IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

The Importance Of Artificial Intelligence In Enhancing Cybersecurity For The Internet Of Things (Iot)

[1]Sachin Suman Vasant Dorage

[1]System Administrator

[1]Information Technology

[1]Oriental College of Pharmacy, Navi Mumbai, India

Abstract: In recent years, the proliferation of the Internet of Things (IoT) has surged dramatically, leading to heightened concerns regarding cybersecurity. At the forefront of cybersecurity advancements is Artificial Intelligence (AI), which plays a crucial role in developing sophisticated algorithms designed to safeguard networks and systems, including those associated with Internet of Things IoT. Nevertheless, cybercriminals have learned to manipulate Artificial Intelligence (AI) technologies and have started employing adversarial Artificial Intelligence (AI) to execute cyberattacks. This review paper consolidates findings from various surveys and research studies related to Internet of Things IoT, Artificial Intelligence (AI), and the attacks that utilize or target Artificial Intelligence (AI), aiming to thoroughly present and summarize the pertinent literature in these interconnected domains.

Keywords - Artificial Intelligence (AI), Internet of Things (IoT), Cybersecurity

1. INTRODUCTION

Cisco Systems characterizes the Internet of Things as the moment when the number of connected "things or objects" surpassed the number of individuals online. They estimate that this phenomenon emerged between 2008 and 2009, with the ratio of devices to people increasing from 0.08 in 2003 to 1.84 by 2010 (Evans, 2011). The expansion of the Internet of Things (IoT) has been remarkable, establishing itself as an integral component of everyday life and finding applications in numerous households and enterprises. Defining IoT can be challenging due to its continuous evolution since its inception; however, it is most effectively characterized as a network comprising both digital and analog devices, equipped with unique identifiers (UIDs), which possess the capability to communicate data autonomously, without the need for human involvement (M., 2020). Typically, this is observed as an individual interacting with a central hub device or application, frequently a mobile application, which subsequently transmits data and commands to one or more peripheral IoT devices (D, 2019). The fringe devices possess the capability to perform functions as needed and transmit data back to the central hub device or application, allowing the user to access this information.

The Internet of Things (IoT) concept has enhanced global connectivity by providing improved accessibility, integrity, availability, scalability, confidentiality, and interoperability among devices (Yang Lu, 2019). The Internet of Things (IoT) faces significant vulnerabilities to cyberattacks, primarily attributed to their numerous attack surfaces and the relative novelty of the technology, which results in insufficient security standardizations and requirements (Chalee Vorakulpipat, 2018). A wide range of cyberattacks can be employed by attackers against Internet of Things (IoT) devices, contingent upon the specific components of the system they aim to exploit and their objectives for the attack. Consequently, there is extensive research focused on cybersecurity

in the context of IoT. This research encompasses the application of Artificial Intelligence (AI) techniques to safeguard IoT systems from potential threats, primarily through the identification of anomalous behaviors that could signify an ongoing attack (A., 2020). In the realm of the Internet of Things (IoT), cyber-attackers possess a significant advantage, as they are required to identify only a single vulnerability, whereas cybersecurity professionals must safeguard numerous targets. This disparity has resulted in a heightened adoption of artificial intelligence by cyber-attackers, enabling them to circumvent the complex algorithms designed to identify unusual activities and evade detection (Pendse, 2019). The expansion of IoT technologies has significantly increased the focus on artificial intelligence. As a result, various AI methodologies, including decision trees, linear regression, machine learning, support vector machines, and neural networks, have been implemented in cybersecurity applications within the IoT domain to detect threats and potential attacks (Francesca Meneghello, 2019).

This document aims to deliver an extensive analysis of the security vulnerabilities associated with Internet of Things (IoT) applications, along with potential mitigation strategies. Additionally, it will compare various IoT technologies based on criteria such as integrity, anonymity, confidentiality, privacy, access control, authentication, authorization, resilience, and self-organization. The authors introduce deep learning models that utilize the CICIDS2017 datasets for detecting Distributed Denial of Service (DDoS) attacks, achieving a notable accuracy rate of 97.16% in enhancing cybersecurity within IoT environments (Monika Roopak). Furthermore, in another study (Janice Canedo, 2016), the authors assess the effectiveness of Artificial Neural Networks (ANN) implemented in a gateway device, which is capable of identifying anomalies in data transmitted from edge devices. The findings indicate that this proposed method significantly bolsters the security of IoT systems. The researchers in (Faezeh Farivar, April 2020) put forward an AI-driven control strategy aimed at the detection, estimation, and mitigation of cyber-attacks within industrial IoT systems. The authors present a comprehensive pervasive detection system tailored for IoT environments. They also devise a range of adversarial attacks and corresponding defense mechanisms, validating their methodology using datasets such as MNIST, CIFAR-10, and SVHN (Shen Wang, 2019). Ultimately, examines strategies for capturing and evaluating cybersecurity risks associated with IoT devices, aiming to standardize these practices to enhance the identification and protection against risks in IoT systems (Petar Radanliev, 2020).

This review article addresses a range of subjects related to cybersecurity, the Internet of Things (IoT), and Artificial Intelligence (AI), exploring their interconnections through three survey-style sections. It offers an extensive analysis of cyberattacks targeting IoT devices and suggests AI-driven approaches for mitigating these threats. The primary objective of this paper is to serve as a valuable resource for researchers engaged in these significant topics by summarizing and linking pertinent works that address various dimensions of these fields.

2. TECHNIQUES FOR ATTACKING INTERNET OF THINGS DEVICES

The inadequate security measures present in numerous IoT devices have enabled cybercriminals to exploit various vulnerabilities across multiple attack surfaces. These attack surfaces typically include the IoT device itself encompassing both its hardware and software the network to which the device is connected, and the applications that interact with it. Collectively, these components represent the primary attack surfaces within an IoT ecosystem. A basic overview of a typical IoT system is depicted in Figure 1; the majority of attacks examined in this paper primarily target the network gateway and/or cloud data server connections, as these areas often exhibit significant security deficiencies.

