Endpoint detection and response using machine learning

Harmionee Kaur¹, Dr Richa Tiwari²

¹,²Computer Science Engineering, Amity School of Engineering and Technology, Amity University
Uttar Pradesh, Noida, India 201313
harmioneekaur@gmail.com

Abstract. The need for cybersecurity has increased manifold over the past decade due to an unprecedented shift towards digital. With the increase in the number and sophistication of threats, cybersecurity experts have been forced to seek out new and efficient ways to secure endpoints on a network. Machine learning provides one such solution. This paper discusses how IoT devices are threatened and the need for endpoint security. It overviews different Machine learning-based intrusion detection systems that are currently in use e.g., STAT, Haystack, etc., and other Endpoint Detection and Response Techniques.

1. Introduction

The use of IoT (Internet of Things) devices such as smart TV, speakers, toys, and appliances, etc. has become increasingly popular in home-automation, environment monitoring, and even in the healthcare industry. IoT devices are an integral part of our day-to-day life. The popularity of BYOD (Bring Your Own Device) has been increasing and hence, the number of individual devices that are connected to an organization's network are also increasing. The number of devices connected to the IP network is estimated to become more than three times the population of the world by 2023 according to the annual internet report by Cisco. The use of IoT devices has significant potential in bringing economic benefits but at the same time it opens the shop floor layer of a company to more threats which can have an adverse operational impact.

Any device connected to a network is known as an endpoint. Each endpoint may cause a threat to the network and could lead to leakage of data. It’s impossible to list all the devices connected to a network and therefore systems are put in place to protect them. It takes both technical and behavioural techniques to secure networks. Protecting computer networks that are connected remotely to client devices is known as Endpoint security. Endpoint security has developed quite a lot in the last few years moving towards a more comprehensive, advanced, and proactive defence instead of antivirus software. There are many security approaches for endpoint security issues. The main amongst them is isolation of systems that hold confidential data or payment terminals, lightweight authentication protocols, real-time detection and filtering, detection, and modification, and other pre-emptive security measures. [1][2]

The term endpoint detection and response (EDR) was coined by A. Chuvakin in 2013. EDR is an endpoint security mechanism that should be able to protect in real-time adequately. Endpoint Detection systems collect data from endpoints and store and process them in a central database. Then these collected events are correlated in real-time to detect behavioural anomalies in the host. Therefore, EDR systems alert the user, and the emergency response teams of cyber threats.[3][4]
2. Threats to IOT

IoT security includes authorization, access control, authentication, integrity, privacy, and anonymity. It is an integral part of designing IoT systems. However, Low-end IoT technology does not uphold good security standards leading to leakage of information, DoS attacks (Denial of Service Attacks), etc. According to the NIST cybersecurity framework, detection is deployed based on potential threat patterns and on the risks identified. The following reports current vulnerabilities of commercial IoT devices and the security mechanisms used by popular IoT communication protocols. Fig. 1 shows the various threats being discussed in this review.[5][6].

2.1. Sensor-based threats

There has been an emergence of attacks that exploit the use of sensors. This is due to the lack of security measures put in place to control the use of sensors by applications. An attacker can extract information, trigger malicious activities, or transfer malware by abusing the sensor on an IoT device. The leakage of information is the most prominent threat where sensor-based threats are concerned. Crucial data like passwords, credit card information, the secret key of a cryptographic system etc. can be leaked due to sensors on an IoT device. This leaked information can provide a base for future attacks or can be used to violate user privacy. LPPMs (Location-Privacy Preserving Mechanisms) provide a solution to reduce the success of getting location data.[7]

2.2. Telnet Attacks

Telnet (a client-server protocol) based attacks have been increasing since 2014. Pa et al. have made use of IoT sandbox and honeypot to attract Telnet-based attacks against IoT devices that are running on different CPU architectures e.g., ARM, MIPS, PPC and are vulnerable to Telnet-based attacks. They have analysed the captured malware samples using honeypot and gathered 5 DDoS malware families which are targeting IoT devices that are Telnet-based. Amongst them, there is one family which can attack as many as 9 CPU architectures.[8]

