# SCADA-Wireshark Testbed Data-Based Exploratory Data Analytics and Intrusion Detection

**Conference Paper** · May 2024

DOI: 10.1109/SmartNets61466.2024.10577709



**2 authors:**

Hillol Biswas

**21** PUBLICATIONS **24** CITATIONS

SEE PROFILE

Muthyala Manoj Kumar

WAPCOS LTD

**4** PUBLICATIONS **5** CITATIONS

SEE PROFILE

# SCADA-Wireshark Testbed data-based Exploratory Data Analytics and Intrusion Detection

Hillol Biswas
*Power Division*
*WAPCOS Limited*
Gurgaon, India
0000-0001-5451-4515

Muthyala Manoj Kumar
*Power Division*
*WAPCOS Limited*
Gurgaon, India
0009-0005-9167-5218

***Abstract*— The modern power grid is active with a bi-directional flow of energy and stream data. The increased size and complexity of the power transmission network landscape inevitably led to vulnerability to cyber-attacks in these critical infrastructures. Since 2007, with the simulation of the Aurora attack, cyber security preparedness traveled a significant way to safeguard the electricity networks with the advent of sophistication in IEC 61850-based communication systems and SCADA. The contemporary wide area monitoring system (WAMS) that monitors electrical power system state health in near real-time entails collecting and managing enormous stream data. This exposes the system to further attack vulnerability. Using an open-source tool like Wireshark in the OT cybersecurity regime provides a real-world scenario based on different protocols. This paper discusses some realistic challenges for SCADA exploratory data analytics based on PNML testbed-generated Wireshark data.**

***Keywords*— Critical Infrastructure, SCADA, cybersecurity, EDA***

## I. INTRODUCTION

The cyberattack landscape of SCADA includes noteworthy incidents, viz. Utility SCADA attack in the USA in 2001 and. However, the Ukraine Grid attack in 2015 provided a broader scope of analysis by CISA [1]. An interagency team consisting of representatives from the Department of Energy, the Federal Bureau of Investigation, the Industrial Control Systems Cyber Emergency Response Team (ICS-CERT), the National Cybersecurity and Communications Integration Center (NCCIC)/Industrial Control Systems Cyber Emergency Readiness Team (US-CERT), and the North American Electric Reliability Corporation investigated the December 2015 cyberattack on the Ukraine power network. The team discovered that power outages affecting approximately 225,000 customers were caused by remote cyber intrusions at three regional electric power distribution companies (Oblenergos). Although power had been restored, all affected the Oblenergos.

Additionally, three other organizations, some from other critical infrastructure sectors, were still operating under restricted capacity for a while. According to Mandiant [2] recently, the Sandworm attack in Ukraine in late 2022 was a unique one reported; given the Sandworm's global threat activity and the widespread deployment of Micro SCADA products, asset owners worldwide should take action to mitigate their tactics and techniques and procedures against IT and OT systems. This attack immediately threatened critical infrastructure environments in Ukraine that leveraged the Micro SCADA supervisory control system. It remained unclear how Sandworm obtained initial access to the victim. Sandworm was initially detected in the victim's environment in June 2022 when the actor installed the Neo-REGEORG web shell on an internet-facing server. This aligns with the group's previous activities of searching for and taking advantage of internet-facing servers for initial access. Approximately one month later, Sandworm installed GOGETTER, a Golang tunneler that proxies communications for its command and control (C2) server through the open-source library Yamux via TLS. When utilizing GOGETTER, Sandworm employed a System service unit to sustain system persistence. This system service unit permits a program to be run under specific conditions; in this case, it was utilized to execute the GOGETTER binary [2]. The attacks could be more sophisticated in contemporary times, taking advantage of available tools and technological advancement due to modern SCADA systems' increased size and complexity.

The Aurora attack was tested in DOE Idaho Lab in 2007 and found potentially dangerous for rotating machinery operating on the electrical grid under specific conditions. The attack involved opening and closing a circuit breaker or breaker, which created an out-of-synchronism condition that could harm rotating equipment attached to the grid. The attack aimed to purposefully open a breaker and close it out of synchronism to inflict damage on the connected generators and motors. The attack also assumed that these relays could be compromised to achieve their intended purpose. When an out-of-synchronism close is started, the high electrical torque results in stress on the rotating equipment's mechanical shaft [3].

