DNP3 Intrusion Detection

Emel Tuğçe Kara
Engineering Department
Düzce University
Düzce, Türkiye
emeltugcekara@gmail.com

Riad Memmedli
Engineering Department
Düzce University
Düzce, Türkiye
riadmammadli2003@gmail.com

Abstract
Smart grids face growing cybersecurity challenges because their complex nature depends on the Distributed Network Protocol (DNP3). This research investigates the DNP3 Intrusion Detection Dataset through an evaluation of machine learning approaches described in the literature for protecting the DNP3 protocol from cyber threats. The research combines a detailed review of current literature with best practices to determine effective security measures for critical infrastructure protection.

I. Introduction

The technology of Supervisory Control and Data Acquisition (SCADA) operates as a fundamental component for critical infrastructure including electricity grids and water plants [6]. The Distributed Network Protocol 3 (DNP3) serves as a popular choice for these systems because it provides flexibility and enables extended-range communication [1], [5]. The original design of DNP3 lacked strong security features which makes the protocol vulnerable to different types of cyberattacks [1], [4], [7]. The research examines an open DNP3 Intrusion Detection dataset while conducting a thematic literature review of machine learning solutions that detect and counter these attacks as described in [2], [3], [6], [8].

II. Dataset Overview

The DNP3 Intrusion Detection Dataset serves as the research dataset which IEEE DataPort provides to the University of Louisiana at Lafayette. The dataset emerged from rising cybersecurity threats against Industrial Control Systems (ICS). The operation of critical infrastructure including power grids and water supply systems and transportation networks depends on fundamental Industrial Control Systems. The adoption of digital communication protocols such as Distributed Network Protocol version 3 (DNP3) by these environments has led to a substantial increase in cyber threats. Researchers can develop and test machine learning models for network intrusion detection through this dataset which simulates DNP3-based network malicious activity.

A. Source and Purpose

The dataset originated from a laboratory simulation of industrial control system environments. The dataset contains normal and attack traffic which demonstrates the communication between Human Machine Interfaces (HMIs), Remote Terminal Units (RTUs), and Programmable Logic Controllers (PLCs). The main purpose of this dataset is to help developers create IDS systems which detect DNP3 communication-based cyberattacks including Man-in-the-Middle (MitM), command injection, reconnaissance and Denial of Service (DoS). The attacks were executed through internal and external threat actor models to demonstrate actual operational risks.

B. Relevance to Cybersecurity

The DNP3 Intrusion Detection Dataset directly addresses the vulnerability of SCADA systems to cyberattacks. Traditional IT security solutions often fail to consider the unique constraints and operational demands of ICS, such as real-time performance and protocol-specific behavior. Therefore, specialized datasets like this are critical for training and validating cybersecurity solutions tailored to industrial settings. By providing labeled data on both normal operations and a variety of realistic attack scenarios, this dataset bridges the gap between academic research and practical security requirements in cyber-physical systems.

C. Structure and Features of the Dataset

The dataset consists of packet-level network traffic data captured using Wireshark and Zeek tools and then converted into CSV format for easier analysis. Each row is a network flow and contains a timestamp, source and destination IP addresses, source and destination ports, protocol type, and other flow-level metadata. It also contains DNP3-specific fields, such as function codes, which are important for understanding the behavior of messages in ICS. The most valuable component is the inclusion of a labeled field that is benign or malicious and if malicious, the type of attack.

D. Technical Specificities and Use for ML

The dataset provides detailed information which makes it appropriate for statistical feature engineering and advanced machine learning approaches. The labeled records enable supervised learning approaches while the temporal and behavioral patterns in the flows support unsupervised anomaly detection. The extensive feature set allows researchers to derive advanced behavioral indicators needed for detecting complex multi-stage attacks. The dataset can be divided into training and testing sets without extensive preprocessing which makes it easy to integrate into experimental pipelines.

III. Literature Review

As the DNP3 protocol becomes increasingly central to Supervisory Control and Data Acquisition (SCADA) systems in critical infrastructures, researchers have explored various Machine Learning (ML) and Deep Learning (DL) techniques to enhance the detection and mitigation of cyber threats targeting DNP3 communication. This section presents a comprehensive review of relevant studies, categorized by approach and highlighting insights, methodologies, and best practices.

A. Study

The research by Darwish et al. (Igbe, Celebi, Saadawi) [1] demonstrated attack scenarios through the DETER testbed to expose DNP3 vulnerabilities from unsolicited message attacks and dataset injections. The authors suggested deploying host-based intrusion detection agents directly on each Intelligent Electronic Device (IED) to address the issue. The research demonstrated actual weaknesses in DNP3 operation yet did not include automatic threat identification features.

Bai together with Hariri and Al-Nashif [7] developed an Autonomic Network Protection Framework which operates on TCP/IP to protect DNP3 networks. The rule-based anomaly detection system performs DNP3 packet parsing to identify both unauthorized and malformed network communications. The system demonstrates strong effectiveness against known attack patterns but it fails to adapt to new threats and faces scalability problems which rule-based systems commonly experience.

Yin together with Liu and Nkenyereye and Ndibanje developed an Intrusion Detection System for IoT-based smart grids [2]. The researchers extracted packet features from DNP3 data and combined them with malware signature information. The system employed supervised learning algorithms to identify regular system behavior against abnormal system behavior. Real-time performance of their system proved promising yet they encountered issues with false-positive rates during implementation.

Fovino et al. [4] developed an IDS system that tracked system state transitions through monitoring of DNP3 and other SCADA protocols. The IDS system identified complex multi-step attacks through the analysis of sequential benign commands. The system required extensive prior knowledge about operational states for its operation yet this dependency restricted its adaptability.

