UDC 004.77.056 **doi: 10.32620/reks.2024.1.12**

Lahcen Idouglid[1], Said Tkatek[1], Khalid Elfayq[1], Azidine Guezzaz[2]

[1] ***Computer Sciences Research Laboratory, Ibn Tofail University Kenitra, Morocco***
[2] ***Higher School of Technology, Cadi Ayyad University, Essaouira, Morocco***

A NOVEL ANOMALY DETECTION MODEL FOR THE INDUSTRIAL INTERNET OF THINGS USING MACHINE LEARNING TECHNIQUES

In recent decades, the pervasive integration of the Internet of Things (IoT) technologies has revolutionized various sectors, including industry 4.0, telecommunications, cloud computing, and healthcare systems. Industry 4.0 applications, characterized by real-time data exchange, increased reliance on automation, and limited computational resources at the edge, have reshaped global business dynamics, aiming to innovate business models through enhanced automation technologies. However, ensuring security in these environments remains a critical challenge, with real-time data streams introducing vulnerabilities to zero-day attacks and limited resources at the edge demanding efficient intrusion detection solutions. ***This study addresses this pressing need*** *by proposing a novel intrusion detection model (IDS) specifically designed for Industry 4.0 environments.* ***The proposed IDS leverages a Random Forest classifier with Principal Component Analysis (PCA)*** *for feature selection. This approach addresses the challenges of real-time data processing and resource limitations while offering high accuracy. Based on the Bot-IoT dataset, the model achieves a competitive accuracy of 98.9% and a detection rate of 97.8%, outperforming conventional methods.* ***This study demonstrates the effectiveness*** *of the proposed IDS for securing Industry 4.0 ecosystems, offering valuable contributions to the field of cybersecurity.*

Keywords: *Industry 4.0 Security; IIoT; IoT; Anomaly Detection; Feature Selection; Random Forest.*

Introduction

In recent years, cloud computing and IoT environments have been widely utilized in various fields [1, 2]. Modern IoT applications are enabling smart cities worldwide [3, 4], providing remote monitoring, management, control, and the extraction of new perspectives from massive amounts of real-time data [5]. The Industrial IoT (IIoT), also known as Industry 4.0, is an innovative technology that enables the integration and utilization of intelligent sensors and actuators to improve industrial operations [6]. Within diverse communications infrastructures, these various smart edge devices exchange data. Real-time monitoring of data, resources, and networks is necessary given the surge in threats and the expansion of the IIoT environments [7].

The IIoT security solutions aim to accurately monitor data and edge devices by providing enhanced approaches [8, 9]. Therefore, multiple techniques are implemented to protect heterogeneous and large amounts of data from threats, such as integrity attacks, denial-of-service (DoS) attacks, distributed DoS (DDoS) attacks, and others that aim to compromise system security [10, 11]. These attacks infect services and smart devices in the IIoT environments. Consequently, securing the IIoT becomes a main priority [12]. However, the poor deployment environment, particularly the limited computational resources often available at the network edge, has raised significant concerns about network security [8]. With the advancement of ML, multiple algorithms have been used to enhance security approaches [6, 10]. The basic idea is to create models that work faster and with higher performance, thus solving more classification, detection, and regression problems [13]. A crucial security concern in such the IIoT context is the detection of intruders [6, 14]. Several IDSs have been created for different types of attacks by integrating machine learning (ML) and deep learning (DL) techniques. Intrusion detection approaches provide numerous methods and techniques for more secure the IoT environments [15, 16].

This study introduces a novel anomaly detection model to address critical security challenges within Industry 4.0 environments. In particular, it addresses the challenges of real-time data processing and resource limitations prevalent in these environments. The approach utilizes the Bot-IoT dataset and proposes an Intrusion Detection System (IDS) based on a Random Forest (RF) classifier for anomaly detection in the IIoT networks. Random Forest is known for its efficiency, which makes it suitable for real-time processing on resource-constrained devices. This system incorporates Principal Component Analysis (PCA) for feature selection, ultimately improving the effectiveness of intrusion detection. By meticulously preprocessing, normalizing, and selecting significant features, this approach strengthens data quality and guarantees the model's robustness.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

This study offers two key contributions. First, it demonstrates the effectiveness of preprocessing, normalization, and feature selection techniques in enhancing the quality of input data, consequently improving the overall performance of the IDS. Second, the implementation of the RF classifier model provides an efficient and reliable method for intrusion detection in the IIoT environments, particularly those with limited computational resources. Through comprehensive experimental evaluations on two distinct datasets, the model consistently delivers reliable results, demonstrating its efficacy in accurately detecting intrusions in real time.

The rest of this paper follows the structure outlined below. The related research on IoT security employing intrusion detection strategies that incorporate ML and DL techniques is presented in section 2. The proposed methods to verify our suggested strategy are described in Section 3. Section 4 discusses the evaluation of performances and their outcomes. The paper ends with a conclusion.