Master Terminal Units (MTUs), Programmable Logic Controllers (PLCs), Remote Terminal Units (RTUs), Intelligent Electronic Devices (IEDs), Human Machine Interfaces (HMIs), and Communication Media are the critical parts of a typical/traditional SCADA system [4]. Increased DER integration to improve grid stability and market efficiency led to the creation of the Transactive Energy System (TES), a new framework for power system operation and control [5]. Malicious activities are directed towards the TES's physical and cyber layers. Three data sets are typically produced: normal, component outages, and attack. Every anomaly scenario is known in a perfect world, making supervised learning an effective way to solve the problem. In reality, though, most cyberattacks remain unknown until they are first observed. Overall, the size and complexity of SCADA and associated cybersecurity is only increasing daily.

The smart grid optimization pattern based on time granularity typically comprises optimal power flow, scheduling, and planning regime, which spreads from minutes to hours to days to months and years [6]. Moreover, the contemporary system uses communication networks with either wired viz power line communication (PLC), broadband, or wireless technology viz ZigBee, which have their respective pros and cons in the critical infrastructure

979-8-3503-8532-8/24/$31.00 ©2024 IEEE

deployment. Moreover, the architecture of SCADA is categorized as monolithic, distributed, network-based, and IOT-based. The corresponding communication protocol is further subdivided as Modbus, DNP3, IEC 6870-5-101, foundation fieldbus, Profitbus, and recent IEC 61850 based communications [7].

A limited approach has been reported for smart grid intrusion detection systems. Critical infrastructures, such as smart grids (SG), will likely be susceptible to more sophisticated unknown attacks shortly. While anomaly-based intrusion detection systems (IDS) can detect novel attacks, their deployment in industry is limited due to high false positive rates and uninterpretable trained models [8].

### *A. Vulnerability in smart grid*

Wasumwa categorized [9] The vulnerabilities in a smart grid that can compromise the integrity, availability, and confidentiality as Insecure communication networks, Weak authentication and access control, Lack of security monitoring and incident response, Vulnerable software and firmware, Physical security risks, Lack of secure firmware and software updates, Distributed energy resources (DERs), Insider threats, Third-party integration, and supply chain risks.

The absence of datasets to validate the functionality of the IDSs' algorithms hinders their research efforts [7]. Moreover, anomaly detection techniques like local outlier factors or one-class support vector machines should be prioritized. Nevertheless, current anomaly detection techniques using supervised techniques/machine learning have several drawbacks, including i) Typically, existing techniques cannot handle high-dimensional data problems (in this case, each training sample is a multivariate time series with thousands of dimensions); ii) Typically, existing techniques cannot distinguish between anomalous events like component outages and cyberattacks, which are both classified as anomalous events) [5].

Siemens s7-based SCADA intrusion detection system using deterministic finite automation was reported in 2014 [10]. Shitharth et al. proposed a work based on selecting optimal parameters to improve the security of SCADA network architecture. Here, the RPCO technique is implemented to select the most suitable parameters by reducing the set of features, which helps reduce the classifier's training time. The main parameters considered during this process are entropy, output weights of sensors, energy properties, trust ratings, Lebesgue measure, probability density of sequence, overall power, likelihood ratio, average sigma parameter, and amount of packets transmitted. Based on the optimized set of features, the BCNN classification technique accurately predicts the attacks from the provided dataset [11]—the preprocessing technique followed by optimization-based selection of features and subsequent classification.

Reviews exclusively for PLC-based attacks are also available wherein attacks, intrusion detection techniques, digital forensics, vulnerabilities, and upcoming projects from the perspectives of the system and the core component levels have been provided [12].

### *B. Indicators of Compromise in SCADA*

The indicator of compromise in industrial control systems (ICS), as Asiri et al. [13] tabulated 13 types bearing different hallmarks. The Control Logic Modification, Unsupported or Unusual Function Code, and Mismatch between the Historian and Control Logic are applicable, especially for SCADA, and could suggest anomaly detection.

### *C. Cyber-attack landscape in SCADA*

One challenge and open issue in the SCADA data cybersecurity study flagged up is further dataset development, which is not widely available. Traditional SCADA systems have a static, centralized architecture. While vulnerability databases like NVD and CVE are helpful, they must pay more attention to the SCADA system [14].

