engineering proceedings

MDPI

Proceeding Paper

Anomaly Detection on Network Traffic for the Healthcare Internet of Things †

Hsiao-Ching Huang, I-Hsien Liu, Meng-Huan Lee and Jung-Shian Li *

Department of Electrical Engineering, Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan 70101, Taiwan; hchuang@cans.ee.ncku.edu.tw (H.-C.H.); ihliu@cans.ee.ncku.edu.tw (I.-H.L.); mhlee@cans.ee.ncku.edu.tw (M.-H.L.)

* Correspondence: jsli@cans.ee.ncku.edu.tw

† Presented at the IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability, Tainan, Taiwan, 2–4 June 2023.

Abstract: The Internet of Things (IoT) has revolutionized technologies in society, including in households, offices, factories, and health centers. Among these, the Healthcare Internet of Things (HIoT) significantly transforms medical assistance for patients. By using wearable devices with remote network connections, caregivers monitor patients' physiological data to gain valuable insights into their health conditions. Despite the many benefits of the HIoT, several security vulnerabilities still exist. Hackers can exploit the internet connection to steal or modify credential information regarding patients, violating the integrity and confidentiality of the security policy. Moreover, they can launch cyberattacks on hospitals or critical life-support systems, further endangering patients' lives. Consequently, it is crucial to implement robust cybersecurity measures to enhance the security of healthcare services. Therefore, we proposed an anomaly detection method based on network traffic for the HIoT, adopting Markov models. Owing to their simplicity, interpretability, and well-developed theory, the Markov models have been applied to network traffic prediction and modeling, serving as a viable approach to cater to our needs. We evaluated the proposed method using the public dataset ToN_IoT and analyzed the results.

Keywords: healthcare internet of things; anomaly detection; network traffic; Markov models

Citation: Huang, H.-C.; Liu, I.-H.; Lee, M.-H.; Li, J.-S. Anomaly Detection on Network Traffic for the Healthcare Internet of Things. *Eng. Proc.* **2023**, *55*, 3. https://doi.org/10.3390/engproc2023055003

Academic Editors: Teen-Hang Meen, Kuei-Shu Hsu and Cheng-Fu Yang

Published: 22 November 2023

1. Introduction

The Healthcare Internet of Things (HIoT) redefines health services for patients. Due to the exponential growth in the network traffic generated by all connected devices in the HIoT, monitoring the network's performance and overcoming its inefficiencies pose a challenge. Network traffic prediction is one of the subfields of Network Traffic Monitoring and Analysis (NTMA) [1], which focuses on analyzing past characteristics of network traffic to predict future trends. This serves as a solution to be addressed, particularly in anomaly detection. Anomaly detection is crucial to cybersecurity and is further integrated into an intrusion detection system (IDS). An anomaly-based IDS, in comparison to its signature-based counterpart, characterizes the normal behavior of a system to differentiate attack traffic, whereas its counterpart searches for features that directly match the attack traffic from its pre-built database. Before the advent of the HIoT and the proliferation of IoT applications, most of the prediction and anomaly detection methods only considered univariate time series. However, network traffic consists of different attributes and statistical contexts, such as packet counts, interarrival time, protocol type, and connection status. Focusing solely on univariate time series may overlook the underlying correlation between the different attributes [2]. Therefore, in this study, we adopted a multivariate analysis to detect anomalies in network traffic in the HIoT environment.

2. Related Works

Anomaly detection methods have been a subject of ongoing research in many fields of study. Recent studies mostly focus on various machine learning and deep learning algorithms for anomaly detection. Wu et al. [3] applied the graph neural network (GNN) for anomaly detection in Industrial Internet of Things (IIoT) scenarios, specifically to the studies on smart factories, smart transportation, and smart energy. They provided a comprehensive investigation of different types of anomalies such as point anomalies, contextual anomalies, and collective anomalies. Chen et al. [4] proposed a transformer-based framework, called GTA, to learn graph structures for multivariate time series' anomaly detection in IoT sensor data. Park et al. [5] addressed the issue of data imbalance in an AI-based network intrusion detection system (NIDS) using a generative adversarial network (GAN) by generating synthetic data for minor attacks, along with an autoencoder-driven model for detection. Furthermore, Liu et al. [6] combined an attention-mechanism-based convolutional neural network long short-term memory (AMCNN-LSTM) model with federated learning to detect edge device failures. Qi et al. [7] introduced a novel approach called MDS_AD, which employed locality-sensitive hashing (LSH), isolation forest, and principal component analysis (PCA) to detect point and group anomalies considering multi-aspect data. Regarding studies of Markov models and network traffic analysis, Aceto et al. [8] applied high-order Markov chains and hidden Markov models (HMM) to predict mobile-app traffic. Liu et al. [9] introduced tensor operations to a multivariate, multi-order Markov chain for network traffic's multi-modal prediction.