2.3. Authentication Attacks

Cloning of tags or signal replaying can enable attackers to get access to sensitive data and services. Usually, organisations have a private and public network. Per access rights, these are exposed to personnel, customers, partners, or suppliers. IoT systems can be reverse engineered to collect information from network and hardware or software parts. This type of attack can be prevented by cryptography, automatic malware detection, resistance to cloning.[9]

2.4. PCN Threats

Organizations need to deploy proactive endpoint protection that is built specifically for PCNs (Primary Care Networks). Most of the EDR tool vendors are corporate IT-based. They are created for IT servers, workstations, routers, and switches. These systems comprise of only 20 percent of assets in PCNs. To detect changes and protect the endpoints configuration, configuration backups, vulnerability management, and inventory should be used. Incident workflows should drive remediation after deviations and potential compromises have been identified. Industrial cybersecurity leaders are majorly concerned with ICS endpoints that control production. If we fail to reduce these attacks it will badly affect the brand, the safety and may lead to unintended configurations, and malicious changes.[10]

2.5. Port Scanning

Port Scanning is the penetration technique used by attackers when trying to steal data. It is a method of determining which ports on a network are open and could be receiving or sending data. In Port Scanning data packets are sent to specific ports to analyse the vulnerabilities in a network. Usually, firewalls and IDS (Intrusion Detection Systems) can detect such activities but now attackers have started dodging these detection measures. To prevent this Port Scan alerts should be used to monitor traffic to your ports and ensure cyber-attackers do not detect potential vulnerabilities to gain unauthorized entry into your network.[11][12]
2.6. APTs
The term advanced persistent threat (APT) is used to describe an attack in which the attacker maintains stealth and long-term persistence on a user’s network which results in impeding detection and event correlation. The main goal is to disrupt services or steal user data. APTs differ from the traditional hit and run attacks and can span up to years. The attacks are launched by high-skilled groups, which are either nation-state or state-sponsored. APTs hold the power to sway politics in some countries. According to a report by George Karantzas, even the best of the EDRs cannot prevent and log a huge bulk of attacks. Attackers also employ methods that could tamper with telemetry providers of EDRs which enables them to attack more stealthily.[13][14]

2.7. Malware
There are 3 million new malware samples detected every hour. Attackers have evolved to use file-less malware, ransomware, and cryptocurrency malware. There has been an emergence of intelligent malware that can bypass endpoint detection systems and uses AI to alter its signature, regulate its activities, generate lures, self-propagate, strategically deliver other malware, and maximize its damage while minimizing its footprint. Traditional malware detection systems cannot keep up with new threats and attacks. Cybersecurity needs a quantum jump forward. AI-based security has also been deteriorating because of silver-bullet vendors which provide fake AI solutions based on imprecise algorithms. Signature or heuristic-based cybersecurity infrastructure can be used to counter intelligent malware.[15]

2.8. Ransomware
According to Sun-Jin Lee there has been an increase in the damage caused by ransomware in the years 2014 to 2018. Ransomware is malware that employs encryption to hold a victim’s information at ransom. A user or organization’s critical data is encrypted so that they cannot access files, databases, or applications. A ransom is then demanded to provide access. It spreads quickly in APT (advanced persistent threat attacks), and its range is increasing to public institutions and government Organisations. In recent times cyberattacks caused by AI using APT have been progressing. Ransomware is evolving to attack several users simultaneously on the same network. EDR tools can extract and respond to ransomware attacks efficiently and quickly. [16][17]

![Figure 1. Threats to IoT versus the no. of articles that provide their respective solutions in this review.](image-url)
3. Machine Learning Based EDR Techniques