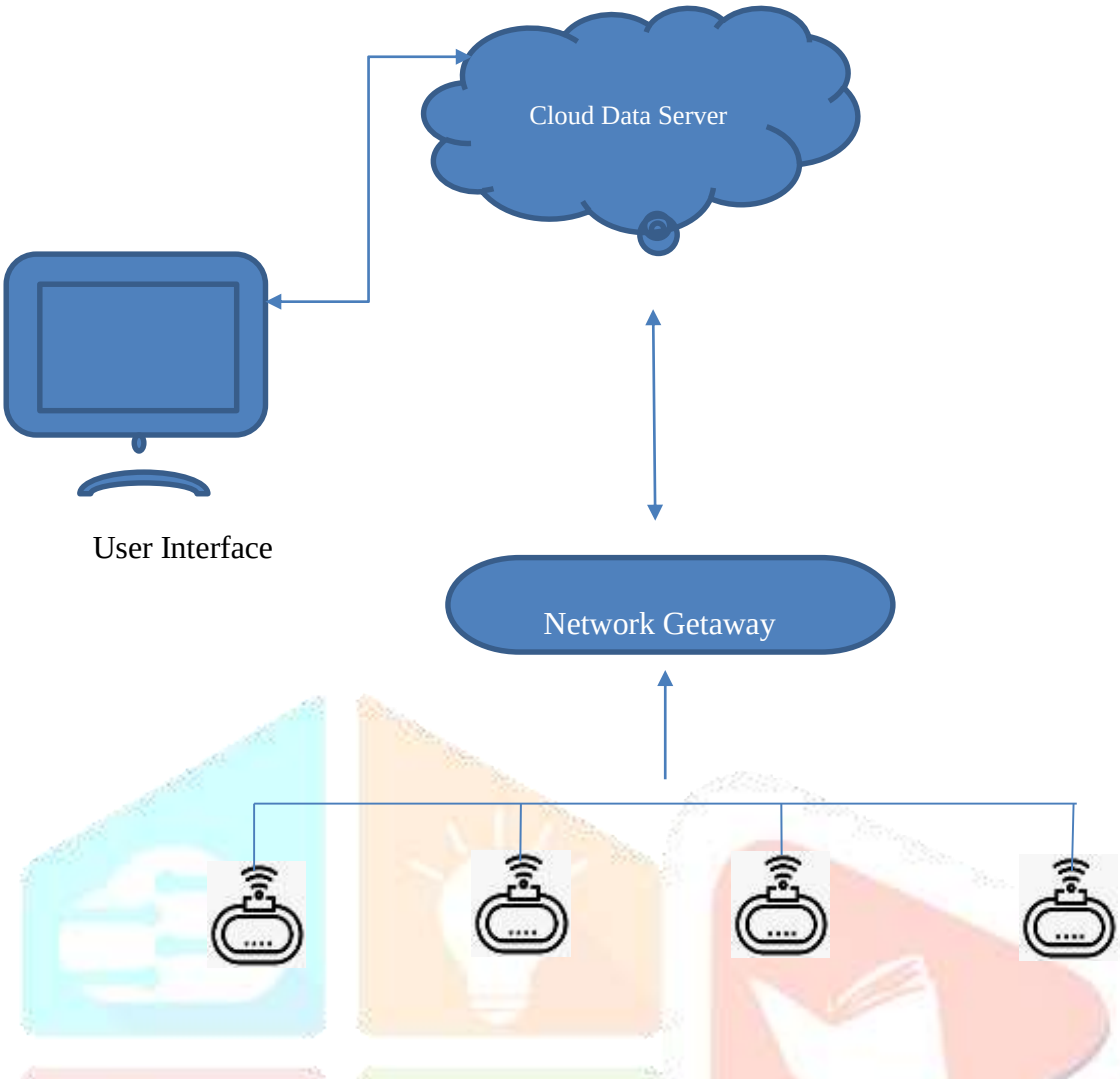


Figure 1. An overview of the standard architecture of IoT systems

2.1 Preliminary Assessment

Prior to launching cyberattacks on Internet of Things (IoT) devices, attackers typically conduct a thorough examination of the target device to uncover potential vulnerabilities. This process often involves purchasing a replica of the IoT device from the market. Subsequently, they engage in reverse engineering to simulate an attack, assessing the possible outputs and identifying various attack vectors. This may include disassembling the device to scrutinize its internal components, such as flash memory, to gain insights into the software, as well as manipulating the microcontroller to extract sensitive data or induce unintended behaviors (Woo, 2019). To mitigate the risks associated with reverse engineering, it is crucial for IoT devices to incorporate hardware-based security measures. The application processor, which encompasses sensors, actuators, power supply, and connectivity, should be housed in a tamper-resistant environment. Additionally, device authentication can be enhanced through hardware-based security, enabling the device to verify its authenticity to the connected server (Woo, 2019).

2.2 Physical Attacks

A frequently encountered category of attacks that is often characterized by low technological sophistication encompasses physical attacks, wherein the attacker in various ways exploits the hardware of the targeted device. There exists a range of physical attack types, including outage attacks, which involve disabling the network to which the devices are connected, thereby disrupting their operations; physical damage, which entails harming devices or their components to hinder their proper functioning; malicious code injection, exemplified by an attacker inserting a USB drive containing a virus into the target device; and object jamming, where signal jammers are employed to obstruct or manipulate the signals emitted by the devices (Hezam Akram Abdul-Ghani, 2018). Permanent denial of service (PDoS) attacks, elaborated upon later in this document, can be executed as a physical assault. For instance, if an Internet of Things (IoT) device is linked to a high voltage power supply, it may experience an overload in its power system, necessitating its replacement (Herberger, 2015).

2.3 Man-in-the-Middle

A prevalent form of attack targeting Internet of Things (IoT) devices is the Man-in-the-Middle (MITM) attack. In the context of computing, an MITM attack involves the interception of communication between two nodes, enabling the attacker to assume the role of an intermediary. Such attacks can occur across various connections, including those between a computer and a router, two mobile phones, and, most frequently, between a server and a client. A basic illustration of an MITM attack between a client and a server is depicted in Figure 2. When it comes to IoT, attackers typically execute MITM attacks between an IoT device and its corresponding application. IoT devices are particularly susceptible to these attacks due to their lack of standard security measures. There are two primary modes of MITM attacks: cloud polling and direct connection. In the cloud-polling scenario, the smart home device maintains continuous communication with the cloud, often to check for firmware updates. Attackers can manipulate network traffic through methods such as Address Resolution Protocol (ARP) poisoning, altering Domain Name System (DNS) settings, or intercepting HTTPS traffic using self-signed certificates or tools like Secure Sockets Layer (SSL) stripping (Zoran Cekerevac, 2017). Numerous Internet of Things (IoT) devices fail to authenticate the legitimacy or trustworthiness of certificates, rendering the self-signed certificate approach particularly advantageous. In scenarios involving direct connections, devices interact with a hub or application within the same network. This enables mobile applications to identify new devices by scanning each IP address on the local network for a designated port. An attacker can replicate this process to uncover devices on the network (Zoran Cekerevac, 2017). A notable instance of a man-in-the-middle (MITM) IoT attack involves a smart refrigerator capable of displaying the user's Google calendar. While this feature may appear innocuous, attackers discovered that the system did not verify SSL certificates, thereby facilitating an MITM attack that compromised the user's Google credentials (Zoran Cekerevac, 2017).

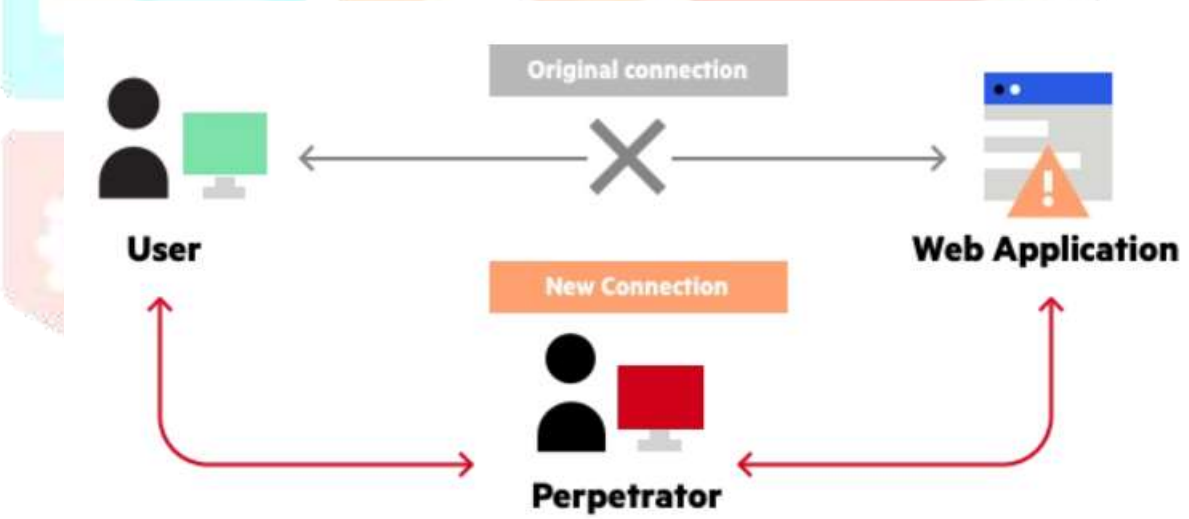


Figure 2. An uncomplicated illustration of a Man-in-the-Middle attack (htt7)

2.4 Bluetooth Man-in-the-Middle Attack

A prevalent method of man-in-the-middle (MITM) attack targeting Internet of Things (IoT) devices occurs through Bluetooth connections. Numerous IoT devices utilize Bluetooth Low Energy (BLE), which is specifically designed for IoT applications to be compact, cost-effective, and energy-efficient (MELAMED, 2018). Nevertheless, BLE is susceptible to MITM attacks. While BLE employs AES-CCM encryption, which is generally regarded as secure, the process of exchanging encryption keys often lacks adequate security. The overall security is contingent upon the pairing method utilized for the exchange of temporary keys between devices. BLE implements a three-phase pairing process: initially, the initiating device transmits a pairing request, and the devices share their pairing capabilities over an unsecured channel; subsequently, the devices exchange temporary keys and confirm that they are utilizing the same temporary key, which is then employed

to generate a short-term key (some newer devices may utilize a long-term key exchanged through Elliptic Curve Diffie-Hellman public-key cryptography, offering significantly enhanced security compared to the conventional BLE protocol); finally, the generated key is transmitted over a secure connection and can be utilized for data encryption (MELAMED, 2018). The three-phase pairing process is illustrated in Figure 3.

