IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 August 2024, accepted 26 September 2024, date of publication 4 October 2024, date of current version 26 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3473895

RESEARCH ARTICLE

A Novel Dataset for Experimentation With Intrusion Detection Systems in SCADA Networks Using IEC 60870-5-104 Standard

M. AGUS SYAMSUL ARIFIN[1], DERIS STIAWAN[2], BHAKTI YUDHO SUPRAPTO[3], SUSANTO[1], TASMI SALIM[3,4], (Member, IEEE), MOHD YAZID IDRIS[5], MOHAMED SHENIFY[6], (Member, IEEE), AND RAHMAT BUDIARTO[6]

[1]Faculty of Engineering, Universitas Bina Insan, Lubuklinggau 31626, Indonesia
[2]Faculty of Computer Science, Universitas Sriwijaya, Palembang 30119, Indonesia
[3]Faculty of Engineering, Universitas Sriwijaya, Palembang 30119, Indonesia
[4]Faculty of Computer Science, Universitas Indo Global Mandiri, Palembang 30129, Indonesia
[5]Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Johor Bahru, Johor 81310, Malaysia
[6]College of Computing and Information, Al-Baha University, Al Aqiq 65779, Saudi Arabia

Corresponding author: Deris Stiawan (deris@unsri.ac.id)

ABSTRACT Supervisory Control and Data Acquisition (SCADA) systems, particularly Remote Terminal Units (RTUs), are vulnerable to cyber-attacks due to their limited computing resources. This study addresses the need for a reliable, publicly available dataset for comprehensive attack detection experiments in SCADA networks. We developed a dataset for SCADA systems operating under the IEC 60870-5-104 protocol in an electricity distribution network. Using a hybrid virtual-physical testbed that simulates SCADA communications, we generated normal and attack scenarios, including port scans, brute force attacks, ICMP floods, SYN floods, Xmas scans, and IEC 104 floods. Snort and Suricata verified the integrity of the dataset. We then evaluated six Intrusion Detection System (IDS) models using different machine learning algorithms, i.e.: Artificial Neural Network, Categorical Naïve Bayes, Decision Tree, K-Nearest Neighbors, Gradient Boosting, and Random Forest. The Decision Tree and Random Forest models achieved the highest accuracy of 93.66%. This dataset aims to support further research and development of robust IDS solutions for SCADA systems.

INDEX TERMS SCADA, intrusion detection system, IEC 60870-5-104, Snort, Suricata.

I. INTRODUCTION

Supervisory control and data acquisition (SCADA) plays an important role in various industries, handling the automatic control and monitoring of different equipment used in critical infrastructure. SCADA systems are designed for closed networks, where data exchange traffic is carried out on a local network [1]. The isolated network protects the SCADA system from cyber-attacks from the internet. However, an isolated network makes the process of monitoring and controlling industrial equipment inflexible. Attacks against critical infrastructure have increased after the Stuxnet tragedy and have attracted attention from various sectors, including researchers. The interconnection of SCADA systems to open networks is required in economic terms to reduce operational costs because the monitoring and controlling of SCADA devices are done remotely using commonly used protocols, such as TCP/IP. The interconnection of SCADA networks to open networks, especially corporate networks, makes them vulnerable to cyber attacks [2], [3].

This study discusses the malicious activity on SCADA IEC 60870-5-104 (IEC 104), which is widely used in the power plant industry to monitor and control distribution lines [4], which is one of the applications of the use of a smart grid in SCADA systems [5]. To execute relevant malicious activity scenarios, we built a testbed for a SCADA system that

The associate editor coordinating the review of this manuscript and approving it for publication was Leandros Maglaras.

allows normal scenarios and attack scenarios to be carried out. We then used Snort and Suricata to reveal and validate the existence of the attacks on the dataset. The scenarios are carried out using Human Machine Interface (HMI) and remote terminal unit (RTU) physical devices on the SCADA system to obtain relevant data; the virtual devices used are the Historian server and malicious machine for carrying out the attacks. The malicious/attack activities that occur on the SCADA IEC 104 network are port scan, brute force, and Denial of Service (DoS). The DoS attack types are SYN flood, ICMP flood, Xmas, and IEC 104 flood. All traffic data were recorded in pcap format.

In recent years, there has been a significant surge in cyber threats targeting SCADA systems, highlighting the necessity for advanced security measures. Traditional network security solutions often fall short in protecting SCADA environments due to their unique communication protocols and operational requirements. Consequently, there is a pressing need for specialized intrusion detection systems (IDS) that can accurately identify and mitigate threats specific to SCADA networks. This study addresses this gap by developing a comprehensive dataset tailored for SCADA systems using the IEC 60870-5-104 protocol, which is critical for enhancing the effectiveness of IDS in these environments.

In the supervised learning method, the dataset is crucial for developing a reliable and robust IDS model. Specifically, specific traffic data are needed in developing IDS on the SCADA network; these traffic data must be taken from data communication on the SCADA system because SCADA network communication is different from that of traditional computer networks [6].

Existing standard IDSs usually were trained on datasets captured from traditional network traffic. Thus, the IDSs do not perform well when we implement them on the SCADA network traffic, due to different characteristics of the traffic [7]. Publicly available SCADA network traffic datasets are very limited. Thus, researchers face difficulty in obtaining proper datasets for conducting experiments on the accuracy performance, and robustness of their proposed IDSs. Therefore, this study creates a dataset as well as the IDSs using six (6) different machine learning-based classification algorithms i.e.: Artificial Neural Network, Categorical Naïve Bayes, Decision Tree, K-Nearest Neighbors, Gradient Boosting, and Random Forest as the core detection engine. The Random Forest, Gradient Boosting, and categorical Naïve Bayes algorithms have been proven to have excellent performance on conventional network traffic datasets [8].

Moreover, the implementation of machine learning (ML) algorithms in IDS has shown promising results in improving detection accuracy and reducing false positives. The findings from this research provide valuable insights into the strengths and weaknesses of these algorithms when applied to SCADA network traffic, offering a robust foundation for developing more effective and reliable IDS solutions for protecting critical infrastructure.

The main contribution of this study is the provision of a new dataset of SCADA network traffic running the IEC 60870-5-104 protocol, allowing researchers to test the robustness of their proposed IDSs. More specifically, this study possesses various contributions in the domain of intrusion detection for SCADA networks running the IEC 60870-5-104 protocol, as follows.

1. Development of an Authentic Dataset: This study successfully generated a dataset derived directly from real testbed traffic and supplemented it with various types of attack traffic. The dataset not only reflects real conditions within SCADA networks operating the IEC 60870-5-104 protocol but also encompasses a wide range of relevant threats. Various attack data on the dataset, i.e., data on port scan, brute force, and different types of DoS.
2. Diversity of Attacks in the Dataset: The generated dataset includes various types of attack data, such as port scans, brute force attacks, and different forms of DoS attacks. This diversity makes the dataset a rich resource for studying and developing more comprehensive and robust IDS models.
3. Identification of Specific Rules and Features for IEC 104 Flood Attacks: This study contributes by identifying relevant rules and features for detecting flood attacks on the IEC 104 protocol within SCADA networks. This is a crucial step in enhancing the accuracy and effectiveness of IDS models in detecting complex and specific attacks.
4. Performance Evaluation of IDS Models: The study also includes a performance evaluation of the various IDS models generated using different machine learning algorithms. This evaluation aims to identify the best-performing model that can be effectively implemented in real-world SCADA environments.

In comparison to other studies focused on building IEC 104 SCADA datasets, this research offers more diverse and realistic attack scenarios. Robles-Durazno et al. [9] emphasize the importance of hybrid testbeds that combine physical and virtual components to accurately replicate the complexities of SCADA networks. Likewise, Crussell et al. [10] note that virtual environments, while cost-effective, often fail to capture the low-level network behaviors critical for effective IDS model development. By generating a dataset using a hybrid testbed that incorporates both physical and virtual elements, this study ensures the authenticity and applicability of the data to real-world SCADA systems. As a result, the dataset produced in this research can serve as a robust learning resource for developing reliable IDS models that leverage artificial intelligence techniques, with greater accuracy and relevance to actual SCADA environments.

The rest of the paper is arranged as follows. Section II provides the background and related work about IDS on SCADA networks. Section III discusses the research methodology

used in this study. Section IV presents the experimental results and analysis, and Section V concludes the work.

II. BACKGROUND AND RELATED WORK

In this section, we briefly review the structure of IEC 104 SCADA data packets, malicious activities on SCADA networks, IDS-related works on SCADA networks, and other research related to SCADA datasets.

A. SCADA IEC 60870-5-104 PROTOCOL

The IEC 104 protocol has become popular in the power plant industry due to its support for automation generation control (ACG) [11]. The IEC 104 protocol is a TCP/IP-based modification of the IEC 60870-5-101 (IEC 101) standard for power system monitoring and telecontrol [12], which are widely used in modern SCADA systems built upon TCP/IP [13]. The basic frame in the IEC-104 protocol is called the application protocol data unit (APDU). The APDU is transmitted as part of the TCP payload [14]. The maximum frame of the APDU is 255 octets. The APDU is divided into two parts: the application service data unit (ASDU) and the application protocol control information (APCI) [15]. APDU may contain only APCI without ASDU [16]. For the IEC 104 protocol, APDU has a start byte value of 0×68 as a header followed by the 8-bit length of the APDU and 4-octet control fields of 8-bit length this value is in the APCI section [15]. The APCI contains basic information, such as APDU length or sender and receiver sequence numbers, and has a fixed packet length of 4 Bytes [17]. An APDU frame can be in U, S, or I format [18].

B. MALICIOUS ACTIVITY ON THE SCADA NETWORK

Malicious activities in a SCADA network become very complicated because the SCADA system becomes open to heterogeneous networks for flexibility to reduce costs purposes. TCP/IP is a protocol that is widely integrated with SCADA protocols, such as distributed network protocol 3 (DNP3), Modbus, and IEC 60870-5-104 [11].

Port scanning is the initial step in computer network attacks [19]. In port scanning, a probe packet is sent to the target port, and based on the target system's response or lack of response, it can be inferred that the target port is in one of the following states: open, closed, or filtered.

A brute force (BF) attack is a common type of attack that may lead to intrusion and control being taken by an attacker [20]. Brute force attacks are carried out to control or retrieve data. This type of attack mostly leads to secure shell (SSH) and file transfer protocol (FTP) services, which will be discussed further in this study. Brute force attacks work by counting every possible combination that can form a password and then testing them to determine the correct password. As the lengths and number of combinations of passwords grow, the amount of time it takes to find the correct password increases exponentially [21]. In devices and systems that have high computing resources, more security methods can be embedded. For example, after usernames and passwords are incorrectly inputted many times, the system will be locked. However, SCADA devices, such as RTUs, have limited computing resources, so this technique cannot be applied to them.

