzmann Model (RBM)							With Refined Boltzmann Model (RBM)						
	Normal	**DDoS1**	**DDoS2**	**DDoS3**	**DDoS4**	**DDoS5**		**Normal**	**DDoS1**	**DDoS2**	**DDoS3**	**DDoS4**	**DDoS5**
Normal	756	0	8	1	9	3	**Normal**	1990	0	1	1	7	1
DDoS1	1	1133	39	7	24	25	**DDoS1**	6	2945	1	19	10	27
DDoS2	29	21	1092	0	1	1	**DDoS2**	38	8	2816	12	12	16
DDoS3	21	33	34	1002	12	15	**DDoS3**	16	15	4	2773	13	15
DDoS4	18	25	30	8	1049	20	**DDoS4**	39	17	14	12	2737	15
DDoS5	14	19	11	11	2	1086	**DDoS5**	14	31	11	12	22	2730

Figure 9: Comparison of classification rates with or without an RBM [3]

From this information, the question of what if more RBMs are added to the solution must be answered in order to conclude whether or not adding more RBM layers can drastically improve the performance of the system as well as adding a single layer did. Unfortunately, figure 10 shows data that suggests adding more layers to the RBM actually reduces the accuracy of the solution. Thus, the optimal solution for this algorithm that currently exists is to have one and only one RBM for interpreting data to be fed into the neural network.

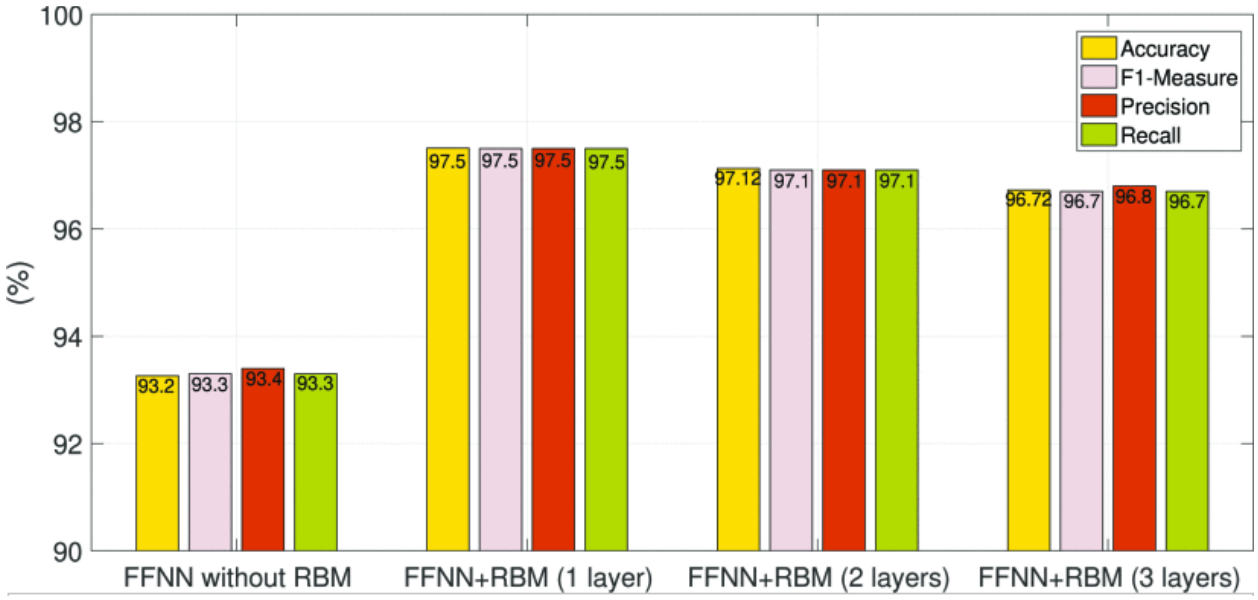


Figure 10: Performance of The Classification Algorithm with Multiple RBM Layers [3]

Machine learning can do more for the security and integrity of network infrastructure than just intrusion detection. Along the TCP/IP model, attacks can be done at every layer. An attack can be done at the application layer, the network level, and interestingly: the physical layer. A malicious actor could skip the virtual battlefront and simply tamper with network switches by rearranging or breaking cables. This is especially true for a fiber optic network which is where the field of networking is shifting to. Fiber optic cables are extremely fragile and can be rendered useless in ways that a standard ethernet cable would not. Ethernet is also not without its own more unique faults because the electrical signals can be confused with electromagnetic waves or some other form of jamming signals. Of course, many other network issues can arise and machine learning can be used to detect what those errors may be and where exactly the problem physically exists so that engineers may be dispatched to resolve the issue.

In the near future, when smart cities are fully realized, their networks will need to be fully autonomous in their diagnostics since the network infrastructure will be much more hectic than in a conventional household or business today. “These [machine learning] techniques have been shown successful in, e.g., detecting unauthorized signals in the network… or identifying jamming and polarization scrambling attacks.” [19]. Since machine learning has shown to be rather successful in the automated diagnosis of network infrastructures, surely it can be applied in a smart city’s optical network, as well. Unfortunately, this is not without its own challenges since applying machine learning can have a non-negligible false-detection rate because it was not manually programmed to have a specific set of rules which can classify a network fault. The

solution must also be very scalable since, as history has shown, the internet can and will continue to rapidly grow, even in a single city. Machine learning also comes with its own unique obstacles such as the complexity of training a neural network, finding how the system creates inferences from its reasonings, and also how well a dynamic system such as a machine learning algorithm can react to network architecture changes.

There are a few options that exist for training the network management machine learning model which are: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning can be appropriately applied when the knowledge of the network is complete with input and output for the model to learn. Unsupervised learning is quite different and is appropriate for when the input does not necessarily have a clearly defined output, nor any known normal or abnormal conditions - This can still be applied when these datasets should be grouped by similarity. Unsupervised learning sounds very unhelpful for a networking application of machine learning; however, considering that data will be grouped by similarity and that abnormal behavior should be less common than normal behavior, it suddenly sounds much more applicable. Semi-Supervised learning is a sort of hybrid of supervised learning and unsupervised learning. Semi-Supervised learning is appropriate for use when labeling the full dataset is impractical. This approach allows for the labeling of some data to be used in a sort of supervised-sense while the rest of the data can be categorized based on the small amount of information given.

Applying a network management system using machine learning can have some significant benefits in the form of autonomously diagnosing network issues. This can be done by finding the subset of devices which are affected by some disruption and finding the possible locations that would allow for that specific subnetwork to be affected by an issue and nowhere else. Of course this detection can be manually made (with some difficulty) but having a machine-learning based network management system is for much more than detecting where some problem may be. Depending on the learning approach used, a management system like this can also identify what exactly the problem is based on the signals recently received.

