

## **NETWORK THREAT DETECTION USING MACHINE/DEEP LEARNING IN SDN-BASED PLATFORMS: A COMPREHENSIVE ANALYSIS OF STATE-OF-THE-ART SOLUTIONS, DISCUSSION, CHALLENGES, AND FUTURE RESEARCH DIRECTION**

**BHANU PRASAD<sup>1</sup>, A. AKHILA<sup>2</sup>, BHUMIKA.P<sup>3</sup>, CHARISHMA<sup>4</sup>**  
**1ASSISTANT PROFESSOR, DEPARTMENT OF CSE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD.**  
**2,3&4UG SCHOLAR, DEPARTMENT OF CSE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN, HYDERABAD**

**ABSTRACT:** A revolution in network technology has been ushered in by software defined networking (SDN), which makes it possible to control the network from a central location and provides an overview of the network's security. Despite this, SDN has a single point of failure that increases the risk of potential threats. Network intrusion detection systems (NIDS) prevent intrusions into a network and preserve the network's integrity, availability, and confidentiality. Much work has been done on NIDS but there are still improvements needed in reducing false alarms and increasing threat detection accuracy. Recently advanced approaches such as deep learning (DL) and machine learning (ML) have been implemented in SDN-based NIDS to overcome the security issues within a network. In the first part of this survey paper, we offer an introduction to the NIDS theory, as well as recent research that has been conducted on the topic. After that, we conduct a thorough analysis of the most recent ML- and DL-based NIDS approaches to ensure reliable identification of potential security risks. Finally, we focus on the opportunities and difficulties that lie ahead for future research on SDN-based ML and DL for NIDS.

**Keywords:** software defined network; intrusion detection systems; machine learning; deep learning; security attacks

**INTRODUCTION** Over the last two decades, network technologies have tremendously improved; at the same time, network security threats have also increased. Web-based security attacks, denial-of-service (DoS), and malicious insiders are a few examples that cause the devastating cybercrimes. With such malicious activities, critical disruptions can occur within a network. To ensure network security, antivirus software, firewalls, and network intrusion detection systems (NIDS) can be deployed. Among these, NIDS is broadly used for detecting intruders within a network by continuously monitoring the network traffic for any suspicious and malicious behavior. NIDS is useful to detect different kinds of network threats, including distributed denial-of-service (DDoS) attacks, worms, and viruses. Reliability, accuracy, and detection speed are the success factors of NIDS. Enormous research work has been done on NIDS, but it still requires improvements in reducing false alarm and increasing detection accuracy. To reduce false alarm rate [1] and increase threat detection accuracy [2], different approaches of machine learning (ML) have been used in NIDS. The advanced type of ML that is deep learning (DL) is also used in developing a more advanced field of NIDS. Software defined networking (SDN) has revolutionized network technology in recent years. In contrast to a traditional network, SDN decouples the control plane and data plane of a network switch. In SDN, the control plane is moved to a remote controller (server), which can add packet forwarding rules in network switches according to a given program. This central control of a network offers more programmability and visibility compared with a traditional network. In addition, it is also attractive from a network security perspective, as

