Survey on the Application of Machine Learning and Blockchains in IoT security

Haneen Alfauri (A paper written under the guidance of Prof. Raj Jain)

abstract

Internet of things (IoT) is the revolutionary approach that is built on the promise of ubiquitous connectivity that will enable billions of devices to connect and exchange data with minimal human intervention. For IoT technology to reach its true potential, security issues and challenges need to be addressed. This requires embedding other technologies (i.e Machine Learning and Blockchains) to support its security. This paper aims at providing a comprehensive list of the most popular Machine Learning methods and blockchains mechanisms and their notable applications for IoT security.

Table of Contents:

Keywords

Internet of Things (IoT), Blockchain for IoT security, IoT Security, IoT intrusion detection systems, Machine Learning (ML) for IoT security, IoT Security Challenges, Deep Learning (DL) for IoT security.

1. Introduction

The estimates for the number of IoT devices are projected to amount to 30.9 billion units by 2025 provided by Statica, to the breath-taking 200 billion projection provided by Intel. However, the security of the data is an issue, which is why the researchers recommended the integration of the presented applications of Machine Learning (ML) and blockchain technology in the support of the IoT systems security.
Typically, IoT devices have limited computing capabilities, small storage, small network capacity, which makes IoT devices hackable and vulnerable. Blockchains can solve problems of the current centralized IoT structure by applying distributed characteristics of blockchain to IoT networks. Whereas ML models and deep learning applications can be utilized as Intrusion detection systems in addition to other applications we discuss in detail later.
A lot of work is published in this area, mainly the surveys that have tried to capture the state and challenges of IoT Security. The authors in [Perrone2017] focused on providing an overview of the security of the IoT systems and devices security. Other surveys, by another authors [Weber2016] and [Roman2013], focused their analyses on specific IoT aspects, such as regulation and legal challenges of the IoT security. This survey is a comprehensive analysis of the applications and trends Machine Learning (ML) and Blockchains play in the area of IoT security.
There are mainly three aspects that are important in securing the end-to-end IoT system. here we should consider the data life cycle, starting from the point where data is created by an entity or end device, and along the path where it travels through many intermediate devices, and subsystems, and finaly ends up somewhere at rest. The trust factor is needed to be spread across the entire flow of that data.
The second part is that the network itself and the devices are of various types, so it's hard for one fixed standard to cover all the variety of the devices under the trust umbrella. The third challenge is that trust is of a multiple kinds, a trust that is related to security, and a trust that is around privacy and policy engine, around how and with whom and when to share that data.
There exists various IoT lightweight application communication protocols such as MQ Telemetry Transport MQTT and Constrained Application Protocol, or Extensible Messaging and Presence Protocol XMPP, or even protocols related to routing as those of RPL and 6LoWPAN. These protocols are designed with a small fingerprint, and by that we mean they require a small bandwidth, small computation, and memory requirements and exchange a small size control messages thus they are not secure by design.
In most frameworks, these protocols are encapsulated within other secure protocols such as Transport Layer Security protocol for messaging, IPSEC or TCP-like protocols where there is a need for a negotiation phase where parties negotiate the encryption keys and other parameters for encryption and hashing to establish the master key sessions. So, we need to handle and exchange

IKE in case of IPSec or PKI certificates at the handshake in case of TLS.
Previous work proposed a centralized public key infrastructure for IoT security, yet by utilizing blockchains, the centralized key management and key distribution are completely eliminated. In blockchains, each IoT device would have his own unique asymmetric key pair, and a globally unique identifier (GUID) given by the blockchain.
These security issues and challenges need to be addressed at multiple levels. That requires embedding other technologies such as machine learning and blockchains to support its security. We start our survey by exploring the background of IoT systems security challenges and provide an overview of the IoT systems attack surface. In Section 3 we discuss the application of Machine Learning and Deep Learning in the context of securing IoT systems and their role in anomaly intrusion detection for the IoT systems. In Section 4 we layout the foundation of applying blockchains for IoT data integrity and access control. Section 5 concludes this work and proposes future research directions.
To the best of our knowledge, this is the first comprehensive survey combining Machine Learning and blockchains applications in the field of IoT security.

---

2. IoT Security challenges

IoT implementation has presented unique challenges that need to be addressed, particularly as this technology continues becoming incorporated in every aspect of our lives.
In this section we discuss the IoT system security challenges. We later classify the IoT attack surface into three categorizations and we discuss each in a subsection.

2.1 Security challenges

The diversity of IoT devices makes them more vulnerable to security challenges in various ways. These challenges ranges challenges range from poor encryption, authentication and physical security to insecure web and mobile interfaces and network services. In this section we discuss the main security issues facing IoT systems.

Many IoT devices is designed in a way that they trust the local network to a level that no further authentication or authorization is required, and that any new device connected to the same network is also trusted. This raise red flags especially when the device is connected to the Internet, that if one component of the system is compromised the whole system gets compromised.

Another common problem with the IoT devices is that all devices of the manufactured by the same entity are sold with the same default password (e.g. Camera123) and other times the password might be given serially, that allows the attacker to retrieve the other passwords very easily. In the IoT systems, there is usually one privilege level with no further access control. This vulnerability is better managed in the IoT cloud where the broker (main server) can revoke users access and control the access level, yet this is a major vulnerability in the case of locally connected IoT system. A typical problem is that most IoT devices transmit their data using a plain-text version of a protocol (i.e., HTTP) although an encrypted version is available (HTTPS).

An attacker or a Man-in-the-Middle can sniff the transmitted information and obtain the credentials and later secretly accesses, or relay communications, possibly altering this communication, without either party being aware. Weaknesses may be present even in the case that the IoT system uses encryption that is if there is a misconfiguration. Most of the IoT devices do not have logging or alerting capabilities to notify the user of any security issues, so typically when a device is compromised, it goes unnoticed from the viewpoint of the user and any additional bandwidth or power usage is usually not detected. The result is that users rarely discover that their device is under attack or has been compromised, preventing them from taking any further mitigating measures.

Many Institutions and groups like the International Society of Automation (ISA) and the National Institute of Standards and Technology (NIST) and have tried to help solve these challenges by issuing IoT cybersecurity standards. Unfortunately, these guidelines lack clear implementation recommendations and very complex and difficult to understand that increases the burden on the IoT devises manufacturers and integrators are left to determine how to align with these guidelines and recommendations, this means that standards are not put into real-world practice because the perception is that they are too complex.