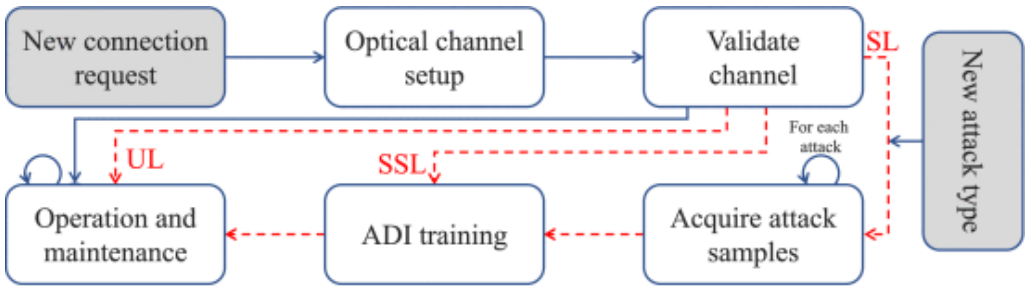


Figure 11: Procedure for Adding a New Node to The Network [19]

Unfortunately, the only learning method that supports attack identification is supervised learning. This is more of an issue because the training complexity of applying this class of learning is much higher than its counterparts. Figure 11 shows the steps taken by the various learning methods when a new device is connected to the network, or a newly discovered attack for supervised and semi-supervised learning. When either of these cases occur, an unsupervised learning approach will allow for a management system that does not need to relearn everything and can go straight to operation and maintenance. This is because unsupervised learning is not given a pruned dataset of any form to find patterns and classifications. Semi-supervised learning will go to the Attack Detection and Identification training stage (ADI) where it will learn from the small dataset provided. Finally, in a supervised learning setup, the system will have to go through every attack type and gather all known (and labeled) attack samples in order to train the neural network sufficiently. For a much more lightweight approach, unsupervised learning seems to be the best option. However, for a more capable management system, supervised learning is best for its ability to identify attack types.

Property	SL	SSL	UL
Requires NOC data	Yes	Yes	Yes
Attack detection	Yes	Yes	Yes
Attack identification	Yes	No	No
Requires attack-specific labeled data	Yes	No	No
Training complexity	High	Low	None
Re-training for new OChs	Yes	Yes	No
Inference complexity	Low	Low	High
Requires prior samples	No	No	Yes
Supports stateless operation	Yes	Yes	Yes

Figure 12: Properties of Learning Methods for Network Management Systems [19]

Figure 12 shows the properties of each learning method for the management neural network. As expected, each method requires network data and is able to detect anomalies (attacks). The point where supervised learning really shines is in its ability to identify attacks rather than just detecting them. This classifying of attack knowledge can help immensely in mitigating the problem and finding a more permanent solution. This attribute does come at a cost; however, because to attain this attack identification capability, the supervised learning method requires labeled and organized data specific to each attack whereas the others do not. Training complexity trends increase as more organized data is required for training, as expected. Thus, supervised learning is very complex for training, unsupervised learning has no

complexity associated with it and is more of a “plug and play” sort of solution, and semi-supervised learning has some training complexity since it is a hybrid of supervised and unsupervised. As discussed previously, supervised and semi-supervised both require re-training for the discovery of attacks and new devices connecting to the network. Inferencing for unsupervised learning can be rather difficult since there are no labeled categories for data to be categorized into while the others do form groups for data to fall under which can help in the decision-making process. Unsupervised learning requires prior samples since it needs a basis for categorizing similar data and thus is somewhat less supportive of stateless operation than its alternatives.

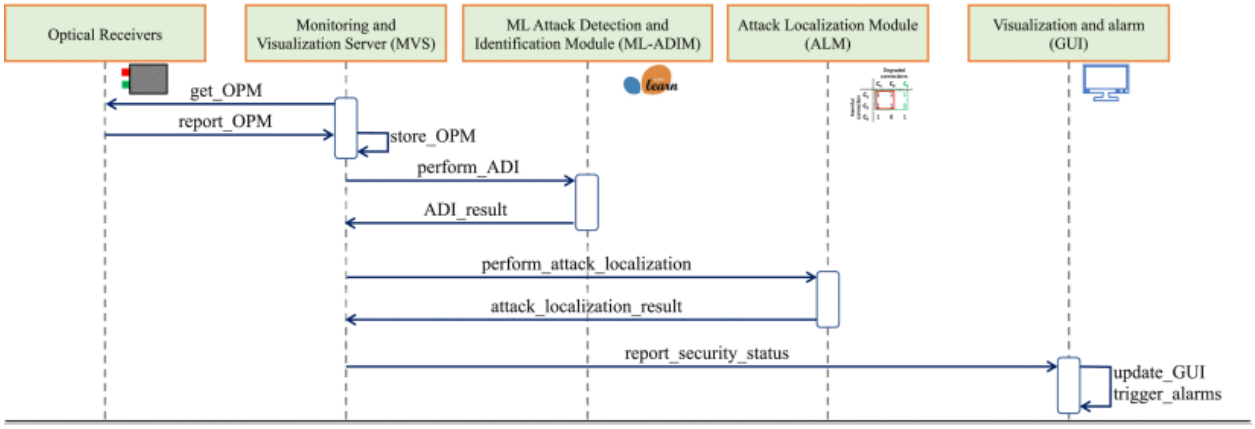


Figure 13: Proposed Framework for a Machine Learning Network Management System [19]

The proposed solution framework, depicted in figure 13 above, applies a Representational State Transfer (REST) architecture in order to form various microservices for the monitoring and visualization server (MVS) to invoke. First, the MVS collects data of the network from optical receivers, then requests identification of any attacks to the ML Attack Detection and Identification Module. This module is the one which actually applies machine learning techniques to discover anomalous behaviors and potentially identify and categorize attacks. Next, the results of the attack detection and identification are interpreted and sent to the Attack Localization Module which discovers the physical location of some issue so that a team of engineers may be dispatched to resolve the problem. Of course, there needs to be some alert system for these engineers to be sent out, so there is a visualization and alarm system for users to view the system in a human-readable format.

3.6 Machine Learning approaches towards Data Security: Malware and Spam Detection

Malware Detection

Non-ML Based Approaches

For android malware detection, there have traditionally been two approaches – static and dynamic or a combination of the two [7].

In static approaches, features distinguishing malware from benign applications are extracted without having to execute the code. However, in dynamic approaches, the same is done by running or executing the code. Since code execution is not carried out, static approaches are usually more efficient compared to dynamic approaches. Features such as APIs, permissions, intent, domain names among others are considered. However, during static analysis, parts of the code that require dynamic loading of libraries cannot be analyzed. On the other hand, dynamic approaches can be more informative as only code that executes is considered [9]. However, with dynamic analysis, all execution paths may not be considered, as is done in static approaches, which can limit the overall view of the code.

Machine Learning Approaches

The success of traditional malware detection approaches is highly dependent on the selection of right features. Even with the selection of right features, the system relies heavily on static features, making it difficult to classify newer and different types of malwares. An attack can consist of variants of any malware which can easily evade the traditional malware detection systems. Having said that, the different types of malwares share similar patterns. Because of these recurring patterns within malware families, applying machine learning approaches to the task of malware detection have yielded increasingly better results, both in terms of accuracy and robustness.

Deep learning approaches in particular are being used to automate the feature extraction. A deep learning model consists of a neural network with multiple layers, each of which transform the input data as it passes through them. Each layer helps extract features in an incremental way. This way of feature extraction helps minimize the reliance on static features for training a classification model.

One such machine learning pipeline for the task of malware detection has been proposed by [14]. Two major issues have been addressed with this approach – extracting the right features and working with a limited and often unlabeled dataset. The pipeline consists of two machine learning models. The first takes care of extracting features from a limited dataset, without the need for it to be labeled. The second model uses this feature representation to

perform the actual classification. Sparse autoencoders have been used for the first task while artificial neural networks have been used for the second. A balanced dataset consisting of a comparable number of malware and benign applications is used. This is created by pulling applications from various sources such as Malgenome, Contagio Minidump, Playdrone datasets, among others.