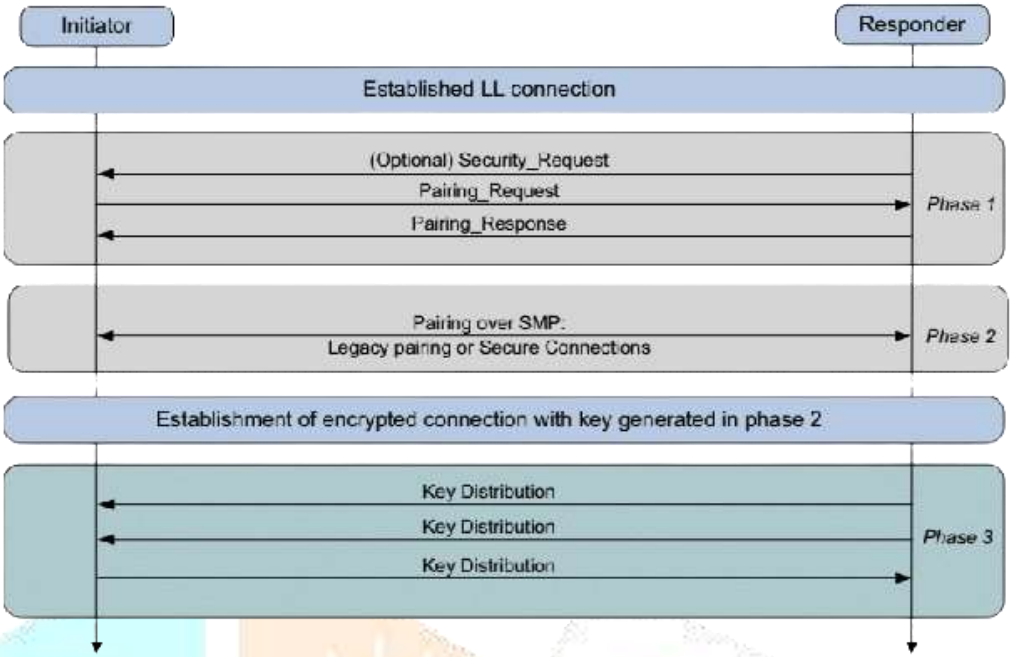


Figure 3. A visual representation depicting the fundamental process of Bluetooth Low Energy (BLE) pairing (Ren, 2016)

The temporary key is established based on the pairing method, which is defined at the operating system level of the device. There are three prevalent pairing methods commonly utilized with IoT devices. The first method, known as Just Works, assigns the temporary key a value of zero, which is inherently insecure. Despite this vulnerability, it remains one of the most widely adopted pairing methods for Bluetooth Low Energy (BLE) devices (MELAMED, 2018). The second method, Passkey, requires the user to manually input a six-digit numerical combination into the device, offering a reasonable level of security, although there are techniques that can circumvent this protection. Lastly, the Out-of-Band pairing method facilitates the exchange of temporary keys through mechanisms such as Near Field Communication. The security of this method is contingent upon the protective measures of the exchange channel against man-in-the-middle (MITM) attacks, thereby ensuring the security of the BLE connection (MELAMED, 2018). However, the Out-of-Band method has not yet gained widespread adoption among IoT devices. Another significant aspect of BLE devices is the Generic Attribute Profile (GATT), which enables communication between devices through a standardized data schema. The GATT outlines the roles, behaviors, and other metadata of devices. Any application that supports BLE and is within range of an IoT device can access its GATT schema, thereby obtaining essential information. For an attacker to execute MITM attacks within BLE networks, they must utilize two connected BLE devices: one device serves as the IoT device to connect with the target mobile application, while a counterfeit mobile application connects to the intended IoT device. Additional tools for conducting BLE MITM attacks include GATT acker, a Node.js package that scans and replicates BLE signals, subsequently running a cloned version of the IoT device, and BtleJuice, which enables MITM attacks on Bluetooth Smart devices that possess enhanced security features compared to BLE (MELAMED, 2018).

2.5 Attacks involving the injection of erroneous Data

When an attacker gains access to one or more devices within an IoT network through a man-in-the-middle (MITM) attack, a subsequent action they may undertake is a False Data Injection (FDI) attack. In an FDI attack, the perpetrator subtly modifies the readings from IoT sensors to avoid detection, subsequently transmitting the altered data (G. R. Mode, 2020). While there are various methods to execute FDI attacks, utilizing MITM techniques is often the most effective approach. These attacks frequently target sensors that relay information to algorithms designed to make predictions or draw conclusions based on the incoming data. Such algorithms, commonly known as predictive maintenance systems, are widely employed to monitor the

condition of mechanical equipment and forecast maintenance needs (G. R. Mode, 2020). These systems are also integral to the infrastructure of smart cities, where FDI attacks could have severe consequences. A pertinent example of an FDI attack involves sensors on an aircraft engine that assess when critical maintenance is required. If attackers manage to infiltrate even a fraction of these sensors, they can introduce minor distortions that evade detection by existing data validation systems, yet are sufficient to mislead the algorithm's forecasts (G. R. Mode, 2020). In experimental scenarios, this manipulation could result in postponing essential maintenance, potentially leading to catastrophic failures during operation, which could incur significant unplanned costs or even endanger lives.

2.6 BOT NETWORKS (BOTNETS)

A prevalent form of attack on IoT devices involves the recruitment of numerous devices to form botnets for executing Distributed Denial of Service (DDoS) attacks. A Denial of Service (DoS) attack is defined by a coordinated effort to obstruct legitimate access to a service, while a DDoS attack employs multiple sources to accomplish this objective. The primary goal of DDoS attacks is to inundate the target service's infrastructure, thereby disrupting its normal data flow. Typically, DDoS attacks unfold in several stages: first, recruitment, where the attacker identifies vulnerable machines to be utilized in the assault; next, exploitation and infection, during which these machines are compromised and malicious code is introduced; then, communication, where the attacker evaluates the infected devices to determine their online status and plans the timing of attacks or upgrades; and finally, the attack phase, where the attacker directs the compromised machines to unleash malicious packets on the target (Michele De Donno, 2017). A prevalent method for acquiring compromised machines and executing DDoS attacks involves utilizing IoT devices, primarily because of their widespread accessibility and typically inadequate security and upkeep. In Figure 4, you can see a typical command structure where the attacker’s main computer issues commands to one or more infected command and control centers, each managing a group of compromised devices that can subsequently launch attacks on the target.