Different testbeds, viz. physical, virtual, virtual-physical, and hybrid, are used for simulation purposes [7]. Wireshark [15],[16]–[18] is an open-source tool often used by cybersecurity researchers in the SCADA environment of the utility sector.

Intrusion detection systems (IDS) are typically separated into two categories: Host-Based Intrusion Detection Systems (HIDS) and Network Intrusion Detection Systems (NIDS) based on the information source. HIDS depends on host activity and state information, such as file system updates and application logs [7]. A signature-based intrusion detection technique, usually used to identify known threats, works with a list of known threats (such as misuse patterns) and a programmed indicator of compromise. It can achieve high accuracy but cannot detect future attacks because different known threats do not exhibit the same attack signs. In anomaly-based intrusion detection techniques, the system compares network traffic regularly and raises alarms if extremely abnormal behavior appears. The distinctive normal model is trained and learned through statistical and mathematical techniques [14]. While the latter looks for activity that deviates from an expected pattern or a preset model of the system's typical behavior, the former looks for an attack whose signature is already known [4]. Because hackers and malicious actors can use AI to create more sophisticated cyberattacks, get around security measures, and take advantage of weaknesses in systems, the use of AI is becoming increasingly risky in terms of security issues. The emergence of AI-driven autonomous weaponry also raises concerns about the risks of rogue states or non-state actors using this technology, particularly when considering the potential loss of human control in crucial decision-making processes [19].

Supervised and unsupervised machine learning techniques are being used to study intrusion detection systems. Compared to unsupervised classification and feature selection approaches, researchers have worked hard to develop advanced supervised classification approaches, which have higher performance and accuracy than unsupervised classification, leading to the widespread use of supervised machine learning techniques across a wide range of SCADA-specific IDS approaches. The number of publications in classification/supervised, clustering/unsupervised, and feature selection has grown over time [4], [20].

Deep learning using normal data [5] has been previously proposed. The model's input includes all the output variables from the simulation, bid prices, and bid quantities. This SAE is trained using 418 days of standard data to model the normal activities, and hyperparameters are tuned by the validation dataset, another group of normal data (103 days). To detect the anomalies, we propose two detection rules. The detection of a component outage is based on the reconstruction error of the

output from 4 generators of the normal data; and the detection of an attack is based on the reconstruction error of the bid prices and quantities of the normal data. The validation dataset determines thresholds for both rules separately, so the percentage of normal data under those two rules is approximately 96%. To estimate the performance of the model, 130 days of normal data was used. Because SAE can receive raw data as input immediately and does not require extra work to generate features, it can be used in various contexts.

SCADA complexities and vulnerabilities associated are manifolds. Serial links are still used by many legacy substations and distribution communication systems for a variety of functions, i.e., communications between distribution field equipment and control centres. Moreover, a lot of the serial protocols that are now in use transmit messages in cleartext and lack measures to safeguard the confidentiality or integrity of the information compared to TCP/UDP. Discovering unexplored vulnerabilities in the system is likely to be one of the key aspects equally appealing to both attackers and defenders of the cyber-physical world of power grids. Handling specialized protocols like Modbus, DNP3, IEC 61850, etc., is one problem unique to power systems, and standardized IDS and security event detection and management models should be developed for these protocols and systems. To be more precise, these models should be able to identify situations in which aberrant orders could have unanticipated and unwanted effects by representing a comprehensive contextual awareness of device function and status [21].

In network intrusion detection, signature-based and anomaly-based detection are usually adopted [22]. In the case of anomaly detection, both classification and clustering approaches have been inculcated among scholars. However, two aspects become prominent when taking cognizance in real-world scenarios, i.e., availability of labelled data and data imbalance. In the cyber security ecosystem, future attacks can be completely unknown, thus limiting the usage of classification, i.e., the scope of the labelled data based supervised technique. Moreover, compared to the vast amount of normal data, the attack data, even if made available, is likely to show far less in numbers, thus limiting the scope of using the technique further. Though synthetic data augmentation, among others, is one of the convenient techniques being used, however, there is a curious question of whether oversampling and under-sampling will fetch the similar result, especially when compared with many machine learning algorithms. In comparison, the clustering of data approach entails training with the normal data to detect any attack data as an anomaly, depending upon the performance. Depending upon the size of training datasets, deep learning techniques generally require extensive time and stand out as a good choice over machine learning.