having central control can offer better network monitoring [3]. In SDN, innovative network applications can be developed to monitor and control the network. In this regard, NIDS is extended for SDN-based architecture. To enhance network security and traffic monitoring, different approaches of ML/DL can be implemented in the controller of SDN. From the past few years, the invention of graphics processor units (GPUs) has increased the popularity of ML/DL approaches in network security. Both ML and DL techniques are very efficient at predicting any malicious or suspicious behavior from network traffic, as they can extract and learn new features from network traffic. ML-based NIDS heavily depend upon the learned features from network traffic, whereas DL-based NIDS automatically learn from the raw data of complex features and do not rely on learned features [4]. Many researchers have worked on ML- and DL-based NIDS in order to improve its performance in detecting network intruders. However, in larger networks, security threats are also increased due to increased network traffic, which affects the efficiency of NIDS in detecting malicious activities. Very few studies have been conducted on developing SDN-based NIDS systems through DL approaches so that there is enough room to deploy these techniques for improving detection efficiency of intrusions within a network. The basic purpose of this review paper is to comprehensively review the current advancements and trends in ML/DL-based NIDS systems and, more significantly, to provide an overview of the work on SDN-based NIDS systems using ML/DL approaches. This paper covers the area of knowledge for people with basic to moderate knowledge related to ML and DL for network threat detection in SDN. The motivation is to provide an overall picture of the existing research outcomes in this area. In addition, we aim to identify future research directions that may be useful for new researchers who are interested in this field of study. We analyzed the scientific research carried out on network threat detection (NTD) in SDN, based on ML and DL mechanisms. We covered the main and sub-parts of the NTD paradigm to efficiently cover threat issues and their protection using ML and DL approaches to avoid adversary attacks and protect sensitive information during storage and transmission on a public network. There were various review papers covering different aspects of this domain, leveraging ML/DL approaches [5–9]. Much research work has been done on NIDS using ML and DL approaches, but there are few studies on SDN-based NIDS and very few on DL-based NIDS systems in SDN. We believe our study is different from existing studies for the following reasons (and thus, are the main contributions of our work):

- First, we conducted a comprehensive review on ML/DL-based network intrusion detection systems;
- Second, we reviewed each study on SDN-based NID systems using ML and DL algorithms;
- We also explored recent advancements and trends in ML/DL approaches for NIDS, followed by the NIDS system leveraging SDN using ML/DL approaches, and research issues in NID systems using ML/DL approaches.

**RELATED WORK** Several papers have investigated the cybersecurity issues in the healthcare sector. Some of them are listed in [1], [9]–[13]. In particular, in [1], T. Yaqoob et al. investigate the vulnerabilities of the smart medical devices and propose appropriate countermeasures. In [9], S. Chentharra et al. discuss the cybersecurity and privacy challenges of the e-health solutions in cloud-computing environments. Similarly, in [10] S. Wolker-Roberts et al. discuss relevant countermeasures against internal threats in healthcare CIs. In [11], P.Vijayakumar et al. provide an anonymous authentication framework for Wireless Body Area Networks (WBANs). Finally, in [12] Y. Sun et al. provide a detailed survey about the IoMT security and privacy issues. Next, we elaborate on some similar works regarding (a) IEC 60870-5-104 threat modelling, (b) detecting intrusions against IEC 60870-5-104 and (c) mitigating or even preventing cyberattacks through SDN. In [5], the authors conduct an abstract threat analysis of the IEC 60870-5-104 industrial systems. Based on a Coloured Petri