1. State of the art

In this section, we explore recent advancements in the use of Machine Learning (ML) and Deep Learning (DL) methods to bolster the effectiveness of Intrusion Detection Systems (IDS) in safeguarding both the IoT and the IIoT environments. Various machine learning techniques have been harnessed to develop robust models, enabling efficient learning from training data and refining intrusion detection methodologies.

Guezzaz [17] introduced the Decision Tree-based Entropy IDS (DTE IDS), a novel approach that integrates decision tree algorithms with entropy-based feature selection techniques. The primary aim of this study is to enhance the accuracy of decision tree classifiers through improved feature selection processes. Evaluation using datasets such as NSL-KDD and CICIDS2017 showcased promising results, with the DTE IDS achieving remarkable accuracy and detection rates. Verma [18] proposed a machine learning model tailored for anomaly detection, specifically targeting Denial-of-Service (DoS) attacks in IoT networks. Leveraging datasets like NSL-KDD, CIDDS-001, and UNSW-NB15, Verma [18] demonstrated impressive performance using Classification and Regression Tree (CART) and AdaBoost (AB) methods, achieving notable accuracy and detection rates.

Bagaa [19] validated a machine learning anomaly detection model based on Support Vector Machines (SVM), emphasizing security enhancements in IoT systems. Through evaluation of the NSL-KDD dataset, the proposed model exhibited high accuracy and detection rates, underscoring its efficacy in anomaly detection tasks. Sai Kirana [20] proposed a model for detecting Main-In-The-Middle attacks in the IoT networks by employing classifiers such as Naive Bayes (NB), SVM, and AdaBoost. Rigorous experimentation and evaluation of sensor data highlighted significant performance improvements, particularly with SVM achieving notable accuracy and detection rates.

Dovbysh [21] focused on enhancing machine learning in cybersecurity education programs, aiming to optimize learning strategies and educational content efficiently. This innovative approach addresses the increasing demand for skilled cybersecurity professionals, thereby contributing to the enhancement of cybersecurity education. Dovbysh [22] also proposed a composite method that integrates signature and anomaly detection approaches for cyber-attack detection, demonstrating high efficiency and reliability. This approach offers a comprehensive solution for detecting various types of cyber threats effectively.

Bobrovnikova [23] introduced a novel approach rooted in control flow graph analysis for the IoT malware detection, demonstrating its effectiveness in safeguarding the IoT devices from cyber threats. This innovative method contributes to the advancement of the IoT security measures. Al Amien [26] presented the Bot_IoT dataset, which combines the IoT, network traffic, and botnet attack data. Through statistical analysis and the use of XGBoost method, exceptional accuracy was achieved, highlighting the potential for further optimization and algorithmic exploration.

Lazzarini [24] explored Federated Learning (FL) as an alternative to centralized models for the IoT attack detection, emphasizing the importance of data privacy in the IoT security measures. This approach offers a decentralized solution while ensuring the confidentiality and integrity of data.

Musleh [25] distinguished malicious from normal traffic using feature extraction techniques and machine learning algorithms, achieving promising accuracy. This research contributes to the development of efficient methods for detecting malicious activities in network traffic.

The preceding literature underscores the diverse methodologies and innovations employed in advancing the field of intrusion detection for the IoT and the IIoT environments. From decision tree-based approaches to sophisticated anomaly detection models, each contribution brings us closer to achieving robust security frameworks tailored to the unique challenges posed by interconnected systems.

However, gaps remain, necessitating further exploration into alternative algorithms, optimizations, and holistic approaches to fortify the IoT and the IIoT security despite evolving threats.

2. Methodology and implementation

This section outlines our methodology, focusing on techniques that support precise classification and feature selection for reliable decision-making in anomaly detection.

2.1. Proposed design

The proposed design outlines a scheme for intrusion detection aimed at monitoring and securing data circulated within various devices in the IIoT environment. Many considerations are considered to better implement our model, such as memory characteristics and processing rate, which allow our model to perform reliable calculations, especially considering integrated techniques that often require efficient storage and high processing power.