The DoS has become one of the most common cyber attacks targeting hosts [22]. DoS attacks are a malicious way to consume users' bandwidth [23]. A DoS attack on a computer network is an attack on the availability of computer resources to prevent legitimate users from accessing those resources over the network [24]. DoS is a simple attack but has a big impact if aimed at devices with limited compute resources, such as RTU on SCADA. Now, DoS can be done easily with Python scripts [25], using the Scapy library to deliver the desired DoS packets. This DoS attack can be in the form of an ICMP flood [26], SYN flood [27], Xmas [28], and IEC 104 flood [29].

C. MALICIOUS ACTIVITY ON THE SCADA NETWORK

SCADA devices, such as RTUs, have low computing resources, making it impossible to implement standard security systems [30]. This limitation means that attacks, which are less severe on traditional network devices may have a major impact on SCADA devices.

Snort and Suricata are open-source IDSs that utilize signature-based techniques to detect attacks based on predefined rules [31] and are designed for more flexibility. Unlike commercial IDSs such as Cisco Secure IDS, CyberSafe, and Network Ice Blackie Devender; Snort and Suricata allow administrators to add signatures or patterns deemed threatening on the network to their respective rules databases. Snort is the most widely deployed IDS worldwide. It relies on a relatively simple language for the specification of misuses and attack signatures.

In recent years, machine learning has emerged as a solution for making IDSs used in detecting attacks on SCADA network communication systems that have limited device capabilities to process complex data. Conventional IDSs rely on manually designed rules. These rules depend heavily on professional experience, thereby making it challenging to represent the increasingly complicated industrial control logic [32]. Table 1 summarizes the studies on SCADA network security.

In our research, we adopted a hybrid testbed combining both physical and virtual components to ensure realistic and practical outcomes [9]. The physical elements, such as the Remote Terminal Units (RTUs), are crucial for obtaining accurate data that reflect real-world scenarios.

It is important to note that we did not compare the IDS model performance of our experiments with models trained on other datasets. The main reason for this is that datasets are generated under different circumstances, which can result in variations in data distribution, attack scenarios, network setups, and testbed environments. Since each dataset

TABLE 1. Summary of SCADA network security works.

Ref. & (Year)	Threats	Methods	Protocol	Pros and Cons
Gumaei *et al.* [7] (2020)	MiTM	KNN, RSL-KNN, CFS KNN	-	Provides a comprehensive comparison of methods for IDS models, but the dataset used is not explained and they do not discuss the attack patterns carried out by attackers nor explain the SCADA protocol, which is significant since SCADA networks are different from general computer networks.
Egger *et al.* [17] (2020)	Port Scan, DoS	Supervised, Semi-Supervised, Unsupervised	IEC 104	Provides a comprehensive comparison of methods for building an IDS but does not discuss attack patterns and attack types on diverse datasets.
Riyadi *et al.* [21] (2021)	MiTM	BRC4 data encryption	DNP3	This paper discusses another perception of data encryption on SCADA DNP3 by securing the data path, which is explained comprehensively. It does not explain the effect of data encryption on increased compute resource requirements even though SCADA devices have limitations on their compute resource capabilities.
Qian *et al.* [27] (2020)	MiTM, Replay, DoS, Zero-Day attack	NHFC, SVN, OCSVM, FCMSVM, GENFIS	Modbus/ TCP	This study used a large amount of data generated in plants, and the accuracy of the IDS model obtained is high. However, the IDS model created cannot determine the type of attack that occured.
Grigoriou *et. al* [4] (2022)	Zero-Day attack	Honeypots	IEC 104	This paper discusses the use of honeypots to protect SCADA assets that use the IEC 104 protocol from Zero-Day attacks by hiding the SCADA device.
Arifin, *et al.* [33] (2022)	**Port Scan, Bruteforce, ICMP flood (ping flood), Xmas, IEC 104 flood.**	**Rule-base, RF, Gradient Boosting, Categorical Naïve Bayes**	**IEC 104**	**The dataset was generated using a physical testbed on an RTU device, various attack data on the dataset were gathered, and a comprehensive explanation of attack pattern recognition for SCADA networks was provided.**

is usually designed for specific conditions, making direct comparisons between them would not be appropriate. Even when datasets use the same standard, such as IEC 60870-5-104, the scenarios executed in each testbed can vary. These differences might include the types of attacks simulated, the system configurations, or the operational conditions, all of which shape the characteristics of the dataset.

D. DATASET ON SCADA IEC 60870-5-104

The dataset is considered an important component in machine learning for creating IDSs. The effectiveness of existing machine learning techniques for cyber-security depends on the characteristics of the datasets [34]. The dataset used to build IDS on traditional computer networks is no longer relevant for building IDS on SCADA networks [28], [29], so a special dataset for IDS SCADA is required. Table 2 shows the testbed characteristics we used to generate the dataset compared to other studies' testbeds.

SCADA dataset 104, generated by Egger et al. [17], contains port scan and SYN flood attacks data in comma-separated values (CSV) format. The attack data in this dataset do not contain a variety of attack types. Compared to the SCADA dataset in their studies, which is limited in attack types, our dataset includes a broader range of attacks, making it more suitable for training advanced IDS models.

The RICSel21 dataset [29] contains DoS, scanning, man-in-the-middle (MiTM), and injection attacks. The dataset is generated using a virtual testbed on the power network,

TABLE 2. Comparison of SCADA testbed's characteristics.

Ref. & (year)	Protocol	Testbed	Threats
Egger *et al.* [17] (2020)	IEC 104	Virtual	MiTM, DoS
P. Maynard et.al [35] (2018)	IEC 104	Virtual	MiTM
C. Y. Lin *et.al* [29] (2021)	IEC 104	Virtual	DoS, scanning, MiTM, injection attacks
Arifin, *et al.* [33] (2022)	IEC104	Hybrid	**Port Scan, Bruteforce, ICMP flood (ping flood), Xmas, IEC 104 flood.**

and the dataset is stored in the form of raw data in pcap format. Research using a virtual testbed was also carried out by Maynard et al. [35] to create an attack dataset on the IEC 104 SCADA network. The attack scenario in this study was the MiTM attack. It must be noted that the resulting dataset was created using a virtual testbed. Although it produces diverse attack data, it does not reflect the real conditions of the IEC 104 SCADA system [6]. Unlike the RICSel21 dataset [29] and the dataset in [35], which were generated solely in virtual testbeds, our dataset leverages a hybrid testbed approach. This allows us to capture both the realistic network behaviors found in physical SCADA systems and the scalability of virtual environments.

Robles-Durazno et al. [9] conducted a comprehensive evaluation of physical, hybrid, and virtual testbeds for the cyber-security analysis of industrial control systems. Their findings suggest that while virtual testbeds offer a cost-effective

solution, they may introduce variances in network behavior that could impact the accuracy of IDS models. The use of hybrid testbeds, combining both physical and virtual components, was shown to mitigate these issues by preserving the fidelity of critical infrastructure, such as Remote Terminal Units (RTUs), which are crucial for obtaining realistic results.

Similarly, Crussell et al. [10] highlighted that although virtual environments can replicate many aspects of SCADA operations, there are significant differences in the low-level network behaviors that can affect the detection capabilities of IDS models. These variations emphasize the necessity of using physical components in testbeds, especially when the goal is to develop robust IDS solutions for real-world SCADA networks.

The dataset [36] is used for health and contains MiTM, DoS, and injection attack data. Because it is used for the health sector, the normal data from this dataset are different from the normal data from the SCADA dataset in the power plant industry.

The traffic dataset in our dataset [33] is generated from real conditions using a physical RTU on the testbed with scenarios in the distribution section in the power generation industry. The use of the TCP/IP protocol as a transmission medium for the IEC 104 protocol allows the attacks that often occur on traditional computer networks, such as port scan, brute force, ICMP flood, SYN flood, and Xmas, to become executable on SCADA networks. Furthermore, IEC 104 flood attacks are simulated by sending ASDU packets with an unknown format to the SCADA IEC 104.

III. METHODOLOGY

The dataset development starts with the setting up of testbed network topology. Then data collection process is carried out with three scenarios.

A. TESTBED TOPOLOGY

This study used a hybrid testbed with physical devices combined with virtual devices. To obtain reliable data, common SCADA devices, such as RTUs and HMIs, were used as physical devices. In this testbed, the virtual devices used were the Historian server, an IDS sensor, and a malicious machine. Fig. 1 shows the testbed topology. The testbed topology was set up by mimicking the network environment of the power distribution network of a regional electricity company. Under the supervision of the company's site engineer, we tune the setup until it runs like a real power distribution network.

The SCADA IEC 60870-5-104 network topology consists of a Historian server with an IDS Sensor, HMI, router, malicious machine, RTU, cubicle, and power meter. RTU 1 and RTU 2 use a wireless connection, and RTU 3 uses a wired connection and then connects to the HMI and Historian server with IDS sensors through switches and routers. We do not use programmable logic controller (PLC) because all PLC functions can be executed by RTU. In addition, the RTU supports wireless data communication and has a large memory [37]. The malicious machine is connected to the router as a MiTM attack. The MiTM allows an attacker to sit in between the communicating parties. The attacker is able to read or modify communications, inject

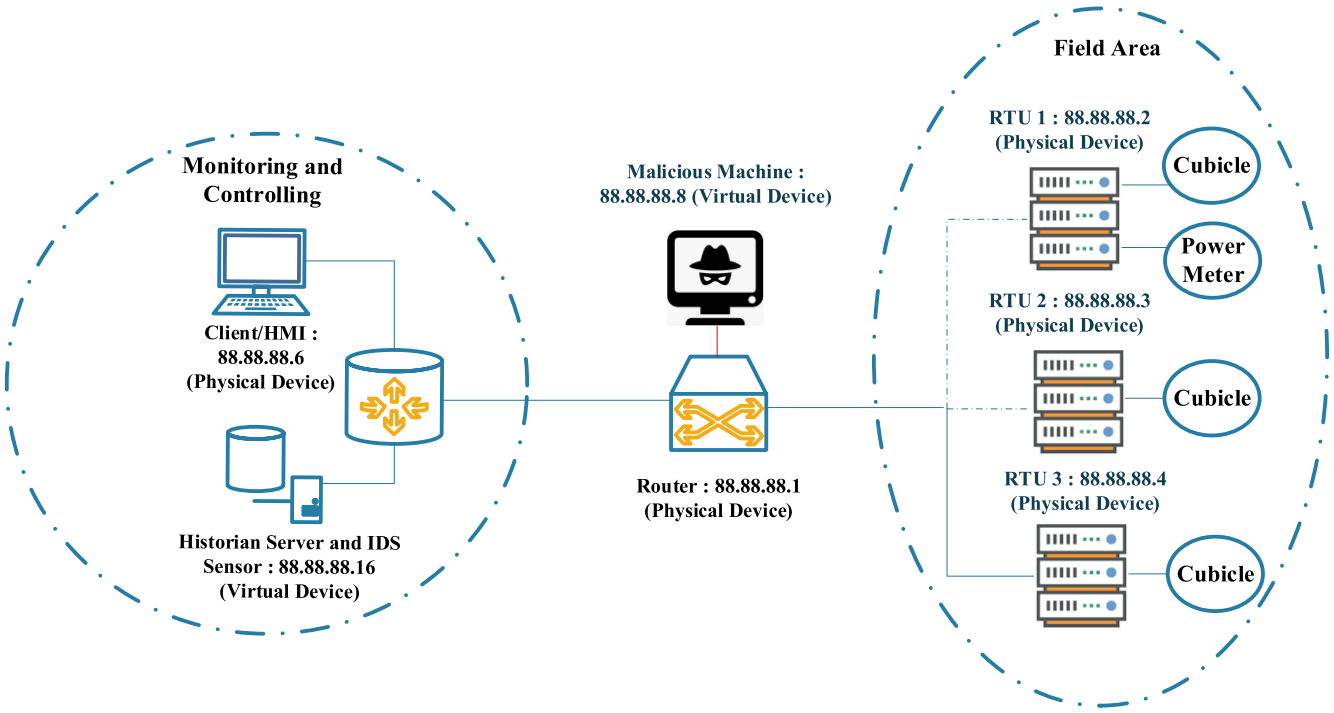

FIGURE 1. The SCADA IEC 60870-5-104 testbed topology.