Net (CPN) analysis, two cyberattack categories are specified: (a) physical attacks and (b) cyberattacks. The first category denotes those activities performed by an attacker having physical access to the target system. On the other side, the cyberattacks refer to those that exploit the IEC 60870-5-104 vulnerabilities. In particular, based on the authors, the second category includes four kinds: (a) unauthorised access, (b) Main-In-The-Middle (MITM), (c) DoS and (d) traffic analysis. Each of the aforementioned cyberattacks is assigned to the CPN transitions. Next, the authors emulate the four IEC 60870-5-104 cyberattacks and quantify their risk based on the AlienVault OSSIM risk model. In [3], E. Hodo et al. adopt various ML algorithms to detect cyberattacks against an emulated industrial environment using the IEC 60870-5-104 protocol. To this end, the authors use a dataset consisting of (a) replay attacks, (b) DoS attacks and (c) Address Resolution Protocol (ARP) spoofing attacks. Thus, they evaluate the classification performance of various ML classifiers, including Random Forest, OneR, J48, IBk and Naive Bayes. According to the evaluation results, J48 achieves the best performance. In [4], Y. Yang et al. create Snort-compliant signature and specification rules to detect IEC 60870-5-104-related cyberattacks. The difference between the signature and specification rules lies in the fact that the former category defines malicious patterns, while the second determines the normal behaviour. The same authors in [7] introduce a specification-based Intrusion Detection System (IDS) capable of recognising IEC 60870-5-104 anomalies. The proposed IDS relies on a Detection State Machine (DSM), which relies on Finite State Machines (FSM). The experimental results confirm the efficiency of the proposed IDS. In [14], H. Lin introduces an SDN-based in-network honeypot, which can mitigate the impact of a cyberattack by (a) isolating the cyberattacker and (b) spoofing the network communication, thereby establishing a connection with a cyberattacker via non-existent nodes, called phantom nodes. This connection allows the defender to mislead the cyberattacker and gather useful information. Initially, the SDN controller quarantines the malicious nodes by corrupting their communication with any legitimate node. Next, the SDN controller uses spoofed IP addresses that communicate with the cyberattacker by adapting appropriately the network packets' content at the network and application layers. To this end, statistic and physical models are utilised, respectively. In [15], T. Xing et al. present an SDN-based Intrusion Prevention System (IPS) called SDNIPS. The SDNIPS architecture consists of four modules: (a) Snort agent, (b) SDNIPS daemon, (c) alert interpreter and (d) rules generator. The Snort agent is responsible for detecting the potential cyberattacks by applying the respective signature rules. Next, the SDNIPS daemon undertakes to transform the detection results into a (JavaScript Object Notation) JSON format, which is transmitted to the SDN controller. The alert interpreter processes the JSON files, thus extracting the appropriate information, such as the IP addresses. Finally, the rule generator produces the OpenFlow entries introduced into the Open vSwitch flow tables. The authors evaluate their IPS with a typical IPS relying on iptables. The evaluation criterion is whether both IPS can generate alerts under tremendous network traffic conditions. To this end, two DoS attacks are emulated. The proposed IPS exceeds the performance of the typical IPS using iptables. Undoubtedly, the aforementioned works provide useful and significant insights. (b) IEC 60870-5-104 anomaly detection, (c) IEC 60870-5-104 cyberattack discrimination and (d) IEC 60870-5-104 cyberattack mitigation. Apart from the aforementioned works, Table I contains also [6] and [8] that provide an IDS and a Security Information and Event Management (SIEM) system for IEC 60870-5-104, respectively. As depicted, most of the current works cannot discriminate the various IEC 60870-5-104 cyberattacks and mitigate them. In particular, they do not consider (a) the various cyberattacks depending on the IEC 60870-5-104 commands and (b) the sensitive nature of the CIs, such as the industrial healthcare systems. Regarding the first key point, this paper provides a quantitative threat model, taking into account the IEC 60870-5-104 commands. Moreover, the proposed IDPS

can discriminate precisely the various cyberattacks with respect to the IEC 60870-5-104 commands. On the other side, although the existing works demonstrate how SDN can mitigate the possible intrusions, they do not take into account that the automated countermeasures (such as the isolation of the compromised assets in a sensitive environment) can lead to more devastating consequences. To this end, in this paper, we formulate the mitigation decision as a MAB problem, which is solved with the TS method.

### **EXISTING SYSTEM**

- Several papers have investigated the cybersecurity issues in the healthcare sector. Some of them are listed in [1], [9]– [13]. In particular, in [1], Yaqoob et al. investigate the vulnerabilities of the smart medical devices and propose appropriate countermeasures. In [9], Chentharat et al. discuss the cybersecurity and privacy challenges of the e-health solutions in cloud-computing environments. Similarly, Wolker-Roberts et al. [10] discuss relevant countermeasures against internal threats in healthcare CIs.
- Vijayakumar et al. [11] provide an anonymous authentication framework for wireless body area networks. Finally, Sun et al. [12] provide a detailed survey about the IoMT security and privacy issues. Next, we elaborate on some similar works regarding 1) IEC 60 870-5-104 threat modeling, 2) detecting intrusions against IEC 60 870-5- 104, and 3) mitigating or even preventing cyberattacks through SDN.
- In [5], the authors conduct an abstract threat analysis of the IEC 60 870-5-104 industrial systems. Based on a colored Petri net (CPN) analysis, two cyberattack categories are specified: 1) physical attacks and 2) cyberattacks. The first category denotes those activities performed by an attacker having physical access to the target system. On the other side, the cyberattacks refer to those that exploit the IEC 60 870-5-104 vulnerabilities. In particular, based on the authors, the second category includes the following four kinds:
  - 1) unauthorized access;
  - 2) man-in-the-middle (MITM);
  - 3) DoS;
  - 4) traffic analysis.