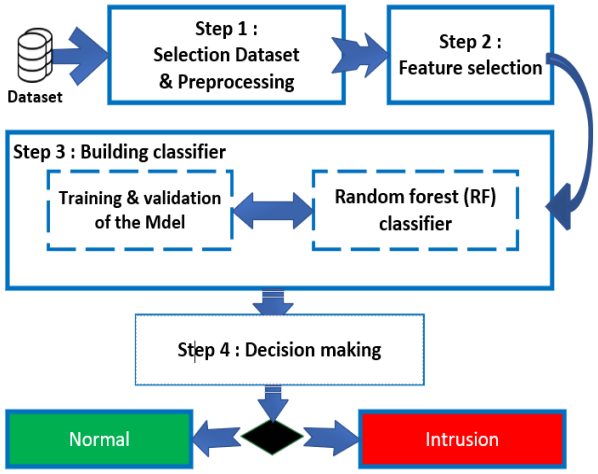

Fig. 1. Architecture of Proposed Model

However, to decrease the IIoT resource utilization and improve time complexity, preprocessing, normalization, and feature selection are performed on the dataset. As illustrated in Figure 1, the proposed design of our anomaly detection model includes four steps: preprocessing and normalization step, feature selection, training and validation, and decision-making.

a. Preprocessing steep

Preprocessing is a crucial step aimed at cleaning up the data and removing noise. To improve data quality and facilitate precise classification, we focus on preprocessing and normalization operations. Based on the min-max normalization method, the data are scaled to fall within the [0, 1] range. This approach mitigates the negative effects of features with high values that could dominate the analysis and potentially lead to biased classification results. Equation (1) is employed to determine the new normalized value for each feature, effectively restricting them to a range between 0 and 1. Here, "min" and "max" denote the lowest and greatest values, respectively, of a feature x. Consequently, every x value is transformed as follows:

$$\frac{(x)-\min}{\max-\min}, \quad (1)$$

where **x:** represents the original feature to normalize;

min: is the minimum value of the feature in the dataset;

max: is the maximum value of the feature in the dataset.

b. Feature selection step

In this stage, we perform feature selection using Principal Component Analysis (PCA). PCA is a statistical approach that reduces the high dimensionality of the original data to a lower-dimensional version while retaining the most significant features for intrusion detection. The reduction task decreases the computational cost and training time without compromising the results. By selecting the most important attributes from the source data through PCA, the feature selection stage improves data quality and enables the creation of a more robust classifier for the proposed model.

c. Building classifier step

To test and verify our model's effectiveness in intrusion detection, we utilize 10-fold cross-validation during the classifier development stage. This technique involves randomly dividing the entire dataset into ten equal-sized portions. Nine parts are used for training the model, and model testing is conducted on the remaining parts. This process is repeated ten times to create a trustworthy classifier that can detect intrusions with high accuracy. The classification process is the final step in which instances are categorized. The classifier model developed during the training and validation stages leverages the learned knowledge to predict the class (normal or intrusive) of new occurrences within the network traffic. We employ the Random Forest (RF) algorithm for this classification task because of its efficiency in real-time processing, making it suitable for resource-constrained environments.

d. Decision-making step

The fourth phase, following the creation of the classifier model, involves deploying the intrusion detection model. Real-world testing plays a crucial role in refining the model's performance and ensuring its robustness.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Based on the classification findings, the proposed system can be further optimized to achieve a high degree of accuracy in categorizing activities as normal or intrusive network traffic.

2.2. Exploration of Machine Learning Algorithms used in Intrusion Detection

This section describes in detail the algorithm statements used in the implementation.

a. Random forest (RF) classifier

Random forest is a widely-used machine learning algorithm employed in Intrusion Detection Systems (IDS). Falling under the ensemble learning category, RF is adept at both classification and regression tasks. In IDS, RF constructs multiple decision trees during training, amalgamating their outputs to predict potential intrusions. Each decision tree is trained on a random subset of the dataset (bagging) and a random subset of features (feature bagging), which mitigates overfitting and enhances the robustness of the IDS model [27].

b. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) [28] is a fundamental dimensionality reduction technique used in various fields, including statistics, data science, and machine learning. PCA identifies and extracts the most significant components from a dataset that are orthogonal to each other.

These components capture the maximum variance in the data and reduce it to a lower-dimensional space. Thus, PCA simplifies data while preserving its essential characteristics. It is widely used for feature selection, data visualization, and noise reduction.

c. K-NN

K-Nearest Neighbors (K-NN) [29] is a supervised machine learning algorithm used for both classification and regression tasks. Classifies data points by identifying the majority class among their k-nearest neighbors in feature space. This algorithm's simplicity and versatility make it a valuable tool in various domains, such as pattern recognition, image analysis, and recommendation systems.

d. Decision Trees

Decision Trees (DT) [30] are versatile machine learning techniques used for classification and regression. They create a hierarchical tree structure to make decisions based on input features, enabling interpretable and transparent models. DTs find applications in various domains, from finance to healthcare, because of their ease of use and comprehensibility.

3. Experiments and Results

The experimental setting and performance evaluation are described in this section. We provide a thorough explanation of the outcomes and contrast the suggested model with alternative approaches.

3.1. Experiments

The research tests and performance assessments were all carried out on a PC running Windows 7 Professional 64-bit with a Core i7 2700K CPU running at 2.50 GHz and 32 GB of DDR3 memory. The implementation of PCA feature selection and RF model training is performed using Python 3.8.0. The datasets play a crucial role in evaluating and validating intrusion detection methods. In this study, the Bot-IoT dataset is implemented to train, evaluate, and validate our model.

