An Anomaly Detection Mechanism for IEC 60870-5-104

Panagiotis Radoglou Grammatikis[†], Panagiotis Sarigiannidis[†], Antonios Sarigiannidis[‡], Dimitrios Margounakis[‡], Apostolos Tsiakalos[‡] and Georgios Efstathopoulos[§]

***Abstract*—The transformation of the conventional electricity grid into a new paradigm called smart grid demands the appropriate cybersecurity solutions. In this paper, we focus on the security of the IEC 60870-5-104 (IEC-104) protocol which is commonly used by Supervisory Control and Data Acquisition (SCADA) systems in the energy domain. In particular, after investigating its security issues, we provide a multivariate Intrusion Detection System (IDS) which adopts both access control and outlier detection mechanisms in order to detect timely possible anomalies against IEC-104. The efficiency of the proposed IDS is reflected by the Accuracy and F1 metrics that reach 98% and 87%, respectively.**

***Index Terms*—Anomaly Detection, Cybersecurity, IEC-60870-5-104, Supervisory Control and Data Acquisition**

I. Introduction

The Critical Infrastructures (CIs) and especially the electrical grid constitute a frequent target of the Advanced Persistent Threats (APTs). In particular, they are composed of legacy technologies characterised by severe security flaws. Moreover, although the rapid advance of the Internet of Things (IoT) introduces new beneficial characteristics to CIs, it increases in parallel the attack surface due to the insecure nature of the Internet and specifically of the respective communication protocols [1].

In this paper, we focus on the Transmission Control Protocol (TCP)-based IEC 60870-5-104 (IEC-104) protocol, which is commonly utilised by Supervisory Control and Data Acquisition (SCADA) systems in Europe. IEC-104 uses the 2404 TCP port and does not include sufficient authorisation mechanisms, thus allowing potential cybercriminals to violate the IEC-104 communications either via unauthorised IEC-104 commands or Man in The Middle (MiTM) attacks [2]. Based on the aforementioned security gaps of IEC-104, in this paper, we provide a relevant Intrusion Detection System (IDS) which relies on essential access control rules and machine learning-based outlier detection mechanisms.

In particular, the rest of this paper is organised as follows. Section II discusses previous works related to the security of IEC-104. In section III, we provide a background about the IEC-104 security and the various machine learning anomaly detection methods. Section IV is devoted to the architecture of the proposed IDS, while Section V evaluates its efficacy. Finally, Section VI concludes this paper.

*This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 787011 (SPEAR).

[†]P. Radoglou-Grammatikis and P. Sarigiannidis are with the Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani 50100, Greece - E-Mail: {pradoglou,psarigiannidis}@uowm.gr

[‡]A. Sarigiannidis, D. Margounakis and A. Tsiakalos are with SIDROCO, Anaximandrou, 5A 3113, Limassol, Cyprus - E-Mail: {asarigia,dmargoun, atsiakalos}@sidroco.com

[§] G. Efstathopoulos is with the 0INF, Imperial Offices, London, UK, E6 2JG - E-Mail: george@0inf.com

II. Related Work

Many authors have investigated the security issues of IEC-104. In particular, in [2], the authors provided a risk assessment model regarding the IEC-104 communications, taking into account a Coloured Petri Net (CPN)-based threat assessment model as well as the risk assessment model of AlienVault OSSIM [3]. In [4], P. Maynard et al. focused on the possible MiTM and replay attacks against IEC-104, covering also the corresponding injection commands. Accordingly, in [5] C.Lin and S. Nadjm-Tehrani analysed IEC-104 traffic patterns, aiming at discovering underlying timing patterns of spontaneous events. In [6], E. Hodo et al. presented an anomaly-based IDS for IEC-104, utilising classification machine learning methods, such as J48, Naive Bayes, OneR and RandomTree. Finally, in [7] Y. Yang et al. provided a set of IEC-104 signature rules, while in [8], Y. Yang et al. introduced a relevant specification-based IDS relying on a Finite State Machine (FSM).