The contribution of this paper is the study of a PNML [23] open-source testbed simulated SCADA-Wireshark dataset that comprises millions of data using Wireshark software. This Wireshark tool has been previously cited for network protocol analysis [24]. However, a specific data-driven approach and further demonstration of context using exploratory data analytics to understand a real-world scenario is a novelty.

## II. Methods

The overall methodology consists of proposing our approach, selecting target data, preprocessing the data, including anomaly detection, scaling the data, balancing the dataset, and presenting data visualization and comparison. The workflow depicts the data preprocessing, model selection, building model, training and testing for validation purposes, evaluation of performance by employing appropriate matrices, and further display of the results for easy visual comparison purposes. Based on the related works, we frame the questions of what kind of dataset and size would be appropriate for specific purpose machine learning applications, what algorithms and corresponding evaluation metrics have been used, and what the outcomes are. We identified the PNNL dataset [23], the day 1 data itself comprises of over five million samples. The results of high-fidelity hardware in the loop experimentation on simulated models of representative electric and natural gas distribution systems with real cyberattack test cases are included in this dataset. These datasets are crucial for assessing new intrusion detection techniques and comprehending the system's behavior under different attack scenarios. The dataset comprises normal-day data as well as various attack test cases. We used the average day data for training and one case attack data for anomaly detection. We selected k-means clustering using scikitlearn [25] to group the normal data to have a flavor of the patterns in the data. After that, the autoencoder model in Keras Google Colab [26] for performing the deep learning of the training dataset. The pcap dataset initially comprises seven columns of variables, viz number, time, source, destination, protocol, frame length, and info. However, during the inspection, it was found that the Source and Destination variables are further feasible to restrict as a port basis. Hence, Source Port and Destination Port columns were added without compromising the data fabric in the Wireshark environment. After that, the CSV files were imported into Colab using the Pandas library.Deep learning autoencoder models are constructed using the Python Keras package in a Google Colab environment, with respective user parameters having predetermined or default values. Recurrent linear Unit (Relu) for encoding and sigmoid for decoding, as activation unit and ADAM (Adaptive Moment Estimation) optimizer have been chosen as the deep learning optimizer from those offered by Keras (Keras Optimizers, 2022). The technique also considers the learning rate, callback, and reduced learning rate (ReduceLROnPlateau). In equation (1), metrics represent the expected, actual, and observed values as mean squared error (MSE), Pj, Aj, and n.

$$\mathrm{MSE} = \frac{1}{n}(Pj - Aj)^2 \tag{1}$$

## III. Results and Discussion

The data visualization of the day one training data is provided in subsequent Figures. Figures 1 and 2 depict the different protocols used for the testing and the corresponding bar chart for the day one normal and selected test case attack data set. Figure 3 provides the bar chart representation of the port-wise segregation group based on the source and destination ports.

TABLE I. COMPARISON OF UNIQUE COUNTS FOR PORTS-PROTOCOLS

| **Uniqueness** | **Protocol and Port** | | |
|---|---|---|---|
| | ***Source*** | ***Destination*** | ***Protocol*** |
| Unique Count | 106 | 86 | 28 |
| | **Source Port** | **Destination Port** | **Frame length** |
| Unique Count | 26017 | 26022 | 351 |

Table I looks at the significant uniqueness of the day 1 data. The unique protocols type are 'TCP,' 'Modbus/TCP,' 'DNP 3.0', 'STP,' 'SSDP,' 'ICMPv6', 'ARP,' 'DHCP,' 'UDP,' 'LLDP,' 'DHCPv6', 'CLNP,' 'ESIS,' 'HTTP,' 'SMB Mailslot,' 'LLMNR,' 'NBNS,' 'DTP,' 'NTP,' 'BROWSER,' 'SSH,' 'TELNET,' 'HTTP/XML,' 'ICMP,' 'IGMPv3', 'MDNS,' 'TLSv1.2', 'FTP.'