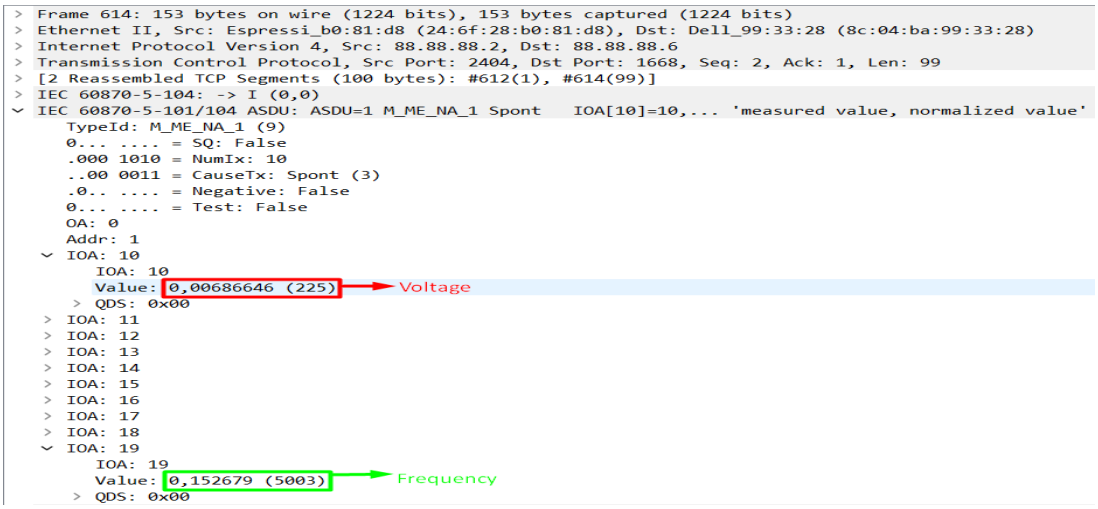

FIGURE 2. Voltage and frequency monitoring in the normal scenario.

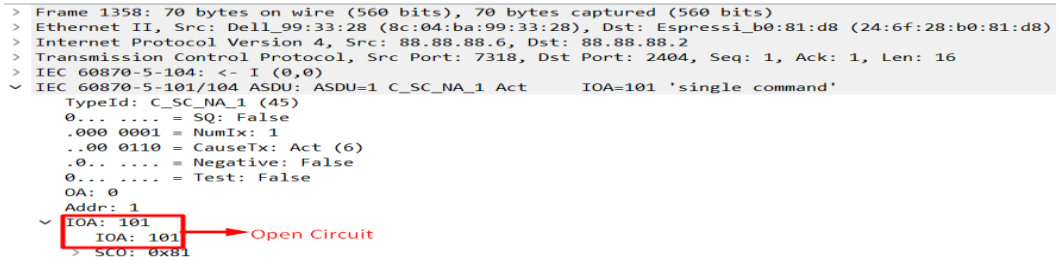

FIGURE 3. HMI instruction to RTU 1 for open circuit in the normal scenario.

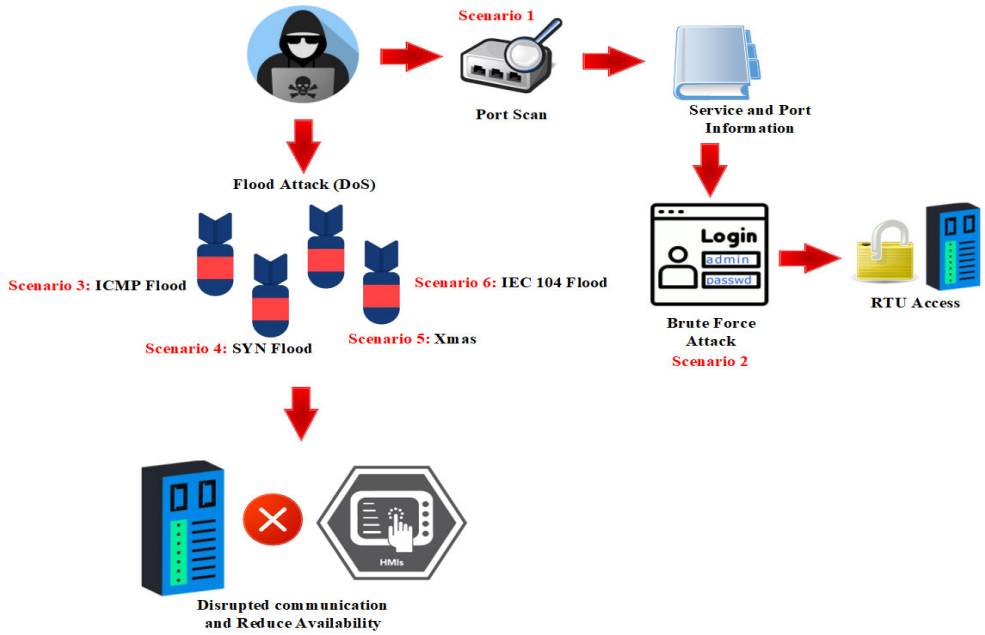

FIGURE 4. The attack scenario on the dataset.

commands, or drop packets [38]. Table 3 shows the list of SCADA IEC 60870-5-104 instructions mentioned in this study.

B. DATA COLLECTION SCENARIO

The process of data collection was divided into three scenarios: normal data collection, attack data collection, and

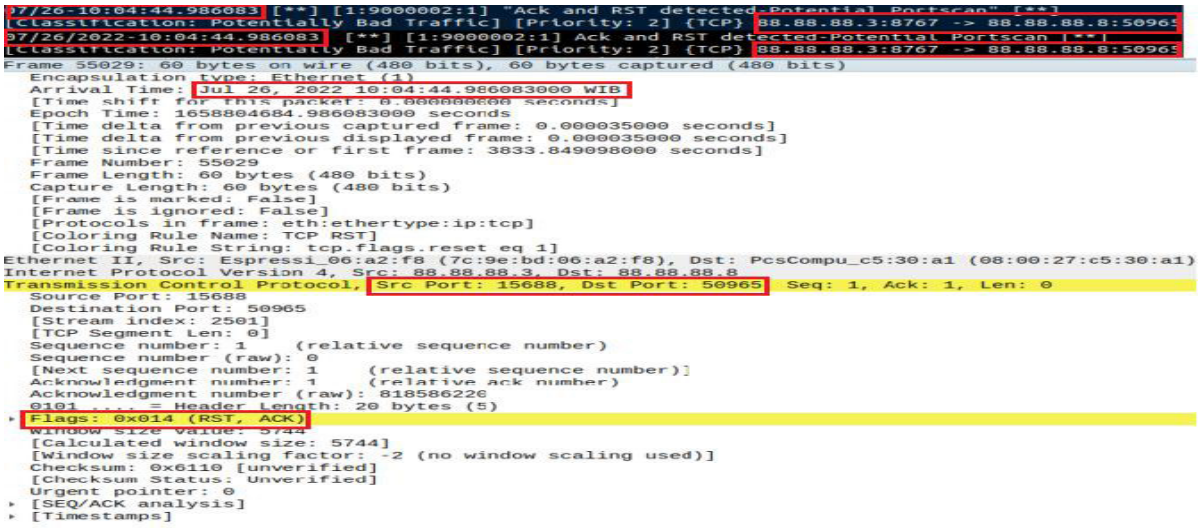

FIGURE 5. The validation of port scan activity using Snort and Suricata.

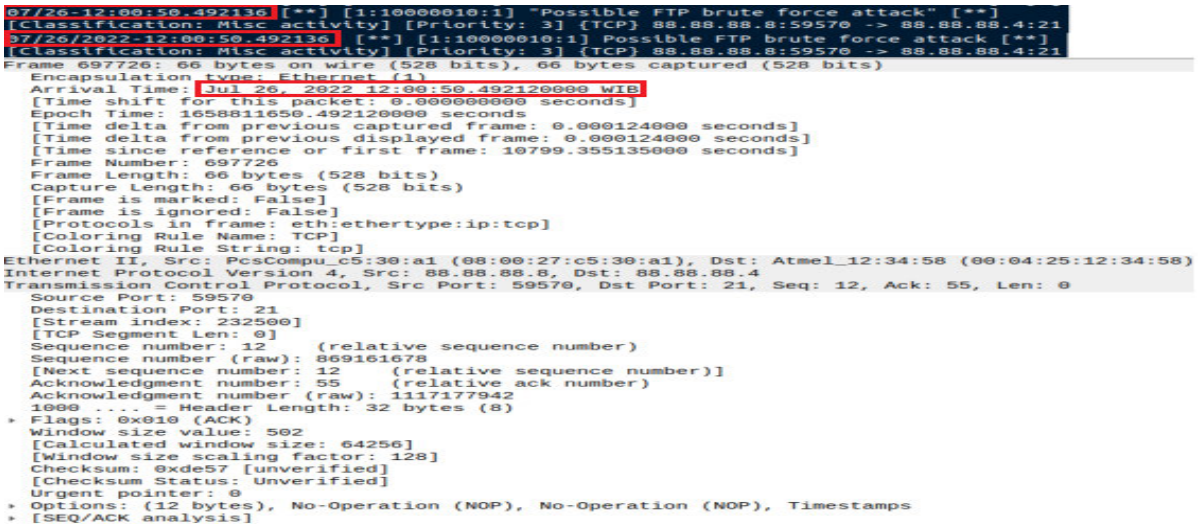

FIGURE 6. The validation of a brute force attack using Snort and Suricat.