3.1. Why do we need Machine Learning based Solutions?

There has been an increase in the use of machine learning in the past few years. It has moved from laboratories and is now an integral part of the industry. Google, Microsoft, Facebook use Machine Learning to improve customer experience, connect people with applications, and forge new personal connections. Cybersecurity is also a field that makes use of machine learning. It can significantly change the dynamic of cyberspace. Attackers set up beachheads for future attacks and prey on sensitive data. They sell network access as a service and transition onto networks whose access they have gathered from the breach after exhausting the data on the network. New techniques involving machine learning can be used to combat this problem. It increases efficiency, lowers cost, and is more reliable. It can be used for recognizing threats and advanced targeting. For example, it could target infrastructure vulnerabilities, interdependent vulnerabilities, exploits, etc. [18]

3.2. Intrusion Detection Systems

IDS (Intrusion detection systems) are becoming more important in recent times due to the increasing number of sophisticated and intelligent threats and malicious activity in cyberspace. The following are some IDS solutions that are currently being used in the industry. A comparison has been drawn between them based on their respective domains and the datasets that have been used to verify them (In table 1).[19]

3.2.1. DIDS (Distributed Intrusion detection system). With an increase in the connectivity of computer systems, networks have become more vulnerable and susceptible to attacks. It has become very easy for attackers to avoid detection. Heterogeneous computer networks only add to this problem. Steven R. Snapps proposed DIDS which aims to combine data reduction (by use of Lan monitors), centralized data analysis, and data monitoring to monitor a heterogeneous network. It tackles NUIP (network user identification problem) which involves tracking a user through a network with a new user-id on each system. Results have shown that DIDS deals with this problem quite efficiently.[20]

3.2.2. USTAT. It is a state transition analysis tool for UNIX. It is a real-time intrusion detection tool developed by A Porras and R.A Kemmerer. It provides a new approach to identify system penetration. STAT identifies a sequence of state changes from the initial state to the compromised state. The very first USTAT prototype made use of SunOS 4.1.1, which uses audit trails collected by its C2 basic model and keeps track of only critical actions that are necessary for successful completion of penetration. It differs from other IDS systems as it employs pattern matching on the audit logs. [21]

3.2.3. STL. Here, STL (self-taught learning) has been used to propose an intrusion detection system. Unlabelled network traffic data from different networks and a good feature representation from these datasets can be obtained by the use of deep learning techniques. Using these features, we can apply supervised classification to labelled traffic dataset consisting of normal as well as anomalous data traffic records. The traffic data for the labelled dataset can be collected in an isolated, confined, and private network. This is the basis for Self-taught learning. This method uses STL on NSL-KDD dataset (the benchmark for network intrusion). This article compares the performance of the proposed system concerning past IDS. It has been compared for precision, accuracy, and f-measure values. [22]

3.2.4. Worm detection. A worm virus is a malicious and self-propagating program that can spread throughout a network without human help. UC Davis et al. have proposed a method to detect such viruses. They discuss the basic architecture of a network-level IDS. Its purpose is to monitor base level information in network packets which include packet size, time, packet source, etc. It learns the normal behaviour of the users in the network and makes use of GA (Genetic Algorithm) to flag anomalies as they occur. They have improved upon current IDS to detect network-level intrusions. [23]
3.2.5. **MINDS (Minnesota Intrusion Detection System).** It makes use of data mining techniques to automatically detect a system or network. MINDS focuses on two main solutions. First, it uses behaviour-based anomaly detection for detecting new and unknown threats. Its algorithm assigns a score to each connection found based on the chances of it being an intrusion. Experimental results in the University of Minnesota indicate successful intrusion detection that makes use of tools such as SNORT in identifying previously unknown threats. These threats were present in the CERT/CC list of recent advisories. Second, it focuses on how association pattern analysis can be used to characterize and summarize anomalous network connections. Characterization of attacks is essential to recognize emerging threats due to the large volume of connections observed. [24]

3.2.6. **Haystack.** Haystack is used by the Air force for multi-user networks. It is an IDS (Intrusion detection system). It was created to detect intrusions and malicious activity (even from the inside i.e., authorized users) for SSO (system security officer). It summarizes voluminous system user audit trails to behavioural, anomalous events and security incidents. Haystack works under behavioural constraints and models that specify behaviour by individual users and groups.[25]