Sakib et al. [3] proposed a hybrid IDS based on LSTM-CNN using timestamped attack data. The model reached 98.3% accuracy in attack type classification when trained with the DNP3 dataset from the University of Western Macedonia. The integration of temporal (LSTM) and spatial (CNN) learning proved highly effective in recognizing complex intrusion patterns. The authors highlighted data preprocessing, feature selection and appropriate train-test-validation splits as best practices for maximizing model performance.

Altaha, Lee, Aslam, and Hong [6] also used deep neural networks for DNP3-specific IDSs. Their model, which extracted 12 protocol-specific features, was able to detect data injection attacks effectively. Their results support the benefit of domain-specific feature engineering for SCADA-related IDS tasks.

Nguyen, Phan, and Song [8] proposed a GAN-based IDS with robust anomaly detection in SCADA networks, particularly in the case of limited labeled attack data. They used GANs to generate synthetic attack data to improve class balance and model generalization. They demonstrated that GANs could successfully reduce both false positives and overfitting, some of the key problems for industrial network intrusion detection.

Kelli et al. [5] conducted eight distinct DNP3 cyberattack experiments and implemented a multi-model IDS using a deep neural network model trained on their experimental data set. Their suggested solution achieved a classification accuracy of 99.0%, outperforming traditional ML models. Their defense design, comprising stacked layers of multiple classifiers, provided insights into ensemble learning approaches for enhanced detection accuracy.

B. Insights and Best Practices

From the above studies, several key insights and practices emerge:

- Feature Engineering is Critical: Studies that derived protocol-specific features (e.g., packet lengths, timestamps, flags) from DNP3 traffic consistently achieved higher accuracy.
- Deep Learning Outperforms Traditional ML: RNNs and CNNs handle temporal and contextual patterns better than decision trees or SVMs in complex SCADA datasets.
- GANs Address Data Imbalance: Adversarial approaches augment training data and improve generalization in scarce-label environments.

- Hybrid and Ensemble Models Yield Superior Results: Combining multiple learning models or integrating rule-based logic with neural networks improves robustness.
- Real-Time Performance and Low False Positives: Systems with real-time detection and low false-positive rates (e.g., Sakib et al. [14], Nguyen et al. [19]) are preferred for industrial applications.

C. Research Gaps and Future Directions

Despite significant progress, gaps remain:

- Limited Availability of Realistic Datasets: Most studies use synthetic or emulated data. More real-world datasets would improve generalizability.
- Explainability of Deep Models: Models like CNNs and GANs often lack interpretability—an area ripe for research.
- Edge Deployment Constraints: Many models assume centralized computation. Lightweight versions are needed for deployment at IEDs or edge nodes.
- Adaptation to New Attack Types: Models should evolve with new data to remain effective against emerging threats.

Acknowledgment

The authors extend their deepest appreciation to Lecturer Dr. Mennan Guder who provided essential guidance and meaningful feedback that supported our work on this assignment throughout its completion. The authors received essential help from his expertise in cybersecurity and data mining which enabled them to grasp the importance of machine learning applications.

We show appreciation to IEEE DataPort because they made the DNP3 Intrusion Detection Dataset freely available for our research and analysis needs.

We acknowledge the research scientists who published their work because their discoveries have expanded both SCADA security and Intrusion Detection Systems knowledge. The authors express gratitude to researchers who conducted rigorous work to share their knowledge which enabled our literature survey and delivered essential insights throughout this report.

References

[1] I. Darwish, O. Igbe, O. Celebi, T. Saadawi, and J. Soryal, “Smart Grid DNP3 Vulnerability Analysis and Experimentation,” in *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*, pp. 141–148, 2015, doi: 10.1109/CSCloud.2015.86.

[2] X. C. Yin, Z. G. Liu, L. Nkenyereye, and B. Ndibanje, “Toward an Applied Cyber Security Solution in IoT-Based Smart Grids: An Intrusion Detection System Approach,” *Sensors*, vol. 19, no. 22, p. 4952, 2019, doi: 10.3390/s19224952.

[3] M. N. Sakib et al., “Securing Critical Infrastructure: A Robust DNP3 Intrusion Detection System Employing Recurrent Neural Networks,” in *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*, 2024, doi: 10.1109/COMPAS60761.2024.10797009.

[4] I. N. Fovino et al., “Modbus/DNP3 State-Based Intrusion Detection System,” in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, pp. 729–736, 2010, doi: 10.1109/AINA.2010.86.

[5] V. Kelli et al., “Attacking and Defending DNP3 ICS/SCADA Systems,” in *2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pp. 183–190, 2022, doi: 10.1109/DCOSS54816.2022.00041.

[6] M. Altaha, J.-M. Lee, M. Aslam, and S. Hong, “Network Intrusion Detection Based on Deep Neural Networks for the SCADA System,” *Journal of Physics: Conference Series*, vol. 1585, no. 1, 2020, doi: 10.1088/1742-6596/1585/1/012038.

[7] J. Bai, S. Hariri, and Y. Al-Nashif, “A Network Protection Framework for DNP3 Over TCP/IP Protocol,” in *2014 IEEE International Conference on Cloud Engineering*, pp. 168–173, doi: 10.1109/IC2E.2014.33.

[8] H. N. Nguyen, T. L. Phan, and C.-J. Song, “Generative Adversarial Network-Based Network Intrusion Detection System for Supervisory Control and Data Acquisition System,” in *2024 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, 2024, doi: 10.1109/ICCE-ASIA63397.2024.10773791.

[9] Dataset Link: https://zenodo.org/records/7348493/files/DNP3_Intrusion_Detection_Dataset_Final.7z?download=1