For instance, the technical documentation for the Trusted Computing Group's TPM standards runs more than 4,000 pages, as guidance for embedding a unique secret key into microchips and firmware to help prove the identity of IoT devices, with such complexity putting these guidelines into practice is difficult. It can be seen that up until this moment there is no one-size-fits-all solution for IoT systems security, therefor there is a need to incorporate dynamic and smart algorithms (i.e., ML) at multiple layers. In addition to incorporating other decentralized technologies (i.e., Blockchains) to achieve the highest level of data integrity, and remove the central authority, to achieve the highest possible IoT security. This research study aims to investigate existing and proposed methods to secure IoT systems, to achieve this objective, firstly, it requires an understanding of the attack surface of IoT systems.

2.2 IoT Attack surface

In this section we present an overview of some of the common IoT attacks. Figure 1 below provides a general overview of the IoT attack surface.
We mainly categorize attacks into three categories, the first is network attacks and the second is the physical attacks. The third is the application attacks. We detail each of the categories in a separate subsection below.

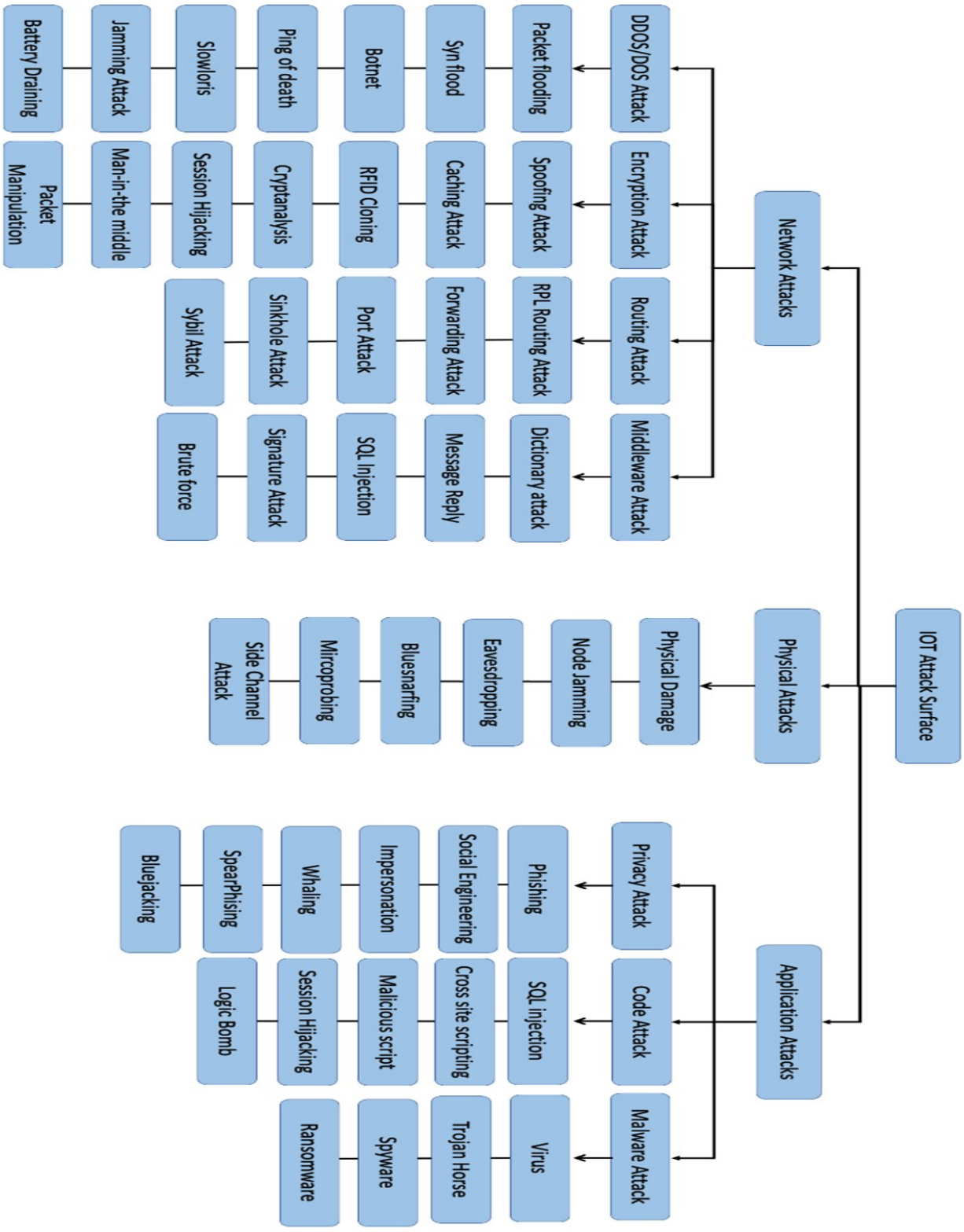

Figure 1: IoT Attack Surface. [Perrone2017]

2.2.1 Network Attacks

In this section we talk about the second type of IoT attacks, these attacks are focused on the network of IoT system. We categorized network attacks mainly into four types. Firstly, the attacks that causes denial of service (DoS) whether it is distributed or not, such as Packets flooding attacks and Sync flood attacks, and Botnet attacks.

Jamming attack is a type of DoS attack, where the attacker transmits radio signals that disrupt the IoT wireless communications by decreasing the Signal-to-Inference-Noise ratio (SINR), researchers proposed solutions to solve this problem, one is the reactive jammer [Alam2020] they used an artificial noise to distract the adversary and accomplishes the transmission task successfully in the presence of an attacker. Slowloris attack is an application layer DDoS attack which uses partial HTTP requests to open connections with a targeted victim's server, and maliciously keeping those connections open for as long as possible, to waste resources and thus overwhelming and slowing down the victim server. Another type of the DoS attack is the Ping of Death (PoD) where an attacker sends a malformed or oversized packets using a simple ping command maliciously to crash, destabilize, and freeze the victims devices.