III. Background

A. IEC 60870-5-104 Security Issues

The functionality of IEC-104 relies on the TCP/IP, which exhibits a number of cybersecurity issues. Although IEC 62351 [9] provides sufficient guidelines that can enhance the security of IEC-104, the industrial nature of SCADA hinders their immediate upgrade. A severe security issue of IEC-104 is the transmission of data without any encryption mechanism, thus making it possible to execute traffic analysis and MiTM attacks. In addition, many IEC-104 commands, such as reset commands, interrogation commands, read commands do not integrate authentication and authorisation procedures, thereby allowing the unauthorised access. This vulnerability is crucial since a cyberattacker is capable of controlling the field devices and possibly, the overall operation of the infrastructure.

B. Machine Learning Algorithms Background

In this section, a short overview of the anomaly detection methods based on machine learning solutions is provided. A more comprehensive literature review can be found in recent surveys [10], [11]. The machine learning methods for anomaly detection can be separated to model, clustering, reconstruction and proximity-based. Model-based approaches include the Gaussian mixture models (GMM) [12] that fit the

This paper has been published in IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/9200285

P. R. Grammatikis, P. Sarigiannidis, A. Sarigiannidis, D. Margounakis, A. Tsiakalos and G. Efstathopoulos, "An Anomaly Detection Mechanism for IEC 60870-5-104," 2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 2020, pp. 1-4, doi: 10.1109/MOCAST49295.2020.9200285.

whole dataset to a mixed Gaussian distribution. The GMM parameters usually are estimated with Expectation-Maximization solutions or deep estimation networks.

The attribute-based approaches for anomaly detection assume that the features of normal examples can be predicted by the rest or in the case of the Isolation Forest algorithm, it finds anomalies by deliberately overfitting models that memorise each data point. Particularly, in this case, outliers have more empty space around them, and therefore they take fewer steps to memorise. Many anomaly detection methods are considered clustering-based detectors, assuming that the normal data are located close to their closest cluster. The methods Principal Component Analysis (PCA), Matrix Factorization (MF), Stochastic Outlier Selection (SOS) and deep Auto-encoders belong to the reconstruction-based approaches. The concept behind these methods is to learn a mapping from a higher to a lower-dimensional space through the compression and decompression stages and identify points with high reconstruction error as anomalies. Regarding SOS, it is an unsupervised anomaly-selection algorithm that takes as input either a feature matrix or a dissimilarity matrix and outputs for each data point an anomaly probability. Intuitively, a data point is considered to be an anomaly when the other data points have an insufficient affinity with it. One-Class Support Vector Machine (OC-SVM) aims to find a hyperplane that can separate the vast majority of data from the origin in the projected high dimensional space without making any assumptions about their distribution. In particular, OC-SVM separates all the data points from the origin (in feature space) and maximises the distance from this hyperplane to the origin. This results in a binary function, which captures regions in the input space where the probability density of the data lives. The idea of OC-SVM for anomaly detection is to find a function that is positive for regions with a high density of points, and negative for small densities.

Proximity-based methods do not require any training or assumptions about the dataset. They consider the rarity of a point, measuring, for example, the distance to K-Nearest Neighbour (KNN) or the ratio of local reachability density.

IV. IEC-104 IDS

Fig. 1 illustrates the architecture of the proposed IDS, which consists of two main components, namely a) *Sensor* and b) *Server*. The *Sensors* consist of three modules, namely *a) Network Traffic Monitoring Module*, *b) Network Packet Access Control* and *c) IEC-104 Flows Extraction Module* responsible respectively for monitoring and analysing the entire network traffic generated in the infrastructure. On the other hand, *Server* constitutes a centralised point where the anomaly detection processes take place, and the security events are stored. In particular, *Server* is composed of an *Elasticsearch database*, the *Anomaly Detection Module* and the *Response Module*. The following subsections analyse in detail each module.

A. Network Traffic Monitoring Module

The *Network Traffic Monitoring Module* relies on the Scapy library [13] and is responsible for monitoring and capturing the overall network traffic based on a predefined frequency which can be defined by the user.

B. Network Packet Access Control Module

This module receives the captured network traffic from the previous module and utilises Scapy [13] in order to apply some initial security controls. In particular, it adopts a whitelist in which all legitimate, Medium Access Control (MAC) and Internet Protocol (IP) addresses are stored. Therefore, if a packet contains a MAC or an IP address which is not included in the whitelist, then a security event is generated and stored in the Elasticsearch database of *Server*. The legitimate MAC and IP addresses should be defined by the system operator or the security administrator. In addition, this whitelist defines also the permitted TCP and UDP ports. Therefore, if a packet includes a non-legitimate port, the corresponding security event is generated.