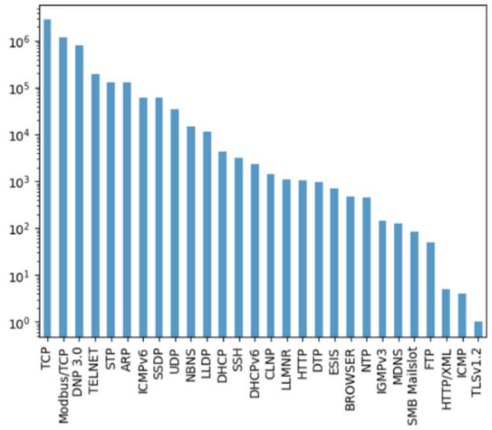


Fig. 1. The Day 1 data protocols bar chart

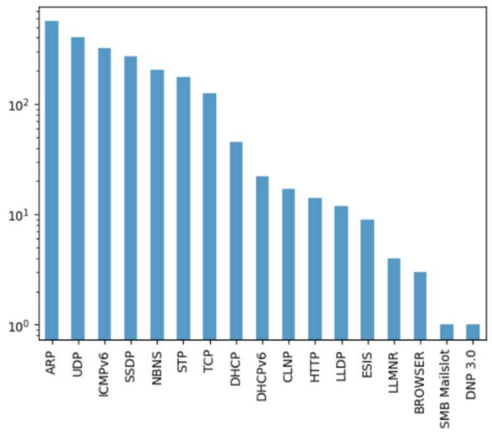


Fig. 2. The test case 5 data protocols bar char

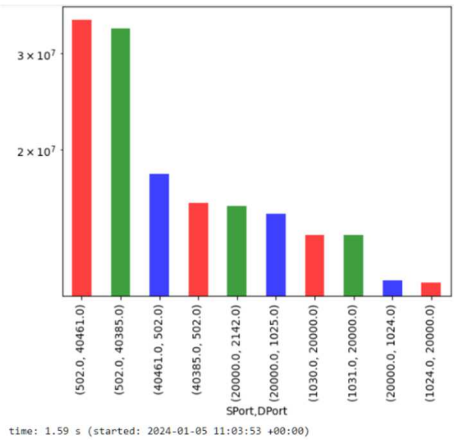


Fig. 3. Source Port to Dest Port frame length bar plot on the day 1 dataset

Figs. 1 and 2 provide an easy comparison of significant variations of protocols for day one and selected test case data.

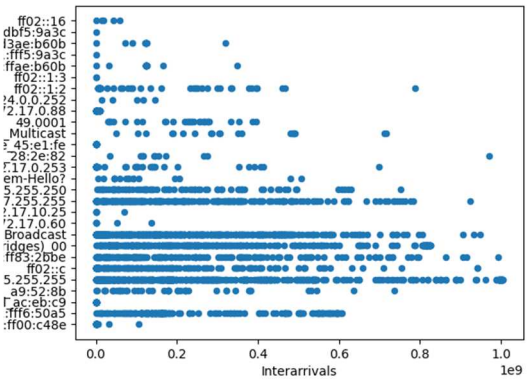


Fig. 4. IP address-wise time series plot for a typical few instances of normal source data

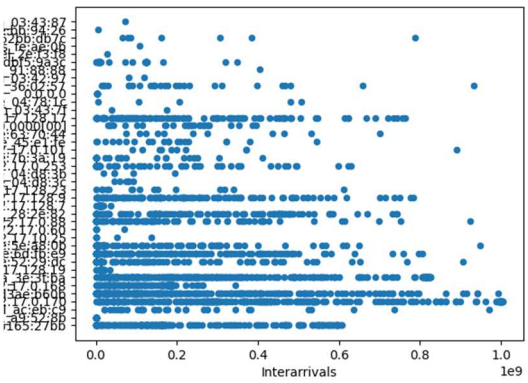


Fig. 5. IP address-wise time series plot for a typical few instances of normal destination data

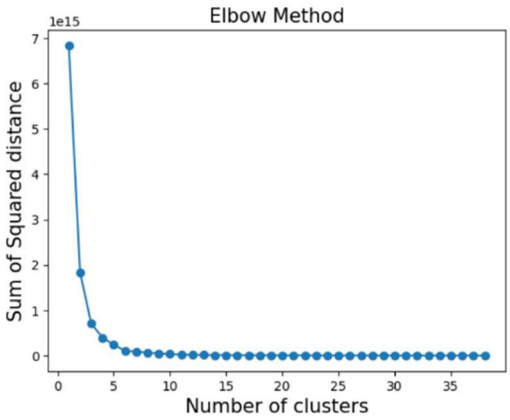


Fig. 6. Various clusters formed on the day 1 dataset

The k-means clustering-based Custer labels and size after a run are provided in Figure 6. Depending upon the size, the various labeling forms by the clustering are labeled 1 - 1357430, 0 – 1248277, 2 – 1139762, 3 - 753317, and 4 689470, respectively.