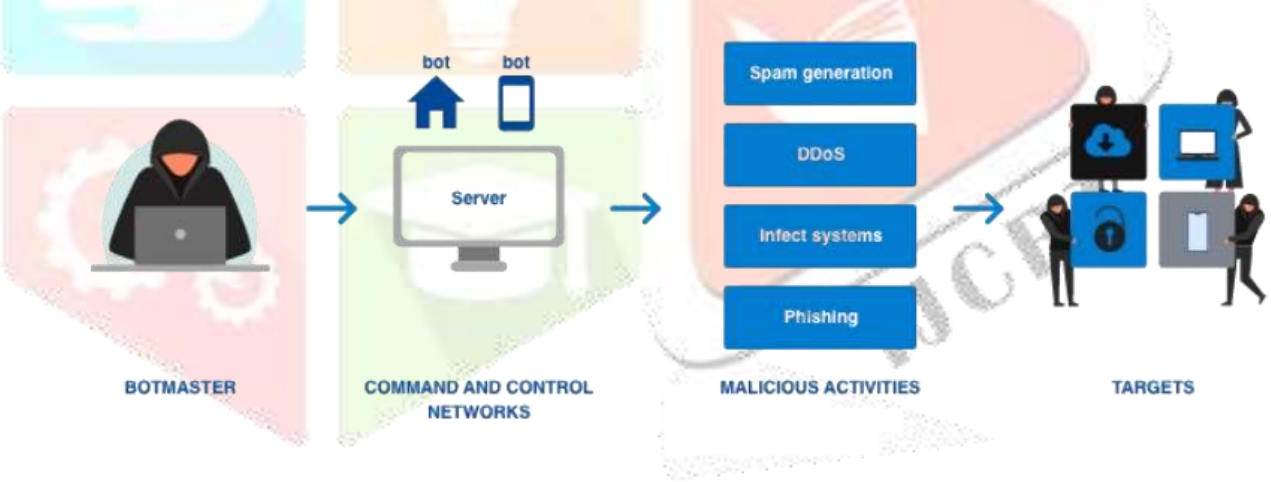


Figure. 4 A visual depiction of a typical botnet structure (Spajic, 2023)

The Mirai worm, recognized as one of the most notorious forms of malware, has been instrumental in executing some of the most significant DDoS attacks recorded in history. It is specifically engineered to infiltrate and take control of Internet of Things (IoT) devices, including network or digital video recorders, surveillance cameras, and residential routers. Once compromised, these devices are integrated into a vast botnet capable of launching various types of DDoS attacks. Mirai was developed to support a range of widely used CPU architectures found in IoT devices, such as x86, ARM, Sparc, PowerPC, and Motorola, thereby maximizing its potential to infect numerous devices (Nguyen, n.d.).

2.7 Service disruption attacks

Internet of Things (IoT) devices can frequently initiate Denial of Service (DoS) attacks; however, they are also vulnerable to such attacks. These devices are especially at risk of permanent denial of service (PDoS) attacks, which can incapacitate a device or system entirely. This incapacitation may occur through the overloading of battery or power systems, or more commonly, through firmware attacks. In a firmware attack, an attacker exploits vulnerabilities to substitute a device's fundamental software, typically its operating system, with a compromised or faulty version, thereby rendering the device inoperative (Herberger, 2015). 1. The legitimate procedure referred to as flashing contrasts with its illegitimate version, termed "Phlashing." When a device undergoes phlashing, the owner is compelled to restore the device by installing a fresh version of the operating system along with any additional content that may have been previously added. In severe instances of attack, the compromised software may cause excessive strain on the device's hardware, rendering recovery unattainable without the replacement of certain components (Herberger, 2015). Assaults on the power system of a device, while not as widely recognized, can be even more catastrophic. A pertinent illustration of such an attack involves a USB device embedded with malware, which, upon being connected to a computer, excessively drains the device's power, ultimately leading to irreversible damage to the hardware that necessitates its replacement (Herberger, 2015).

An illustrative instance of PDoS malware is referred to as BrickerBot. This malware employs brute force dictionary attacks to infiltrate Internet of Things (IoT) devices. Upon successfully accessing the device, it executes a sequence of commands that lead to irreversible damage. These commands involve altering the device's storage and kernel settings, disrupting internet connectivity, impairing device functionality, and erasing all data stored on the device (I. Gulatas, 2023). This assault is so destructive that it frequently necessitates either the reinstallation of hardware or the total replacement of the device. In cases where the hardware endures the attack, the software is typically compromised and would require reflashing, resulting in the loss of all data that may have been stored. Notably, BrickerBot was engineered to target the same devices as the Mirai botnet, which it would utilize as bots, and it employs a similar or identical dictionary for its brute force attacks. Ultimately, BrickerBot was designed with the intention of incapacitating those devices that Mirai could potentially recruit, as part of a strategy to counteract the botnet (I. Gulatas, 2023).

The architecture of IoT systems presents numerous potential vulnerabilities; however, the most prevalent method of compromising these systems is through their connections, which are often the most susceptible points. Moving forward, it is essential for IoT developers to implement robust security measures to safeguard their products against such threats. Additionally, the establishment of IoT security standards would assist consumers in avoiding the purchase of insecure devices. Furthermore, ensuring the security of the network hosting the IoT system can significantly reduce the risk of common attacks. Maintaining a degree of separation between the IoT system and other critical infrastructures, along with having contingency plans in place, will also help to minimize the impact of any successful attacks.

3. ARTIFICIAL INTELLIGENCE IN THE FIELD OF CYBERSECURITY

To effectively safeguard systems against cyber threats, numerous cybersecurity professionals are increasingly utilizing Artificial Intelligence (AI). Artificial Intelligence (AI) is predominantly employed for intrusion detection within the field of cybersecurity by examining traffic patterns and identifying behaviors that are indicative of an attack**.**