First, the information from the training data is condensed into sparse matrices, referred to as “API-Images”. The temporal sequence of API calls made during the execution of the application in the training set are captured in the sparse matrices. Looking at these API-Images helps in describing the behavior of the application over time. Also, capturing the information in this way helps standardize it so it can be consumed by any type of machine learning model later in the pipeline.

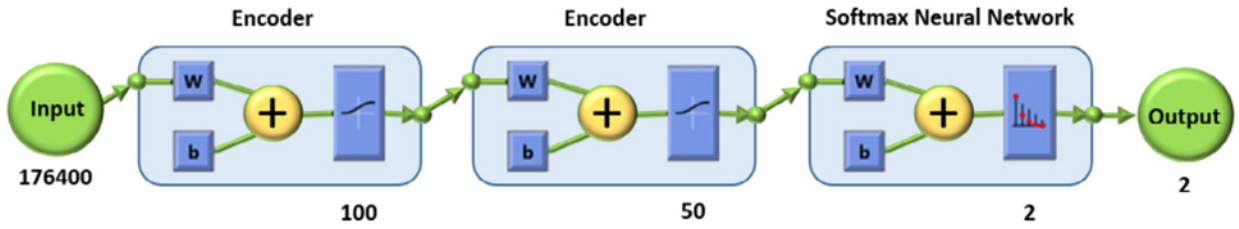


Figure 14: Malware Detection Framework [14]

Second, autoencoders are used to extract useful information, features from these API-Images. Autoencoders have traditionally been used for dimensionality reduction, a process through which less useful features in a dataset are removed. The output is a compressed, information rich representation of input data while also preserving any complex relationships among features. This learned representation can be used in downstream tasks. In the proposed approach, sparse autoencoders have been used. At a given time, a minimal number of hidden neurons in the layers are active, mimicking brain function. This helps to disregard the less helpful features of a dataset, resulting in a representation consisting of discriminating features.

Third, an artificial neural network is used to classify an application as either malware or benign. A softmax activation function is added as a layer just before the output layer, with binary cross entropy as the loss function. Such an architecture aids in multi-category classification of malwares.

The results of this autoencoder based architecture is compared with other traditional machine learning algorithms. The results comparison is shown in Table 2.

Classifier	Acc.	Sens.	Spec.	Prec.	MCC	AUC	F-measure
J48	0.85	0.95	0.25	0.88	0.28	0.60	0.91
Naive Bayes	0.83	0.93	0.25	0.87	0.22	0.59	0.90
MPL	0.83	0.96	0.12	0.86	0.12	0.54	0.90
API-Images	0.94	0.96	0.88	0.98	0.79	0.92	0.97

Table 2: Malware Detection: Comparison of DL approach with other ML models [14]

The proposed approach deals with the issue of a limited dataset as well as dynamic feature representation useful for malware detection and classification.

Spam Detection

Spam refers to unwanted messages, often sent in bulk, through various digital mediums such as email, SMS, et cetera. Spam messages are a threat to data security as they act as a medium for malware, especially among mobile devices. Spam detection can be applied to both incoming and outgoing messages.

Several spam filters exist, the most common of which is content based spam filtering. In this approach, the emphasis is on the textual content of the spam message. Translated to the machine learning domain, spam detection becomes a text classification problem. Content based filtering is further divided into two types – direct content filtering and collaborative content filtering. Verbatim words and phrases form the training dataset in direct content-based filtering. In collaborative content filtering, a group of users share information regarding spam messages, usually resulting in a library of spam signatures. Any incoming message is compared against this library to be classified as either spam or not spam.

SMS spam detection has unique problems when compared to email spam detection. Unlike the latter, text messages tend to be shorter, which makes the task of differentiating between spam and non-spam texts harder. In addition, the text in SMS tends to be informal, which can be influenced by several factors. Variations in phrases, idioms, abbreviations etc. are commonplace and would affect the reliability of any spam detection system. SMS spam datasets are also not as widely available as email spam datasets.

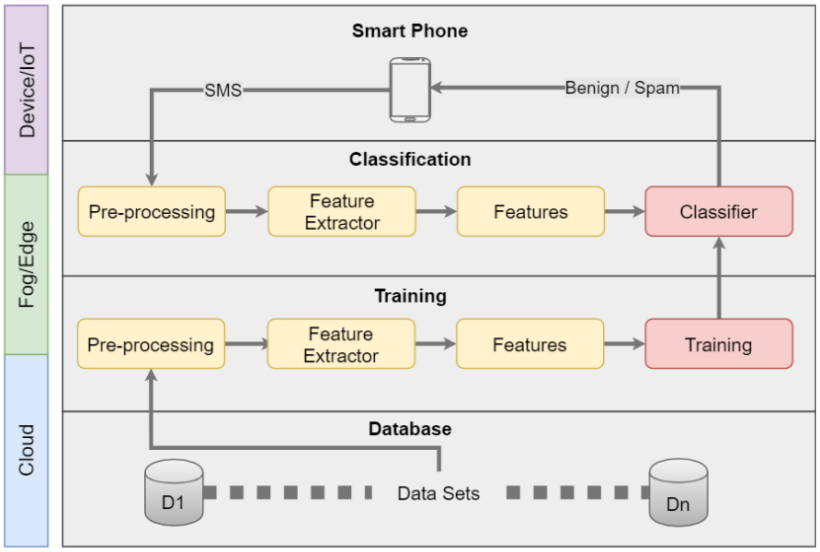


Figure 15: The System Architecture for ML Based Spam Detection [23]

Several text classification algorithms are used for spam detection. [23] proposed an architecture comparing several of these algorithms. They experimented with multiple filters and classifiers, coming up with novel combinations of preprocessing, feature extraction and text classification algorithms. Preprocessing steps involved tokenization, stop word removal and stemming or a combination of these. The different filters and preprocessing steps are shown in the table below. Three machine learning classifiers have been studied – Naïve Bayes, Naïve Bayes Multinomial and Support Vector Machine.

Name	Description
PF1	Word Tokenization, No Stemmer or Stop word, words to keep: 1000
PF2	Lower Case Tokens, evaluator: Stop word (Rainbow), stemming (Lovins Stemmer) and words to keep: 500
PF3	Pre-processing: Lower Case Tokens, evaluator: Stop word (Rainbow), stemming (Lovins Stemmer) and words to keep: 500. Feature Extractor: InfoGainAttributeEval-search: Ranker with a threshold=0
PF4	Lower Case Tokens, and words to keep: 500. Feature Extractor: InfoGainAttributeEval-search: Ranker with a threshold=0
PF5	Adding a new Attribute msgLength which stores the message length values and then performs PF4 filter on it

Table 3: Preprocessing and Feature Extraction Filters [23]

They studied how the variations in training datasets, preprocessing and feature extraction steps impact the classifiers' performance. The performance was measured using criteria like accuracy, precision, TPR, FPR, TNR, FNR. It is seen that the best performance is attained with the PF5 preprocessing technique in conjunction with the SVM classifier.