Each of the aforementioned cyberattacks is assigned to the CPN transitions. Next, the authors emulate the four IEC 60 870- 5-104 cyberattacks and quantify their risk based on the Alien-Vault OSSIM risk model.

- Hodo et al. [3] adopt various ML algorithms to detect cyberattacks against an emulated industrial environment using the IEC 60 870-5-104 protocol. To this end, the authors use a dataset consisting of 1) replay attacks, 2) DoS attacks, and (c) address resolution protocol spoofing attacks. Thus, they evaluate the classification performance of various ML classifiers, including Random Forest, OneR, J48, IBk, and Naive Bayes.
- According to the evaluation results, J48 achieves the best performance. Yang et al. [4] create Snort-compliant signature and specification rules to detect IEC 60 870-5-104-related cyberattacks. The difference between the signature and specification rules lies in the fact that the former category defines malicious patterns, while the second determines the normal behavior. The same authors in [7] introduce a specification-based intrusion detection system (IDS) capable of recognizing IEC 60 870-5-104 anomalies. The proposed IDS relies on a detection state machine, which relies on finite state machines. The experimental results confirm the efficiency of the proposed IDS.

### **Disadvantages**

- The system is not implemented SDN-BASED MITIGATION: PROBLEM FORMULATION AND METHODOLOGY.
- The system is not implemented enough method for testing and training for large datasets.

## PROPOSED SYSTEM

- The proposed IEC 60 870-5-104 threat modeling combines both ADT and CVSS that determine the cyberattack paths and their risks, respectively. In particular, an ADT [16] comprises two antagonistic nodes: 1) attacking nodes and 2) defending nodes. The attacking nodes describe the goal and the actions that a cyberattacker may adopt in order to compromise the security of the target system. The defending nodes correspond to the defences that can be used by the defender in order to address or mitigate a cyberattack.
- Each node can have one or more children of the same type (i.e., attacking node or defending node), thus reflecting a refinement into specific subgoals and actions. If a node does not have any refinement (i.e., children of the same type), then it constitutes a nonrefined node, which indicates a basic action. Moreover, a node can have children of the opposite type, thus defining a countermeasure.
- A refinement can be classified into two types: 1) conjunctive and 2) disjunctive. In the first case (i.e., conjunctive refinement), the goal of a refined node is achieved, whether all of its children accomplish their goals. Thus, a conjunctively refined node is characterized by an AND operator. On the other side, a disjunctively refined node is characterized by an OR operator, i.e., its goal is achieved if at least one of its children achieves its goal. On the other side, CVSS is an open vulnerability assessment framework, which quantifies the severity of each vulnerability or attack between 0 and 10 [17].

### Advantages

- The proposed system implemented 1) detection performance and 2) mitigation performance which are enough operations on datasets.
- The proposed system developed notification and response module (NRM) for datasets prediction.

**Machine Learning and Deep Learning in NIDS** In the areas of artificial intelligence (AI), there are numerous powerful ML techniques evolved and commonly used in data mining, and useful structural patterns and models from the training set are learnt by the system with the help of ML [33]. There are mostly two phases included in the ML technique: (1) Training phase and (2) Decision-making phase (Figure ). In the training stage of the ML technique, training data is used by the applied methods of ML to learn the system model. In the second stage, estimated output can be obtained by the system with the help of a trained model for every new input [34]. There are basically four types of ML algorithms: supervised learning, reinforcement learning, unsupervised learning, and semi-supervised learning, as given in Figure The commonly used techniques of ML [35] are described in the following subsections. Figure shows the processing procedure of ML methods.