3.1 The field of machine learning

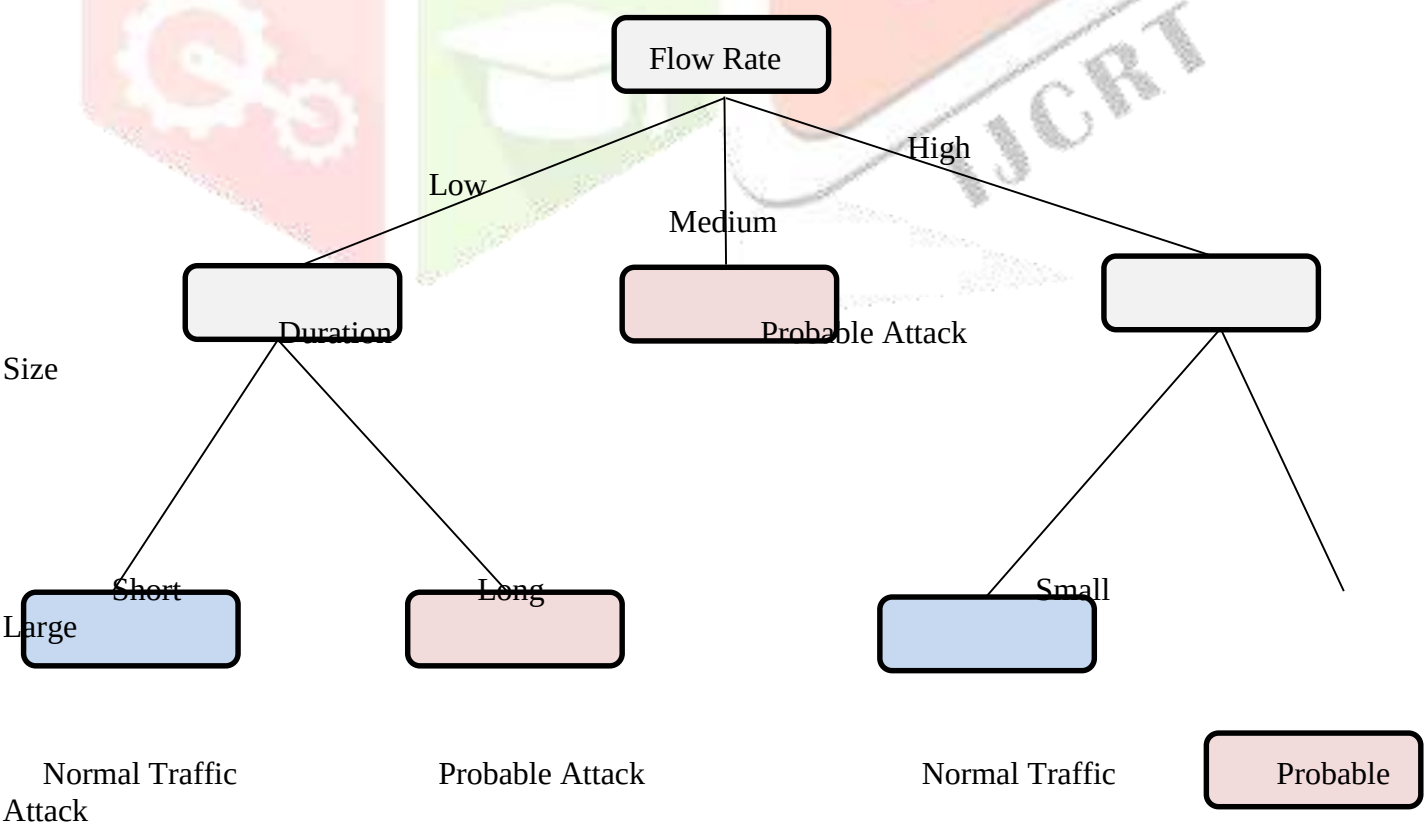
Machine learning can be categorized into two primary types: supervised learning and unsupervised learning. In supervised learning, human experts label the training data as either malicious or legitimate, which is subsequently fed into the algorithm to develop a model that identifies distinct "classes" of data for comparison with the analyzed traffic. Conversely, unsupervised learning does not rely on pre-labeled training data; instead, the algorithm autonomously organizes similar data points into classes and classifies them based on the coherence of data within each class and the modularity of data between different classes (Sherali Zeadally, 2020). A widely utilized machine-learning algorithm in the field of cybersecurity is naïve Bayes. This algorithm aims to classify data by applying the principles of the Bayesian theorem, under the assumption that anomalous activities stem from independent events rather than a single attack. As a supervised learning

algorithm, naïve Bayes, after undergoing training and establishing its classifications, evaluates each activity to assess the likelihood of it being anomalous (Sherali Zeadally, 2020). Machine learning algorithms may also be employed to develop the additional models addressed in this section.

3.2 Tree-based decision models

A decision tree is a form of artificial intelligence that establishes a series of rules derived from its training data samples. It employs a process of iterative division to identify a classification (commonly referred to as "attack" or "normal") that most accurately categorizes the traffic under examination. In the realm of cybersecurity, this method can be utilized to identify DoS attacks by evaluating the flow rate, size, and duration of the traffic. For instance, if the flow rate is low while the traffic duration is extended, it is probable that an attack is occurring, leading to its classification as such *(*Sherali Zeadally, 2020*)*. Decision trees serve as a valuable tool for identifying command injection attacks in robotic vehicles by classifying metrics such as CPU usage, network traffic, and data write volume, as illustrated in Figure 5. This method is favored for its straightforwardness, allowing developers to understand what the AI recognizes as normal versus anomalous traffic. Furthermore, once a robust set of rules is established, the AI is capable of monitoring traffic in real-time, enabling prompt notifications when any irregular activity is observed (Sherali Zeadally, 2020). An alternative method for constructing decision trees is the Rule-Learning technique, which identifies a collection of attack characteristics during each iteration while optimizing a score that reflects the classification quality, specifically the count of misclassified data samples *(*Sherali Zeadally, 2020*)*. The key distinction between conventional decision trees and rule-learning methods lies in their approach: traditional decision trees focus on identifying characteristics that facilitate classification, while rule-learning techniques aim to establish a comprehensive set of rules that define a class. This approach can be beneficial as it allows for the incorporation of human insights in the rule generation process, resulting in a more refined set of rules *(*Sherali Zeadally, 2020*)*.

Figure 5. A sample decision tree utilized for the classification of network traffic.



3.3 K-nearest neighbor's algorithm

The k-nearest neighbor (k-NN) method utilizes data samples to form classifications by evaluating the Euclidean distance between a new data point and existing classified data points, thereby determining the appropriate class for the new data point in a straightforward manner *(*Sherali Zeadally, 2020*)*. For instance, when the number of nearest neighbors, k, is set to three (3), the new data point would be categorized into class two (2). However, if k is increased to nine (9), the same data point would then fall into class one (1), as illustrated in Fig. 6. The k-NN method is particularly appealing for intrusion detection systems due to its ability to rapidly adapt to new traffic patterns, enabling the identification of previously unrecognized threats, including zero-day attacks. Cybersecurity researchers are actively exploring the use of k-NN for the real-time detection of cyber threats *(*Sherali Zeadally, 2020*)*. This technique has been successfully applied to identify various types of attacks, such as false data injection attacks, and it demonstrates effectiveness when data can be modeled in a way that facilitates distance measurement, such as through Gaussian distribution or vector representation.

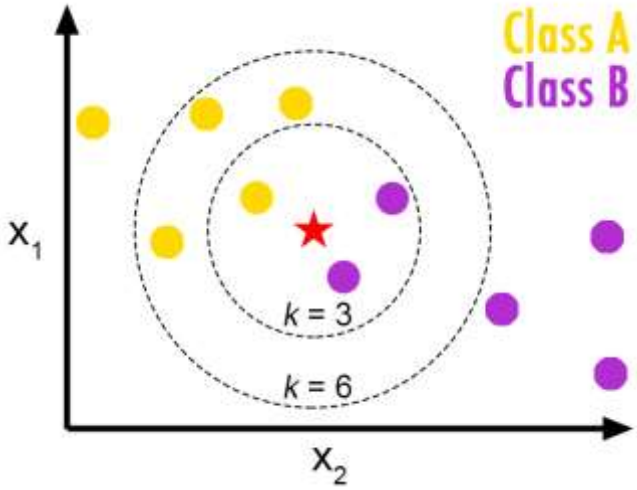


Figure 6. The k-NN method can categorize a data point in various ways depending on the chosen k values *(Sawla, 2018)*

3.4 Support Vector Machines (SVM)

Support vector machines (SVMs) represent an advancement of linear regression models, designed to identify a hyperplane that divides data into two distinct categories (Sherali Zeadally, 2020). This hyperplane may take various forms, including linear, non-linear, polynomial, Gaussian, or sigmoid, contingent upon the function employed within the algorithm. Furthermore, SVMs possess the capability to categorize data into multiple classes by utilizing several hyperplanes. In the realm of cybersecurity, this methodology is employed to scrutinize Internet traffic patterns, classifying them into various types such as HTTP, FTP, SMTP, among others (Sherali Zeadally, 2020). As a supervised machine learning approach, SVM is frequently applied in scenarios where attack simulations are feasible, such as utilizing network traffic generated from penetration testing as training data.

3.5 Artificial Neural Networks

Artificial neural networks (ANNs) represent a computational approach inspired by the interactions of neurons in the brain, facilitating the transmission and interpretation of information. Within ANNs, a neuron functions as a mathematical model that processes input data and generates an output value, which is subsequently transmitted to the next neuron based on its computed value. The ANN algorithm undergoes multiple iterations until the output value aligns sufficiently with the target value, enabling the neurons to learn and adjust their weights by evaluating the discrepancy between the anticipated value and the prior output. Upon completion of this iterative process, the algorithm yields a mathematical model that produces a value suitable for data classification (Sherali Zeadally, 2020). One significant advantage of artificial neural networks (ANNs) is their capacity to modify their mathematical frameworks in response to new data. In contrast, traditional

mathematical models may become outdated as novel traffic patterns and attack vectors emerge. This adaptability enables ANNs to effectively identify previously unknown and zero-day attacks, as they prioritize new information more than static models can. Consequently, ANNs serve as robust intrusion detection systems and have demonstrated strong performance against attacks like Denial of Service (DoS) (Sherali Zeadally, 2020).