Here we highlight the importance of utilizing Machine Learning models as an anomaly intrusion detection system (AIDS) for the IoT network. AIDS play a remarkable role by protecting IoT devices from DDOS attacks more details on this is in Section 3. The second type of the network attack is the encryption related attacks such as the Spoofing attacks that gives false information which seems to be correct and that the system accepts, one example of that is the RFID Spoofing attack where the attacker spoofs RFID signals. Then it captures the information which is transmitted from a RFID tag.

Cryptanalysis Attacks is the encryption attack where the adversary obtains the encryption key by using either plaintext or ciphertext. Based on methodology used, there are different types of cryptanalyses attacks more details can be found in [Ahmad2021].
The third type of IoT network attack is the IoT routing attacks, this attack targets the IoT routing protocols such as the routing protocol for low power and lossy networks (RPL) that has become the standard routing protocol (RPL) for the Internet of things.

A 6LoWPAN network is a Wireless Sensor Network (WSN) utilizes another protocol (IEEE 802.15.4) as a data-link and physical layer protocol and a compressed IPv6 protocol for routing and networking. This combination allows the constrained devices in the IoT system to be accessible remotely through the internet. IoT devices are exposed to attacks both from the Internet and from within the network. One example of an attack on the RPL routing protocol is the rank attack [Andrea2015]. In a rank attack or sometimes this type of attack is called the sinkhole attack, a malicious attacking node advertises itself with a fake rank in the RPL control messages and fake route across the root node to deceive its neighbors to forward their packets through it. This type of attack causes performance degradation, low Packet Delivery Ratio (PDR), delaying and generation of non-optimal paths and loops. Another type of attack on RPL is the version number attack (VNA) that increases the overhead control and degrades the network performance and results in a high end-to-end delay. IoT Port attacks are very common, figure 2 below shows the commonly used ports [Almusaylim2020] by IoT device. These attacks were

collected from the following top 19 services shown in table 1 below. Many IoT devices have started using SSH for remote administration instead of telnet. Note that SSH port 22 brute force attacks are the number one attack type, followed by http web traffic on port 80, then telnet on port 23 and SIP port 5060, and then the alternate http port 8080.

Service	Port	IoT Device Type
Telnet	Port 2323 / 23	ALL
Applications	Port 37777	DVRs
Applications	Port 8291	SOHO routers
SIP	Port 5060	ALL VoIP phones, video conferencing
HTTP_Alt	Port 8080	SOHO routers, smart sprinklers, ICS
TR069	Port 7547	gateways, CCTV
HTTP_Alt	Port 8081	DVRs
Secure SIP	Port 5061	VoIP phones, video conferencing
HTTP	Port 81	Such as IoT: Wifi-cameras
SMTP	Port 25	IoT: Wifi-cameras, Game consoles
Rockwell	Port 2222	ICS
UPnP	Port 52869	Wireless chipsets
WSP	Port 9200	WAPs
HTTP_Alt	Port 8090	Webcams
HTTP	Port 80	Includes common IoT devices, ICS and gaming consoles
Rockwell	Port 2223	ICS
SSH	Port 22	*Includes IoT
UPnP	Port 37215	SOHO Routers
Applications	Port 2332	Cellular gateways

Table 1: Top 20 ports used by IoT devices [Ahmad2021]

IoT devices use all of the above ports for different applications. One example is the small office or home office (SOHO) routers and gaming consoles that have been using port 80 for a while. The fourth and last type of IoT network attack is the middleware attack, this includes Structured Query Language (SQL) injection attack that occurs when the attacker impersonates a legit user

who has privilege access, where the attacker inputs a valid request and add in new instructions that also get executed.
The reason why we highlighted this type of attack is because the risks associated with this type are serious; attackers could use it to trick a web application to modify other users' access levels and allow authentication without a valid password. Other types of middleware attacks are brute force attacks, and signature attacks, message reply attacks, and Dictionary attacks.

2.2.2 Physical Attacks

We classify the application attacks into three categories below we detail each type.
The first type is the privacy attack. In the privacy attack the attacker extracts information about the IoT systems structure, IoT device information, this type of attacks causes a private data to leak. Such as social engineering, Spear-phishing and Bluejacking.
Whaling is another type of highly targeted privacy attack where the attacker aims at senior executives, masquerading as a legitimate email, typically the management and human resources teams are frequent targets of whaling attacks because they have access to sensitive and personal data. The main difference between whaling and Spear-phishing attack is that whaling target specific, high ranking victims within a company, whereas Spear-phishing attack can be used to target any individual at any level.
The second type is the Code attack has different types; one common type is the Code injection that is the exploitation of a software bug that is caused by processing invalid data. The attacker injects a malicious code that causes security breaches to drop the IoT system communication or to steal and leak private information.
The third type is the Malware attack. We define the malware attack as a malicious software that executes unauthorized actions on the targeted victim's system. This malicious software is also called a virus that can encompasses many specific types of attacks such as ransomware, spyware, command and control, Code and programing attacks are impossible to avoid completely when developing software. However, we can reduce the possibility of vulnerabilities by applying best practices to avoid application vulnerabilities, such as consistently performing input validation.

2.2.3 Software Attacks

Physical attacks have a low and limited impact on a single device since this type of attack requires physical interaction and requires close proximity to the IoT device. So it is not possible to perform these attacks en-masse from the Internet, therefore we do not recognize this as one of the biggest security problems, but it is nevertheless included.
If hacker have physical access to a device, they can add a piece of hardware for spying or can directly access the SD card installed in the IoT device so that any protecting software can be bypassed.
A physical attack is usually a small scale attack, yet can be impactful in case the attacker uncovers a device key that is shared amongst all devices of the same model, we talked about this common vulnerability in section 1. This allows the attacker to compromise a wide range of similar IoT devices.

2.3 Summary

The main challenge is that IoT devices vary widely in terms of functionality, size, computational power, capacity storage, and energy, therefore there is a need to incorporate other technologies in an effort to provide defense in depth for the IoT systems against various types of the attacks.
In this section, we covered the main IoT systems attack surface, which lays the basis for the coming sections on how the Machine learning models and blockchains can solve these issues.