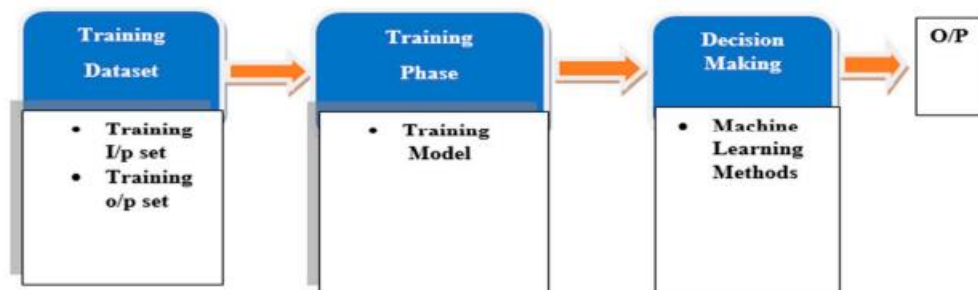


Fig: Processing procedure of the ML method

**Supervised Learning** To anticipate unknown cases, representations are learnt by the algorithms from labelled input data in supervised learning. Examples of this learning include support vector machine (SVM) and random forest methods, where SVM is used for classification-related problems and problems related to classification and regression are dealt

with random forest [36]. In NIDS research, the most commonly used algorithm is SVM, as it is practical in computation and powerful classification abilities. In the case of high dimensional data, SVM algorithms are mostly suitable, but it is very difficult to select an accurate kernel function. Memory and processing units of computation are demanded by SVM [37], so it is resource hungry. In case of uneven data, an effective supervised ML approach involves a random forest algorithm, but it can suffer from the problem of overfitting. **Random Forest** Random decision forest is another name of the random forest method, and tasks related to classification and regression are dealt with this method [39]. Different decision trees are included in the random forest model. Problems of overfitting from the decision-tree method can be mitigated with the construction of decision trees by randomly selecting from the feature support subset; in this way, the accuracy of the model can also be improved. The following steps are included in the random forest method for the classification of a new data sample: (1) Put the sample of data to each tree in the forest. (2) Classification results given by each tree is known as the vote of the tree. (3) The sample of data will be part of the class that receives the most votes.

**Support Vector Machine** Another important technique of supervised learning is SVM, usually used in tasks related to classification and recognition of patterns. Vapnik and others [40] invented SVM; mapping of the  $i/p$  vector into high-dimensional feature support is the main idea behind the SVM. Various kernel functions are applied to achieve mapping, such as radial, polynomial, and linear-based functions. In SVM, selection of kernel function is critical, as it affects the accuracy of classification. The training dataset is the base for the selection of kernel function. If training data is linearly separable, a linear kernel function will be more accurate. Kernel functions for example, Radial based function and polynomial is commonly used in the case when training data is not linearly separable. Generally, accuracy of RBF function is much higher as compared to polynomial and linear kernel functions [41,42]. Further detailed discussion related to SVM classifier is given in [43,44]

**k-Nearest Neighbor** In this supervised learning technique, a sample of data is classified in terms of knearestneighbors (k-NN) of the unclassified sample. The classification is performed on the basis of number k-NN; for example, if a sample has more related k-NN, the sample will be classified in that class. A simple and easy example of k-NN is shown in Figure 5, where the working principle of k-NN is also explained. When the value of k is one, it will be the nearest neighbor algorithm. There is a reduced effect of noise on the classification with a higher value of k. In k-NN algorithms, the main factor is distance, so distance can be defined by various functions between the sample of data that is not labelled and its neighbors, including Euclidean, Chebyshev, Euclidean squared, and City-Block functions. A detailed discussion of k-NN can be found in

**CONCLUSIONS** Despite the necessary digitisation of the healthcare ecosystem, the IoMT progression and mainly the insecure nature of the legacy healthcare systems increase the attack surface. In this paper, we pay our attention to the IEC 60870-5-104 protocol, which is widely adopted by the industrial systems in the healthcare sector. In particular, first, we introduce a quantitative threat model, which evaluates the severity of the possible cyberattacks with respect to the corresponding IEC 60870-5-104 commands. Next, we provide an IDPS system, which combines ML and SDN in order to detect and mitigate the IEC 60870-5-104 cyberattacks. The intrusion detection relies on a CART classifier that uses the TCP/IP network flow statistics and IEC 60870-5-104 payload flow statistics. On the other side, the SDN-based mitigation is transformed into a MAB problem solved with the TS method. The evaluation results demonstrate the efficiency of the proposed IDPS. Our future plans related to this work are focused on enhancing the proposed IDPS so that it can detect multi-step cyberattacks related to IEC 60870-5-104 and other industrial and IoMT protocols

utilised in the healthcare sector, such as Modbus, MQTT and EtherCAT. To this end, ML-based association rules techniques will be adopted.