3.2.7. **NetSTAT.** Network-based attacks are constantly increasing and are becoming more intelligent and sophisticated. Giovanni Vigna et al. discuss a different approach to IDS which is STAT (State Transition Analysis Technique). IDS are slowly moving towards more comprehensive network security instead of security for the host itself. Security in a network is more difficult because of the voluminous data that comes with auditing due to several events being linked to one intrusion. In STAT network-based intrusions are modelled by a state-transition diagram in which state and transition depend on the network environment. The target network is generally represented by a hypergraph. With the help of this graph, we can determine which network needs to be protected and which attacks have been detected and accordingly network intrusion detection components are deployed automatically.[26]

3.2.8. **DT-SVM.** An intrusion detection system monitors a network or a system for intrusion and malicious activities. Here we combine Decision trees (DT) and support vector machines (SVM) and create an intelligent intrusion detection having an ensemble approach where we combine the base classifiers. This algorithm maximizes detection accuracy and minimizes complexity by combining individual base classifiers and other hybrid machine learning techniques. Results show that the proposed method for IDS is highly accurate.[27]

3.2.9. **GA-IDS.** Genetic Algorithm (GA) has been applied for IDS to detect variable network intrusions. In recent times it has become imperative to maintain a high level of security between organizations. But secure communication has been threatened by network intrusions. Therefore, Intrusion detection systems are integral to every organization. So far, no intrusion detection system has been developed that is completely flawless. Here it has been approached by GA and its parameters and features have been discussed. To reduce the complexity of the algorithm used, Evolution theory of information has been used to filter the data traffic. KDD-99 benchmark dataset was used which provided satisfying results and a good detection rate. [28]

3.2.10. **NIDS-HIDS.** Another solution to the diversity of cyber-attackers and the frequently changing data is using an efficient neural-network-based hybrid IDS framework that combines both HIDS (Host-level IDS) and NIDS (Network-level IDS). A network intrusion detection system (NIDS) is a tool that analyses network traffic patterns to identify intrusions for an entire network. It needs to be present at a point where all network traffic is concentrated. A viable place for this is the DMZ. An HIDS usually collects operating system information on a local host computer. HIDS analyses the system logs and command histories to extract useful information. AutoEncoders (AE) has been applied to NIDS and HIDS using word embedding and a convolutional neural network. To evaluate this method experiments have been performed using the NSL-KDD and ADFA datasets.[29]

3.2.11. **AutoEncoder and ID-CNN.** Malicious behaviour like Advanced Persistent Threat (APT) attack does not occur immediately but occurs over a long period. Chanwoong Hwang et al. conducted an experiment where they train a dataset that has been collected over a month in a commercial
endpoint environment provided by G* corporation. In this paper, they propose a technique that detects an unknown attack using an event log without prior knowledge, although the initial response failed with anti-virus. This IDS uses a combination of AutoEncoder and 1D CNN (1-Dimension Convolutional Neural Network) which is predicated on semi-supervised learning. 37 previously unknown attacks were detected within the commercial endpoint environment. 26 of these threats were declared to be malicious by Virus Total (VT). Its future scope includes application to EDR technology to form a secure endpoint environment and reduce time and labour that are required to effectively detect unknown attacks. [30]

### Table 1. Intrusion Detection Systems

| S. No. | Name     | Domain | Dataset               |
|--------|----------|--------|-----------------------|
| 1      | DIDS     | Industry | Sun C2                |
| 2      | USTAT    | Industry | Sun C2                |
| 3      | STL      | Industry | NSL-KDD               |
| 4      | MINDS    | -       | Netflow version 5     |
| 5      | Haystack | Air Force | Audit Trails/Session Histories |
| 6      | DT-SVM   | -       | KD CUP-99             |
| 7      | GA-IDS   | Industry | NSL-KDD               |
| 8      | AutoEncoder 1D-CNN | Industry | G* Corp Event Logs    |

3.3. **Event Management**

Traditional Cybersecurity mainly focuses on taking proactive measures and deploying threat prevention mechanisms, and event management is usually neglected.