---

3. Machine learning solutions for IoT security

In this section we discuss the different models for Machine Learning, and their applications in the context of securing IoT systems.
Machine Learning can provide great value and encourage building and training models by learning from traffic patterns and providing a heuristic solution that can effectively implement a network intrusion detection system.
During the last few years, workable solutions based on Machine learning were developed, such a building classifiers that can learn difficult and complicated patterns from the data to detect many types of attack on the IoT systems such as (i.e. DDOS) attacks.
Anomaly Intrusion Detection Systems (AIDS) dynamically monitors the events taking place within the IoT system and decides whether these events are malicious (intrusion) or a legitimate use of the system.
There are four different types of Machine Learning models, that can be used as an anomaly intrusion detection system. Figure 2 below illustrate these four types in addition to their subcategories based on the work in [Khraisat2021].
Numerous related studies [Mothukuri2021] [Chaabouni2019] [Rasheed2021] [Tanzila2021] applied Machine Learning models to IoT AIDS, but we are in need for a comprehensive comparison between these models to be able to relate on which is the most effective for building an efficient IoT AIDS, and which one is more effective through various datasets, all these questions will be answered in this section.

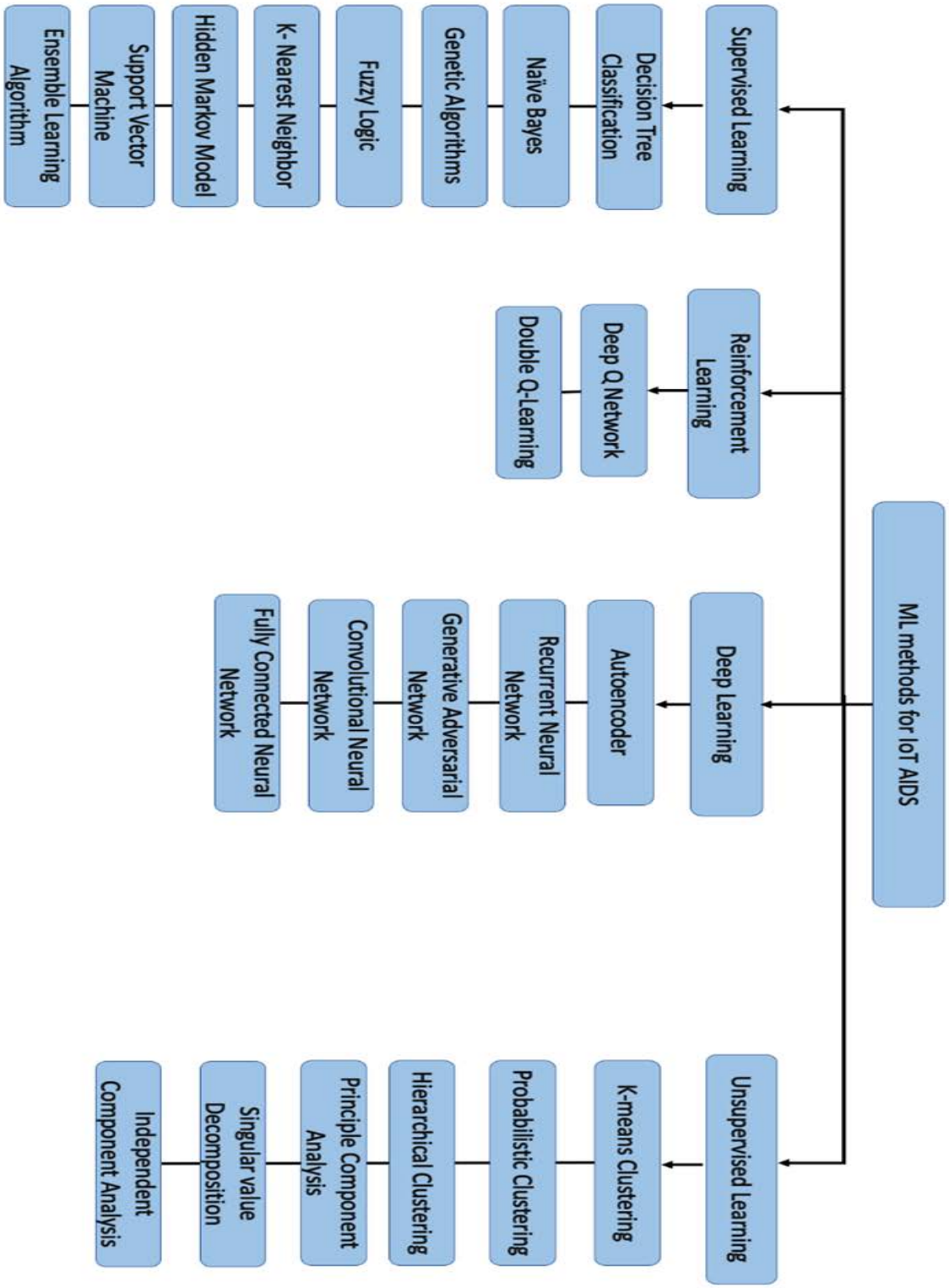

Figure 2: Classification of ML AIDS Methods for IoT Security.

In general for machine learning models [Susilo2020] we need to make sure that the training process is correct and try to avoid memorization and generalization, as memorizing the training dataset is called over-fitting that leads to poor performance as it could memorize relations and structures that are noise or coincidence and may degrade the overall models prediction accuracy. This problem is common to many machines learning algorithms. ML methods are primarily classified into the following four categories [Khraisat2021], we detail each in a subsection below.

3.1 Deep learning

Deep learning model (DL) is a branch of ML that uses many hidden layers to form a neural network (NN). One its main strength over ML is its ability to process massive data. It continues to perform higher than ML algorithms as data size continues to grow. For that its best to implement DL as an IoT intrusion detection system, hence large data volume generated by these IoT devices.
There are different types of DL models, some are famous for sequential data like recurrent neural networks (RNN), and some are specialized in image processing like convolutional neural network (CNN), and some is used in speech recognition.
Researchers have used these and many other powerful models [Annachhatre2015] individually and in an ensemble setting to find optimal solutions for large scale attacks (i.e., DDOS) in IoT systems.

3.2 Unsupervised/ Supervised Learning

Supervised learning-based AIDS techniques detect the anomaly (intrusions) by using labeled training data. This approach usually consists of two parts, namely, training and testing.