TABLE 3. Important instructions of Scada IEC 104 network security.

Instruction	Description	CauseTx
C_SC_NA_1	Single Command	6
C_IC_NA_1	Setpoint Command, Normalized Value	6
M_ME_NA_1	Measured Value, Normalized Value	3
M_SP_NA_1	Single Point Information	3

collection of normal and attack data combinations. The scenarios are as follows:

1. Normal data retrieval was performed on a SCADA system running without an attack. In this scenario, the HMI and RTU communicate to receive voltage and frequency monitoring data. Fig. 2 shows the monitored voltage seen from Wireshark in a normal scenario. The HMI also sent a command to the RTU to perform the switching on the cubicle shown in Fig. 3 on the monitoring process of voltage and frequency. The HMI received ASDU packets M_ME_NA_1 on Information Object Address (IOA) 10, IOA 11, IOA 13 for voltage, and IOA 19 for frequency in the normal scenario.In Fig. 2, IOA 10 indicates a voltage value of 225 volts, and IOA 19 indicates an electric current frequency of 50 Hz.
2. Attack data retrieval was carried out by attacking each RTU. The attacks are port scan, brute force, and DoS. The data were then stored in separate recording files for each type of attack in pcap format. Kali Linux was used by malicious machines to carry out attacks. The port scan applications, i.e.: Nmap and Masscan were used to find vulnerabilities in the SCADA system. Brute

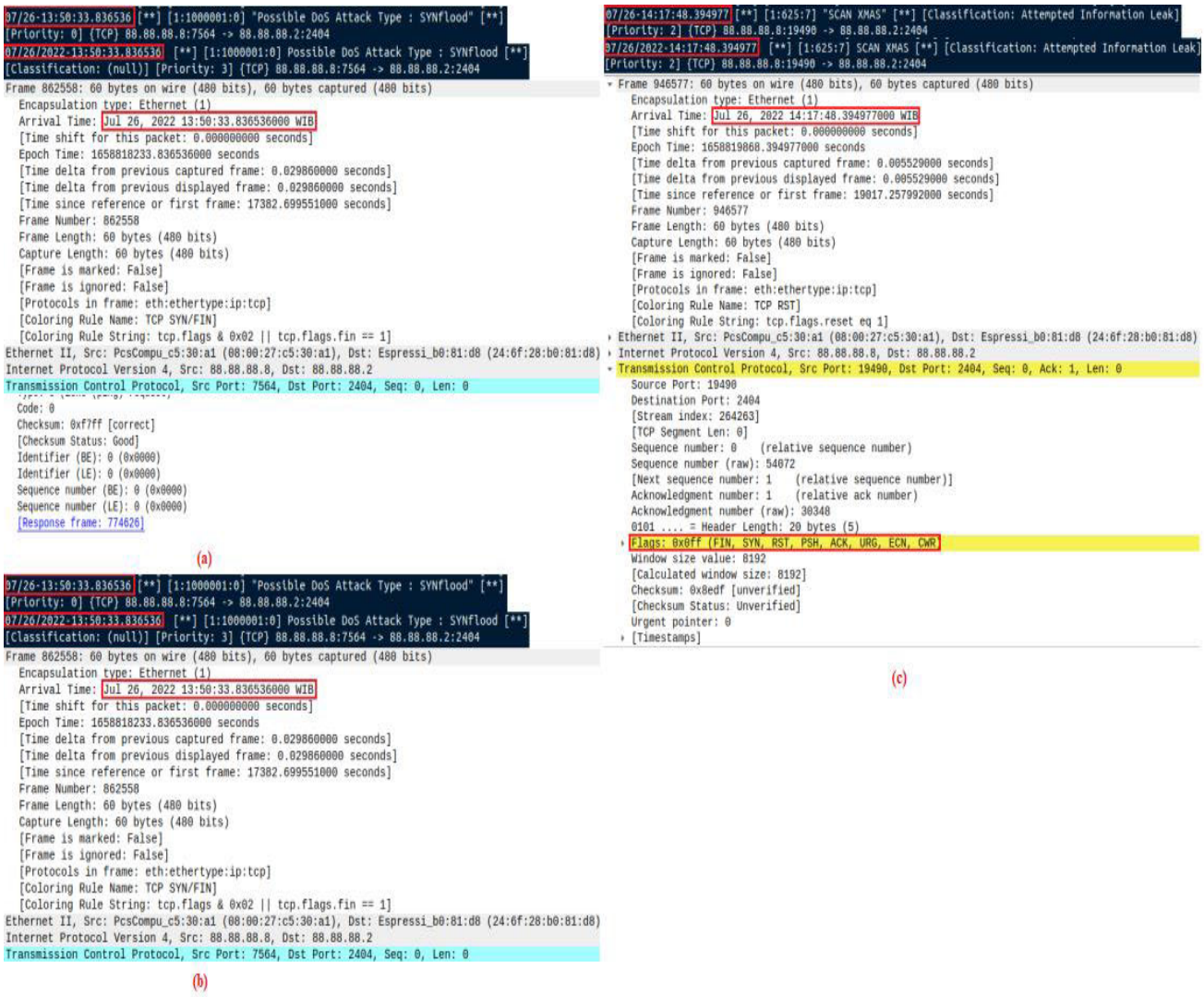

FIGURE 7. The validation of DoS attack using Snort and Suricata.

force attacks were executed to gain access to RTU 3. Hydra and Medusa were used for launching brute force attacks, i.e.: SYN flood, ICMP flood, Xmas attacks while IEC 104 flood. IEC 104 flood attack was done by flooding the RTU using unknown ASDU (104) packets.

3. Normal-attack data retrieval was done by combining the two previous scenarios. The steps taken before launching an attack were: first, to scan all devices in the testbed to find vulnerabilities. Then, after finding the vulnerabilities, a brute force attack aimed at RTU 3 was carried out to take over the device. DoS attacks are carried out afterward to overwhelm and hinder communication between the HMI and the RTU.

Six (6) scenarios are prepared for conducting the attacks. The first scenario is to perform port scanning to reveal which services and ports are open. After this information is obtained, various attacks can be executed, leading to various scenarios. In Scenario 2, a brute force attack targeting the SSH and FTP services was carried out. In Scenario 3, an ICMP flood is used to disrupt the communication between the HMI and RTU with ICMP packets. In Scenario 4, a SYN flood attack is done by sending TCP data packets with the SYN flag massively. The Xmas attack in Scenario 5 is performed by sending all flags on TCP data packets, i.e., FIN, SYN, RST, PSH, ACK, and URG [39]. In the last scenario, IEC 104 attacks are performed by sending unknown ASDU packets massively to disrupt the communication between the HMI and RTU. All types of DoS attacks are directed at port 2404, which in this study is the IEC 104 service port. Fig. 4 shows the scenarios for creating the dataset.

IV. RESULT AND ANALYSIS

This section presents the attack patterns of each scenario along with their analyses. Captured traffic attributes of Wireshark are juxtaposed with the detection results of Snort and Suricata to explain the attacks in the dataset and to reveal

```
07/26-14:46:11.425909 [**] [1:6666617:1] "IEC 104 Flood Attack with Reqco3 Cause of Transmission (40)" [**]
[Classification: Potentially Bad Traffic] [Priority: 3] {TCP} 88.88.88.8:65233 -> 88.88.88.4:2404
07/26/2022-14:46:11.425909  [**] [1:6666617:1] IEC 104 Flood Attack with Reqco3 Cause of Transmission (40) [**]
[Classification: Potentially Bad Traffic] [Priority: 3] {TCP} 88.88.88.8:65233 -> 88.88.88.4:2404
▾ Frame 982230: 182 bytes on wire (1456 bits), 182 bytes captured (1456 bits)
    Encapsulation type: Ethernet (1)
    Arrival Time: Jul 26, 2022 14:46:11.425909000 WIB
    [Time shift for this packet: 0.000000000 seconds]
    Epoch Time: 1658821571.425909000 seconds
    [Time delta from previous captured frame: 0.014026000 seconds]
    [Time delta from previous displayed frame: 0.000000000 seconds]
    [Time since reference or first frame: 20720.288924000 seconds]
    Frame Number: 982230
    Frame Length: 182 bytes (1456 bits)
    Capture Length: 182 bytes (1456 bits)
    [Frame is marked: False]
    [Frame is ignored: False]
    [Protocols in frame: eth:ethertype:ip:tcp:iec60870_104:iec60870_asdu]
    [Coloring Rule Name: TCP SYN/FIN]
    [Coloring Rule String: tcp.flags & 0x02 || tcp.flags.fin == 1]
▸ Ethernet II, Src: PcsCompu_c5:30:a1 (08:00:27:c5:30:a1), Dst: Atmel_12:34:58 (00:04:25:12:34:58)
▸ Internet Protocol Version 4, Src: 88.88.88.8, Dst: 88.88.88.4
▸ Transmission Control Protocol, Src Port: 65233, Dst Port: 2404, Seq: 0, Ack: 1, Len: 128
▾ IEC 60870-5-104: <- I (13364,13364)
    START
    ApduLen: 104
    .... ...0 = Type: I (0x00)
    Tx: 13364
    Rx: 13364
▾ IEC 60870-5-101/104 ASDU: ASDU=26728 <TypeId=104> Reqco3_NEGA IOA[104]=6842472,... '<Unknown TypeId>'
    TypeId: Unknown (104)
    0... .... = SQ: False
    .110 1000 = NumIx: 104
    ..10 1000 = CauseTx: Reqco3 (40)
    .1.. .... = Negative: True
    0... .... = Test: False
    OA: 104
    Addr: 26728
    IOA: 6842472
    Raw Data: 6868686868686868686868686868686868686868686868686868…
```

FIGURE 8. The validation of IEC 104 flood attack using Snort and Suricata.

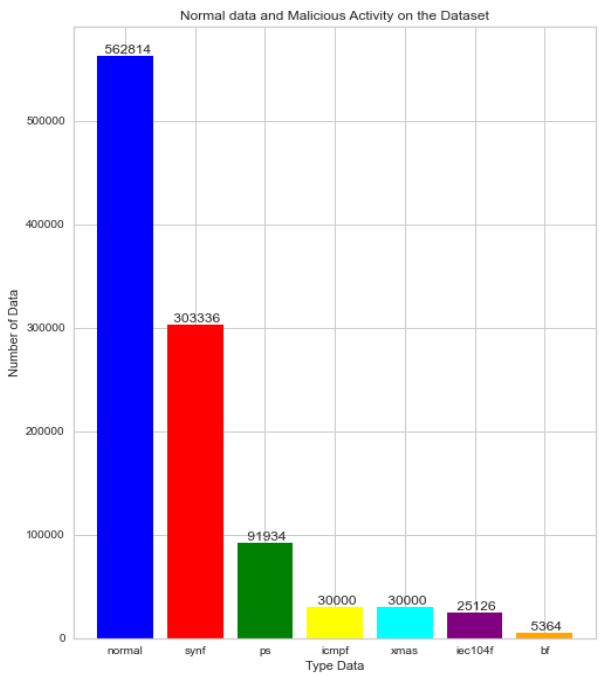

FIGURE 9. Number of classes in the dataset.

the attack pattern. Furthermore, the testing results of the IDS models using the machine learning algorithms are presented to justify the reliability of the dataset.