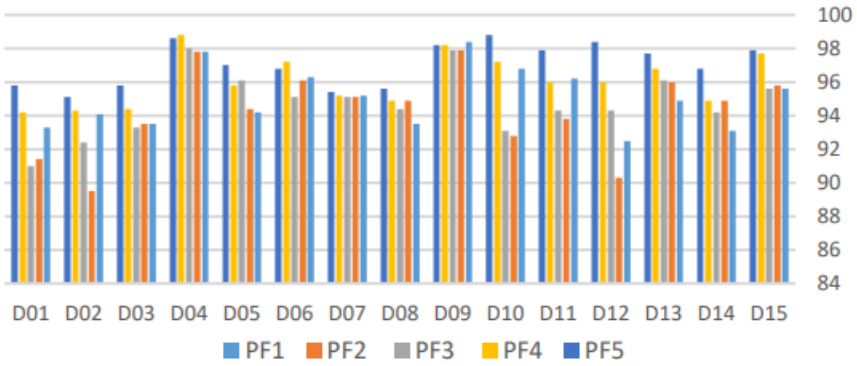


Figure 16: SVM Accuracy for all Preprocessing Filters [23]

The feature engineering step in the approach described above takes a lot of manual effort as well as experimentation to be used efficiently by the classification algorithm. Deep learning is usually used to automate the feature extraction as was seen in the case of malware detection. Deep learning algorithms have been applied for the task of spam detection as well. One such approach used a Convolutional Neural Network model. Extracting features from text using methods like Bag of Words, Word2Vec etc. plays a huge role in the success of CNNs and other deep learning models for text classification. The CNN itself consists of a series of convolution and pooling layers followed by a fully connected layer to perform the classification. Such an architecture allows automation of the feature engineering step. Spam messages can also be embedded within images and an approach to detect these spam texts has been discussed in [25]. The performance of SVMs, MLPs and CNNs on three image spam datasets have been studied. It is observed that CNNs give better performance in two of the three datasets and SVMs and CNNs give comparable results in one dataset.

3.7 Machine Learning approaches towards Data Security: Intrusion and Fraud Detection

A key variable that makes an ordinary infrastructure smart would be the strategic employment of the data involved in the systems. Therefore, the literature recognized that the data layer is a commonly shared critical element of smart city infrastructures [12]. Most of the critical infrastructures in smart cities generate a large volume of data. E-commerce, smart

building, smart healthcare systems, and smart transportation systems are a few examples where we see such a bulk of data collected every day. Therefore, intuitively a special focus and a budget should be allocated for data security in smart cities to protect the infrastructures.

Nowadays, Machine Learning (ML) applications show promising results in many application areas where there exists an abundant amount of data. Therefore, ML approaches come naturally into the picture when discussing automatic data security in smart cities. A recent survey provides a comprehensive overview of how ML techniques come into action in protecting infrastructures in smart cities [12]. By meticulously analyzing the above survey we depicted Figure 17 as the general ML pipeline for data security in smart cities. At a high level we recognized three components; 1) Threat modeling, 2) Machine Learning modeling, and 3) deployment. The main focus of our work is to explore deep into part 2 of the pipeline.

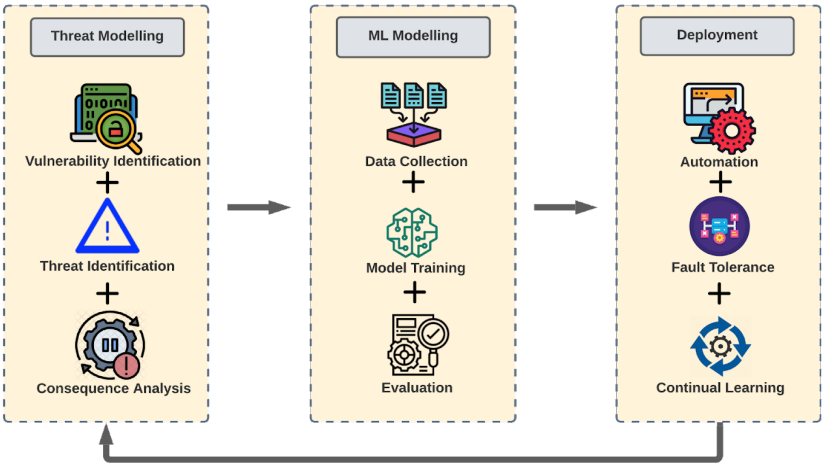


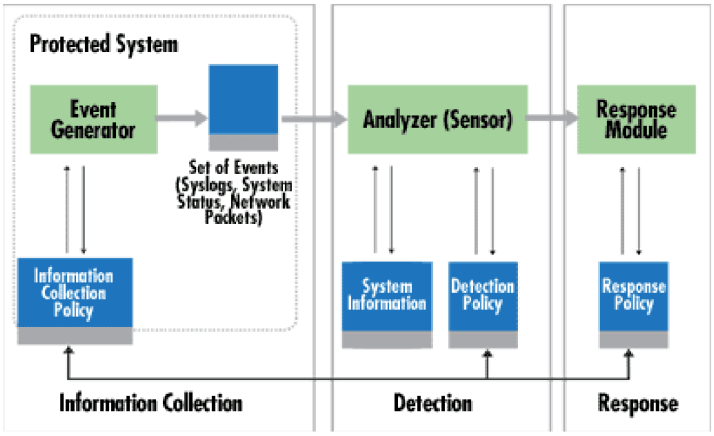
Figure 17: Components in Machine-Learning-Based Applications for Data Security in Smart Cities

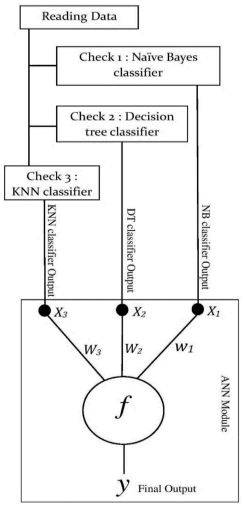
Under this section, we look into two server threat categories for data in smart cities and how ML-based applications have been employed for detecting and mitigating those threats. Namely, these threats are 1) intrusion and 2) fraud. We could identify data intrusion as an attempt to disrupt confidential data in a system in a way that could lead to compromised integrity and availability of the system. On the other hand, we discuss data fraud which can be described as a malicious act of deception in a system that leads to unfair and mistaken gain for a given fraudulent party.

First, we discuss the ML application for data intrusion detection. An intrusion detection

system (IDS) is a quite popular term in cybersecurity due to its prominent usage in Network-based cyberattack mitigation. In section 3.5 we mentioned ML applications for Intrusion detection in smart city networks. However, in this section, we will be looking into intrusion detection at a system level rather than a network level. Basically the discussion would be how to utilize system-level data in ML-based approaches to identify intrusions.

An example of a vulnerable smart city data system would be the E-commerce infrastructures such as banking systems or credit-card systems. [28] did a comprehensive study and proposed an ML-based approach to analyze the customer transaction data in a banking system to detect intrusions. Figure 18 (a) shows how intrusion detection is deployed at a system level.





(a) Proposed IDS Framework

(b) Proposed ML-Based Ensemble for Detection

Figure 18: ML-based Banking Intrusion Detection [28]

They followed an anomaly-based intrusion detection mechanism where the information for the analyzer component would be a yearly snapshot of the user transactions and this is used to identify whether there's an intrusion at the moment of the snap taken. They proposed an ensemble architecture that consists of a multi-level analysis as shown in figure 18 (b); Naive Bayes, Decision Tree, and K- Nearest Neighbor algorithms are used to get the anomaly scores

for the yearly snapshot of deposit and withdrawal transactions. Then these anomalous scores are fused using an artificial neural network to get the final detection score.

They conducted an ablation study to analyze the performance (accuracy) of each component, these results (accuracy) are summarized in the table 4. The authors have presented the results for both the anomalous and non-anonymous classes and we can see that the final fusion method has the best overall performance across both classes.