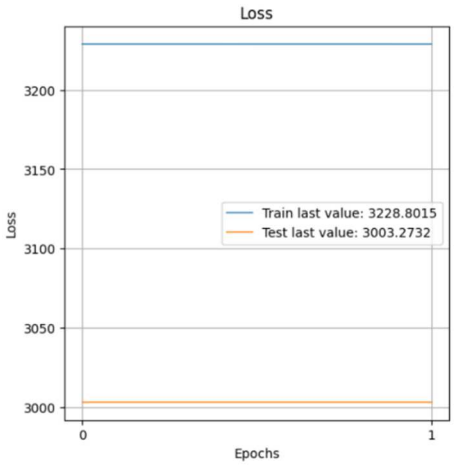


Fig. 7. The history plot of loss and epoch during training of the dataset

During training by an autoencoder model comprising an encoder and decoder for the unlabelled data, the history command in Keras keeps track of the loss and epoch-wise performance and the learning rate, as provided in Figures 7 and 8. Whereas, the fig 9 depicts a mean squared error table yielded in python kernel.

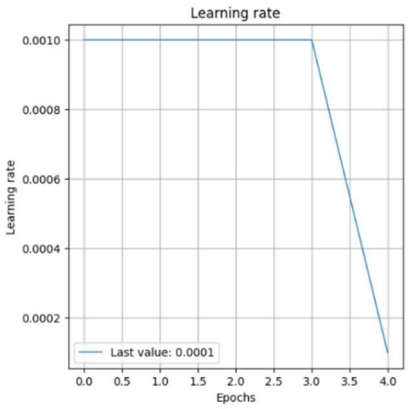


Fig. 8. Learning rate during the training process

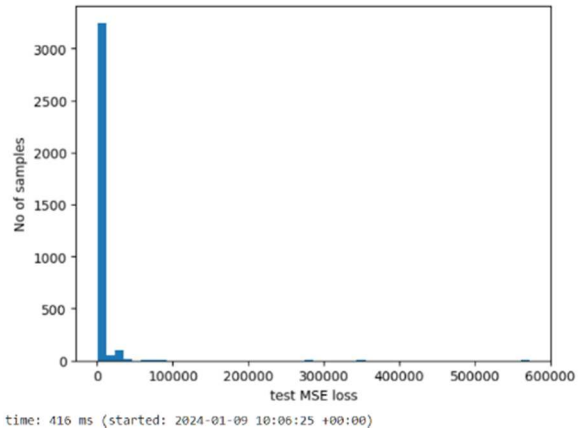


Fig. 9. The test case means squared error plot.

```
Number of anomaly samples:  21
Indices of anomaly samples:  (array([  27,   28,   29,  345,  346,  347,  674,  675,  676,  991,  993,
        994, 1314, 1315, 1316, 1674, 1675, 1676, 2058, 2059, 2060]),)
time: 49.8 ms (started: 2024-01-09 05:42:23 +00:00)
```

Fig. 10. The number of anomalies based on a threshold of MSE on the test data

It appears from Figure 10 that the number of anomalies detected in the test case after training the autoencoder model is 21, and the position corresponds to the detected events. Typically, in network analysis viz. python-based forensics, the source port, destination port, source and destination IP address, and protocol are usually considered. However, in real-world scenarios like SCADA-based communication through IEC 61850 for power grid operation, the complexity entailing manufacturer message specification (MMS), generic object-oriented substation events (GOOSE), application service data unit (ASDU), cause of transmission (COT), time synchronization further increases the complexity at tandem with growing sophistication in recent OT malware viz. INDUSTROYER [27], INDUSTROYER2 [28], and COSMIC ENERGY[29].