3. Methodology

The multivariate high-order Markov chain with Hellinger distance (MHMC-HD) was proposed for detecting anomalies in network traffic for HIoT scenarios in this study.

3.1. Problem Formulation

Let $X = \{X_1, X_2, X_3, \ldots, X_{t-1}, X_t, X_{t+1}, \ldots\}$ be a set of consecutive random variables that describe the state of each network traffic flow at time t. The finite state set is denoted as

$$S \equiv \{1, 2, 3, \ldots, I\} \tag{1}$$

where I represents the total number of states. Similar to network traffic prediction, our objective is to obtain the transition probability of the state at the next time step, given the states at k's preceding ones.

3.2. MHMC

In a classical first-order Markov chain, the current state is determined solely by the previous state.

$$\begin{aligned} &\mathrm{P}(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \ldots X_0 = i_0) \\ &= \mathrm{P}(X_{t+1} = j | X_t = i) \\ &= p_{i,j} \end{aligned} \tag{2}$$

where state $j, i, i_0, \ldots i_{t-1} \in S$. The assumption of temporal homogeneity is made, meaning that the transition probability does not depend on time t. Hence, the transition probability matrix can be expressed as follows:

$$p_{i,j} = \mathrm{P}(X_{t+1} = j | X_t = i) \tag{3}$$

$$P' = (p_{i,j}) \tag{4}$$

where $P \in \mathbb{R}^{I \times I}, \sum_{i=0}^{I} p_{i,j} = 1$.

Compared to the classical Markov chain, a k-order Markov chain not only depends on the previous state but also takes into consideration the k's preceding states.

$$\begin{aligned} &\mathrm{P}(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \ldots X_0 = i_0) \\ &= \mathrm{P}(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \ldots, X_{t-k+1} = i_{t-k+1}) \\ &= p_{i_{t-k+1}, \ldots, i_{t-1}, i, j} \end{aligned} \tag{5}$$

where $i_{t-k+1}, \ldots, i_{t-1}, i, j \in S$. Furthermore, for a network traffic flow, multiple attributes can be obtained, such as interarrival time, packet length, and others. If we apply separate univariate Markov chains to different attributes, the hidden correlation between these attributes may be neglected. Therefore, it is important to retain the underlying correlation between the different variables. In an m-variate k-order Markov chain, the state space can be defined as follows:

$$S' \equiv \{(1, 1, \ldots, 1), (1, 1, \ldots, 1, 2), \ldots, (I_1, I_2 \ldots, I_m)\}. \tag{6}$$

Thus, the transition probability can be represented as follows:

$$\begin{aligned} &p_{i_{t+1,1}, i_{t+1,2}, \ldots, i_{t+1,m}, i_{t,1}, i_{t,2}, \ldots, i_{t,m}, \ldots, i_{t-k+1,1}, i_{t-k+1,2}, \ldots, i_{t-k+1,m}} \\ &= \mathrm{P}(X_{t+1,1}, X_{t+1,2}, \ldots, X_{t+1,m} = i_{t+1,1}, i_{t+1,2}, \ldots, i_{t+1,m} | \\ &X_{t,1}, X_{t,2}, \ldots, X_{t,m} = i_{t,1}, i_{t,2}, \ldots, i_{t,m}, \ldots, X_{t-k+1,1}, X_{t-k+1,2}, \\ &\ldots, X_{t-k+1,m} = i_{t-k+1,1}, i_{t-k+1,2}, \ldots, i_{t-k+1,m}) \end{aligned} \tag{7}$$

where

$$i_{t+1,1}, i_{t+1,2}, \ldots, i_{t+1,m}, i_{t,1}, i_{t,2}, \ldots, i_{t,m}, \ldots, i_{t-k+1,1}, i_{t-k+1,2}, \ldots, i_{t-k+1,m} \in S'$$

and the transition matrix can then be converted to a tensor

$$P'' \in \mathbb{R}^{I_1, I_2 \ldots, I_m \times \ldots \times I_1, I_2 \ldots I_m}.$$