3.3.1. **SIEM (Security Information and Event Management).** Heterogeneous organizations need network access through cloud services (for example Amazon Web Services), extranet and home working, etc. Security information and event management (SIEM) technology provides threat detection and incident management through collecting and analysing (both real-time and historical) security events, as well as other event logs and contextual data sources. Its attributes include log event collection and management, analysing log events and other data across disparate sources, and operational capabilities (such as incident management, dashboards, and reporting).[31]

3.4. **Human Security**

In recent times, researchers have started focusing on User Behaviour Analytics (UBA). While developing their security against external threats, organizations must also defend themselves against threats that are present within the organisation. Most organizations are unaware of the threats active in their networks while only some detected breaches. Threats can be caused, for instance, by an employee or an external actor, capable of causing damage in some capacity due to negligent behaviour. Data leaks have become very prominent in the industry. Endpoints used by the employees (mobiles). Training sessions and awareness about protecting data have not proved to be effective. Using third-party cloud services also makes it difficult to secure data. Such threats are challenging to detect and cause damage to an organization’s property, its intangible assets and consumer confidence and the organization’s brand or reputation.[32]

**Conclusion**

This paper has examined different methodologies that are currently being used for Endpoint Detection and response. It discusses how behavioural and pattern analysis can be used to characterize and summarize incidents. Machine Learning based Intrusion detection systems have been proved to be effective against APTs, DIDS, Ransomware, etc. However, as cybercriminals grow more intelligent, current security techniques are being classified as inadequate and researchers are looking for new ways to combat these rapidly emerging threats by employing more sophisticated ML algorithms.

**References**
[1] Slate S, "Endpoint Security: An Overview and a Look into the Future." (2018).
[2] Chandel S, Sun Y, Tang Y, Zhou Z, and Huang Y. "Endpoint protection: Measuring the effectiveness of remediation technologies and methodologies for insider threat." In 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 81-89. IEEE, 2019.
[3] Sjarif NN, Chuprat S, Maharin MN, Ahmad NA, Ariffin A, Senan FM, Zamani NA and Saupi A, 2019, October. Endpoint Detection and Response: Why Use Machine Learning? In 2019 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 283-288). IEEE.
[4] Bresniker K, Gavrilovska A, Holt J, Milojicic D and Tran T, "Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity," in Computer, vol. 52, no. 12, pp. 45-52, Dec. 2019.
[5] Fraley JB, and Cannady J. "The promise of machine learning in cybersecurity." SoutheastCon 2017. IEEE, 2017.
[6] Meneghelli F, Calore M, Zucchetto D, Polese M, and Zanella A. "IoT: Internet of threats? A survey of practical security vulnerabilities in real IoT devices." IEEE Internet of Things Journal 6, no. 5 (2019): 8182-8201.
[7] Thompson EC. "Detection and Identification of Events." In Cybersecurity Incident Response, pp. 87-98. Apress, Berkeley, CA, 2018.
[8] Sikder AK, Petracca G, Aksu H, Jaeger T, and Selcuk Uluagac A. "A survey on sensor-based threats to internet-of-things (IoT) devices and applications." arXiv preprint arXiv:1802.02041 (2018).
[9] Pa YM, Suzuki S, Yoshioka K, Matsumoto T, Kasama T, and Rossov C. "IoTPOT: A novel honeypot for revealing current IoT threats." Journal of Information Processing 24, no. 3 (2016): 522-533.
[10] Tedeschi S, Emmanouilidis C, Mehnen J, and Roy R. "A design approach to IoT endpoint security for production machinery monitoring." Sensors 19, no. 10 (2019): 2355.
[11] Hollis S, and Zahn D. "ICS Cybersecurity: Protecting the Industrial Endpoints That Matter Most." (2017).
[12] Kumar MS, Ben-Othman J, Srinivasagan KG and Krishnan GU. "Artificial Intelligence Managed Network Defense System against Port Scanning Outbreaks." In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (VITECoN), pp. 1-5. IEEE, 2019.
[13] Avast <https://www.avast.com/en-in/business/resources/what-is-port-scanning#mac>
[14] Karantzas G, and Patsakis C. "An Empirical Assessment of Endpoint Detection and Response Systems against Advanced Persistent Threats Attack Vectors." Journal of Cybersecurity and Privacy 1, no. 3 (2021): 387-421.
[15] Liggett T. "Evolution of endpoint detection and response platforms." PhD diss., Utica College, 2018.
[16] Scott J. "Signature based malware detection is dead." Institute for Critical Infrastructure Technology (2017).
[17] Lee SJ, Shim HY, Lee YR, Park TR, Park SH, and Lee IG. "Study on Systematic Ransomware Detection Techniques." In 2021 23rd International Conference on Advanced Communication Technology (ICACT), pp. 297-301. IEEE, 2021
[18] Mcfee 1977 <www.mcafee.com/enterprise/en-in/security-awareness/ransomware.html#:~:text=Ransomware%20is%20malware%20that%20employs,a%20victims%20information%20at %20ransom.&text=A%20ransom%20is%20then%20demanded,quickly%20paralyze%20an %20entire%20organization.>
[19] Liao HJ, et al. "Intrusion detection system: A comprehensive review." Journal of Network and Computer Applications 36.1 (2013): 16-24.
[20] Snapp SR, Brentano J, Dias G, Goal TL, Heberlein LT, Ho CL and Levitt KN. "DIDS (distributed intrusion detection system)-motivation, architecture, and an early prototype."(1991).
[21] Ilgun K. "USTAT: A real-time intrusion detection system for UNIX." In Proceedings 1993 IEEE Computer Society Symposium on Research in Security and Privacy, pp. 16-28. IEEE, 1993.
[22] Javaid A, Niyaz Q, Sun W, and Alam M. "A deep learning approach for network intrusion detection system." EAI Endorsed Transactions on Security and Safety 3, no. 9 (2016): e2.