A. DATA COLLECTION SCENARIO

This study uses Snort 3.1.17.0 and Suricata 6.0.6 to prove that there was an attack on the dataset. The rules used to detect attacks on Snort and Suricata were the same because some rules in Snort can be also used in Suricata. For the detection of the port scan, brute force, SYN flood, ICMP flood, and Xmas attacks, we used standard rules from Snort and Suricata, while custom-made rules were used for IEC 104 flood attack detection. To demonstrate the attack detection, the dataset containing normal-attack traffic was used.

In the port scan activity, the attacker sends TCP packets with the SYN flag (0×002) to the target port then the target host replies with SYN, ACK (0×012), which indicates that the destination port is open [40], or RST, ACK (0×014) [41], which indicates the destination port is closed. Table 4 presents the results of a port scan using Nmap.

TABLE 4. Port information on the RTU in the testbed after port scan activity.

RTU	PORT	SERVICE	STATUS
RTU 1	2404	iec-104	Open
RTU 2	2404	iec-104	Open
	21	ftp	Open
	22	ssh	Open
RTU 3	80	http	Open
	1234	hotline	Open
	2404	iec-104	Open
	20000	dnp	Filtered

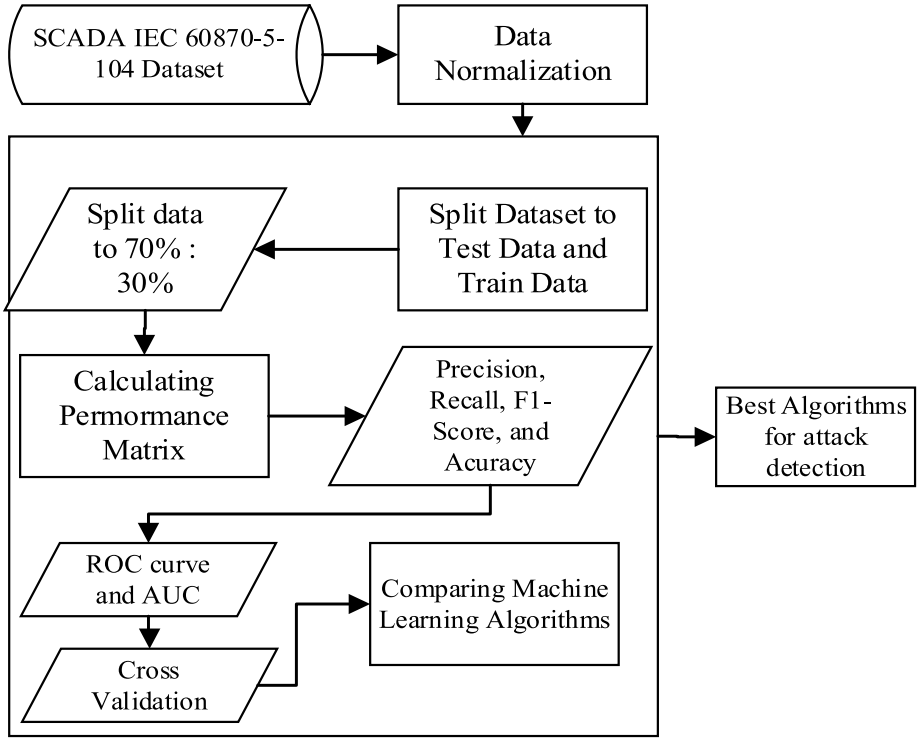

FIGURE 10. Number of classes in the dataset.

Fig. 5 shows the results of the detection of port scan activity on Snort and Suricata along with the correlation of the payload. The results of the port scan activity show that port 50965 is closed because the RTU 3 device replies with TCP flags 0 × 014 (RST, ACK).

The information inside the boxes with the red line in Fig. 5 shows the timestamp, port information, and IP address of the warning message from Snort and Suricata, which are the same as the information obtained from Wireshark. The port scan scenario revealed that ports on the RTU devices were either open or closed based on the TCP flags captured (SYN, ACK, RST). The correlation between the warning messages from Snort and Suricata with the data captured in Wireshark confirmed the presence of port scanning activities within the traffic dataset, as shown in Table 3 and Fig. 5. This fact indicates that the dataset accurately represents real-world attack scenarios, which is critical for its effectiveness in IDS testing.

A brute force attack involves the submission of many username and password combinations by an attacker, intending to access data and resources. In this study, brute force attacks lead to FTP and SSH services on RTU 3. The correlation between Snort, Suricata, and Wireshark data, as shown in Fig. 6, confirmed the presence of these attacks in the dataset. The results demonstrated that the dataset could effectively capture brute force attacks, which are common threats to network security. In Fig. 6, the Snort and Suricata timestamp indicates a brute force attack in the traffic dataset.

SCADA systems under DoS attacks can face severe performance issues. DoS attacks are very flexible and can take different forms in different network settings [42]. In this study, a DoS attack was performed to disrupt and inhibit communication between the HMI and the RTU. In some conditions, when the DoS attack worsens, the RTU becomes slow to process sensor results and fails to send sensor data to the HMI. DoS attacks were also simulated in several forms, including ICMP flood, SYN flood, Xmas flood, and IEC 104 flood. These attacks were targeted at disrupting communication between the HMI and RTU. The successful detection of these attacks by both Snort and Suricata, as corroborated by Wireshark data (Fig. 7), further supports the validity of the dataset. The DoS attacks were aimed at all RTUs. The timestamp in the warning messages from Snort and Suricata is in accordance with the information captured by Wireshark as shown in Fig. 7. These facts indicate there is a DoS attack in the dataset. (a) shows the correlation between the Snort and Suricata detection results with the payload on Wireshark for ICMP flood attack detection, (b) shows the correlation between the Snort and Suricata detection results with the payload on Wireshark Xmas attack detection, and (c) shows the correlation between the Snort and Suricata detection results with the payload on Wireshark for SYN flood detection.

In the IEC 104 flood attack, ASDU was sent using an unknown typeid (104) with the cause of transmission (CauseTx) 40 with NumIx 104. This attack successfully disrupted RTU and HMI communication. IEC 104 flood is

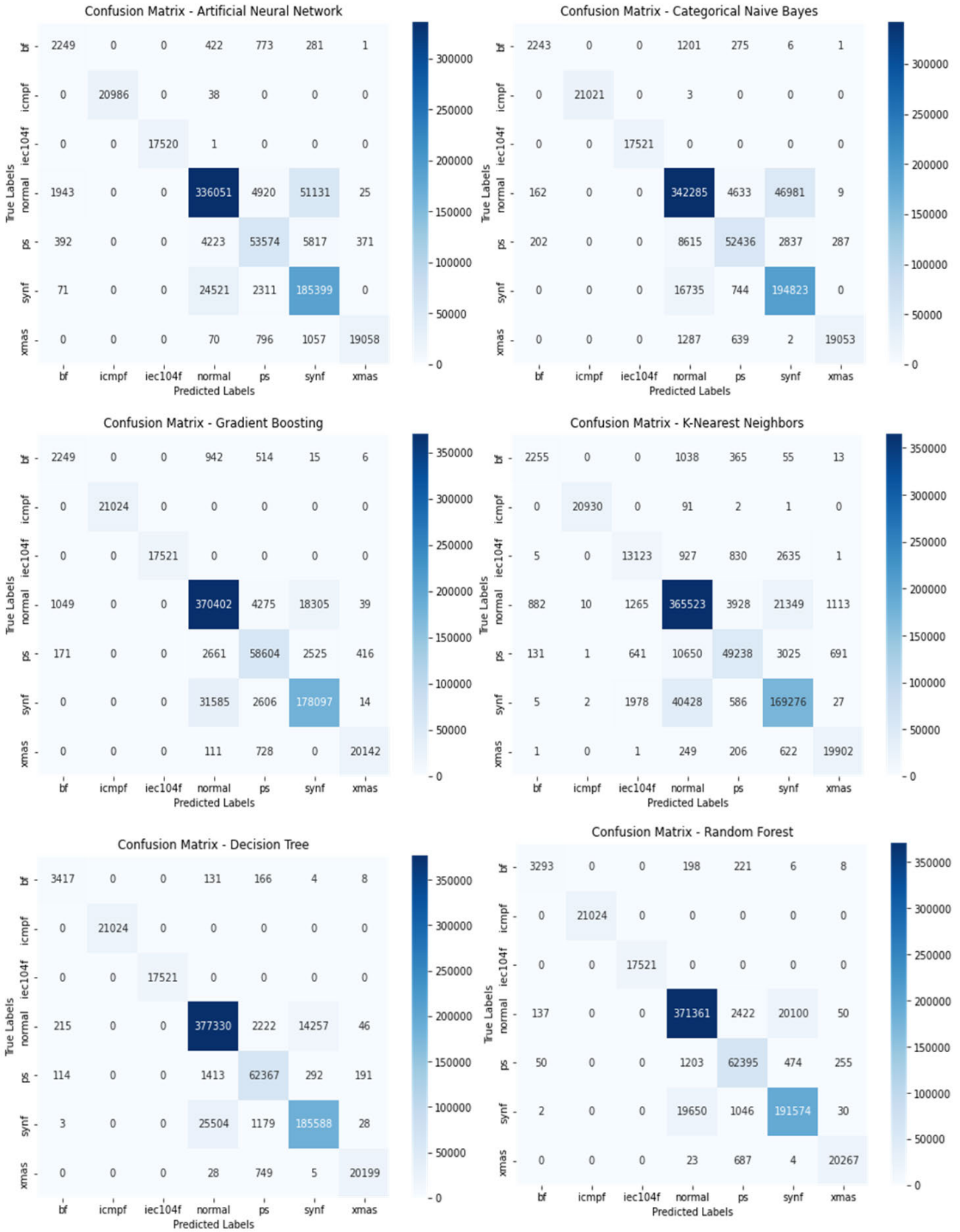

FIGURE 11. The confusion matrix for each IDS model.