In the training stage, the observations in the training set to form the experience that the algorithm uses to learn. In supervised learning problems, each observation consists of an observed output variable and one or more observed input variables. The task of the learning function is mapping input to an output based on a training input-output set each set is consisting of an input object (called a vector) and the desired output value (called the supervisory signal). There are mainly eight types of supervised learning, first is a Naive Bayesian. A Naive Bayesian Model is used for large finite datasets, assigning class labels using a direct acyclic graph. The graph comprises one parent node and multiple children nodes (Tree). In the tree representation, the leaf nodes correspond to class labels, and the internal nodes represent the attributes. The ensemble learning method is also sometimes called a Random Forest Model, where It operates by constructing a multitude of decision trees that is a chart-like model that contains conditional control statements and outputs a classification of the individual trees.

The hidden Markov model (HMM); is a statistical Markov model that requires that there be an observable process A whose outcomes are influenced by the outcomes of B in a known way. HMM analysis can be applied to identify particular kinds of malware [Denning1987], once the model is trained against known malware features such as the operation code sequence the trained model is applied to evaluate and score the incoming network traffic, with the lower score it means the traffic is identified as normal.

Fuzzy logic; this technique allows a degree of uncertainty, as it permits an instance/object to belong, possibly partially, to multiple classes at the same time. Support Vector Machines (SVM) is a novel and old machine learning approach that is widely deployed in anomaly intrusion detection [Denning1987] [Wang2009]. The main reasons are its good generalization performance, absence of local minimal, and fast execution time.

The k-nearest neighbors (KNN) algorithm; assumes that similar patterns exist in close proximity. In other words, similar things or types are near to each other. It is a simple and easy to implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

Figure 3 below illustrates a K-Nearest Neighbors classifier where k = 7. There are four similar patterns from the class Intrusion and three from the class Normal. Point M represents an instance of unlabeled data that needs to be classified. Amongst the seven nearest neighbors of m. Taking a majority vote enables the assignment of M to the Intrusion class.

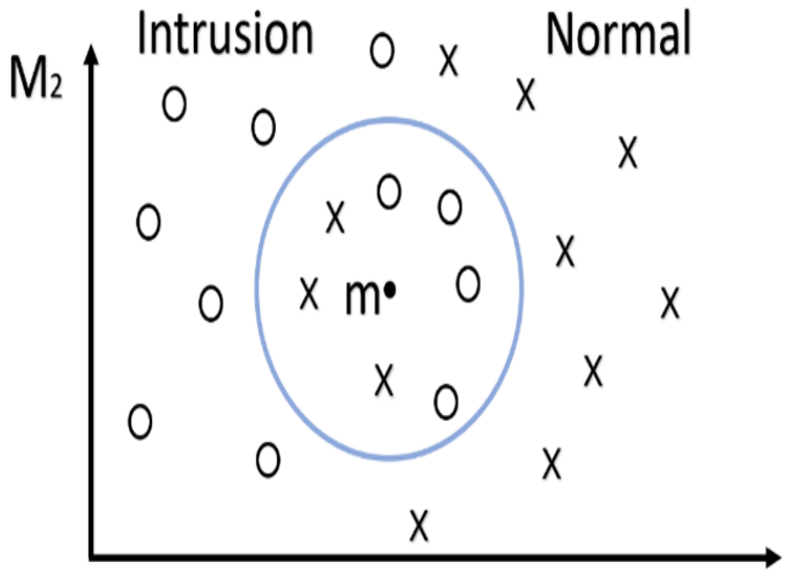

Figure 3: An example of classification by k-Nearest Neighbor for k = 7.

3.3 Reinforcement Learning

Reinforcement learning brings the full power of Machine Learning to anomaly/intrusion detection. It is defined by a 5 parameters consisting of State at time t (St), Action in state s (At), Discount factor for future Rewards (Gamma), State Transition Probabilities P(St+1|St, At) , and Reward function (Rt). The objective is to come up with an optimal policy (I *) that defines what action should be taken at each state so that the agent achieves maximum cumulative rewards over a long period of time. By reinforcement, we mean that we are feeding our rewards function into the network input so that we can learn to associate what actions produce positive results given a specific state of the environment.
Before introducing the two types of reinforcement learning, first, it is important to define what is Q-learning. Q-learning is a type of reinforcement learning algorithm to learn the value of an action in a particular state.
The first type of reinforcement learning Is Double Q-learning It is an off-policy reinforcement learning algorithm that utilizes double estimation to counteract overestimation problems with traditional Q-learning.[Pu2021]
The second type of reinforcement learning is Deep Q learning, where the target network predicts Q values for all actions that can be taken from the next state, and selects the maximum of those Q values. Use the next state as input to predict the Q values for all actions [Hsu2020].
The authors in [Deokar2012] proposed an intrusion detection system by combining methods of log correlation, reinforcement learning, and association rule, the main role of reinforcement learning in this research is that it helps to detect the unknown attack by motivating the rewards activities to identify anomalies.
Another context-adaptive AIDS [Sethi2019] uses multiple independent deep reinforcement learning agents distributed across the network to enhance detection accuracy for the new and complex attacks.

3.4 Unsupervised Learning

The unsupervised machine learning algorithm makes use unlabeled training data, in this case the model distinguishes when it is right or when it has made a mistake so that it can adjust its parameters and training weights. This is a key difference compared to supervised learning that tries to learn a function that will allow us to make predictions given some new unlabeled data, unsupervised learning tries to learn the basic structure of the data to give us more insight into the data. Unsupervised machine learning is used when the information used to train is neither labeled nor classified.
The task of the machine model is to group unsorted information according to patterns, similarities, and differences without any prior training data.
One of the most important assumptions for an unsupervised anomaly intrusion detection algorithm is that the dataset used for the training purpose is assumed to have all non-anomalous training examples (clean and legit) dataset, hence the model will learn the normal baseline from this dataset, thus we must ensure that it contains no anomalies.[Zoppi2020].