Accuracy is a measure of the overall accuracy of a system or model. Represents the ratio of correctly predicted instances to the total number of instances.

The ACC metric is calculated using Eq. (2):

$$\frac{TP+TN}{TP+TN+FP+FN}, \quad (2)$$

where **TP (True Positives):** Instances correctly identified as positive;

TN (True Negatives): Instances correctly identified as negative;

FP (False Positives): Instances incorrectly identified as positive;

FN (False Negatives): Instances incorrectly identified as negative.

Detection Rate (DR) measures the ability of a system or model to correctly identify positive instances. It is also known as Sensitivity or Recall. The DR metric is calculated using Eq. (3):

$$\frac{TP}{TP+FN}, \quad (3)$$

where **TP (True Positives):** Instances correctly identified as positive;

FN (False Negatives): Instances incorrectly identified as negative.

The False Positive Rate (FPR) is a measure of the system's tendency to incorrectly accept an unauthorized

instance as legitimate. The FPR metric is calculated using the formula Eq. (4):

$$\frac{FP}{FP+TN}, \quad (4)$$

where **FP (False Positives):** Instances incorrectly identified as positive;

TN (True Negatives): Instances correctly identified as negative.

3.2. Bot-IoT dataset

The Bot-IoT dataset was meticulously crafted within the UNSW Canberra Cyber's Cyber Range Lab to emulate a realistic network environment. It combines both regular and botnet traffic and provides data in diverse formats, such as pcap, argus, and CSV files, sorted by attack category and subcategory for improved labeling. The dataset covers various attack types, including DDoS, DoS, OS and Service Scan, Keylogging, and Data exfiltration attacks, with further categorization of DDoS and DoS attacks by protocol. For ease of use, a representative 5% subset was extracted using select MySQL queries, comprising four files totaling about 1.07 GB and approximately 3 million records.

The BoT-IoT dataset encompasses various attributes crucial for network traffic analysis and intrusion detection in the IoT scenarios. Key attributes include details like source and destination IP address ports, communication protocols, flow duration, packet counts, packet lengths, data rates, and inter-arrival times. These attributes are essential for characterizing data flows, identifying anomalies, and detecting potential intrusions in complex the IoT network environments.

.

3.3. Results discussion

Based on the analysis of the Bot-IoT dataset, Figure 2 illustrates the consistent outperformance of our proposed model compared with the K-NN and DT models in terms of Accuracy (ACC). Similarly, Figure 3 showcases the superiority of our model in terms of Detection Rate (DR) compared with the aforementioned models. In addition, Figure 4 highlights the False Alarm Rate (FAR) of each model, further emphasizing the effectiveness of our proposed approach. These results underscore the robustness and reliability of our intrusion detection model, especially in the context of the IoT and the IIoT security challenges.

The obtained results are demonstrated in Tables 1 and 2. They demonstrate that the suggested model outperforms the K-NN, DT, and DTE models in terms of performance. The ACC of the new model provides 98.9%, whereas the RF model gives 98.1%. Our model presents 97.8% DR and 2.6% FAR, whereas the RF model presents DR of 96.7% and FAR of 2.8%.