C. IEC-14 Flows Extraction Module

This module receives the captured network packets and exports the corresponding bi-directional IEC-104 flows, utilising the CICFlowMeter software [14]. In particular, CICFlowMeter generates for each flow 83 features that are stored in a different index of the Elasticsearch database. Also, it is noteworthy that different flow-timeout thresholds can be used for extracting the corresponding IEC-104 flows, thus affecting proportionally the 83 features [14].

D. Anomaly Detection Module

The *Anomaly Detection Module* constitutes the core of the proposed IDS. First, it receives the captured IEC-104 flows from the Elasticsearch database and applies outlier detection models in order to detect which of them are anomalies. The efficacy of these models is discussed in Section V. Finally, it stores the corresponding security events (i.e., anomalous IEC-104 flows) in a different index of the Elasticsearch database.

E. Response Module

The *Response Module* undertakes to inform the user about the various security events via Kibana of the Elastic Stack. Moreover, it provides statistic charts that assist the user in understanding better the security status of the infrastructure. Regarding the security events, the format of AlienVaut OSSIM [3], [15] was utilised. In particular, the security events detected by the proposed IDS are related to the controls of Network Packet Access Control and Anomaly Detection Modules.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

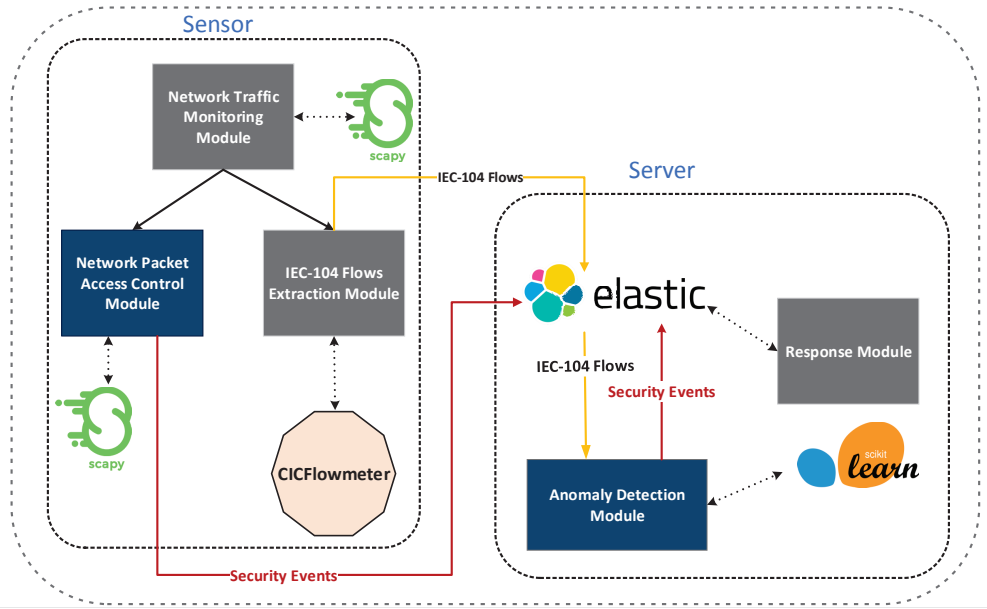

Fig. 1: IEC-104 IDS Architecture

V. Evaluation Analysis

This section is devoted to the efficacy of the outlier detection models of the *Anomaly Detection Module*. In particular, three outlier detection algorithms were evaluated, namely a) OC-SVM, b) Isolation forest and c) LOF to detect anomalous IEC-104 flows under four different flow-timeout thresholds: 15, 30, 60 and 120 seconds. In order to train the corresponding models, we combined normal IEC-104 data stemming from a real substation as well as IEC-104 malicious data of [16]. Moreover, utilising the PCA method, we chose only seven features from the 83 ones generated by CICFlowMeter, including a) the total packets in the forward direction, b) the total size of the packets in the backward direction, c) the standard deviation size of the packets in the forward direction, d) the number of the flow bytes per second, e) the maximum time between two packets sent in the flow, f) the minimum length of a packet, g) the average number of bytes in a sub-flow in the backward direction and h) the maximum time where a flow was active before becoming idle. It is worth mentioning that the previous features are related only to the IEC-104 packets since Scapy [13] and CICFlowMeter were configured to capture and extract only IEC-104 flows, respectively.