3.5 Evaluation metrics

There are mainly three Machine Learning Anomaly Intrusion detection evaluation metrics, each one has its own use cases.[Rasheed2021]
For example in the case of evaluating an anomaly detection algorithm, for the supervised or unsupervised machine learning models, we need to pay attention that the number of occurrence of anomalies is relatively very small when compared to normal data points in the training dataset, therefore we can't use accuracy as an evaluation metric because for a model that predicts everything as non-anomalous, the accuracy will be greater than 99.99% and we wouldn't have captured any anomaly. Therefore the appropriate metric is precision, which will allow us to evaluate how many malicious packets did we detect and how many did we miss.
The goal in all anomaly detection algorithms is to reduce as many false negatives as possible, hence lowering the number of false negatives, better is the performance of the anomaly detection algorithm or increasing the true positives for better accuracy.
The first evaluation metric is the sensitivity that shows how certain we are that all the positive instances have been predicted positive. In other words, it provides a proportion of malicious packets correctly identified. The formula for the sensitivity is as follows:
Sensivity= TP/(TP+FN)
We define a True Positive (TP) occurs where the model correctly predicts the positive class (i.e., legit packets as legit). Similarly, a true negative (TN) occurs when the model correctly predicts the negative class correctly (malicious packets as malicious). On the other hand, a false negative (FN) is an outcome where the model incorrectly predicts the negative class (malicious packets as legit)
The second evaluation metric is the accuracy, which is a percentage that is calculated as the ratio of correctly classified intrusion attempts to the total number of packets inspected, this metric provides a good indication of how well a model is trained.
Accuracy= (Correctly classified packets)/(Total number of packets) x 100%
The third evaluation metric is the precision: it is the ratio of correctly classified packets over the total number of packets. Namely, it is about how certain the truly positive results are, the following formula illustrates how to calculate the precision for a ML model.
Precision= TP/(TP+FP)
Where the True Positives (TP) are the correct predictions, whereas a False Positive (FP) occur where the model incorrectly predicts the positive class (legit packets as anomalous).

3.5 Summary

In the recent years, advancement in artificial intelligence technology such as machine learning and deep learning techniques has been used to improve IoT AIDS (anomaly Intrusion Detection System).
When comparing different models of Machine Learning, we observe that supervised learning is a more trustworthy method compared to unsupervised learning; that can be computationally complex and less accurate in some cases. However, supervised learning comes with limitations, one may encounter difficulty in classifying big data, so concrete training dataset (examples) are required for training classifiers, and at times decision boundaries can be over trained in the absence of the right training set
On the other hand, the Unsupervised learning algorithms can (learn) the typical pattern of the network and can report anomalies without any labeled dataset. It can detect new types of intrusions but is very prone to false positive alarms.

More recently, deep learning has gained prominence in regard to intrusion detection when compared to traditional machine learning methods as deep learning out perform other techniques if the data size is large.

---

4. Blockchain solutions for IoT security

Utilizing blockchain for the security of the IoT systems comes with a lot of advantages. First, it allows us to create a fully transparent and open to all databases. Secondly, it provides a strong protections against data tampering. In addition to the two fundamental costs that blockchains will fundamentally address in the context of IoT communication that we will discuss below.
Blockchain is defined as a distributed ledger technology characterized by decentralized operations, where all data is stored as blocks are immutable once joined and authenticated in the chain. Blockchains are efficient in supporting computing solutions that is suitable for a range of applications. In Section 4.1 we discuss some of these applications in the context of IoT Security.
Blockchain was first popularized through bitcoin by Satoshi Nakamoto in 2008 [Satoshi2008]. Initially, bitcoin was created with blockchain technology to avoid double-spending. But now researcher demonstrated that blockchain can be used effectively for other purposes, one is for the security of IoT systems.
One advantage of utilizing blockchain for the security of the IoT systems is reducing the cost of verification, where society spends a lot of money on identity verifications and third-party authentication so that transactions or data exchange can be executed securely.
The second cost that this technology is affecting is the cost of networking, which is the cost of running and operating any digital platform by building up the infrastructure that will support that platform.
These digital platforms are centralized in their nature and have a lot of control, which raises consumers' data privacy concerns and broadly to censorship risk. blockchains remove the need for such infrastructure, where no single authority is needed, in addition blockchains are censorship-resistant and more resilient to attacks because its distributed, and that distribution of computing power and resources or storage.
So the goal ultimately is to lift up trust and to rely less and less on trust and rely more and more on a consensus-based verification.

4.1 Blockchain Applications

In this section, we will discuss the current trends for applying blockchains in the support of IoT security.
Blockchain offers encryption, validation, and verification, in blockchain technology data is encrypted, not modified and it is provable as every single bit of the data in a blockchain is fully encrypted and digitally signed. In addition to providing resource management that is vital given the limited capability of IoT devices. It is important for enforcing access control delegation to edge IoT devices.

The Authors [Satoshi2008] presented an IoT access control model based on blockchain. Their design provides a lightweight and decentralized secure access control framework for enforcing

access control permissions using smart contracts. Their aim is to provide a secure communication and trustworthy policy between the edge IoT devices. Another blockchain-based design presented by [Sultan2021] provides distributed access control and access rights. The authors identified three requirements that are essential for their design one is the secure data storage, the second is the IoT compatibility, and third is the decentralized access rights management.

Blockchain can offer enhanced security for IoT as the data inside the chain is hashed and cannot be altered and the new information which is recently added to the chain is shared across a consensus peers throughout the network and thus making it tamper-proof technology. Any tampering with the data within a block can be tracked as the resulting hash will not be equal to the previously generated hash within that block, it is also difficult to falsify a block across billions of participating peers.
Figure 4 summarizes the previously mentioned benefits of utilizing blockchains in the context of IoT systems [Pal2021]. Such as resiliency, Immutability, anonymity, auditability, blockchains allows scalability, its cost effective and fast, and most importantly its decentralized and protected by the encryption and hashing mechanisms.