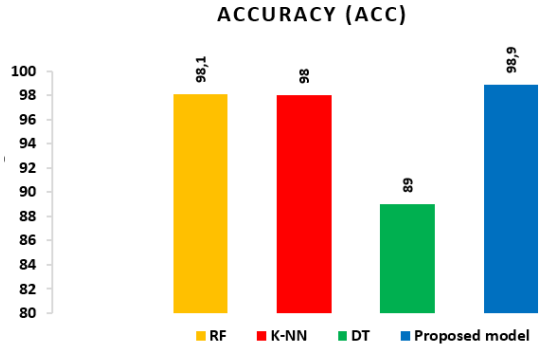

Fig. 2. The ACC of models on Bot-IoT dataset

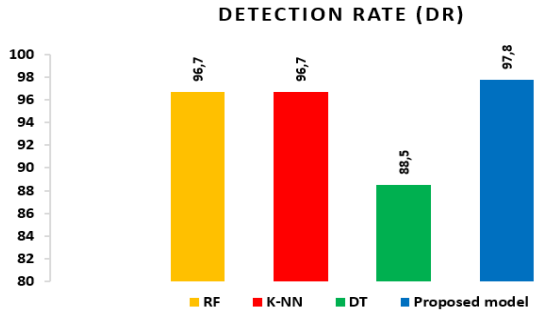

Fig. 3. The DR of models on Bot-IoT dataset

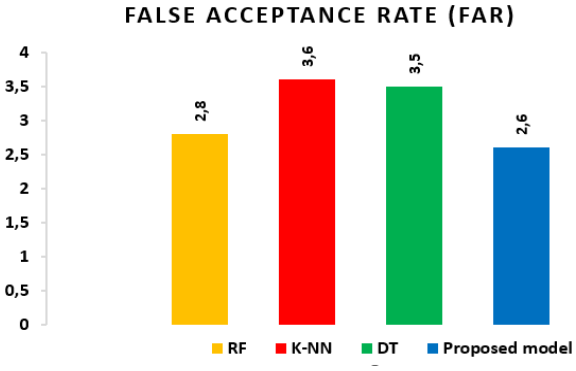

Fig. 4. The FAR of models on Bot-IoT dataset

Table 1

Performance evaluation of the proposed model and RF model on the Bot-IoT dataset

Model	ACC (%)	DR (%)	FAR (%)
RF	98.1	96.7	2.8
Proposed model	98.9	97.8	2.6

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 2 presents a comparison of various recent intrusion detection models obtained using the Bot-IoT dataset. The DT model gives ACC 89%, DR 88.5%, and FAR 3.5%. While the K-NN model reports ACC 98%, DR 96.7%, and FAR 3.6%, the DTE model reports ACC 99.42%, DR 98.20%, and FAR 2.64%. Additionally, using tables, we demonstrate that the suggested model has strong performance metrics for ACC, DR, and FAR. These metrics are 98.9%, 97.8%, and 2.6%, respectively.

Table 2

Performance evaluation of the proposed model and K- NN, DT models on the Bot-IoT dataset

Model	**ACC (%)**	**DR (%)**	FAR **(%)**
K-NN	98	96.7	3.6
DT	89	88.5	3.5
Proposed model	98.9	97.8	2.6

A detailed comparison with existing intrusion detection models reveals the significant advancements offered by our proposed approach. For instance, while the DT model shows respectable performance, its ACC and DR fall notably below those of our model. Similarly, the K-NN model, although competitive, exhibits lower accuracy and detection rates than our proposed solution. Moreover, the DTE model, despite its high ACC, fails to match the detection capabilities of our approach.

Through comprehensive analyses and comparisons, we substantiate the robustness and reliability of our proposed anomaly detection model. The consistently high ACC, DR, and low FAR metrics validate the efficacy of our approach in accurately identifying anomalies within the IoT environments. Moreover, the superior performance of our model compared with established methods like K-NN, DT, and DTE underscores its potential for widespread adoption in real-world applications.

Undertaking a thorough comparative analysis, our study evaluates our proposed intrusion detection model against recent advancements based on the Bot-IoT dataset. We prioritize a balanced approach that achieves competitive accuracy while emphasizing efficiency, interpretability, and scalability. This approach leverages machine learning techniques such as Random Forest (RF) and Principal Component Analysis (PCA), offering versatility and robustness (as shown in Table 3).

Central to the developed methodology is a multifaceted approach aimed at enhancing intrusion detection efficacy. By harnessing the power of Random Forest and Principal Component Analysis, we focus on robust feature selection and model training. Our innovative preprocessing techniques meticulously optimize data quality and reduce noise.

In addition, PCA for dimensionality reduction ensures efficiency without sacrificing crucial information. These choices strengthen the accuracy and reliability of our intrusion detection system, making it suitable for real-world IoT environments.

Table 3

Performance evaluation of the Bot-IoT dataset in comparison with prior studies

Model	**Method**	**ACC (%)**	**DR (%)**
Verma [18]	RF, XGB, CART, AB	96.74	-
Bagaa [19]	SVM	99.71	98.8
Sai Kiran [20]	NB, SVM, Adabost	98	98
Guezzaz [14]	KNN, PCA	99.10	98.4
Al Amien [26]	XGBoost	100	100
Lazzarini[24]	FL	98.17	98.31
Musleh[25]	RF, KNN, DT	98.3	96.2
The proposed model	RF, PCA	98.9	97.8

While acknowledging slight discrepancies in accuracy and detection rates compared with select approaches, it is important to highlight the distinct advantages of our model. Our Random Forest-based model boasts significantly faster training times than XGBoost. In addition, the inherent interpretability of Random Forest allows for a deeper understanding of the decision-making process behind intrusion detection.

Furthermore, its modular design facilitates seamless scalability, which is a critical attribute for handling the vast amount of data generated by the IIoT environments. Ongoing refinement efforts are poised to further elevate our model's performance and competitiveness, solidifying its potential as a valuable solution for intrusion detection in the Industrial Internet of Things (IIoT) ecosystems.

4. Conclusion

In conclusion, this study introduces a novel intrusion detection model tailored for real-world the Industrial Internet of Things (IIoT) environments, emphasizing efficiency, interpretability, and scalability. Constructed using Random Forest (RF) and Principal Component Analysis (PCA), the model achieves competitive accuracy (98.9%) and detection rate (97.8%), surpassing traditional methods like K-NN and DT. Its balanced approach ensures robust detection capabilities, faster training times, and clearer decision-making insights.