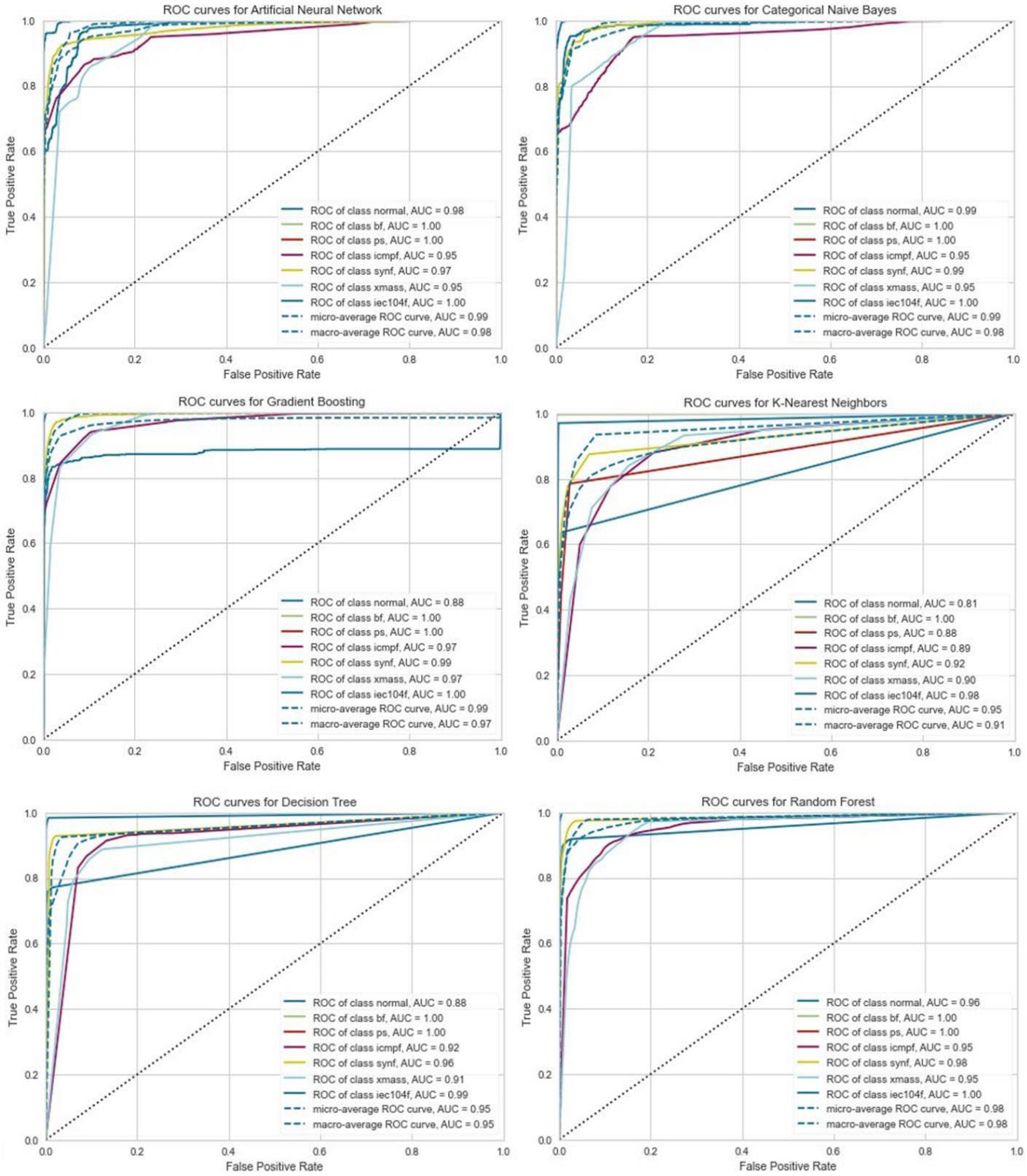

FIGURE 12. The AUC and ROC curves for each model.

done by modifying the IEC 60870-5-104 protocol packet and sending an unknown ASDU format (104) then using Reqco3 (40) for the causeTX value. Object address (OA) is modified by using address 104. Fig. 8 shows the correlation between the Snort and the Suricata detection results with the payload on Wireshark for IEC 104 flood detection.

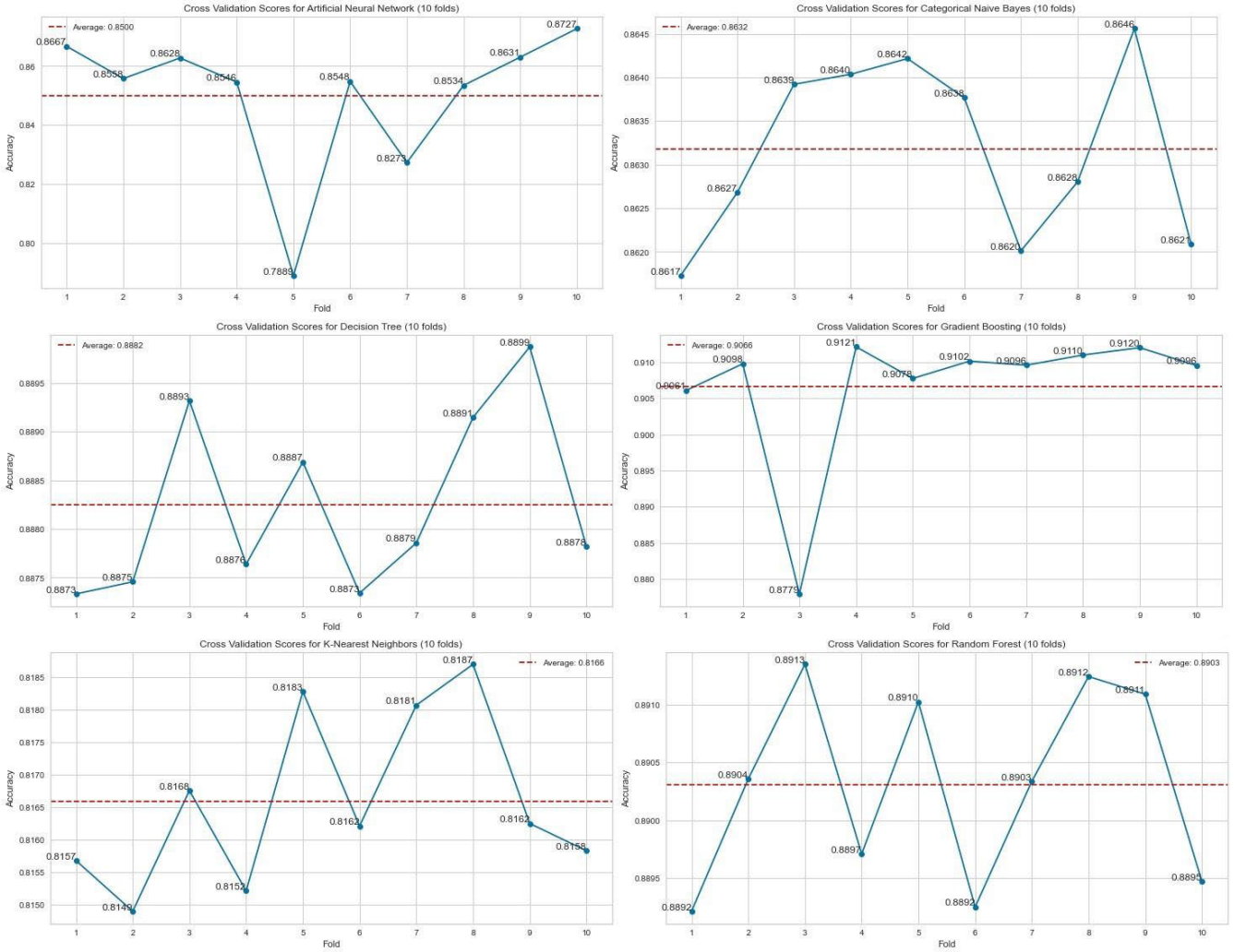

FIGURE 13. The cross-validation results for each model.

In the attack scenario, we sent a total of 485,760 attack packets. There are false alarms from the intrusion detection results generated on Snort and Suricata. Table 5 shows the results of intrusion detection from Snort and Suricata.

TABLE 5. Snort and Suricata detection result.

IDS	Detection	False Detection	Undetected
Snort	364,940	1,199	120,621
Suricata	364,658	3,156	117,946

False detection refers to normal communication that is detected as a port scan; the activity is a three-way handshake activity, which is a standard of TCP communication protocol. From the performance measurements carried out on Snort, 0.25% were false alarms and undetected packets accounted for 24.8% of the total detected ones. For Suricata, 0.65% were false alarms, and undetected packets accounted for 24.2% of the total suspected activity. The use of the same rule in Snort can detect more suspected activity than in Suricata.

B. MACHINE LEARNING APPROACH FOR DATASET TESTING

Based on the known attack patterns, preprocessing is carried out to create training materials for the IDS models using machine learning techniques. In this study, we use classification algorithms such as Artificial Neural Network, Categorical Naïve Bayes, Decision Tree, Gradient Boosting, K-Nearest Neighbors, and Random Forest to detect attacks in the dataset. Fig. 9 presents a comparison of normal data and attack data on the dataset after preprocessing.

The total dataset used after preprocessing is 1,048,574 packets. Normal data comprises the majority of the dataset with a total of 562,814 packets; SYN flood attacks totaled 303,336 packets, ICMP floods totaled 30,000 packets, Xmas 0,000 totaled packets, IEC 104 flood totaled 25,126 packets, port scan totaled 91,934 packets, and brute force attacks totaled 5,364 packets. Fig. 10 shows the workflow used in this study to determine the best IDS model.

TABLE 6. The result of the performance evaluation.