3.3. Maximum Likelihood Estimation (MLE)

After acquiring the network traffic flows, the unknown transition probability tensor is estimated based on these observations. The MLE is a common technique used for this purpose. For a classical first-order Markov chain, the transition probability matrix can be constructed as follows:

$$\hat{p_{i,j}} = \frac{n_{ij}}{\sum_k n_{ik}} \tag{8}$$

$$n_{ij} = \sum_{t=0}^{N_t} 1_{\{X_t = i\}} 1_{\{X_{t+1} = j\}} \tag{9}$$

where $\hat{p_{i,j}}$ and n_{ij} are the estimated transition probability and the count of transitions from state i to state j, respectively. N_t denotes the total number of time steps in the training data, and $1_{\{.\}}$ is an indicator function.

3.4. MHMC with Hellinger Distance (MHMC-HD)

Initially, in the approach of this study, we applied the Hellinger distance to determine whether testing data samples exhibited a similar underlying probability distribution to the training data. The Hellinger distance is a measure to quantify the dissimilarity between two discrete probability distributions. According to [8], we considered two hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, to investigate whether the two datasets were represented by the same Markov model or by different ones. For a first-order Markov chain, the two probability distributions were described through $S \times S$ matrices $\mathbf{\Pi}^x$ and $\mathbf{\Pi}^y$, where each Π was one-to-one mapped to the corresponding transition matrix, with S denoting the finite state set.

Similar to estimating unknown transition matrices, the matrix Π was obtained using MLE.

$$\hat{\pi} = \frac{n_{ij}}{n} \tag{10}$$

Subsequently, we applied the Hellinger distance to measure the dissimilarity between the two matrices $\mathbf{\Pi}^x$ and $\mathbf{\Pi}^y$.

$$H\left(\hat{\mathbf{\Pi}}^x, \hat{\mathbf{\Pi}}^y\right) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{S} \sum_{j=1}^{S} \left(\sqrt{\hat{\pi}_{i,j}^x} - \sqrt{\hat{\pi}_{i,j}^y}\right)^2} \tag{11}$$

Given a threshold γ, if $H\left(\hat{\mathbf{\Pi}}^x, \hat{\mathbf{\Pi}}^y\right) > \gamma$, the two datasets belonged to different probability distributions ($\mathcal{H}_0$), whereas $H\left(\hat{\mathbf{\Pi}}^x, \hat{\mathbf{\Pi}}^y\right) < \gamma$ suggested that the two datasets used the same Markov model ($\mathcal{H}_1$). Following that, we applied MHMC to assess the probability of the generation of each testing data considering k.

4. Experiment and Result

4.1. Network Traffic Data

To evaluate our approach for anomaly detection on HIoT network traffic, we used the ToN_IoT datasets. The datasets comprised IoT/IIoT telemetry data from sensors, operating system data from Windows and Linux systems, as well as network traffic data collected during normal operations and under various attack interferences [10–12]. The network traffic datasets were derived from pcap and log files with Zeek logs. Among the various attributes in the dataset, we specifically selected source payload (src_bytes), destination payload (dest_bytes), and connection state (conn_state) to evaluate the proposed MHMC-HD approach. The testing data included different types of attack techniques, including DoS, injecting, ransomware, password attacks, and more [13–15].

4.2. Performance Metrics

We adopted the following common performance metrics to evaluate the results using the confusion matrix [16]. The confusion matrix provided an overview of the outcomes of predictive analytics and classification studies, presenting four different cases. Each case represented the number of testing data samples falling into one of the following categories as shown in Table 1: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN).

Table 1. Confusion Matrix.