Currently, the application of artificial intelligence in the field of cybersecurity is a relatively small yet swiftly expanding domain. This approach tends to be costly and resource demanding, making it impractical for the protection of smaller systems. Conversely, organizations with extensive networks stand to gain significantly from these solutions, particularly if they are contemplating or have already implemented Internet of Things (IoT) devices within their infrastructure. The integration of AI in cybersecurity would also prove advantageous in large-scale systems typical of smart cities, where rapid response times are crucial, especially in areas such as traffic management. Looking ahead, AI-driven cybersecurity could potentially be adapted for smaller systems, including autonomous vehicles and smart home technologies. Furthermore, it is important to note that many AI cybersecurity strategies focus on detecting or mitigating ongoing attacks rather than preventing them outright, underscoring the necessity for additional preventive security measures to be established.

4. ARTIFICIAL INTELLIGENCE TARGETING THE INTERNET OF THINGS

Not all artificial intelligence is employed for cybersecurity; in fact, cybercriminals are increasingly leveraging harmful AI to facilitate their attacks. This often involves circumventing intrusion detection systems, particularly in the realm of IoT, or manipulating beneficial AI to turn it against its own framework (Murat Kuzlu, 2021).

4.1 Streamlining the process of identifying vulnerabilities

Machine learning serves as a powerful tool for uncovering vulnerabilities within systems. This capability is advantageous for security professionals aiming to intelligently identify and address weaknesses that require remediation. However, malicious actors also leverage this technology to pinpoint and exploit vulnerabilities in their targets. As the adoption of technology increases, particularly in low-security environments like IoT devices, the number of exploitable vulnerabilities has surged, including zero-day vulnerabilities. To swiftly identify these weaknesses, attackers frequently employ AI, allowing them to exploit vulnerabilities at a pace that outstrips developers' ability to resolve them. While developers can utilize similar detection tools, they face a significant challenge: they must identify and rectify every potential vulnerability, whereas attackers only need to discover one. This disparity highlights the critical advantage that automated detection provides to those with malicious intent (Murat Kuzlu, 2021).

4.2 Fuzz testing

Fuzzing is fundamentally a testing technique that creates random inputs such as numbers, characters, metadata, binary data, and particularly "known-to-be-dangerous" values like zeros, negative or excessively large numbers, SQL queries, and special characters that can lead to the failure of the target software (Jeesoo Jurn, 2015). Fuzzing can be categorized into two types: dumb fuzzing and smart fuzzing. Dumb fuzzing operates by randomly altering input variables, which allows for rapid execution due to its simplicity. However, its effectiveness in identifying defects is limited because it achieves only narrow code coverage (Jeesoo Jurn, 2015). In contrast, smart fuzzing creates input values that are tailored to the specific software, taking into account its structure and potential error points. This analytical approach provides a significant advantage, as it enables the fuzzing algorithm to pinpoint where errors are likely to arise. Nevertheless, crafting an effective smart fuzzing algorithm requires specialized knowledge and careful adjustments (Jeesoo Jurn, 2015).

4.3 Symbolic analysis

Symbolic execution is a method akin to fuzzing that identifies vulnerabilities by assigning symbolic values to input variables rather than actual values (Jeesoo Jurn, 2015). This approach is typically categorized into offline and online symbolic execution. Offline symbolic execution focuses on exploring a single path at a time, generating new input variables by resolving the path predicate (Jeesoo Jurn, 2015). Consequently, each time a new path is to be examined, the algorithm must restart, which presents a drawback due to the

considerable overhead associated with re-executing the code. In contrast, online symbolic execution duplicates states and produces path predicates at each branch statement (Jeesoo Jurn, 2015). While this technique incurs minimal overhead, it necessitates substantial storage capacity to maintain all state information and to manage the concurrent processing of the numerous states it generates, resulting in significant resource utilization.

4.4 Input Attacks

An input attack occurs when an assailant modifies the input of an artificial intelligence system in a manner that leads to a malfunction or erroneous output. Such attacks involve the introduction of an attack pattern into the input, which can range from placing tape over a physical stop sign to mislead autonomous vehicles, to introducing subtle noise into an image that remains undetectable to the human eye yet confounds the AI (Comiter, 2019). Importantly, the integrity of the AI's algorithm and its security does not need to be compromised for an input attack to be effective; it is sufficient for the attacker to alter the specific input intended to disrupt the output. For instance, in the scenario involving tape on a stop sign, the attacker may not require any technological means. Conversely, more advanced attacks can be entirely imperceptible to humans, where the attacker meticulously modifies a minuscule portion of an image to mislead the algorithm. Input attacks are frequently classified according to two dimensions: perceivability and format.

The visibility of an input attack refers to the extent to which the attack can be detected by the human eye, while the format indicates the degree to which the attack is digital as opposed to physical. At one extreme of the visibility spectrum are perceivable attacks (Comiter, 2019). These include modifications to targets, such as deforming, removing portions of, or altering colors, as well as additions to the target, like applying physical tape or incorporating digital markings. Although perceivable attacks are detectable by humans, individuals may overlook minor alterations, such as tape on a stop sign, or may not regard them as significant (Comiter, 2019). A human driver is likely to recognize a stop sign, even if it is obscured by tape or scratches, whereas an autonomous vehicle might not. This characteristic enhances the effectiveness of perceivable attacks, enabling them to often remain unnoticed. In contrast, imperceptible attacks are undetectable to the human eye. Examples include "digital dust," which consists of a minimal amount of noise added to an image that remains invisible to humans but can significantly influence an AI's output, or an undetectable pattern on a 3D printed object that can be recognized by AI (Comiter, 2019). Additionally, imperceptible attacks can occur through audio, such as sounds played at frequencies beyond human hearing that can still be captured by a microphone. Generally, imperceptible attacks pose a greater security threat, as there is a minimal likelihood that a human would identify the attack prior to the AI algorithm producing an erroneous response.

Attacks typically manifest in either digital or physical formats, with few instances combining both modalities. In the realm of physical attacks, the patterns tend to be more overt, as physical objects must undergo digitization for processing, which can result in the loss of finer details (Comiter, 2019). Nevertheless, certain attacks remain challenging to detect despite this detail loss, such as those involving 3D printed items where the pattern seamlessly integrates with the object's structure, rendering it undetectable to the human eye. In contrast, digital attacks target digital inputs, including images, videos, audio files, and other data. Since these inputs are inherently digitized, there is no risk of detail loss during processing, enabling attackers to execute highly precise attacks that can be less noticeable than their physical counterparts (Comiter, 2019). However, digital attacks are not always imperceptible; for instance, altering an image of a celebrity by overlaying a peculiar pattern may lead an AI to misidentify the individual, yet the image still depicts a person. A pertinent example of input attacks can be observed in IoT smart vehicles and, more broadly, smart city infrastructures. As previously noted, strategically placing tape on a stop sign can prevent an algorithm from recognizing it, potentially misclassifying it as a green light. This poses significant risks to passengers if the vehicle disregards the stop sign and could disrupt traffic pattern detection systems within smart cities. Furthermore, noise-based input attacks may induce malfunctions in smart assistants, resulting in unintended actions (Comiter, 2019).