Class	**Method**			
	Naive Bayes	**Decision Tree**	**KNN**	**ANN Ensemble**
Normal	0.84	0.89	0.86	0.96
Anomaly	0.90	0.93	0.94	0.91
Combine	0.87	0.91	0.91	0.93

Table 4: Intrusion Detection in Banking System: Comparison of Different Machine Learning Applications [28]

Another vital part of smart city infrastructure that could easily be vulnerable to data intrusion attacks is the Cyber-Physical Systems (CPS) [15]. CPS can be defined as a computer-based system that controls a real-world physical system. Therefore, it would be catastrophic if an intrusion occurs at a CPS that contains critical and confidential data and controls a significant part of the smart-city ecosystem. [21] discuss how ML-based methods could be utilized to prevent intrusions in such CPS at a system network level. They considered a use-case of a smart-home system where all the smart devices in the home communicate to an IoT gateway. If an intruder hacks into this gateway, the intruder could either control the home systems or could get private and confidential data of the homeowners. For example: if the home contains a smart assistant such as google-assistant or Alexa the intruder could even access the homeowner's commercial services and manipulate those adversely. [21] followed a supervised machine learning approach to create an anomaly-based smart-home intrusion detection system by analyzing the communication through the IoT gateway. They incorporated bidirectional long short-term memory (LSTM) with a convolutional neural network (CNN) to detect gateway packet anomalies. Figure 19 depicts the complete ML architecture used for this task; it contains 11

layers consisting of BiLSTMs and CNN as the principal processing units to identify anomalous patterns in the IoT gateway data packets.

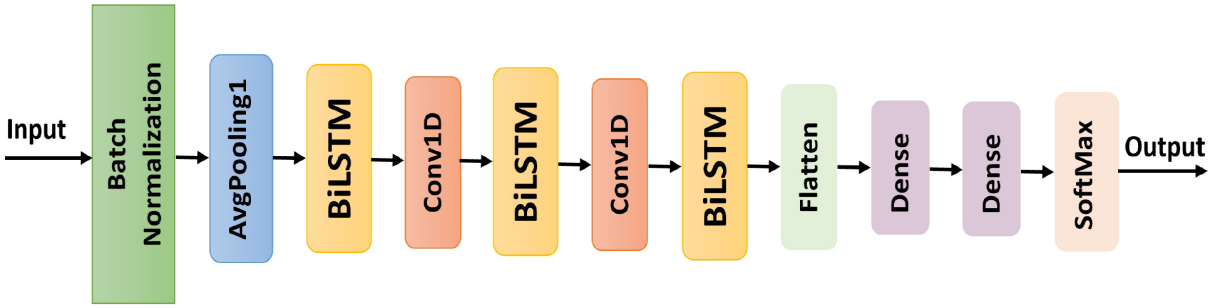


Figure 19: Components in Machine-Learning-Based Applications for Data Security in Smart Cities [21]

Authors evaluated their BiLSTM model in a hypothetical smart-home environment by using an IoT intrusion dataset. This environment consists of two smart Wi-Fi cameras that generate data packets and send those to a gateway. Table 5 shows how BiLSTM performs on the data and table 6 compares the model performance with current state-of-the-art baseline models. It is evident that the BiLSTM +CNN outperforms all the other baselines.

Metrics	Testing Result
Accuracy	98.93%
Precision	98.20%
Recall	99.61%
F1-Score	98.90%
# train parameters	42,180
# all parameters	42,182

Table 5: Intrusion Detection in Smart Houses: Performance of BiLSTM + CNN model [21]

Methodology	Accuracy
K-Means	96.3%
ANN	98.27%
LSTM	97.94%
RF-ET	98.01%
BiLSTM-RNN	98.48%
BiLSTM-CNN	**98.93%**

Table 6: Intrusion Detection in Smart Houses: Comparison of Different Machine Learning Approaches [21]

Similar to data intrusion detection, we could see ML models have been used for fraud detection in smart city applications. Fraudulent activities are quite prevalent in online e-Commerce systems. With the notion of smart cities, more and more services have become virtual and online e-Commerce systems. Consequently, users tend to utilize credit-card more frequently for these online transactions. In other words, credit-cards-based transactions start to dominate online E-commerce platforms. Therefore, it is inevitable for credit card fraud to take place in smart cities. [8] conducted a comprehensive analysis on ML-based credit-card fraud

detection in smart city applications. The authors have identified that the task of classifying a given fraudulent credit-card transaction would be the key to a proper credit card fraud detection model. Moreover, as shown in table 7 they summarize the current state-of-the-art models that have been used for effective autonomous fraudulent credit card transaction detection.

Technique	Advantage	Disadvantage
Decision tree	Powerful detection technique	Tree structure prone to sampling
Genetic algorithm	Supports multi-objective	May require a lot of time
Neural networks	Can handle substantial amounts of data	No clear rules to set up parameters
Bayesian networks	Suitable for small and incomplete data sets	Collecting and structuring expert knowledge
Hidden Markov model	Strong statistical foundation	Have large numbers of unstructured parameters
Parenclitic networks reconstruction	Time efficient	False positive alarms can occur
Meta-learning	High accuracy	Low speed of detection

Table 7: Summary of Existing Credit-Card Fraud detection Models [8]

The main contribution of their paper is to propose a novel approach that overcomes the weakness in the current state-of-the-art model. Especially false positive alarms, i.e. predicting a correct credit-card transaction as a fraudulent one. The authors proposed a new way of optimizing learned parameters of a given ML model known as Multi-Verse Features Extraction (MVFEX). They used a Support Vector Machine (SVM) as their base ML model for classification. Here the proposed multi-verse optimizer is a meta-heuristic evolutionary algorithm that mimics the laws of the multiverse theory. The resilience of this technique comes from the capacity of searching agents to interact to obtain the optimal answer (in the case of fraudulent credit card activity, the best set of features that identify credit-card fraudulent sources and the parameter values of the SVM model used in classification).

The overall detection system is built in multiple stages. For fraudulent transaction pattern detection in credit cards, the suggested multi-verse optimization algorithm is employed for feature extraction and SVM parameter optimization. Then a fitness function is defined in the second step, to identify the fraudulent source. The full framework is shown in figure 20.