Tables I-IV and Fig. 2 depict the efficacy of the aforementioned outlier detection algorithms under the different flow-timeout values, in terms of Accuracy, F1 and True Positive Rate (TPR). These metrics are defined by the Equations 1-3 respectively. In particular, True Positives (TP) denotes the number of the correct classifications that detected the malicious flows as successfully. Accordingly, True Negatives (TN) denotes the amount of the correct classifications that recognized the normal flows as normal. On the other hand, False Positives (FP) indicates those classifications that detected the normal flows as anomalous, while False Negatives (FN) defines the incorrect classifications that wrongfully recognized the malicious flows as normal. According to the evaluation results, when the flow-timeout value is equal to 120s, the Isolation Forest method achieves the highest Accuracy, Precision, TPR and F1 that reach 0.982, 0.990, 0.777 and 0.875 respectively.

TABLE I: The Outlier Detection Evaluation Results for flow-timeout 15s.

Model	Accuracy	Precision	TPR	F1
OS-SVM	0.519	0.509	0.993	0.673
LOF	0.65	0.98	0.30	0.46
Isolation Forest	0.536	0.519	0.992	0.6817

TABLE II: The Outlier Detection Evaluation results for flow-timeout 30s.

Model	Accuracy	Precision	TPR	F1
OS-SVM	0.805	0.943	0.650	0.769
LOF	0.783	0.886	0.650	0.750
Isolation Forest	0.811	0.964	0.647	0.774

TABLE III: The Outlier Detection Evaluation for flow-timeout 60s.

Model	Accuracy	Precision	TPR	F1
OS-SVM	0.811	0.964	0.647	0.774
LOF	0.790	0.941	0.620	0.747
Isolation Forest	0.812	0.964	0.647	0.775

TABLE IV: The Outlier Detection Evaluation results for flow-timeout 120s.

Model	Accuracy	Precision	TPR	F1
OS-SVM-Linear	0.812	0.962	0.647	0.774
LOF	0.812	0.964	0.647	0.775
Isolation Forest	0.982	0.990	0.777	0.875

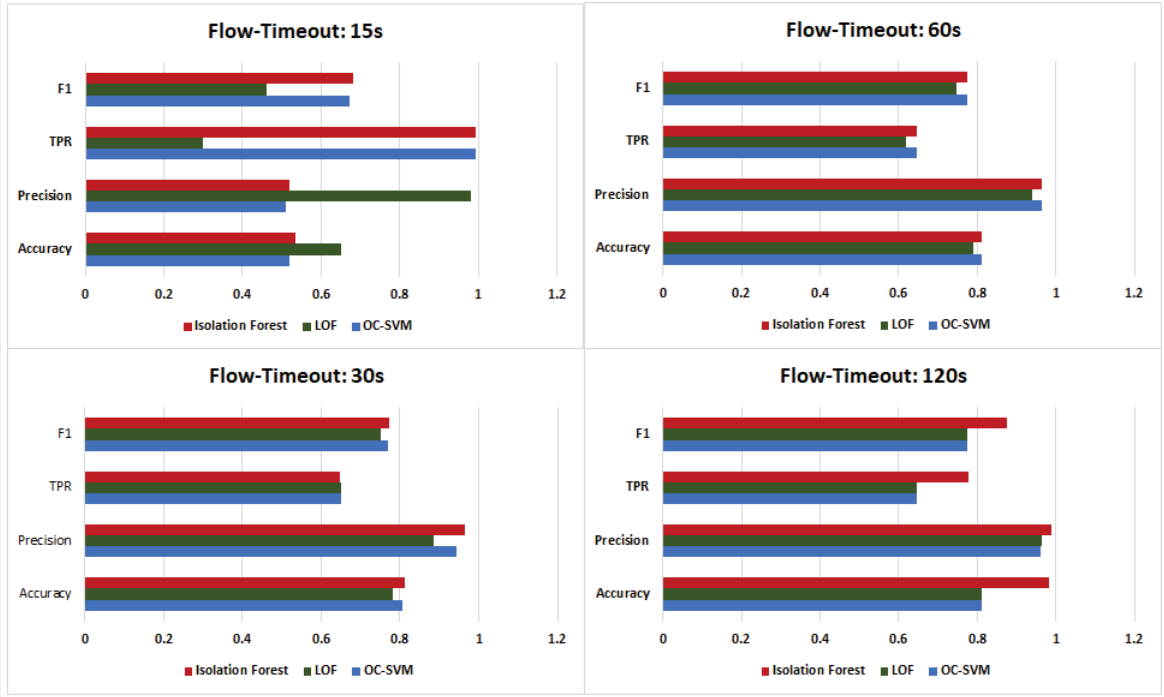

Fig. 2: The overall performance of the outlier detection methods with the different flow-timeout values.