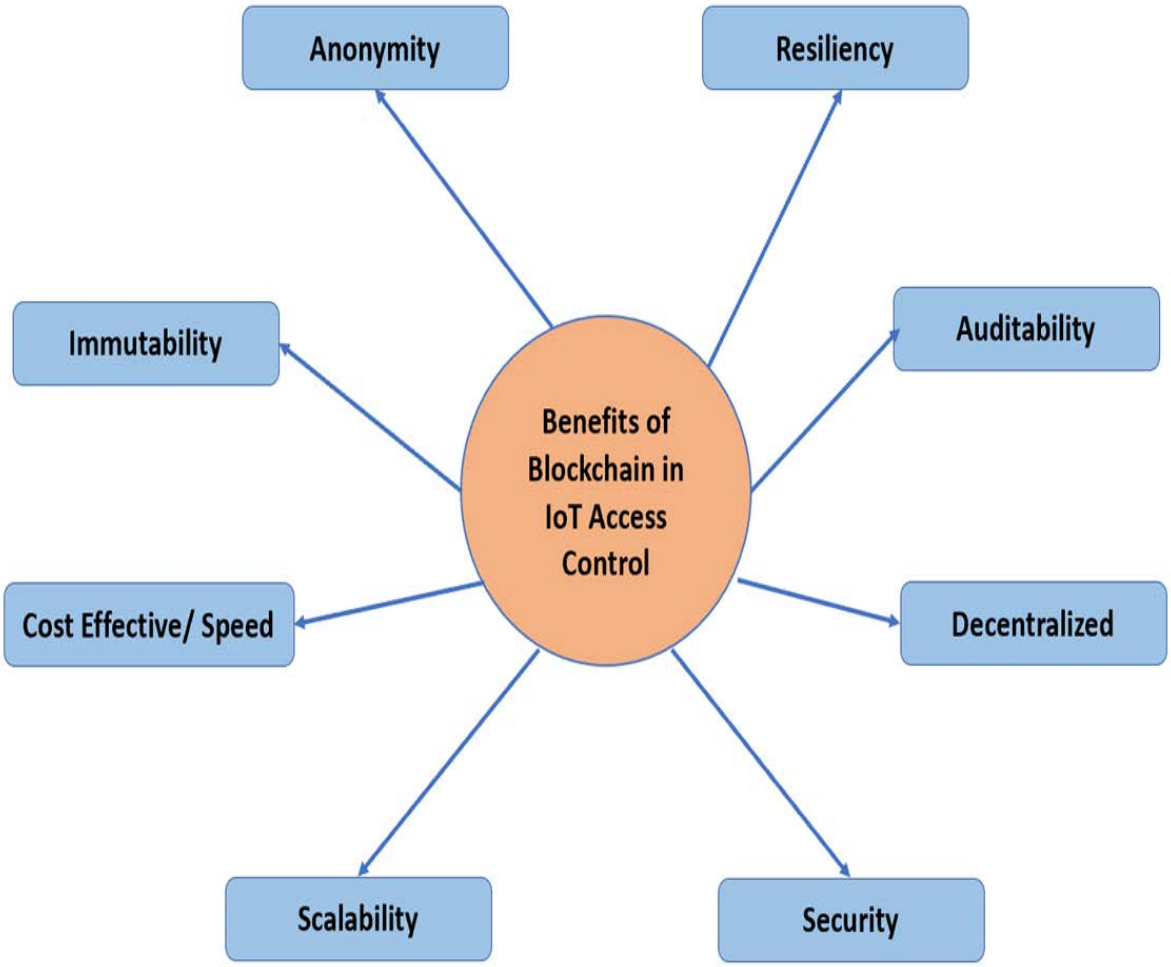

Figure 4: Benefits of using blockchain for IoT access control.

4.2 Summary

The advantage of using blockchains is that they can work at multiple layers, starting from the application layer, and ending in the lower layer of the communications, thus enabling the wide range use of the mechanism across layers and domains of the IoT system.
Recall, that blockchain removes the control from a centralized node and provides more flexibility in resource management for a number of scenarios including supply chain, transportation, and energy sectors, consisting of a vast amount of IoT devices. Yet, many of the proposed frameworks are implemented using simulators and not in a real environment, this change is needed to measure the achievement of the fundamental security goals for the IoT systems.
Based on research literature, and multiple experiments proved that utilizing blockchain technology in IoT networks can solve security problems that arise in communication between IoT devices because it has a higher level of security than the commonly used standalone IoT communication. Hence, blockchain guarantees data integrity.
More work is needed in the field of blockchains to enable efficient blockchain-based security mechanisms and enhance how the distributed ledgers (databases) can be optimally implemented. Also, establishing which of the implementations of blockchains are best suited for a given practical application or in a given platform.

---

5. Conclusion and Future Work

In conclusion, blockchains solve the problem of trust where we can exchange data securely and communicate without having to rely on an intermediary that increase the costs in addition to all sorts of communication and verification challenges. Yet continuous research in this field is needed to optimize Blockchain for the IoT network, such as research to make the existing blockchain lighter, there is also a need for presenting new blockchain algorithms that can compensate for all IoT network problems.
Future research should be directed, among other efforts to incorporate blockchain technology with other Security mechanisms, such as Machine Learning (i.e. anomaly intrusion detection systems), firewalling, and encryption to support the philosophy of Multi-Layer protection.
We hope that this comprehensive survey joining blockchains and Machine learning technology will be the first building block in the ongoing effort in this research field.

References

Perrone, Giovanni, Massimo Vecchio, Riccardo Pecori, and Raffaele Giaffreda. "The Day After Mirai: A Survey on MQTT Security Solutions After the Largest Cyber-attack Carried Out through an Army of IoT Devices." In IoTBDS, pp. 246-253. 2017."