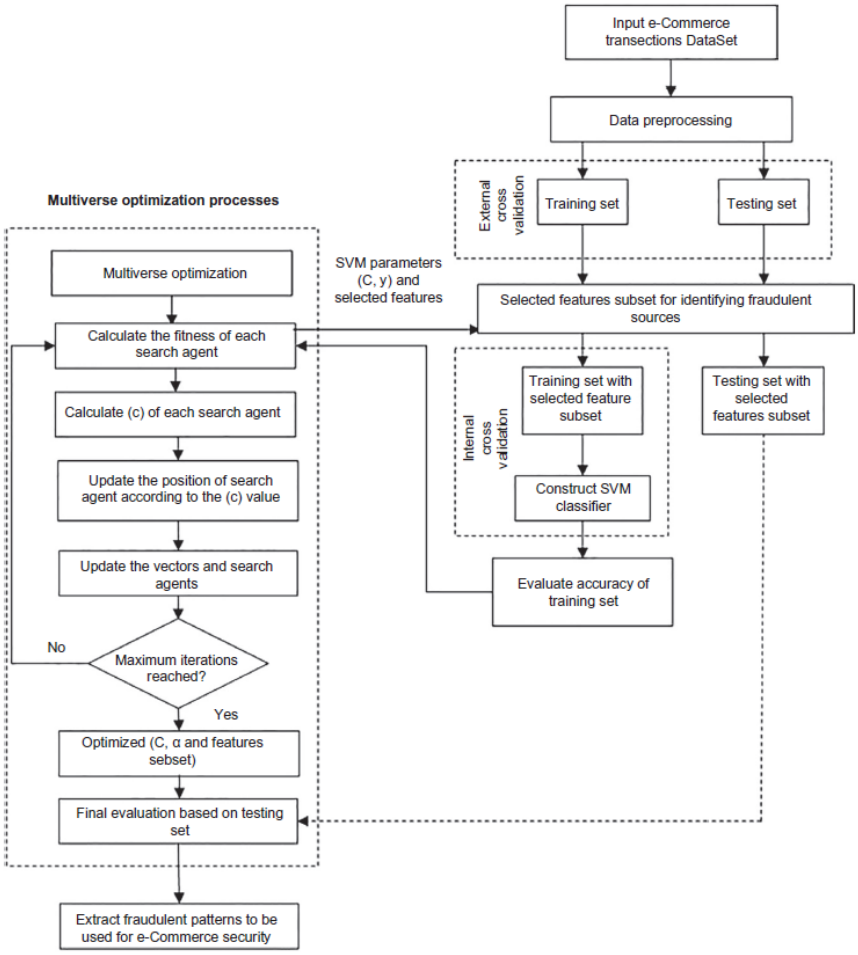


Figure 20: MVFEX + SVM-based Credit-Card Fraud Detection [8]

Authors extracted e-Commerce data from online merchants in Germany, Taiwan, Australia, and Japan. They evaluated their proposed method with the current state-of-the-art credit-card fraud detection model as shown in table 8**.** It is evident that the proposed MVFEX augments the SVM performance in almost all the datasets.

Dataset	CCFD-NB Accuracy	CCFD-MLP Accuracy	CCFD-K-NN Accuracy	CCFD-RbfNetwork Accuracy	CCFD-SVM+Grid search Accuracy	CCFD-MVFEX Accuracy
Germany	72.70	66.00	69.50	73.20	71.47	**74.00**
Australia	77.54	83.77	82.03	80.29	83.33	**85.65**
Taiwan	67.43	80.91	76.92	79.10	**81.80**	79.11
Japan	77.68	83.04	81.01	80.00	85	**86.09**

Table 8: Credit Card Fraud detection in e-Commerce: Comparison of Different Machine Learning Applications [8]

4 Conclusion and Recommendations

4.1 Jacob's conclusion

Just like the world has started to use smart devices and computers that are mostly connected to the internet at all times, it is essential that state-of-the-art technologies must be used to protect them. While enterprise organizations are primarily concerned with data breaches that could adversely affect their operations and reputation, those of critical infrastructures are concerned mostly with vulnerabilities. A vulnerability that is not addressed in Critical Infrastructure Networks can cause expensive, deadly, and long-term repercussions that not only affect the particular organization but the whole country.

Methods such as Multi-Factor Authentication using hardware security keys, one-time passwords (OTP), and stronger passwords can to some extent prevent data from being accessed by unauthorized entities. However, once access is gained the ability to restrict the damage that can be made is very limited. Moreover, if not evaluated closely it might never be known that these systems' integrity has been compromised. The hashing property of blockchains proves to be very useful since even minor changes can cause the hash to be different. Blockchains maintain data integrity by rejecting hashes that are different and can thereby ensure integrity of data stored in them. In a way, blockchains can be said to be protecting the system, even in the case of an intrusion. With a considerable number of users and devices now using open source applications, integrity of data is given more priority in these cases, something which blockchain can facilitate.

At the present it cannot be asserted thoroughly that blockchain technology is at its apogee. Compared to the developments that have happened in computer technology, blockchains could be said to be in a nascent stage. However, research has been promising. The Light Scalable Blockchain (LSB) is an evident proof. Furthermore, as computing hardware and computing power are getting less expensive and more powerful day by day, blockchains are only going to be more widely used Presently an increasing number of organizations have started using blockchains. This includes Critical Infrastructure Networks of many countries. The United States Department of Defense (DoD) has emphasized the benefits of using blockchains in different applications that can benefit the US. Hence, just like encryptions are now used in almost all network communications, blockchains could also become frequent in computer networks.

4.2 Tristan's conclusion

Blockchain is used to store, read, and verify transactions in a distributed database system and has been increasingly popular as a measure of security. It is an effective and applicable method to increase security in smart cities' infrastructure, and IoT environments. The IoT environment is rapidly expanding and the push for an increase in security measures continues to become more prevalent. The IoT network across smart cities is enormous, however, each device within the network is small and limited by its hardware leading to some risks and vulnerabilities. As a result, blockchain has become one of the primary possible solutions for solving some of these vulnerabilities to create a more secure network.

Although blockchain is an effective technology for ensuring security in IoT, its application in the IoT context presents several significant challenges including complexity, overhead costs, and scalability. In order to address these challenges, researchers have proposed a blockchain-based architecture for IoT that is efficient and lightweight, but still offers decentralized security and privacy.

Lightweight Scalable Blockchain (LSB) is optimized for low-resource devices, such as those in the IoT environment. LSB uses simpler algorithms that eliminate the need for solving any hash puzzle prior to appending a block to the blockchain, ultimately using less hardware for computations. In addition, LSB uses a distributed trust method to help with validating new blocks in an efficient manner by utilizing overlay block managers. Both features are beneficial to the IoT environment due to its lack of computational resources and hardware. In-depth studies and security analyses have also found that these applications are secure against a broad range of attacks, ranging from denial of service (DoS) attacks to blockchain modification. This is essential and continuous security testing will be required if a LSB were to be implemented.

Lightweight Scalable Blockchain has opened the door for blockchain technology to be applied in computationally limited fields across smart cities. It is an effective and efficient way to implement security measures that does not require a large number of resources that a standard blockchain application normally would. The IoT network in the infrastructure of smart cities is an ideal situation in which LSB can be applied and effectively used to increase security.

4.3 Nikhil's conclusion

To conclude, this proposal will help in making an accurate recognition of blockchain usage for healthcare safety. Blockchain technology has the potential to transform multiple health sectors. In the healthcare field, electronic contracts enabled by smart agreements are among Blockchain's greatest significant uses. Smart agreements will save expenses by eliminating

brokers from the manipulation process. Utilizing device monitoring, clinics can track their operations using only a Blockchain architecture, over the complete lifecycle. Blockchain technology has the potential to promote healthcare history administration, particularly surveillance and also the healthcare arbitration procedure, allowing medical activities to be completed faster using enhanced information preservation. Overall, this innovation would strengthen and potentially transform how patients and healthcare providers manage while using medical records, as well as better health services.

Transfers can be verified and documented utilizing Blockchain within the next few years, with the agreement of network participants. Even as the basis of a younger breed of healthcare information exchange, blockchain can guarantee quantitative safety towards the clinical level through secret keys cryptography. The technology aims to handle patient information, minimize violation, increase connectivity, rationalize operations, and regulate medicines and prescriptions, and treatment and cure and supply lines.