Confusion Matrix		**Actual Condition**	
		Positive	**Negative**
Predicted Condition	Positive	TP	FP
	Negative	FN	TN

- Precision: The number of correctly detected anomaly samples over the total number of samples predicted as anomalies.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

- Recall: The number of correctly detected anomaly samples over the total number of actual anomaly samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

- F1-score: The harmonic mean of precision and recall providing a balance measure for the model's performance.

$$\mathrm{F1} = 2\frac{(\mathrm{Precision})(\mathrm{Recall})}{\mathrm{Precision} + \mathrm{Recall}} \tag{14}$$

- True Negative Rate (TNR): A metric for evaluating the false alarm rate. The number of correctly predicted normal traffic samples over the total number of normal traffic samples, as follows:

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \tag{15}$$

When evaluating the testing dataset consisting of 80,000 flows, we implemented the MHMC-HD with a threshold of $\gamma = 0.5$ for the Hellinger distance measure and an order of $k = 4$ for the MHMC. Next, we compared the MHMC-HD with three other approaches, including a four-order MHMC, the Hellinger distance measure without MHMC, and an ML-based long short-term memory (LSTM) with an autoencoder (AE). The results are presented in Table 2 and Figure 1. It was found that the MHMC_HD performed the best in terms of precision, F1, and TNR, while the LSTM-AE achieved the highest score in the recall metric.

Table 2. Comparison of the results with different methods.

Methods	Precision	Recall	F1	TNR
Four-order MHMC	0.9815	0.9951	0.9882	0.9436
Hellinger Distance	0.8894	0.9875	0.9359	0.6316
LSTM + AE	0.8104	**1**	0.8995	0.3
MHMC-HD	**0.9940**	0.9908	**0.9924**	**0.9821**

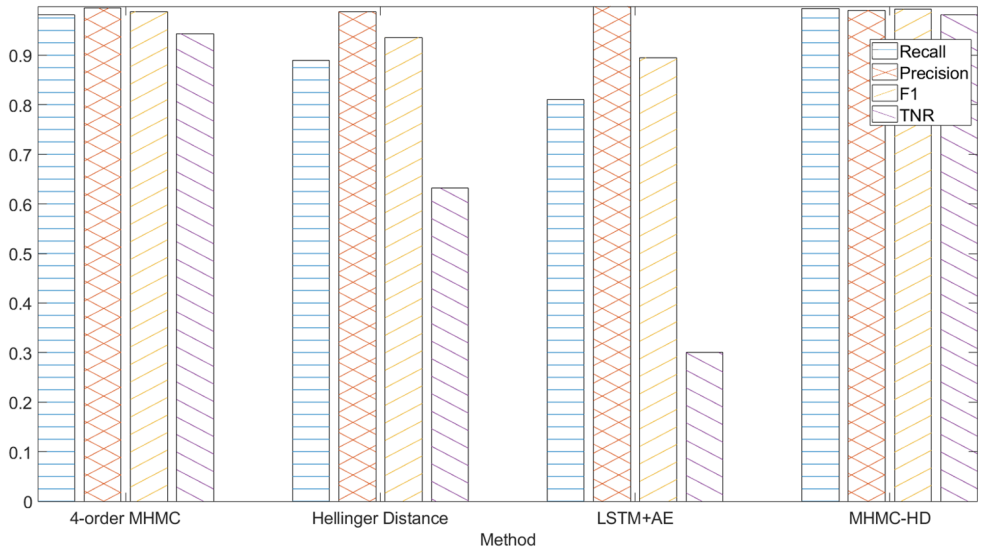

Figure 1. Comparison of the results with different methods.

To investigate the influence of a different order (k) on the MHMC-HD in terms of its performance in anomaly detection, we conducted experiments from order one to ten. The results are presented in Table 3 and Figure 2. The recall and F1 scores were improved at a higher order but the TNR metric dropped as the order increased. In Figure 3, a significant improvement of the TNR is shown, after integrating the Hellinger distance measure into the MHMC. This improvement indicated a reduction in the false alarm rate when implementing anomaly detection in IoT network traffic. Overall, the results suggested that the enhancement of the recall and F1 scores through the implementation of

a higher order in the MHMC-HD undermined the TNR metric. Moreover, the integration of the Hellinger distance effectively reduced the false alarm rate in anomaly detection in IoT traffic.

Table 3. Evaluation of the MHMC-HD with different orders of *k*.