A key strength lies in meticulous data preprocessing and PCA-based dimensionality reduction, optimizing data quality while preserving crucial information. Additionally, RF interpretability aids in model debugging and improvement. Despite slight accuracy disparities with existing approaches, the proposed model excels in addressing dynamic the IIoT security needs. Its modular design enables seamless scalability for handling growing data volumes, while faster training ensures adaptability to evolving threats.

Future endeavors refine the model further, establishing it as a premier solution for IIoT intrusion detection. Implementation in Industry 4.0 settings holds promise for enhanced security protocols, alongside advancements in real-time threat detection capabilities to combat evolving cyber threats in the IoT domain.

Contribution of the authors:

Conducted a comprehensive review and analysis of references, contributed to literature analysis, participated in the paper's writing, formulated conclusions, and played a key role in model development and research result analysis – **Lahcen IDOUGLID**; Laid the foundation for the research purpose and tasks – **Said TKATEK;** Led the development of methods – **Khalid El Fayq;** Handled the selection and application of software and hardware tools for modeling and presenting results – **Azidine Guezzaz.**

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

This study was conducted without any financial support.

Data availability

The manuscript contains no associated data.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current study.

All authors have reviewed and approved the final version of the manuscript for publication.

References

1. Azrour, M., Mabrouki, J., & Chaganti, R. New Efficient and Secured Authentication Protocol for Remote Healthcare Systems in Cloud-IoT. *Security and Communication Networks*, 2021, vol. 2021, article no. 5546334, pp. 1-12. DOI: 10.1155/2021/5546334.

2. Čolaković, A., & Hadžialić, M. Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Computer Networks*, 2018, vol. 144, pp. 17-39. DOI: 10.1016/j.comnet.2018.07.017.

3. Batool, T., Abbas, S., Alhwaiti, Y., Saleem, M., Ahmad, M., Asif, M., & Elmitwally, N. S. Intelligent Model of Ecosystem for Smart Cities Using Artificial Neural Networks. *Intelligent Automation & Soft Computing*, 2021, vol. 30, iss. 2, pp. 513-525. DOI: 10.32604/iasc.2021.018770.

4. Tkatek, S., Belmzoukia, A., Nafai, S., Abouchabaka, J., & Ibnou-ratib, Y. Putting the world back to work: An expert system using big data and artificial intelligence in combating the spread of COVID-19 and similar contagious diseases. *Work*, 2020, vol. 67, iss. 3, pp. 557-572. DOI: 10.3233/WOR-203309.

5. King, J., & Awad, A. I. A Distributed Security Mechanism for Resource-Constrained IoT Devices. *Informatica*, 2016, vol. 40, iss. 1, pp. 133-143. Available at: https://www.informatica.si/index.php/informatica/article/view/1046 (accessed 12/12/2023)

6. Yao, H., Gao, P., Zhang, P., Wang, J., Jiang, C., & Lu, L. Hybrid Intrusion Detection System for Edge-Based IIoT Relying on Machine-Learning-Aided Detection. *IEEE Network*, 2019, vol. 33, iss. 5, pp. 75-81. DOI: 10.1109/MNET.001.1800479.

7. Azrour, M., Mabrouki, J., Guezzaz, A., & Kanwal, A. Internet of Things Security: Challenges and Key Issues. *Security and Communication Networks*, 2021, vol. 2021, article no. 5533843, pp. 1-11. DOI: 10.1155/2021/5533843.

8. Chanal, P. M., Kakkasageri, M. S. Security and Privacy in IoT: A Survey. *Wireless Personal Communications*, 2020, vol. 115, pp. 1667-1693. DOI: 10.1007/s11277-020-07649-9.

9. Yu, X., & Guo, H. A Survey on IIoT Security. *2019 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS)*, Singapore, 2019, pp. 1-5. DOI: 10.1109/VTS-APWCS.2019.8851679.

10. Idhammad, M., Afdel, K., & Belouch, M. Semi-supervised machine learning approach for DDoS detection. *Applied Intelligence*, 2018, vol. 48, pp. 3193-3208. DOI: 10.1007/s10489-018-1141-2.

11. Yan, Q., Huang, W., Luo, X., Gong, Q., & Yu, F. R. A Multi-Level DDoS Mitigation Framework for the Industrial Internet of Things. *IEEE Communications*

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Magazine, 2018, vol. 56, iss. 2, pp. 30-36. DOI: 10.1109/MCOM.2018.1700621.