The scope for this study will be useful and profitable in recognizing the numerous benefits that blockchain technology may provide to various clinical sectors in terms of preserving, transmitting, and securing patient data. The need for innovation in the healthcare area is growing at a breakneck pace. There is a growing demand for high-quality health-care facilities that are supported by cutting-edge technology. In this situation, blockchain has the potential to revolutionize the healthcare industry. In addition, the health-care environment is evolving toward a patient-centered strategy that emphasizes two essential aspects: always-available solutions and effective health supplies. Medical institutions' ability to give acceptable clinical outcomes and rising medical centers is improved by the Blockchain.
A time-consuming & repeated activity that contributes to increased health-care expenses can be resolved rapidly with this innovation. Citizens can participate in healthcare research programmes employing Blockchain. Additionally, improved data and research sharing in public well-being would improve therapy for so many populations.

4.4 Sasanka’s conclusion

It can be concluded that the essential security and resilience are not updated in the power grid to prevent cyberattacks on DERs, devices of grid edge, and infrastructure. The challenges like cyber vulnerabilities and interoperability can also increase behind the meter for automation of buildings and control systems. The application of Blockchain might help in increasing the fidelity and building to grid communication security. Numerous consumers can use Blockchain to verify the data signature to create a distributed trust mechanism. Several

optimizations and reliability related issues can be resolved by Blockchain that assists in grid modernization. Still, the solution for energy value chain related costs is not covered by Blockchain.

Customers generate sales and purchases which are executed efficiently by blockchain technology. In DERs power flows in real-time, visibility is not managed efficiently. At the substation level, the data of different network and residual energy recording can be optimized efficiently with the application of blockchain. The data fidelity and control increment help settle the bulk systems and negotiate future contracts. Blockchain technology can be improved for modernizing and securing the grid. It is vital for simulating real-world applications to improve speed and scale.

It can be recommended that blockchain applications can help in increasing the modern grid speed, scale, and control. Also, cost consumption and security is being maintained. But thorough research is required for controlling power flows in real-time. Further, for increasing the data integrity and security new technology is required or needs more work on the blockchain which helps in improvised techniques. Also, due to the changes in the technology the chance of cyberattacks are also increased which can be controlled with new and advanced software. It helps in maintaining the system efficiently and helps individuals to improve speed and scale.

4.5 Derek's conclusion

Machine learning has a lot of applications in a lot of different fields but one place where it truly shines is within network security. Machine learning techniques drastically improve the success of intrusion detection by allowing for a much simpler implementation of anomaly-based detection engines. Recently, many developments have been made in researching these anomaly-based detection engines by experimenting with different data representation and interpretation techniques. One notable example was performed by A. Elsaeidy, et al. where a Restricted Boltzmann Model (RBM) was used to organize data in a meaningful way for a Feed-Forward Neural Network (FFNN) to learn from and decide what classifies as normal traffic and also what type of signals can be classified as different types of attacks. Another significant use case of machine learning in network management is from the research of M. Furdek, et al. where various learning techniques were applied to detect, identify, and locate attacks or network deficiencies.

When creating a very effective intrusion detection system, following in the footsteps of A. Elsaeidy, et al. is a very good idea since they found that applying one RBM to a neural network nearly tripled the accuracy of the detection system. This might be applicable in several other

scenarios which could, in turn, drastically improve the performance of other machine learning applications. In addition to just improving the detection accuracy, multiple categories can be defined preemptively which can ease the stress on the team in charge of repairing any damage dealt by some attack.

Moreover, since machine learning can be used to create an autonomous network management system, the world will likely be moving towards solutions such as this. This is because as the world ages, the internet becomes more popular and more widespread. Soon society will reach a point where humans can no longer reasonably monitor and maintain networks under a budget. With the help of machine learning techniques, issues such as broken wires, malicious devices entering a network, devices powering down, or anything else, can all be detected, identified, located, and resolved.

When it comes to applying machine learning to network engineering and network security, there is almost no reason not to. Of course maintaining a standard signature-based intrusion detection system is fine and can be very effective alongside an anomaly-based detection system to create a demilitarized zone (DMZ) within the network which only serves to make the network more secure. Since these methods are still experimental, applying them in any network within a DMZ would likely be the best way to go since the old security would be contained within the new security.

4.6 Bishnupriya's conclusion

Malware and spam-based attacks have become serious threats to data security within smart city infrastructures. Classifying malware and spam quickly and efficiently has become critical for any robust security framework in place, especially considering the growth in malware variants and spam volume.

Traditional as well as several machine learning-based approaches for malware and spam detection have been studied. With traditional approaches, the main goal in malware and spam detection is to create a reliable "fingerprint" of a malware application or spam message so that any new application or message can be checked against it. Such approaches usually fail in the face of more sophisticated attacks. Several machine learning approaches have been discussed and it has been seen that using machine learning approaches drastically increases the accuracy and performance of malware and spam detection. Also, by using machine learning approaches, the solutions do not need to be limited by dataset size and do not need to depend on manual feature engineering.

Using deep learning methods, relevant features can be automatically extracted and used

to train any classification algorithm. Hence, the need to hard-code the features is eliminated. Also, since features are no longer static, the same methods or machine learning pipeline can be used for any malware and spam dataset. This is especially useful as malwares and spam can be replicated and differentiated from existing ones by making a change in their structure.

That being said, several challenges still exist before any such malware or spam detection framework can be put in place within a smart city infrastructure. The machine learning model once deployed needs to be updated to reflect the changes in the training data so that the performance of the system can be maintained or improved. Newer versions of malware which are significantly different from previous versions come into the picture regularly and can cause serious changes in data distribution used while training the model, causing detection rate to degrade over time. The detection framework should be robust enough to handle such cases, allowing the model to be updated or retrained on the fly. This usually comes under the purview of MLOps.

Machine Learning is a great way to develop robust security frameworks against an ever-changing threat landscape. Malware and spam detection frameworks based on machine learning approaches can help detect these faster and more accurately.

4.7 Tharindu's conclusion

It is evident that the data layer is a vital organ in any given smart city infrastructure. Threats such as Intrusions, fraud on the data layer could have catastrophic and cascading effects and compromise the integrity, availability of the smart-city ecosystem. Machine Learning applications show promising results in many application areas where there exists an abundant amount of data, thus ML-based approaches come naturally into the picture when discussing automatic data security in smart cities.

In this report, we discussed various ML-based approaches that are quite effective in detecting these threats against the data layer in smart infrastructures. In particular, we presented research work done in e-Commerce, smart building infrastructures as case studies to prove the significance of incorporating ML approaches to protect the data layer of various smart city infrastructures. We discussed in detail two main threats faced by digital infrastructures in smart cities; data intrusion and fraud. At the moment most of the data intrusion and fraud detection approaches seen in the literature are based on supervised learning. There exist svm classifier approaches and also ensemble approaches that aggregate the power of multiple ML models such as random forest, artificial neural networks. These supervised learning approaches are quite intuitive given that we could find an extensive amount of data for the threat cases in

these smart infrastructures. Therefore, high performance would be achieved by incorporating supervised learning with labeled datasets. However, it would require more real-time and end-to-end frameworks when more complex smart infrastructures get realized in smart cities.

Therefore, going forward we see there are a couple of challenges for ML applications for data security in smart cities. 1) Newly emerging threats with new smart infrastructures with fewer to no historical data and 2) updating existing ML models with new data from existing infrastructures. To address these challenges, first, we recommend and envision the utilization of low resource machine learning to tackle the problem of newly emerging threats. Second, we envision employing continual learning approaches in the future to make the current state-of-the-art models up to date.

5 References

[1] A. Alsaafin, I. Ahmed Qasse, M. Abu Talib and Q. Nasir, "Lightweight Blockchain-Based System for Internet of Things Security," 2020 SoutheastCon, 2020, pp. 1-8, doi: 10.1109/SoutheastCon44009.2020.9368272.