Order k	Precision	Recall	F1	TNR
1	**0.9979**	0.5164	0.6806	**0.9968**
2	0.9968	0.9387	0.9669	0.9910
3	0.9952	0.9816	0.9884	0.9858
4	0.9940	0.9908	**0.9924**	0.9821
5	0.9933	0.9950	0.9942	0.9799
6	0.9929	0.9968	0.9949	0.9786
7	0.9926	0.9979	0.9952	0.9775
8	0.9923	0.9993	0.9958	0.9766
9	0.9919	0.9996	0.9957	0.9756
10	0.9917	**0.9997**	0.9957	0.9748

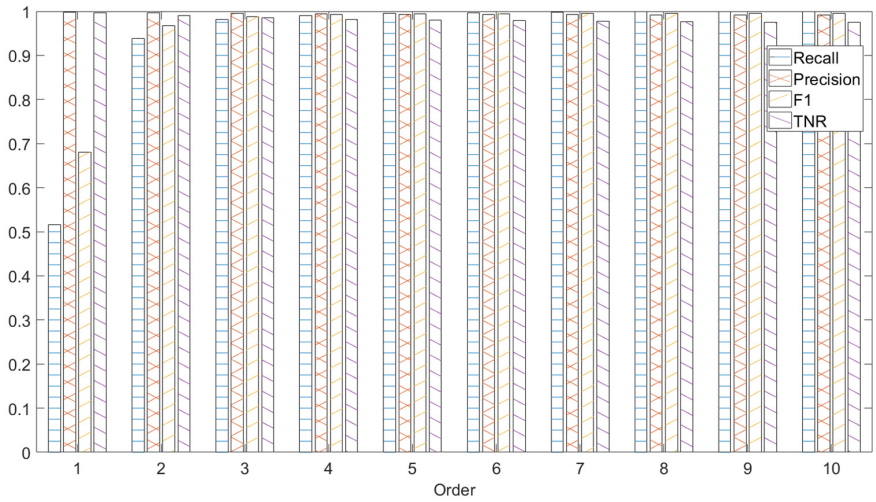

Figure 2. Evaluation of *k*-order Markov chain with Hellinger distance concerning *k*.

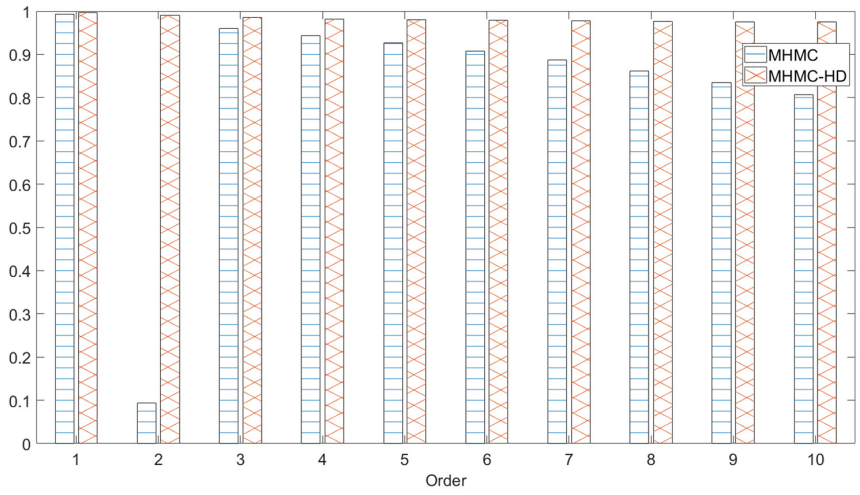

Figure 3. Comparison of the TNR metric depending on the existence of the Hellinger distance.

5. Conclusions

We developed the MHMC-HD method to detect anomalies specifically for attacking traffic flows on the IoT's network traffic. The impact of a higher order on the performance

of the MHMC-HD was investigated. The results showed a considerable improvement in anomaly detection. The results also demonstrated that integrating the Hellinger distance into the MHMC produced a low false alarm rate, thereby enhancing the reliability of anomaly detection in HIoT network traffic analysis.