[23] Heady R, Luger G, Maccabe A, and Servilla M. The architecture of a network level intrusion detection system. No. LA-SUB-93-219. Los Alamos National Lab., NM (United States); New Mexico Univ., Albuquerque, NM (United States). Dept. of Computer Science, 1990.

[24] Ertöz L, Elertson E, Lazarevic A, Tan PN, Kumar V, Srivastava J and Paul Dokas. "Minds-Minnesota intrusion detection system." Next generation data mining (2004): 199-218.

[25] Smaha SE. "Haystack: An intrusion detection system." In Fourth Aerospace Computer Security Applications Conference, vol. 44. 1988

[26] Vigna G and Kemmerer RA "NetSTAT: A network-based intrusion detection system." Journal of computer security 7, no. 1 (1999): 37-71.

[27] Peddabachigar S, Abraham A, Grosan C and Thomas J. "Modeling intrusion detection system using hybrid intelligent systems." Journal of network and computer applications 30, no. 1 (2007): 114-132.

[28] Hoque SM, Mukit B and Naser A. "An implementation of intrusion detection system using genetic algorithm." arXiv preprint arXiv:1204.1336 (2012).

[29] Liu J, Xiao K, Luo L, Li Y and Chen L. "An intrusion detection system integrating network-level intrusion detection and host-level intrusion detection." In 2020 IEEE 20th International Conference on Software Quality, Reliability and Security (QRS), pp. 122-129. IEEE, 2020.

[30] Hwang C, Kim D and Lee T. "Semi-supervised based Unknown Attack Detection in EDR Environment." KSII Transactions on Internet & Information Systems 14, no. 12 (2020).

[31] Li Q, Zheng J, Tsai A and Zhou Q. "Robust endpoint detection and energy normalization for real-time speech and speaker recognition." IEEE Transactions on Speech and Audio Processing 10, no. 3 (2002): 146-157.

[32] Vähäkainu P and Lehto M. "Artificial intelligence in the cyber security environment." In ICCWS 2019 14th International Conference on Cyber Warfare and Security: ICCWS 2019, p. 431. Academic Conferences and publishing limited, 2019.