12. Malik, P. K., Sharma, R., Singh, R., Gehlot, A., Satapathy, S. C., Alnumay, W. S., Pelusi, D., Ghosh, U., & Nayak, J. Industrial Internet of Things and its Applications in Industry 4.0: State of The Art. *Computer Communications*, 2021, vol. 166, pp. 125-139. DOI: 10.1016/j.comcom.2020.11.016.

13. Buczak, A. L., & Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*, 2016, vol. 18, iss. 2, pp. 1153-1176. DOI: 10.1109/COMST.2015.2494502.

14. Guezzaz, A., Azrour, M., Benkirane. S., Mohy-Eddine, M., Attou, H., & Douiba, M. A Lightweight Hybrid Intrusion Detection Framework using Machine Learning for Edge-Based IIoT Security. *The International Arab Journal of Information Technology*, 2022, vol. 19, iss. 5, 822-830. DOI: 10.34028/iajit/19/5/14.

15. Alanazi, M., & Aljuhani, A. Anomaly Detection for Internet of Things Cyberattacks. *Computers, Materials & Continua*, 2022, vol. 72, iss. 1, pp. 261-279. DOI: 10.32604/cmc.2022.024496.

16. Guezzaz, A., Asimi, A., Sadqi, Y., Asimi, Y., & Tbatou, Z. A New Hybrid Network Sniffer Model Based on Pcap Language and Sockets (Pcapsocks). *International Journal of Advanced Computer Science and Applications*, 2016, vol. 7, iss. 2. DOI: 10.14569/IJACSA.2016.070228.

17. Guezzaz, A., Benkirane, S., Azrour, M., & Khurram, S. A Reliable Network Intrusion Detection Approach Using Decision Tree with Enhanced Data Quality. *Security and Communication Networks*, 2021, vol. 2021, pp. 1-8. DOI: 10.1155/2021/1230593.

18. Verma, A., & Ranga, V. Machine Learning Based Intrusion Detection Systems for IoT Applications. *Wireless Personal Communications*, 2020, vol. 111, iss. 4, pp. 2287-2310. DOI: 10.1007/s11277-019-06986-8.

19. Bagaa, M., Taleb, T., Bernabe, J. B., & Skarmeta, A. A Machine Learning Security Framework for Iot Systems. *IEEE Access*, 2020, vol. 8, pp. 114066-114077. DOI: 10.1109/ACCESS.2020.2996214.

20. Sai Kiran, K. V. V. N. L., Devisetty, R. N. K., Kalyan, N. P., Mukundini, K., & Karthi, R. Building a Intrusion Detection System for IoT Environment using Machine Learning Techniques. *Procedia Computer Science*, 2020, vol. 171, pp. 2372-2379. DOI: 10.1016/j.procs.2020.04.257.

21. Dovbysh, A. S., Shelekhov, I. V., Khibovs′ka, Yu. O., & Matyash, O, V. Informatsiyno-analitychna systema otsinyuvannya vidpovidnosti suchasnym vymoham navchal′noho kontentu spetsial′nosti kiberbezpeka [Information and analytical system for assessing the compliance of educational content specialties ciber security with modern requirements]. *Radioelectronic and Computer Systems*, 2021, no. 1, pp. 70-80. DOI: 10.32620/reks.2021.1.06. (In Ukrainian)

22. Dovbysh A, Liubchak V, Shelehov I, Simonovskiy J, Tenytska A. Information-extreme machine learning of a cyber attack detection system. *Radioelectronic and Computer Systems*, 2022, no. 3, pp. 121-131. DOI: 10.32620/reks.2022.3.09.

23. Bobrovnikova, K., Lysenko, S., Savenko, B., Gaj, P., & Savenko, O. Technique for IoT malware detection based on control flow graph analysis. *Radioelectronic and Computer Systems*, 2022, no. 1, pp. 141-153. DOI: 10.32620/reks.2022.1.11.

24. Lazzarini, R., Tianfield, H., & Charissis, V. Federated Learning for IoT Intrusion Detection. *AI*, 2023, vol. 4, iss. 3, pp. 509-530. DOI: 10.3390/ai4030028.

25. Musleh, D., Alotaibi, M., Alhaidari, F., Rahman, A., & Mohammad, R. M. Intrusion Detection System Using Feature Extraction with Machine Learning Algorithms in IoT. *J. Sens. Actuator Netw.*, 2023, vol. 12, iss. 2, article no. 29. DOI: 10.3390/jsan12020029

26. Al Amien, J., Ab Ghani, H., Md Saleh, N. I., Ismanto, E., & Gunawan, R. Intrusion detection system for imbalance ratio class using weighted XGBoost classifier. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 2023, vol. 21, iss. 5, article no. 1102. DOI: 10.12928/telkomnika.v21i5.24735.

27. Cutler, A., Cutler, D. R., & Stevens, J. R. Random Forests. In: Zhang C, Ma Y, editors. *Ensemble Machine Learning*, New York, NY: Springer New York; 2012, pp. 157-175. DOI: 10.1007/978-1-4419-9326-7_5.