Model	Evaluation	Performance			
		Accuracy	Precision	Recall	F1-Score
Artificial Neural Network	Normal	86.49%	0.92	0.85	0.89
	Port scan		0.86	0.83	0.85
	Brute force		0.48	0.60	0.54
	ICMP flood		1.00	1.00	1.00
	Syn flood		0.76	0.87	0.81
	Xmas		0.98	0.91	0.94
	IEC 104 flood		1.00	1.00	1.00
Categorical Naïve Bayes	Normal	88.47%	0.92	0.87	0.90
	Port scan		0.89	0.81	0.85
	Brute force		0.86	0.60	0.71
	ICMP flood		1.00	1.00	1.00
	Syn flood		0.80	0.92	0.85
	Xmas		0.98	0.91	0.94
	IEC 104 flood		1.00	1.00	1.00
Decision Tree	Normal	93.66%	0.93	0.96	0.95
	Port scan		0.94	0.97	0.95
	Brute force		0.91	0.92	0.91
	ICMP flood		1.00	1.00	1.00
	Syn flood		0.93	0.87	0.90
	Xmas		0.99	0.96	0.97
	IEC 104 flood		1.00	1.00	1.00
Gradient Boosting	Normal	91.01%	0.91	0.94	0.93
	Port scan		0.88	0.91	0.89
	Brute force		0.65	0.60	0.63
	ICMP flood		1.00	1.00	1.00
	Syn flood		0.90	0.84	0.87
	Xmas		0.98	0.96	0.97
	IEC 104 flood		1.00	1.00	1.00
K-Nearest Neighbors	Normal	87.23%	0.87	0.93	0.90
	Port scan		0.89	0.76	0.82
	Brute force		0.69	0.61	0.64
	ICMP flood		1.00	1.00	1.00
	Syn flood		0.86	0.80	0.83
	Xmas		0.92	0.95	0.93
	IEC 104 flood		0.77	0.75	0.76
Random Forest	Normal	93.66%	0.95	0.94	0.94
	Port scan		0.95	0.97	0.95
	Brute force		0.95	0.88	0.91
	ICMP flood		1.00	1.00	1.00
	Syn flood		0.90	0.90	0.90
	Xmas		0.98	0.97	0.97
	IEC 104 flood		1.00	1.00	1.00

The normalization of the dataset was performed, leaving only the required data related to the attack data. We split the dataset into 30% testing data and 70% training data. The features used to create the IDS model using machine learning are *tcp.flags, tcp.srcport, tcp.dstport, protocol, frm.len, icmp.type, tcp.ckcsum, start.frame, typeid, causetx, ioa, and addr*. *tcp.flags*. The features are the ones that contain special marks in the TCP protocol. The *tcp.srcport* is the source port of the sender, *tcp.dstport* is the destination port for the sent data, *protocol* is the type of protocol used in SCADA communication, *frm.len* is the length of the data frame that was sent or received, *icmp.type* is the type of icmp packet transmitted, *tcp.ckcsum* is the integrity of data portions for data transmission, *start.frame* is the header of the SCADA IEC 60870-5-104 data packet with a value of 0×68, *typeid* is a feature that describes the command to monitor the condition of the RTU as well as the command by the HMI to execute the program to the RTU device, *causetx* is the cause of transmission in the SCADA system IEC 60870-5-104 that contains a code or value to validate the received ASDU packet, *ioa* is used to uniquely identify each item on the device and is transmitted in each ASDU that includes information about that particular input or output, and *addr* is the field of the ASDU address.

Matrix values such as false1 positive (FP), true negative (TN), and false negative (FN) are used to evaluate the performance of the testing model. The precision, recall, and F-measure (F1 score) values are considered for validating the accuracy result. The performance metrics are determined by (1)–(4).

$$\textbf{Accuracy} = \frac{(\textbf{TN} + \textbf{TP})}{(\boldsymbol{TN} + \boldsymbol{TP} + \boldsymbol{FN} + \boldsymbol{FP})} \tag{1}$$

$$\textbf{Precision} = \frac{\textbf{TP}}{(\boldsymbol{TP} + \boldsymbol{FP})} \tag{2}$$

$$\textbf{Recall} = \frac{\textbf{TP}}{(\boldsymbol{TN} + \boldsymbol{FP})} \tag{3}$$

$$\textbf{F1 Measure} = 2\frac{(\textbf{Precision} \times \textbf{Recall})}{(\textit{Precision} + \textit{Recall})} \quad (4)$$

By displaying the various ways through which a classification model makes errors, the confusion matrix provides valuable information about the type and frequency of these errors. This knowledge is essential for understanding the performance of the classifier and identifying areas for improvement. Table 6 presents the results of the performance evaluation. Fig. 11 presents the confusion matrix of the IDS models for each algorithm and Fig. 12 presents the area under the curve-receiver operating characteristic (AUC–ROC) for each model. The ROC is used to visualize the trade-offs between the true positive rate (TPR) and the false positive rate (FPR). Further, to define the capability of the model to differentiate the classes, we use the AUC to present the degree of separability of the trained model.

Performance metrics such as accuracy, precision, recall, and F1-score were calculated to assess the IDS models. The Random Forest and Decision Tree algorithm achieved the highest accuracy at 93.66%, followed by Gradient Boosting at 91.01%, Categorical Naïve Bayes at 88.47%, K-Nearest Neighbors at 87,23%, and Artificial Neural Network at 86,49%.

Validated results are shown in the AUC–ROC curve that measures the overall model performance. The curves validate the model's ability to classify normal and attack data. In this study, cross-validation was also conducted to ensure that the model was not overfitting, with the results confirming the robustness of the model (Fig. 13). The high performance across different algorithms and the validation against known attack patterns suggest that the dataset is highly reliable for training IDS models.

V. CONCLUSION AND FUTURE WORK

This study has made a significant contribution to the domain of cybersecurity for SCADA systems by creating a comprehensive dataset using physical SCADA devices that operate within the electrical distribution process, utilizing the IEC 60870-5-104 protocol. The dataset includes various types of malicious activity data, such as port scan, brute force, ICMP flood, SYN flood, Xmas, and IEC 104 flood attacks. Notably, the IEC 104 flood attack was simulated by modifying the IEC 104 packets and sending a large volume of them to each RTU. The traffic data in the dataset are categorized into normal, attack, and combined normal-attack types, ensuring a robust foundation for developing and evaluating Intrusion Detection Systems (IDS).

Experimental results indicate that all DoS attacks commonly executed in traditional computer networks can also be launched on the SCADA IEC 60870-5-104 network, with potentially more severe consequences due to the limited computing resources of SCADA devices. This highlights the critical need for specialized security measures in SCADA environments.

The open-source IDSs employed in this study yielded mixed results. Suricata generated more false alarms than Snort but had a higher detection rate for attacks within the dataset. The reliability of the created dataset is further supported by the detection results obtained using machine learning models, with Decision Tree and Random Forest algorithms achieving the best accuracy of up to 93.66%. The similarity in accuracy between these two algorithms is attributed to their shared use of decision trees as the basis for decision-making. The Random Forest algorithm, in particular, benefits from the ensemble approach, combining multiple tree predictors to enhance overall performance. Random Forest is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.

Given SCADA's inherent limitations in device capabilities, especially as these systems are increasingly connected to open networks to reduce operational and monitoring costs, there is a clear need for specific IDS solutions tailored to these environments.

Future research stemming from this study could explore the application of advanced machine learning techniques, such as deep learning algorithms, for anomaly and attack detection in SCADA networks. These approaches have the potential to leverage the extensive dataset created in this study to enhance detection accuracy and adaptability to emerging attack methods. Additionally, future work could focus on developing efficient algorithms and architectures for real-time anomaly and attack detection in SCADA networks, with an emphasis on optimizing model inference speed and scalability. This is particularly important for managing the high volume and velocity of network traffic in operational SCADA environments, ensuring timely and accurate threat detection.

REFERENCES

[1] G. Yadav and K. Paul, "Architecture and security of SCADA systems: A review," *Int. J. Crit. Infrastruct. Protection*, vol. 34, Sep. 2021, Art. no. 100433, doi: 10.1016/j.ijcip.2021.100433.

[2] A. Volkova, M. Niedermeier, R. Basmadjian, and H. de Meer, "Security challenges in control network protocols: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 619–639, 1st Quart., 2019, doi: 10.1109/COMST.2018.2872114.

[3] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Comput. Netw.*, vol. 169, Mar. 2020, Art. no. 107094, doi: 10.1016/j.comnet.2019.107094.

[4] E. Grigoriou, A. Liatifis, P. R. Grammatikis, T. Lagkas, I. Moscholios, E. Markakis, and P. Sarigiannidis, "Protecting IEC 60870-5-104 ICS/SCADA systems with honeypots," in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, Jul. 2022, pp. 345–350, doi: 10.1109/CSR54599.2022.9850329.

[5] I. Cindrić and T. Hadjina, "An analysis of IEC 62351 implementations for securing IEC 60870-5-104 communication," in *Proc. IEEE Belgrade PowerTech*, vol. 54, Jun. 2023, pp. 1–6, doi: 10.1109/powertech55446.2023.10202923.

[6] X. Wang and E. Foo, "Assessing industrial control system attack datasets for intrusion detection," in *Proc. 3rd Int. Conf. Secur. Smart Cities, Ind. Control Syst. Commun. (SSIC)*, Oct. 2018, pp. 1–8, doi: 10.1109/SSIC.2018.8556706.

[7] A. Gumaei, M. M. Hassan, S. Huda, M. R. Hassan, D. Camacho, J. Del Ser, and G. Fortino, "A robust cyberattack detection approach using optimal features of SCADA power systems in smart grids," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106658, doi: 10.1016/j.asoc.2020.106658.

[8] N. R. Rodofile, K. Radke, and E. Foo, "Framework for SCADA cyber-attack dataset creation," in *Proc. Australas. Comput. Sci. Week Multiconference*, Jan. 2017, pp. 1–10, doi: 10.1145/3014812.3014883.