VI. Conclusions

The continuous progression and involvement of IoT in the industrial domain and especially in the electrical grid requires the presence of appropriate cybersecurity measures. In this paper, we focused our attention on the security of the IEC-104 protocol, which is commonly utilised by SCADA systems. In particular, after investigating IEC-104 security issues, we provided a relevant IDS, which applies access control and outlier detection mechanisms in order to detect IEC-104 anomalies. The performance of the proposed IDS is demonstrated through the evaluation analysis, where Accuracy and F1 score reach 98% and 87%, respectively.

VII. Acknowledgement

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 787011 (SPEAR).

References

[1] P. I. R. Grammatikis, P. G. Sarigiannidis, and I. D. Moscholios, "Securing the internet of things: Challenges, threats and solutions," *Internet of Things*, vol. 5, pp. 41 – 70, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2542660518301161

[2] P. Radoglou-Grammatikis, P. Sarigiannidis, I. Giannoulakis, E. Kafetzakis, and E. Panaousis, "Attacking iec-60870-5-104 scada systems," in *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642-939X, July 2019, pp. 41–46.

[3] S. G. Zarzosa, "D2.1 in-depth analysis of siems extensibility," DiSIEM Project, Tech. Rep. 1, 2017.

[4] P. Maynard, K. McLaughlin, and B. Haberler, "Towards understanding man-in-the-middle attacks on iec 60870-5-104 scada networks," in *2nd International Symposium for ICS & SCADA Cyber Security Research 2014 (ICS-CSR 2014) 2*, 2014, pp. 30–42.

[5] C.-Y. Lin and S. Nadjm-Tehrani, "Understanding iec-60870-5-104 traffic patterns in scada networks," in *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security*, 2018, pp. 51–60.

[6] E. Hodo, S. Grebeniuk, H. Ruotsalainen, and P. Tavolato, "Anomaly detection for simulated iec-60870-5-104 trafiic," in *Proceedings of the 12th International Conference on Availability, Reliability and Security*, 2017, pp. 1–7.

[7] Y. Yang, K. McLaughlin, T. Littler, S. Sezer, B. Pranggono, and H. Wang, "Intrusion detection system for iec 60870-5-104 based scada networks," in *2013 IEEE power & energy society general meeting*. IEEE, 2013, pp. 1–5.

[8] Y. Yang, K. McLaughlin, S. Sezer, Y. Yuan, and W. Huang, "Stateful intrusion detection for iec 60870-5-104 scada security," in *2014 IEEE PES General Meeting— Conference & Exposition*. IEEE, 2014, pp. 1–5.

[9] R. Schlegel, S. Obermeier, and J. Schneider, "A security evaluation of iec 62351," *Journal of Information Security and Applications*, vol. 34, pp. 197 – 204, 2017.

[10] C. C. Aggarwal, "An introduction to outlier analysis," in *Outlier analysis*. Springer, 2017, pp. 1–34.

[11] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2013.

[12] X. Yang, L. J. Latecki, and D. Pokrajac, "Outlier detection with globally optimal exemplar-based gmm," in *Proceedings of the 2009 SIAM International Conference on Data Mining*. SIAM, 2009, pp. 145–154.

[13] R. R. S, R. R, M. Moharir, and S. G, "Scapy- a powerful interactive packet manipulation program," in *2018 International Conference on Networking, Embedded and Wireless Systems (ICNEWS)*, 2018, pp. 1–5.

[14] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of tor traffic using time based features." in *ICISSP*, 2017, pp. 253–262.

[15] Alienvault ossim security events. [Online]. Available: https://cybersecurity.att.com/documentation/usm-appliance/events/event-details-fields.htm

[16] P. Maynard, K. McLaughlin, and S. Sezer, "An open framework for deploying experimental scada testbed networks," in *5th International Symposium for ICS & SCADA Cyber Security Research 2018 5*, 2018, pp. 92–101.