28. Yeung, K. Y., & Ruzzo, W. L. Principal component analysis for clustering gene expression data. *Bioinformatics*, 2001, vol. 17, iss. 9, pp. 763-774. DOI: 10.1093/bioinformatics/17.9.763.

29. Kramer, O. K-Nearest Neighbors. *Dimensionality Reduction with Unsupervised Nearest Neighbors. Intelligent Systems Reference Library*, vol. 51, Berlin, Heidelberg: Springer Berlin Heidelberg; 2013, pp. 13-23. DOI: 10.1007/978-3-642-38652-7_2.

30. Brodley, C. E., & Utgoff, P. E. Multivariate Decision Trees. *Machine Learning*, 1995, vol. 19, iss. 1, pp. 45-77. DOI: 10.1023/A:1022607123649.

Received 25.12.2023, Accepted 20.02.2024

НОВА МОДЕЛЬ ВИЯВЛЕННЯ АНОМАЛІЙ ДЛЯ ПРОМИСЛОВОГО ІНТЕРНЕТУ РЕЧЕЙ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Лахсен Ідуглід, Саїд Ткатек, Халід Ель Файк, Азідін Гуеззаз

В останні десятиліття повсюдна інтеграція технологій інтернету речей (IoT) зробила революцію в різних секторах, включаючи Індустрію 4.0, телекомунікації, хмарні обчислення та системи охорони здоров'я. Додатки Індустрії 4.0 змінили глобальну динаміку бізнесу, спрямовану на інноваційні бізнес-моделі завдяки вдосконаленим технологіям автоматизації. Однак забезпечення безпеки IoT залишається критично важливим завданням як для дослідників, так і для практиків галузі, оскільки конфіденційність і безпека стають першочерговими питаннями для захисту цих технологій, що еволюціонують. Це дослідження спрямоване на вирішення нагальної потреби в посиленні заходів безпеки в середовищі Індустрії 4.0, пропонуючи нову модель виявлення вторгнень, оцінену на основі набору даних Bot-IoT. Основною метою дослідження є розробка системи виявлення вторгнень (IDS), здатної відстежувати, виявляти і приймати надійні рішення у відповідь на вторгнення в режимі реального часу. Використовуючи різні методи глибокого навчання (DL) та машинного навчання (ML), запропонована модель має на меті покращити як швидкість виявлення (DR), так і точність (ACC) IDS. Наші експериментальні результати демонструють, що запропонована модель перевершує традиційні методи, такі як K-NN і DT, досягаючи значної точності 98,9% і швидкості виявлення 97,8%. Порівняння з існуючими підходами до виявлення вторгнень підтверджує стійкість та надійність запропонованої моделі. Отже, дослідження представляє перспективне рішення для захисту середовищ Індустрії 4.0 та є певним внеском у сфері кібербезпеки. Демонструючи більшу продуктивність у порівнянні з існуючими підходами з використанням набору даних Bot-IoT, запропонована IDS має потенціал для впровадження і посилення безпеки екосистем Індустрії 4.0.

Ключові слова: безпека Індустрії 4.0; IIoT; IoT; виявлення аномалій; вибір ознак; випадковий ліс.

Лахсен Ідуглід – аспірант, Дослідницька лабораторія комп'ютерних наук, Університет Кенітра, Марокко.

Саїд Ткатек – д-р комп'ютерних наук, професор, Дослідницька лабораторія комп'ютерних наук, Університет Кенітра, Марокко.

Халід Ель Файк – аспірант, Дослідницька лабораторія комп'ютерних наук, Університет Кенітра імені Ібн Тофаїла, Марокко.

Азідін Гуеззаз – д-р комп'ютерних наук, професор, Вища школа технологій, Університет Каді Айяд, Ес-Сувейра, Марокко.

Lahcen Idouglid – PhD, Computer Sciences Research Laboratory, Ibn Tofail University Kenitra, Morocco,
e-mail: lahcen.idouglid@uit.ac.ma, ORCID: 0009-0008-6570-9869, Scopus Author ID: 57916861600.

Said Tkatek – Doctor of Computer Sciences, Professor, Computer Sciences Research Laboratory, Ibn Tofail University Kenitra, Morocco,
e-mail: saidtkinfo@yahoo.fr, Scopus Author ID: 56968120800.

Khalid Elfayq – PhD, Computer Sciences Research Laboratory, Ibn Tofail University Kenitra, Morocco,
e-mail: khalid.elfayq@uit.ac.ma, Scopus Author ID: 57916861500.

Azidine Guezzaz – Doctor of Computer Sciences, Professor, Higher School of Technology, Cadi Ayyad University, Essaouira, Morocco,
e-mail: a.guzzaz@gmail.com, Scopus Author ID: 57194492918.