## REFERENCES

- [1] T. Yaqoob, H. Abbas, and M. Atiquzzaman, "Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices: a review," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3723–3768, 2019.
- [2] M. Conti, D. Donadel, and F. Turrin, "A survey on industrial control system testbeds and datasets for security research," *arXiv preprint arXiv:2102.05631*, 2021.
- [3] E. Hodo, S. Grebeniuk, H. Ruotsalainen, and P. Tavolato, "Anomaly detection for simulated iec-60870-5-104 traffic," in *Proceedings of the 12th International Conference on Availability, Reliability and Security*, 2017, pp. 1–7.
- [4] Y. Yang, K. McLaughlin, T. Littler, S. Sezer, B. Pranggono, and H. Wang, "Intrusion detection system for iec 60870-5-104 based scada networks," in *2013 IEEE power & energy society general meeting*. IEEE, 2013, pp. 1–5.
- [5] P. Radoglou-Grammatikis, P. Sarigiannidis, I. Giannoulakis, E. Kafetzakis, and E. Panaousis, "Attacking iec-60870-5-104 scada systems," in *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642. IEEE, 2019, pp. 41–46.
- [6] P. R. Grammatikis, P. Sarigiannidis, A. Sarigiannidis, D. Margounakis, A. Tsiakalos, and G. Efstathopoulos, "An anomaly detection mechanism for iec 60870-5-104," in *2020 9th International Conference on Modern Circuits and Systems Technologies (MOCASST)*. IEEE, 2020, pp. 1–4.
- [7] Y. Yang, K. McLaughlin, S. Sezer, Y. Yuan, and W. Huang, "Stateful intrusion detection for iec 60870-5-104 scada security," in *2014 IEEE PES General Meeting— Conference & Exposition*. IEEE, 2014, pp. 1–5.
- [8] P. Radoglou-Grammatikis, P. Sarigiannidis, E. Iturbe, E. Rios, S. Martinez, A. Sarigiannidis, G. Efstathopoulos, Y. Spyridis, A. Sesis, N. Vakakis et al., "Spear siem: A security information and event management system for the smart grid," *Computer Networks*, vol. 193, p. 108008, 2021.
- [9] S. Chentharu, K. Ahmed, H. Wang, and F. Whittaker, "Security and privacy-preserving challenges of e-health solutions in cloud computing," *IEEE access*, vol. 7, pp. 74 361–74 382, 2019.
- [10] S. Walker-Roberts, M. Hammoudeh, and A. Dehghantanha, "A systematic review of the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure," *IEEE Access*, vol. 6, pp. 25 167–25 177, 2018.
- [11] P. Vijayakumar, M. S. Obaidat, M. Azees, S. H. Islam, and N. Kumar, "Efficient and secure anonymous authentication with location privacy for iot-based wbans," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2603–2611, 2019.
- [12] Y. Sun, F. P.-W. Lo, and B. Lo, "Security and privacy for the internet of medical things enabled healthcare systems: A survey," *IEEE Access*, vol. 7, pp. 183 339–183 355, 2019.
- [13] S. Meng, W. Huang, X. Yin, M. R. Khosravi, Q. Li, S. Wan, and L. Qi, "Security-aware dynamic scheduling for real-time optimization in cloud-based industrial applications," *IEEE Transactions on Industrial Informatics*, 2020.
- [14] H. Lin, "Sdn-based in-network honeypot: Preemptively disrupt and mislead attacks in iot networks," *arXiv preprint arXiv:1905.13254*, 2019.
- [15] T. Xing, Z. Xiong, D. Huang, and D. Medhi, "Sdnips: Enabling softwaredefinednetworking based intrusion prevention system in clouds," in *10th International Conference on Network and Service Management (CNSM) and Workshop*. IEEE, 2014, pp. 308–311