The study discusses the complexity of applying EDA and any deep learning technique vis-à-vis the trustworthiness of SCADA data in a real-world scenario. By its merit, unlike IT enviornment, SCADA data for power grid critical infrastructures comprises many protocols, ports, and IP addresses, which Wireshark captures along with the respective packets. At this juncture, EDA provides a reasonable scope for first understanding the extent of domain knowledge in OT endeavour. Depending on the complexity, any potential cyberattack can be novel and overwhelming, which might fall well outside the scope of any pre-trained data. Hence, anomaly detection based on normal data is justifiable as a baseline study reasonable candidate in this ever-growing complex albeit vulnerable crossover domain.

## IV. Conclusion

Anomaly detection is one of the vitally desirable phenomena in streaming data, especially where SCADA is put into operation on the electricity grid. While the academic comparison of the advantages of supervised and unsupervised machine/deep learning techniques is an inculcative approach, unlabeled data entails the practical, real-world scenario coupled with the basic properties of big data, i.e., volume, velocity, variety, and veracity. One key constraint is dataset availability and the need for updated data. Our approach of EDA followed by auto-encoder-based anomaly detection on SCADA-Wireshark data research data appears encouraging for further contemplation of future research direction for deployment and subsequent machine learning operation stages to gain meaningful insight into the real-world scenario of one of the critical infrastructures. As one of the key aspects highlighted by various scholars in cyber security research is the availability of updated data. While some suitable datasets are available, the scope becomes further narrower in case of energy networks coupled with SCADA communication. The PNML dataset is quite significant in volume and further research in this direction is encouraging in near future. This study will also be extended in the future with other machine learning and deep learning techniques to compare any trade-off for performance vis-a-vis computation time.

## Acknowledgment

We acknowledge the support and encouragement of everyone in the Power division who enabled us to develop this article.

## REFERENCES


[1] "Cyber-Attack Against Ukrainian Critical Infrastructure | CISA." Accessed: Jan. 07, 2024. [Online]. Available: https://www.cisa.gov/news-events/ics-alerts/ir-alert-h-16-056-01

[2] "Sandworm Disrupts Power in Ukraine Using a Novel Attack Against Operational Technology | Mandiant." Accessed: Dec. 14, 2023. [Online]. Available: https://www.mandiant.com/resources/blog/sandworm-disrupts-power-ukraine-operational-technology

[3] D. Salmon, M. Zeller, A. Guzman, V. Mynam, and M. Donolo, "Mitigating the aurora vulnerability with existing technology," *36th Annu. West. Prot. Relay Conf.*, no. October 2009, pp. 1–7, 2009.

[4] J. Suaboot *et al.*, "A Taxonomy of Supervised Learning for IDSs in SCADA Environments," *ACM Comput. Surv.*, vol. 53, no. 2, 2020, doi: 10.1145/3379499.

[5] Y. Zhang *et al.*, "Cyber Physical Security Analytics for Transactive Energy Systems," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 931–941, 2020, doi: 10.1109/TSG.2019.2928168.

[6] K. A. Abdulsalam, J. Adebisi, M. Emezirinwune, and O. Babatunde, "An overview and multicriteria analysis of communication technologies for smart grid applications," *e-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 3, no. January, p. 100121, 2023, doi: 10.1016/j.prime.2023.100121.

[7] G. Yadav and K. Paul, "Architecture and security of SCADA systems: A review," *Int. J. Crit. Infrastruct. Prot.*, vol. 34, 2021, doi: 10.1016/j.ijcip.2021.100433.

[8] Q. Liu, V. Hagenmeyer, and H. B. Keller, "A Review of Rule Learning-Based Intrusion Detection Systems and Their Prospects in Smart Grids," *IEEE Access*, vol. 9, pp. 57542–57564, 2021, doi: 10.1109/ACCESS.2021.3071263.

[9] Sharmwey A. Wasumwa, "Safeguarding the future: A comprehensive analysis of security measures for smart grids," *World J. Adv. Res. Rev.*, vol. 19, no. 1, pp. 847–871, 2023, doi: 10.30574/wjarr.2023.19.1.1387.

[10] A. Kleinmann and A. Wool, "Accurate Modeling of the Siemens S7 SCADA Protocol for Intrusion Detection and Digital Forensics," *J. Digit. Forensics, Secur. Law*, vol. 9, no. 2, 2014, doi: 10.15394/jdfsl.2014.1169.