Author Contributions: Writing—original draft, H.-C.H.; Validation, M.-H.L.; Writing—review & editing, I.-H.L.; Supervision, J.-S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Science and Technology Council (NSTC) in Taiwan under contract numbers 111-2221-E-006-079- and 112-2634-F-006-001-MBK.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The public data used in this work is mentioned and cited in the text. Link to this dataset: https://research.unsw.edu.au/projects/toniot-datasets (accessed on 1 May 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lohrasbinasab, I.; Shahraki, A.; Taherkordi, A.; Jurcut, A.D. From statistical- to machine learning-based network traffic prediction. *Trans. Emerg. Telecommun. Technol.* **2021**, *33*, e4394. [CrossRef]
2. Sha, W.; Zhu, Y.; Chen, M.; Huang, T. Statistical Learning for Anomaly Detection in Cloud Server Systems: A Multi-Order Markov Chain Framework. *IEEE Trans. Cloud Comput.* **2018**, *6*, 401–413. [CrossRef]
3. Wu, Y.; Dai, H.-N.; Tang, H. Graph Neural Networks for Anomaly Detection in Industrial Internet of Things. *IEEE Internet Things J.* **2022**, *9*, 9214–9231. [CrossRef]
4. Chen, Z.; Chen, D.; Zhang, X.; Yuan, Z.; Cheng, X. Learning Graph Structures with Transformer for Multivariate Time-Series Anomaly Detection in IoT. *IEEE Internet Things J.* **2022**, *9*, 9179–9189. [CrossRef]
5. Park, C.; Lee, J.; Kim, Y.; Park, J.-G.; Kim, H.; Hong, D. An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks. *IEEE Internet Things J.* **2023**, *10*, 2330–2345. [CrossRef]
6. Liu, Y.; Garg, S.; Nie, J.; Zhang, Y.; Xiong, Z.; Kang, J.; Hossain, M.S. Deep Anomaly Detection for Time-Series Data in Industrial IoT: A Communication-Efficient On-Device Federated Learning Approach. *IEEE Internet Things J.* **2021**, *8*, 6348–6358. [CrossRef]
7. Qi, L.; Yang, Y.; Zhou, X.; Rafique, W.; Ma, J. Fast Anomaly Identification Based on Multiaspect Data Streams for Intelligent Intrusion Detection Toward Secure Industry 4.0. *IEEE Trans. Ind. Inform.* **2022**, *18*, 6503–6511. [CrossRef]
8. Aceto, G.; Bovenzi, G.; Ciuonzo, D.; Montieri, A.; Persico, V.; Pescapé, A. Characterization and Prediction of Mobile-App Traffic Using Markov Modeling. *IEEE Trans. Netw. Serv. Manag.* **2021**, *18*, 907–925. [CrossRef]
9. Liu, H.; Yang, L.T.; Chen, J.; Ye, M.; Ding, J.; Kuang, L. Multivariate Multi-Order Markov Multi-Modal Prediction with Its Applications in Network Traffic Management. *IEEE Trans. Netw. Serv. Manag.* **2019**, *16*, 828–841. [CrossRef]
10. Moustafa, N. A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets. *Sustain. Cities Soc.* **2021**, *72*, 102994. [CrossRef]
11. Ashraf, J.; Keshk, M.; Moustafa, N.; Abdel-Basset, M.; Khurshid, H.; Bakhshi, A.D.; Mostafa, R.R. IoTBoT-IDS: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities. *Sustain. Cities Soc.* **2021**, *72*, 103041. [CrossRef]
12. Booij, T.M.; Chiscop, I.; Meeuwissen, E.; Moustafa, N.; Hartog, F.T.H.D. ToN_IoT: The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in IoT Network Intrusion Data Sets. *IEEE Internet Things J.* **2022**, *9*, 485–496. [CrossRef]
13. Alsaedi, A.; Moustafa, N.; Tari, Z.; Mahmood, A.; Anwar, A. TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems. *IEEE Access* **2020**, *8*, 165130–165150. [CrossRef]
14. Moustafa, N.; Keshky, M.; Debiez, E.; Janicke, H. Federated TON_IoT Windows Datasets for Evaluating AI-Based Security Applications. In Proceedings of the 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, China, 29 December 2020–1 January 2021.

15. Moustafa, N.; Ahmed, M.; Ahmed, S. Data Analytics-Enabled Intrusion Detection: Evaluations of ToN_IoT Linux Datasets. In Proceedings of the 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, China, 29 December 2020–1 January 2021.
16. Fathi-Kazerooni, S.; Rojas-Cessa, R. Countering Machine-Learning Classification of Applications by Equalizing Network Traffic Statistics. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 3392–3403. [CrossRef]