[2] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, “LSB: A lightweight scalable blockchain for IOT security and anonymity,” Journal of Parallel and Distributed Computing, vol. 134, pp. 180–197, 2019

[3] A. Elsaeidy, K. S. Munasinghe, D. Sharma and A. Jamalipour, "A Machine Learning Approach for Intrusion Detection in Smart Cities," 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), 2019, pp. 1-5, doi: 10.1109/VTCFall.2019.8891281.

[4] A. Farouk, A. Alahmadi, S. Ghose, en A. Mashatan, “Blockchain platform for industrial healthcare: Vision and future opportunities”, *Computer Communications*, vol 154, bll 223–235, 2020.

[5] A. Hayes, “Blockchain explained,” Investopedia, 05-Mar-2022. [Online]. Available: https://www.investopedia.com/terms/b/blockchain.asp.

[6] A. Haleem, M. Javaid, R. P. Singh, R. Suman, en S. Rab, “Blockchain technology applications in healthcare: An overview”, *International Journal of Intelligent Networks*, vol 2, bll 130–139, 2021.

[7] A. Kapratwar, F. Di Troia, en M. Stamp, “Static and Dynamic Analysis of Android Malware”, 01 2017, bll 653–662.

[8] A.S. Sadiq, H. Faris, A.Z. Ala'M, S. Mirjalili and K.Z, Ghafoor, “Fraud detection model based on multi-verse features extraction approach for smart city applications,” In Smart cities cybersecurity and privacy, 2019, pp. 241-251.

[9] B. Yuan, J. Wang, D. Liu, W. Guo, P. Wu, en X. Bao, “Byte-level malware classification based on markov images and deep learning”, Computers & Security, vol 92, bl 101740, 2020.

[10] C. Frakes, “Why are wire transfers expensive?: Modern Treasury Journal,” RSS, 03-Jun-2021. [Online]. Available: https://www.moderntreasury.com/journal/why-wire-transfers-are-expensive#toc-what-are-wire-transfers-.

[11] C. Martin, “What is blockchain email?,” SparkPost, 08-Jul-2019. [Online]. Available: https://www.sparkpost.com/blog/what-is-blockchain-email/.

[12] D. Chen, P. Wawrzynski and Z. Lv, “Cyber security in smart cities: a review of deep

learning-based applications and case studies," Sustainable Cities and Society, 2021, 66, p.102655.

[13] D. Enwood, "How blockchain is revolutionising food supply chains," Blockhead Technologies, 05-Mar-2021. [Online]. Available: https://blockheadtechnologies.com/how-blockchain-is-revolutionising-food-supply-chains/ .

[14] G. D'Angelo, M. Ficco, en F. Palmieri, "Malware detection in mobile environments based on Autoencoders and API-images", Journal of Parallel and Distributed Computing, vol 137, 11 2019.

[15] H. Habibzadeh, B. H. Nussbaum, F. Anjomshoa, B. Kantarci, en T. Soyata, "A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities", Sustainable Cities and Society, vol 50, bl 101660, 2019.

[16] H. Liu, R. G. Crespo, en O. S. Martínez, "Enhancing Privacy and Data Security across Healthcare Applications Using Blockchain and Distributed Ledger Concepts", *Healthcare*, vol 8, no 3, 2020

[17] J. Hathaliya, P. Sharma, S. Tanwar and R. Gupta, "Blockchain-Based Remote Patient Monitoring in Healthcare 4.0," 2019 IEEE 9th International Conference on Advanced Computing (IACC), 2019, pp. 87-91, doi: 10.1109/IACC48062.2019.8971593.

[18] M. Antwi, A. Adnane, F. Ahmad, R. Hussain, M. Habib ur Rehman, en C. A. Kerrache, "The case of HyperLedger Fabric as a blockchain solution for healthcare applications", *Blockchain: Research and Applications*, vol 2, no 1, bl 100012, 2021.

[19] M. Furdek, C. Natalino, F. Lipp, D. Hock, A. D. Giglio and M. Schiano, "Machine Learning for Optical Network Security Monitoring: A Practical Perspective," in Journal of Lightwave Technology, vol. 38, no. 11, pp. 2860-2871, 1 June, 2020, doi: 10.1109/JLT.2020.2987032.

[20] M. Mylrea and S. N. G. Gourisetti, "Blockchain: A path to grid modernization and cyber resiliency," 2017 North American Power Symposium (NAPS), 2017, pp. 1-5, doi: 10.1109/NAPS.2017.8107313.

[21] N. Elsayed, Z. S. Zaghloul, S. W. Azumah and C. Li, "Intrusion Detection System in Smart Home Network Using Bidirectional LSTM and Convolutional Neural Networks Hybrid Model," 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2021, pp. 55-58, doi: 10.1109/MWSCAS47672.2021.9531683.

[22] R. Krishnamurthy and A. Kumar, "Off-chain storage for block chain," *SNIA*, 22-Sep-2020. [Online]. Available:

https://www.snia.org/educational-library/chain-storage-block-chain-2020.

[23] S. Bosaeed, I. Katib and R. Mehmood, "A Fog-Augmented Machine Learning based SMS Spam Detection and Classification System," 2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC), 2020, pp. 325-330, doi: 10.1109/FMEC49853.2020.9144833.

[24] S. Figueroa, J. Añorga, en S. Arrizabalaga, “An Attribute-Based Access Control Model in RFID Systems Based on Blockchain Decentralized Applications for Healthcare Environments”, *Computers*, vol 8, no 3, 2019.

[25] T. Sharmin, F. D. Troia, K. Potika, en M. Stamp, “Convolutional neural networks for image spam detection”, Information Security Journal: A Global Perspective, vol 29, no 3, bll 103–117, 2020.

[26] Morgan, Steve. “Cybercrime to Cost the World $10.5 Trillion Annually by 2025.” *Cybercrime Magazine*, 27 Apr. 2021, https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025/.

[27] T.R. Vance and A. Vance, "Cybersecurity in the Blockchain Era : A Survey on Examining Critical Infrastructure Protection with Blockchain-Based Technology," 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), 2019, pp. 107-112, doi: 10.1109/PICST47496.2019.9061242.

[28] W.S. Mahdi, and A.T. Maolood, “Banking Intrusion Detection Systems based on customers behavior using Machine Learning algorithms: Comprehensive study,” *Journal of Al-Qadisiyah for computer science and mathematics*, 2020, 12(4), pp.Page-1.

[29] V.A. Gasimov and S. K. Aliyeva, "Using blockchain technology to ensure security in the cloud and IoT environment," 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021, pp. 1-5, doi: 10.1109/HORA52670.2021.9461397.

[30] V.K. Chattu, A. Nanda, S. K. Chattu, S. M. Kadri, en A. W. Knight, “The Emerging Role of Blockchain Technology Applications in Routine Disease Surveillance Systems to Strengthen Global Health Security”, *Big Data and Cognitive Computing*, vol 3, no 2, 2019.

[31] Z. Dong, F. Luo and G. Liang, "Blockchain: a secure, decentralized, trusted cyber infrastructure solution for future energy systems," in Journal of Modern Power Systems and Clean Energy, vol. 6, no. 5, pp. 958-967, September 2018, doi: 10.1007/s40565-018-0418-0.

[32] “What is Data Security? data security definition and overview,” IBM. [Online]. Available: https://www.ibm.com/topics/data-security. [Accessed: 30-Mar-2022].