4.5 Data poisoning attack

Data poisoning represents a form of cyberattack wherein an attacker deliberately undermines a training dataset utilized by an artificial intelligence (AI) or machine learning (ML) model, aiming to alter or control the model's functionality. This malicious act can be executed through various methods, including - The deliberate insertion of inaccurate or deceptive information into the training dataset, Alteration of the existing

dataset, Removal of specific segments of the dataset By tampering with the dataset during the training process, the attacker can instill biases, generate incorrect outputs, introduce vulnerabilities (such as backdoors), or otherwise affect the model's decision-making and predictive abilities. Data poisoning is classified under a broader category of cyberattacks referred to as adversarial AI. Adversarial AI or adversarial ML encompasses any actions intended to disrupt the efficacy of AI/ML systems through manipulation or deception (Lenaerts-Bergmans, 2024).

4.6 Dataset poisoning attacks are effective in their operation

Data poisoning attacks take advantage of the extensive and varied training datasets utilized by artificial intelligence and machine learning models by introducing inaccurate or deceptive information, which can profoundly influence their functionality. This manipulation may occur in a subtle manner, such as making slight modifications to data inputs that gradually diminish the model's performance, or it can be more overt and damaging, designed to create immediate and significant disruptions (Ballejos, 2024).

Cybercriminals employ a range of strategies to introduce inaccuracies into AI models, resulting in compromised decision-making capabilities. Below are several examples of data poisoning attacks: Backdoor (label) poisoning: This method involves inserting data into the training dataset to establish a 'backdoor,' enabling attackers to control the model's output under specific conditions. The attack can be either targeted, where the goal is to induce a particular behavior in the model, or non-targeted, which generally undermines the model's overall performance.

a) Availability attack: This type of attack aims to disrupt the accessibility of systems or services by impairing their performance or functionality, potentially causing system failures or generating erroneous positive or negative results.

b) Model inversion attack: This attack leverages the outputs of the model to deduce or reconstruct the training dataset. It is often perpetrated by insiders who have access to the system's responses.

c) Stealth attack: This approach involves the gradual modification of the training dataset or the covert injection of harmful data to evade detection, resulting in subtle yet significant biases in the model over time.

In addition to these examples of data poisoning attacks, attackers utilize numerous other methods to exploit AI systems, underscoring the necessity of integrating security measures throughout all phases of AI development (Ballejos, 2024).

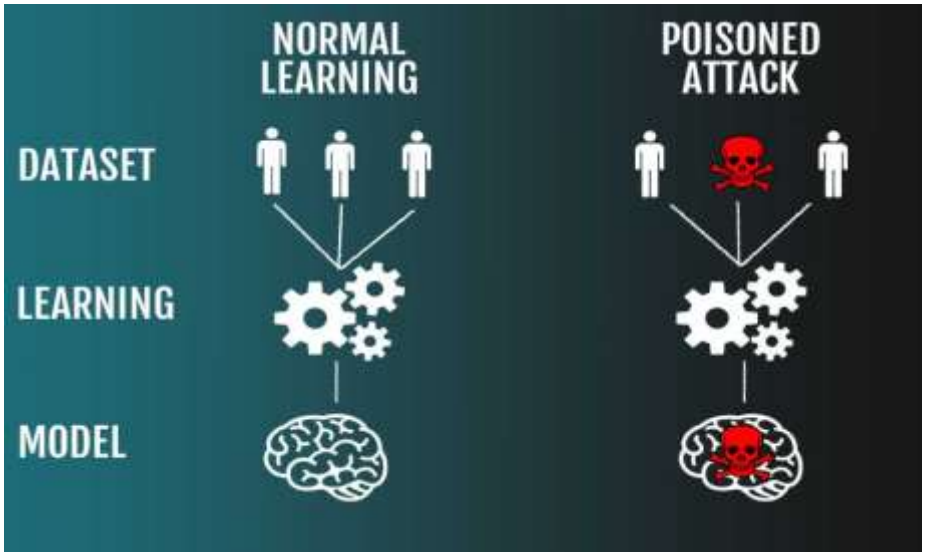


Figure 7. Data Poisoning The Newest Threat in Artificial Intelligence and Machine Learning (Burge, 2024)

5. OVERVIEW OF ASSAULTS AND THEIR COUNTERMEASURES

The different types of attacks examined in this paper are presented in Table 1, along with corresponding strategies for safeguarding an IoT system against these threats. Although achieving comprehensive protection for an IoT system can be difficult due to the multitude of potential attack vectors, several of the methods outlined are effective against various attack types. For instance, since many of the attacks identified begin with man-in-the-middle (MITM) attacks, securing the network that hosts the IoT system will provide defense against numerous prevalent threats.

Table 1.Strategies for safeguarding against IoT attacks and the various techniques employed in such attacks.

IOT ATTACK	METHODS OF PROTECTING AGAINST THE ATTACK
PHYSICAL ATTACKS	Implement tamper-proof hardware, utilize hardware-based security trust anchors (Woo, 2019), incorporate kill commands, and enable self-destruction mechanisms (Hezam Akram Abdul-Ghani, 2018).
MAN-IN-THE-MIDDLE	Regularly perform software updates, ensure appropriate firewall settings, implement robust encryption, and refrain from using unsecured WiFi networks (Zoran Cekerevac, 2017).
BLUETOOTH MITM	Ensure that devices are set to non-discoverable mode, perform regular software updates (Hezam Akram Abdul-Ghani, 2018), restrict access to unknown devices, implement two-factor authentication, and utilize robust pairing techniques, such as Elliptic Curve Diffie-Hellman public-key cryptography or the Out-of-Band method (G. R. Mode, 2020).
FALSE DATA INJECTION	It is essential to conduct regular software updates, implement firewalls with appropriate configurations, refrain from using insecure WiFi networks (Zoran Cekerevac, 2017), utilize anomaly detection techniques, and monitor for any unusual outputs (G. R. Mode, 2020).
BOTNETS	Conduct routine antivirus scans, refrain from opening dubious email attachments or download links, and ensure that regular updates are performed (Boyd, 2024).
MIRAI BOTNET	Conduct routine updates and modify the default login credentials of Internet of Things devices (Hendrickson, 2019).
DENIAL OF SERVICE ATTACKS	Utilize a Denial of Service (DoS) protection service, and implement both antivirus software and a firewall.
BRICKERBOT	Conduct routine updates, reduce the exposure of IoT devices to networks and the internet, implement firewalls, and establish authentication protocols (BrickerBot Malware Emerges, Permanently Bricks IoT Devices, 2017).
DATASET POISONING	Utilize outlier detection techniques, including data sanitization and anomaly detection, and implement micromodels (I, n.d.).

6. CONCLUSION

The inherent characteristics of IoT systems result in numerous potential vulnerabilities, leading to a wide array of attacks targeting these systems, with new threats emerging as the popularity of IoT continues to rise. It is imperative to safeguard these systems against such attacks with the utmost effectiveness. As the frequency and sophistication of attacks increase, professionals are increasingly leveraging artificial intelligence to provide intelligent and real-time protection for these systems. However, it is important to note that attackers are also developing methods to circumvent these AI defenses and may even employ AI technologies to launch their own attacks. This paper examines prevalent techniques used to disrupt or compromise IoT systems and offers a basic overview of how these attacks are executed, supplemented by examples where relevant to enhance understanding. 1. Subsequently, various artificial intelligence algorithms are presented, and their roles in the realm of cybersecurity are examined. In numerous instances, these models have yet to gain traction in commercial applications, as they remain in the stages of research and development or face implementation challenges, rendering them relatively uncommon. However, the models under discussion show significant promise and could potentially evolve into widely adopted attack detection systems within a few years.