https://pdfs.semanticscholar.org/c118/faf091c1a2a538c66c5959efe6ecbafe222e.pdf
Weber, Rolf H., and Evelyne Studer. "Cybersecurity in the Internet of Things: Legal aspects." Computer Law & Security Review 32, no. 5 (2016): 715-728.
https://https://www.sciencedirect.com/science/article/pii/S0267364916301169 Roman, Rodrigo, Jianying Zhou, and Javier Lopez. "On the features and challenges of security and privacy in distributed internet of things." Computer Networks 57, no. 10 (2013): 2266-2279.
https://www.sciencedirect.com/science/article/pii/S1389128613000054
Alam, Mansaf, A. K. Skahil, and Samiya Khan. "Internet of Things (IoT)." Concepts and Applications. Cham, Switzerland (2020). https://www.researchgate.net/profile/Mansaf-Alam/publication/341741439_Internet_of_Things_IoT/links/5ed15e15299bf1c67d274288/Internet-of-Things-IoT.pdf
Ahmad, Rasheed, and Izzat Alsmadi. "Machine learning approaches to IoT security: A systematic literature review." Internet of Things (2021): 100365.
https://www.sciencedirect.com/science/article/pii/S2542660521000093
I. Andrea, C. Chrysostomou and G. Hadjichristofi, "Internet of Things: Security vulnerabilities and challenges," 2015 IEEE Symposium on Computers and Communication (ISCC), pp.180-187,Larnaca, 2015. https://www.researchgate.net/profile/George-Hadjichristofi-2/publication/304408245_Internet_of_Things_Security_vulnerabilities_and_challenges/links/598188270f7e9b7b524b92ac/Internet-of-Things-Security-vulnerabilities-and-challenges.pdf
A Almusaylim, Zahrah, N. Z. Jhanjhi, and Abdulaziz Alhumam. "Detection and mitigation of RPL rank and version number attacks in the internet of things: SRPL-RP." Sensors 20, no. 21 (2020): 5997. https://www.mdpi.com/1424-8220/20/21/5997
Link: https://www.f5.com/labs/articles/threat-intelligence/the-hunt-for-iot--multi-purpose-attack-thingbots-threaten-intern https://www.f5.com/labs/articles/threat-intelligence/the-hunt-for-iot--multi-purpose-attack-thingbots-threaten-intern
https://www.f5.com/labs/articles/threat-intelligence/the-hunt-for-iot--multi-purpose-attack-thingbots-threaten-intern: 1-27. https://www.f5.com/labs/articles/threat-intelligence/the-hunt-for-iot--multi-purpose-attack-thingbots-threaten-intern:
Federated Learning-based Anomaly Detection for IoT Security Attacks, Mothukuri, Viraaji, et al. "Federated Learning-based Anomaly Detection for IoT Security Attacks." IEEE Internet of Things Journal (2021). https://ieeexplore.ieee.org/abstract/document/9424138
Network intrusion detection for IoT security based on learning techniques, Chaabouni, Nadia, et al. "Network intrusion detection for IoT security based on learning techniques." IEEE Communications Surveys & Tutorials 21.3 (2019): 2671-2701.
https://ieeexplore.ieee.org/abstract/document/8629941
Machine learning approaches to IoT security: A systematic literature review, Ahmad, Rasheed, and Izzat Alsmadi. "Machine learning approaches to IoT security: A systematic literature review." Internet of Things (2021): 100365.
https://www.sciencedirect.com/science/article/pii/S2542660521000093
A Machine-Learning-Based Approach for Autonomous IoT Security [Article] Saba, Tanzila, et al. "A Machine-Learning-Based Approach for Autonomous IoT Security." IT Professional 23.3 (2021): 69-75. https://ieeexplore.ieee.org/abstract/document/9464120
Intrusion detection in IoT networks using deep learning algorithm [Paper/Conference] Susilo, Bambang, and Riri Fitri Sari. "Intrusion detection in IoT networks using deep learning algorithm." Information 11.5 (2020): 279. https://www.mdpi.com/2078-2489/11/5/279
Annachhatre C, Austin TH, Stamp M (2015) Hidden Markov models for malware classification.

J Comput Virol Hack Technique 11(2) 59-73 Axelsson S (2000) "Intrusion detection systems: A survey and taxonomy"
https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/httpsredir=1article=1329context=etd_projects
D. Denning, "An intrusion detection model", IEEE Transactions on Software Engineering, vol. 13(2), 1987, pp. 222-232 https://ieeexplore.ieee.org/abstract/document/1702202
Wang, X. Hong and R. Ren, T. Li, "A real-time intrusion detection system based on PSO-SVM", in Proc. of the International Workshop on Information Security and Application 2009 (IWISA 2009), Qingdao, China, 2009, pp. 319-321 https://link.springer.com/chapter/10.1007/3-540-39945-3_7
J. Pu, Y. Li, L. Xiao and X. Dong, "A Detection Method of Network Intrusion Based on SVM and Ant Colony Algorithm " in Proc. National Conference on Information Technology and Computer Science(CITCS 2012), Lanzhou, China, 2012, pp.153-156.
Hsu, Ying-Feng, and Morito Matsuoka. "A Deep Reinforcement Learning Approach for Anomaly Network Intrusion Detection System." In 2020 IEEE 9th International Conference on Cloud Networking (CloudNet), pp. 1-6. IEEE, 2020.
https://ieeexplore.ieee.org/abstract/document/9335796
B. Deokar and A. Hazarnis, "Intrusion Detection System using Log Files and Reinforcement Learning," International Journal of Computer Applications, vol. 45, no. 19, 2012.
K. Sethi, E. S. Rupesh, R. Kumar, P. Bera and Y. V. Madhav, "A contextaware robust intrusion detection system: a reinforcement learning-based approach," International Journal of Information Security, 2019.
https://www.academia.edu/45064097/A_Model_to_Detect_Network_Intrusion_using_Machine_Leaning
T. Zoppi, A. Ceccarelli and A. Bondavalli, "Into the Unknown: Unsupervised Machine Learning Algorithms for Anomaly-Based Intrusion Detection," 2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S), 2020, pp. 81-81, doi: 10.1109/DSN-S50200.2020.00044.
https://ieeexplore.ieee.org/abstract/document/9159168
Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." Decentralized Business Review (2008): 21260. 21260-bitcoin-a-peer-to-peer-electronic-cash-system%20(1).pdf
Algarni, Sultan, Fathy Eassa, Khalid Almarhabi, Abduallah Almalaise, Emad Albassam, Khalid Alsubhi, and Mohammad Yamin. "Blockchain-based secured access control in an IoT system." Applied Sciences 11, no. 4 (2021): 1772. https://www.mdpi.com/2076-3417/11/4/1772
Pal, Shantanu, Ali Dorri, and Raja Jurdak. "Blockchain for IoT Access Control: Recent Trends and Future Research Directions." arXiv preprint arXiv:2106.04808 (2021).
https://arxiv.org/abs/2106.04808

List of Acronyms

ML	Machine Learning
DL	Deep Learning
NN	Neural Network
GUID	Global Unique Identifier

DDoS	Distributed Denial of Service
DoS	Denial of Service
RNN	Recurrent Neural Networks
CNN	Convolutional Neural Network
IoT	Internet of Things
ISA	International Society of Automation
NIST	National Institute of Standards and Technology
TPM	Trusted Computing Groups
RPL	Routing Protocol for Low power and Lossy networks
UPnP	Universal Plug and Play
SOHO	Small Office or Home Office
PDR	Packet Delivery Ratio
VNA	Version Number Attack
SQL	Structured Query Language
PoD	Ping of Death
HMM	Hidden Markov model
SVM	Support Vector Machines
AIDS	Anomaly Intrusion Detection Systems

Last modified on December 15, 2021