[11] S. Shitharth, K. M. Prasad, K. Sangeetha, P. R. Kshirsagar, T. S. Babu, and H. H. Alhelou, "An Enriched RPCO-BCNN Mechanisms for Attack Detection and Classification in SCADA Systems," *IEEE Access*, vol. 9, pp. 156297–156312, 2021, doi: 10.1109/ACCESS.2021.3129053.

[12] Z. Wang, Y. Zhang, Y. Chen, H. Liu, B. Wang, and C. Wang, "A Survey on Programmable Logic Controller Vulnerabilities, Attacks, Detections, and Forensics," *Processes*, vol. 11, no. 3, pp. 1–28, 2023, doi: 10.3390/pr11030918.

[13] M. Asiri, N. Saxena, R. Gjomemo, and P. Burnap, "Understanding Indicators of Compromise against Cyber-attacks in Industrial Control Systems: A Security Perspective," *ACM Trans. Cyber-Physical Syst.*, vol. 7, no. 2, 2023, doi: 10.1145/3587255.

[14] M. Alanazi, A. Mahmood, and M. J. M. Chowdhury, "SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues," *Comput. Secur.*, vol. 125, 2023, doi: 10.1016/j.cose.2022.103028.

[15] U. Banerjee, A. Vashishtha, and M. Saxena, "Evaluation of the Capabilities of WireShark as a tool for Intrusion Detection," *Int. J. Comput. Appl.*, vol. 6, no. 7, pp. 1–5, 2010.

[16] P. Saxena and S. K. Sharma, "Analysis of network traffic by using packet sniffing tool: Wireshark," *Int. J. Adv. Res. Ideas Innov. Technol*, vol. 3, no. 6, pp. 804–808, 2017.

[17] H. Iqbal and S. Naaz, "Wireshark as a tool for detection of various LAN attacks," *Int. J. Comput. Sci. Eng*, vol. 7, no. 5, pp. 833–837, 2019.

[18] P. Goyal and A. Goyal, "Comparative study of two most popular packet sniffing tools-Tcpdump and Wireshark," in *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, 2017, pp. 77–81.

[19] "The 15 Biggest Risks Of Artificial Intelligence." Accessed: Jan. 09, 2024. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=293f34cb2706&is=3276fff00cc835609bda5b2eef99399740dcc4413006949c20153bdfd6e79b65

[20] S. J. Pinto, P. Siano, and M. Parente, "Review of Cybersecurity Analysis in Smart Distribution Systems and Future Directions for Using Unsupervised Learning Methods for Cyber Detection," *Energies*, vol. 16, no. 4, 2023, doi: 10.3390/en16041651.

[21] NISTIR 7628, "NISTIR 7628 Guidelines for Smart Grid Cyber Security, Revision 1," *Nist*, vol. 3, p. 187, 2014, [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7628r1.pdf

[22] K. Scarfone and P. Mell, "Guide to Intrusion Detection and Prevention Systems (IDPS)," *Natl. Inst. Stand. Technol.*, vol. 800–94, no. February, p. 127, 2007, [Online]. Available: http://csrc.ncsl.nist.gov/publications/nistpubs/800-94/SP800-94.pdf

[23] "Electricity and Gas IDS | Datahub." Accessed: Jan. 07, 2024. [Online]. Available: https://data.pnnl.gov/group/nodes/dataset/13470

[24] "Wireshark · Go Deep." [Online]. Available: https://www.wireshark.org/

[25] "Scikit learn." [Online]. Available: https://scikit-learn.org/stable/

[26] "Keras Optimizers."

[27] A. Cherepanov, "WIN32/INDUSTROYER: A new threat for industrial control systems," *Eset*, p. 17, 2017, [Online]. Available: https://www.welivesecurity.com/wp-content/uploads/2017/06/Win32_Industroyer.pdf

[28] "Industroyer2: Industroyer reloaded." Accessed: Feb. 12, 2024. [Online]. Available: https://www.welivesecurity.com/2022/04/12/industroyer2-industroyer-reloaded/

[29] "COSMICENERGY: New OT Malware Possibly Related To Russian Emergency Response Exercises | Mandiant." Accessed: Feb. 12, 2024. [Online]. Available: https://www.mandiant.com/resources/blog/cosmicenergy-ot-malware-russian-response