[9] A. Robles-Durazno, N. Moradpoor, J. McWhinnie, G. Russell, and J. Porcel-Bustamante, "Implementation and evaluation of physical, hybrid, and virtual testbeds for cybersecurity analysis of industrial control systems," *Symmetry*, vol. 13, no. 3, p. 519, Mar. 2021, doi: 10.3390/sym13030519.
[10] J. Crussell, T. M. Kroeger, A. Brown, and C. Phillips, "Virtually the same: Comparing physical and virtual testbeds," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2019, pp. 847–853.
[11] K. Mai, X. Qin, N. Ortiz Silva, and A. A. Cardenas, "IEC 60870-5-104 network characterization of a large-scale operational power grid," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2019, pp. 236–241, doi: 10.1109/SPW.2019.00051.
[12] A. Baiocco and S. D. Wolthusen, "Causality re-ordering attacks on the IEC 60870-5-104 protocol," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5, doi: 10.1109/PESGM.2018.8586010.
[13] C.-Y. Lin and S. Nadjm-Tehrani, "Protocol study and anomaly detection for server-driven traffic in SCADA networks," *Int. J. Crit. Infrastruct. Protection*, vol. 42, Sep. 2023, Art. no. 100612, doi: 10.1016/j.ijcip.2023.100612.
[14] Y. Shih, "Measured wind data in digital: Develop and optimize offshore wind farm SCADA by IEC 60870-5-104 protocol and DMZ," *Energy Rep.*, vol. 8, pp. 1231–1242, Apr. 2022, doi: 10.1016/j.egyr.2021.11.182.
[15] P. György and T. Holczer, "Attacking IEC 60870-5-104 protocol," in *Proc. CEUR Workshop*, vol. 2874, 2021, pp. 140–150.
[16] R. Hareesh, "Passive security monitoring for IEC-60870-5-104 based SCADA systems," *Int. J. Ind. Control Syst. Secur. (IJICSS)*, vol. 3, no. 1, pp. 90–99, 2020, doi: 10.20533/ijicss.9781.9083.20346.2020.0010.
[17] M. Egger, G. Eibl, and D. Engel, "Comparison of approaches for intrusion detection in substations using the IEC 60870-5-104 protocol," *Energy Informat.*, vol. 3, no. S1, pp. 2–17, Oct. 2020, doi: 10.1186/s42162-020-00118-4.
[18] C.-Y. Lin and S. Nadjm-Tehrani, "Understanding IEC-60870-5-104 traffic patterns in SCADA networks," in *Proc. 4th ACM Workshop Cyber-Phys. Syst. Secur.*, May 2018, pp. 51–60, doi: 10.1145/3198458.3198460.
[19] B. Hartpence and A. Kwasinski, "Combating TCP port scan attacks using sequential neural networks," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2020, pp. 256–260, doi: 10.1109/ICNC47757.2020.9049730.
[20] K. Hynek, "Refined detection of SSH brute-force attackers using machine learning," in *Proc. IFIP Int. Conf. ICT Syst. Secur. Privacy Protection*, vol. 2, 2020, pp. 49–63, doi: 10.1007/978-3-030-58201-2.
[21] E. H. Riyadi, A. E. Putra, and T. K. Priyambodo, "Improvement of nuclear facilities DNP3 protocol data transmission security using super encryption BRC4 in SCADA systems," *PeerJ Comput. Sci.*, vol. 7, p. e727, Nov. 2021, doi: 10.7717/peerj-cs.727.
[22] N. Almasalmeh, F. Saidi, and Z. Trabelsi, "A dendritic cell algorithm based approach for malicious TCP port scanning detection," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 877–882, doi: 10.1109/IWCMC.2019.8766461.
[23] S. Dwivedi, M. Vardhan, and S. Tripathi, "Defense against distributed DoS attack detection by using intelligent evolutionary algorithm," *Int. J. Comput. Appl.*, vol. 44, no. 3, pp. 219–229, Mar. 2022, doi: 10.1080/1206212x.2020.1720951.
[24] K. Soni and S. Singh, "A proposed DoS detection scheme for mitigating DoS attack," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 10, no. 4, pp. 172–179, 2020.
[25] C. Birkinshaw, E. Rouka, and V. G. Vassilakis, "Implementing an intrusion detection and prevention system using software-defined networking: Defending against port-scanning and denial-of-service attacks," *J. Netw. Comput. Appl.*, vol. 136, pp. 71–85, Jun. 2019, doi: 10.1016/j.jnca.2019.03.005.
[26] J. D. Gadze, A. A. Bamfo-Asante, J. O. Agyemang, H. Nunoo-Mensah, and K. A.-B. Opare, "An investigation into the application of deep learning in the detection and mitigation of DDOS attack on SDN controllers," *Technologies*, vol. 9, no. 1, p. 14, Feb. 2021, doi: 10.3390/technologies9010014.
[27] J. Qian, X. Du, B. Chen, B. Qu, K. Zeng, and J. Liu, "Cyber-physical integrated intrusion detection scheme in SCADA system of process manufacturing industry," *IEEE Access*, vol. 8, pp. 147471–147481, 2020, doi: 10.1109/ACCESS.2020.3015900.
[28] R. Banu, T. Jyothi, M. Amulya, K. N. Anju, A. Raju, and S. N. Kashyap, "MONOSEK—A network packet processing system for analysis & detection of TCP Xmas attack using pattern analysis," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 952–956, doi: 10.1109/ICCS45141.2019.9065325.
[29] C.-Y. Lin, A. Fundin, E. Westring, T. Gustafsson, and S. Nadim-Tehrani, "RICSel21 data collection: Attacks in a virtual power network," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2021, pp. 201–206, doi: 10.1109/SmartGridComm51999.2021.9632328.
[30] A. H. Dakheel, O. N. Ucan, O. Bayat, and H. H. Jasim, "Cyber attack detection in remote terminal unit of scada systems," *Int. J. Comput. Sci. Mobile Comput.*, vol. 8, no. 3, pp. 193–203, 2019.
[31] G. K. Bada, W. K. Nabare, and D. K. K. Quansah, "Comparative analysis of the performance of network intrusion detection systems: Snort, suricata and bro intrusion detection systems in perspective," *Int. J. Comput. Appl.*, vol. 176, no. 40, pp. 39–44, Jul. 2020, doi: 10.5120/ijca2020920513.
[32] M. Sun, Y. Lai, Y. Wang, J. Liu, B. Mao, and H. Gu, "Intrusion detection system based on in-depth understandings of industrial control logic," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2295–2306, Mar. 2023, doi: 10.1109/TII.2022.3200363.
[33] M. A. S. Arifin, D. Stiawan, Susanto, R. Budiarto, and M. Y. Idris, "Dataset for network intrusion detection system on SCADA IEC 60870-5-104," Zenodo, Palembang, Indonesia, Aug. 2022, doi: 10.5281/ZENODO.7034534.
[34] H. Ahmetoglu and R. Das, "A comprehensive review on detection of cyber-attacks: Data sets, methods, challenges, and future research directions," *Internet Things*, vol. 20, Nov. 2022, Art. no. 100615, doi: 10.1016/j.iot.2022.100615.
[35] P. Maynard, K. McLaughlin, and S. Sezer, "An open framework for deploying experimental SCADA testbed networks," in *Proc. Electron. Workshops Comput.*, 2018, pp. 89–98, doi: 10.14236/ewic/ics2018.11.
[36] P. Radoglou-Grammatikis, K. Rompolos, P. Sarigiannidis, V. Argyriou, T. Lagkas, A. Sarigiannidis, S. Goudos, and S. Wan, "Modeling, detecting, and mitigating threats against industrial healthcare systems: A combined software defined networking and reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 2041–2052, Mar. 2022, doi: 10.1109/TII.2021.3093905.
[37] M. Aamir, J. Poncela, M. A. Uqaili, B. S. Chowdhry, and N. A. Khan, "Optimal design of remote terminal unit (RTU) for wireless SCADA system for energy management," *Wireless Pers. Commun.*, vol. 69, no. 3, pp. 999–1012, Apr. 2013, doi: 10.1007/s11277-013-1060-9.
[38] M. Conti, D. Donadel, and F. Turrin, "A survey on industrial control system testbeds and datasets for security research," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2248–2294, 4th Quart., 2021, doi: 10.1109/COMST.2021.3094360.
[39] S. G. Abbas, F. Hashmat, G. A. Shah, and K. Zafar, "Generic signature development for IoT botnet families," *Forensic Sci. Int., Digit. Invest.*, vol. 38, Sep. 2021, Art. no. 301224, doi: 10.1016/j.fsidi.2021.301224.
[40] M. S. Kumar, J. Ben-Othman, K. G. Srinivasagan, and G. U. Krishnan, "Artificial intelligence managed network defense system against port scanning outbreaks," in *Proc. Int. Conf. Vis. Towards Emerg. Trends Commun. Netw. (ViTECoN)*, Mar. 2019, pp. 1–5, doi: 10.1109/ViTECoN.2019.8899380.
[41] E. V. Ananin, A. V. Nikishova, and I. S. Kozhevnikova, "Port scanning detection based on anomalies," in *Proc. Dyn. Syst., Mech. Mach. (Dynamics)*, Nov. 2017, pp. 1–5, doi: 10.1109/DYNAMICS.2017.8239427.
[42] A. Cetinkaya, H. Ishii, and T. Hayakawa, "An overview on denial-of-service attacks in control systems: Attack models and security analyses," *Entropy*, vol. 21, no. 2, p. 210, Feb. 2019, doi: 10.3390/e21020210.

M. AGUS SYAMSUL ARIFIN received the Ph.D. degree in computer from the Faculty of Engineering, Universitas Sriwijaya. He is currently a Senior Lecturer with the Department of Computer System Engineering, Faculty of Engineering, Universitas Bina Insan, Indonesia. His research interests include computer networks, intrusion detection/prevention systems, blockchain for data and information security, and SCADA network security.

DERIS STIAWAN received the Ph.D. degree in computer engineering from Universiti Teknologi Malaysia. He is currently a Professor with the Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya. His research interests include computer networks, intrusion detection/prevention systems, and heterogeneous networks.

BHAKTI YUDHO SUPRAPTO received the Ph.D. degree in electrical engineering from Universitas Indonesia. He is currently an Associate Professor with the Electrical Department, Faculty of Engineering, Universitas Sriwijaya, Indonesia. His professional profile has derived to robotic and control, which focused on, fuzzy logic and neural networks. His research interest includes control systems.

SUSANTO received the Ph.D. degree in computer from the Faculty of Engineering, Universitas Sriwijaya. He is currently a Senior Lecturer with the Department of Computer System Engineering, Faculty of Engineering, Universitas Bina Insan, Indonesia. His research interests include cryptography, information technology, information security, and network security.

TASMI SALIM (Member, IEEE) received the master's degree in computer science from Universitas Sriwijaya, Indonesia, where he is currently pursuing the Ph.D. degree with the Faculty of Engineering. He is a Lecturer with the Faculty of Computer Science, Universitas Indo Global Mandiri. His research interests include computer networks, network security, intrusion detection systems, SCADA security, and machine learning.

MOHD YAZID IDRIS received the M.Sc. degree in software engineering and the Ph.D. degree in information technology (IT) security, in 1998 and 2008, respectively. In software engineering, he focuses on the research of designing and development of mobile and telecommunication software. His main research activity in IT security is in the area of intrusion prevention and detection (IPD). He is currently an Associate Professor with the School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia.

MOHAMED SHENIFY (Member, IEEE) received the B.Sc. degree in computer science from Indiana State University, Terre Haute, IN, USA in May 1990, the M.Sc. degree in computer science from Ball State University, Muncie, IN, USA, in December 1991, and the Ph.D. degree in computer science from Illinois Institute of Technology, Chicago, IL, USA, in May 1998. He is currently an Associate Professor with the College of Computing and Information, Al-Baha University. He is also the Supervisor of the Administration for International Cooperation and Knowledge Exchange, Al-Baha University. His research interests include natural language processing, information retrieval, healthcare systems, fuzzy logic, and computing education. He is an active member of the Association for Computing Machinery (ACM) and the Association for Computational Linguistics (SIGDAT). He has been the Steering Committee Member of the International Conference on Learning and Teaching in Computing and Engineering (LaTiCE), since its establishment in the year 2013.

RAHMAT BUDIARTO received the M.Eng. and Dr.Eng. degrees in computer science from Nagoya Institute of Technology, Japan, in 1995 and 1998, respectively. He was chairing the APAN Security Working Group, from 2006 to 2009, and established the IPv6 Research Center (NAv6 Center), Universiti Sains Malaysia (USM), in 2005, where he was appointed as the Deputy of the Center, from 2005 to 2009. He is currently a Full Professor with the Department of Computer Science, Al-Baha University, Saudi Arabia. His research interests include intelligent systems, brain modeling, IPv6, network security, wireless sensor networks, and MANETs.