The discourse also encompasses methods of attacking AI and employing AI for offensive purposes, particularly within the framework of Internet of Things (IoT) systems. The expansion of IoT systems is likely to exacerbate these types of threats, especially as extensive networks like smart cities begin to experiment; such large-scale networks present greater challenges for protection due to their numerous attack surfaces, and the reliance on AI for daily life and safety necessitates a high degree of reliability. Following this, a chart summarizes the threats addressed in this paper, alongside common or recommended strategies for mitigating each type of attack. By addressing these subjects, this paper aspires to serve as a valuable resource for researchers and cybersecurity experts studying IoT in relation to cybersecurity and AI, with the goal of securing IoT systems. Furthermore, it seeks to highlight the implications of emerging technologies and the reciprocal effects these fields will have on one another. It is crucial to evaluate all potential ramifications of technological advancements both prior to and following their public release, as cybercriminals are perpetually seeking to exploit new technologies for their advantage, whether by repurposing the technology or utilizing it as a means to facilitate other attacks. This paper illustrates instances where IoT and AI have been exploited for illicit purposes or where vulnerabilities have been taken advantage of, thereby aiding readers in comprehending current risks and fostering an awareness that these vulnerabilities must be addressed in the future to avert cyberattacks.

REFERENCES

(n.d.). Retrieved from https://www.imperva.com/learn/application-security/man-in-the-middle-attack-mitm/

A., L. (2020). The role of artificial intelligence in IoT and OT security.

Ballejos, L. (2024, May 28). *Data Poisoning: The Newest Threat in Artificial Intelligence and Machine Learning*. Retrieved from https://www.ninjaone.com/blog/data-poisoning/#:~:text=Data%20poisoning%20in%20AI%20and,predictive%20or%20decision%2Dmaking%20capabilities.

Boyd, S. (2024, July 19). *What Is a Botnet? And How to Protect Yourself in 2024*. Retrieved from https://www.safetydetectives.com/blog/what-is-a-botnet-and-how-to-protect-yourself-in/#review-2

BrickerBot Malware Emerges, Permanently Bricks IoT Devices. (2017, April 19). Retrieved from https://www.trendmicro.com/vinfo/us/security/news/internet-of-things/brickerbot-malware-permanently-bricks-iot-devices

Burge, S. (2024, March 7). *Data Poisoning – A Security Threat in AI & Machine Learning*. Retrieved from https://securityjournalamericas.com/data-poisoning/

Chalee Vorakulpipat, E. R. (2018). Recent challenges, trends, and concerns related to IoT security: An evolutionary study. *Conference: 2018 20th International Conference on Advanced Communication Technology (ICACT)*, (pp. 405–10). Chuncheon-si Gangwon-do, Korea (South);.

Comiter, M. (2019, Augest). *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It*. Retrieved from https://www.belfercenter.org/publication/AttackingAI

D, L. (2019, Nov 15). App nirvana: when the internet of things meets the API economy. Retrieved from https://techbeacon.com/app-dev-testing/app-nirvana-when-internet-things-meets-api-economy

Evans, D. (2011, April). The Internet of Things. *How the Next Evolution of the Internet Is Changing Everything*.

Faezeh Farivar, M. S. (April 2020). Artificial Intelligence for Detection, Estimation, and Compensation of Malicious Attacks in Nonlinear Cyber Physical Systems and Industrial IoT. *IEEE Transactions on Industrial Informatics, vol. 16, no. 4*, (pp. 1551-3203).

Francesca Meneghello, M. C. (2019). IoT: Internet of Threats? A Survey of Practical Security Vulnerabilities in Real IoT Devices. *IEEE Internet of Things Journal*, (pp. 99, 1-1).

G. R. Mode, P. C. (2020). Impact of False Data Injection Attacks on Deep Learning Enabled Predictive Analytics. *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, (pp. 1-7). Budapest, Hungary.

Hendrickson, J. (2019, March 19). *What Is the Mirai Botnet, and How Can I Protect My Devices?* Retrieved from https://www.howtogeek.com/408036/what-is-the-mirai-botnet-and-how-can-i-protect-my-devices/

Herberger, C. (2015, October 16). *DDoS Fire & Forget: PDoS – A Permanent Denial of Service*. Retrieved from https://www.radware.com/blog/security/2015/10/ddos-fire-forget-pdos-a-permanent-denial-of-service/

Hezam Akram Abdul-Ghani, D. K. (2018). A Comprehensive IoT Attacks Survey based on a Building-blocked Reference Model. *International Journal of Advanced Computer Science and Applications Vol. 9(No. 3)*.

I, M. (n.d.). *Poisoning attacks on machine learning. Towards data science, medium*.

I. Gulatas, H. H. (2023). Malware Threat on Edge/Fog Computing Environments From Internet of Things Devices Perspective. *in IEEE Access, 11*, 33584-33606.

Janice Canedo, A. S. (2016). Using machine learning to secure IoT system. *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, (pp. 219-222). Auckland, New Zealand.

Jeesoo Jurn, T. K. (2015). An Automated Vulnerability Detection and Remediation Method for Software Security. 1652.

Lenaerts-Bergmans, B. (2024, March 20). *Data Poisoning The Exploitation of Generative AI*. Retrieved from https://www.crowdstrike.com/cybersecurity-101/cyberattacks/data-poisoning/

M., R. (2020). What is IoT (Internet of Things) and how does it work? IoT Agenda.

MELAMED, T. (2018). An active man-in-the-middle attack on bluetooth smart devices. *8*, p. 11. Retrieved from https://www.witpress.com/elibrary/sse-volumes/8/2/2120

Michele De Donno, N. D. (2017). Analysis of DDoS-Capable IoT Malwares. *2017 Federated Conference on Computer Science and Information Systems*, (pp. 807–816).

Monika Roopak, G. Y. (n.d.). Deep Learning Models for Cyber Security in IoT Networks. *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, (pp. 0452-0457). Las Vegas, NV, USA.

Murat Kuzlu, C. F. (2021, February 24). Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity. 8.

Nguyen, H. (n.d.). *Understanding the Mirai Botnet Attack Type*. Retrieved from https://www.corero.com/mirai-botnet-ddos-attack-type/

Pendse, A. (2019, Feb 11). *Transforming cybersecurity with AI and ML: View*. Retrieved from https://ciso.economictimes.indiatimes.com/news/transforming-cybersecurity-with-ai-and-ml/67899197: https://ciso.economictimes.indiatimes.com/news/transforming-cybersecurity-with-ai-and-ml/67899197

Petar Radanliev, D. C. (2020). Future developments in standardisation of cyber risk in the Internet of Things (IoT). 169.

Ren, K. (2016, March 29). *Bluetooth Low Energy*. Retrieved from Bluetooth Pairing Part 1 –Pairing Feature Exchange: https://www.bluetooth.com/blog/bluetooth-pairing-part-1-pairing-feature-exchange/

Sawla, S. (2018, June 8). *K-Nearest Neighbors*. Retrieved from https://medium.com/@srishtisawla/k-nearest-neighbors-f77f6ee6b7f5

Shen Wang, Z. Q. (2019). Robust Pervasive Detection for Adversarial Samples of Artificial Intelligence in IoT Environments. *7*, pp. 88693-88704.

Sherali Zeadally, E. A. (2020). Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity. *in IEEE Access, 8*, 23817-23837.

Spajic, D. J. (2023, July 14). What is a Botnet? Cyberattack Technique Explained.

Woo, S. (2019, June 13). *The Right Security For IoT: Physical Attacks and How to Counter Them*. Retrieved from https://iot.electronicsforu.com/headlines/the-right-security-for-iot-physical-attacks-and-how-to-counter-them/

Yang Lu, L. D. (2019, April). Internet of Things (IoT) Cybersecurity Research: A Review of Current Research Topics. *IEEE Internet of Things Journal*.

Zoran Cekerevac, Z. D. (2017, July). INTERNET OF THINGS AND THE MAN-INTHE-MIDDLE ATTACKS – SECURITY AND ECONOMIC RISKS. pp. 15-5.