





Advancing Security in SDN-IoT Networks: DL-Based Autonomous Anomaly Detection with Enhanced Cross-Validation for Poisoning Attack Detection

Tharindu Lakshan Yasarathna^(✉) and Nhien-An Le-Khac^{}

School of Computer Science, University College Dublin, Belfield, Dublin, Ireland
computerscience@ucd.ie, tharindu.yasarathna@ucdconnect.ie

Abstract. The convergence of Software-Defined Networking (SDN) and the Internet of Things (IoT) is the emergence of highly dynamic and heterogeneous SDN-IoT networks vulnerable to various cyber threats. In response, Autonomous Anomaly Detection (AAD) systems leveraging deep learning (DL) techniques have become crucial for securing SDN-IoT networks. However, DL-based AAD systems are susceptible to adversarial attacks, particularly in continual learning settings, where models must adapt to evolving threats and changing network conditions. This paper proposes an enhanced cross-validation strategy for poisoning attack detection in DL-based AAD systems deployed in SDN-IoT networks. By integrating advanced cross-validation techniques with anomaly detection algorithms, the framework aims to maintain DL model robustness against poisoning attacks and enhance overall security. Evaluations of popular baseline datasets have provided insights into the effectiveness of detection, highlighting strengths and limitations. The discussion emphasizes the challenges and improvements in existing detection methods and contributes to advancing DL-based AAD systems for SDN-IoT networks. In addition, Future research directions aim to enhance the proposed detection mechanism and optimize scalable detection algorithms.

Keywords: SDN · IoT · Network security · Continual learning · Autonomous anomaly detection · Poisoning attacks

1 Introduction

Integrating Software-Defined Networking (SDN) and the Internet of Things (IoT) has significantly reshaped the landscape of networked systems, presenting both opportunities and challenges. SDN enables centralized control and management of network resources, while IoT devices introduce many interconnected endpoints, resulting in highly dynamic and heterogeneous network environments [9]. However, this convergence also raises the complexity and security challenges in

SDN-IoT networks. The expanded attack surface, diverse communication protocols, and varying levels of device security make these networks susceptible to a wide array of cyber threats, including unauthorized access, data breaches, and distributed denial-of-service (DDoS) attacks [6].

In response to these challenges, Autonomous Anomaly Detection (AAD) systems have emerged as a critical line of defence for securing SDN-IoT networks. AAD systems leverage advanced machine learning (ML) and deep learning (DL) techniques to monitor network traffic autonomously and identify anomalous behaviour indicative of potential security incidents [7, 10]. However, the effectiveness of DL-based AAD systems depends on the robustness and resilience of the underlying DL models. Adversarial attacks exploit the inherent vulnerability of DL models, posing a significant threat to the robustness and resilience of DL-based AAD security systems developed for SDN-IoT networks [21]. Despite the emergence of various adversarial attacks, including the white box, black box, and grey box attacks, they can be further explored as three attack dimensions: data, model, and hybrid (data and model) level attacks. Data-level Black box Adversarial attacks, such as poisoning attacks, pose a significant threat to the reliability of DL models by injecting maliciously crafted data to manipulate their behaviour. Detecting poisoning attacks in DL-based AAD systems in SDN-IoT is crucial, particularly in continual learning settings where models must adapt to evolving threats and changing network conditions [18].

This continual learning process makes DL models vulnerable to adversarial attacks, as attackers exploit vulnerabilities in the model's architecture and training process [14]. Adversarial attacks in SDN-IoT networks can manifest in various forms, including data poisoning attacks aimed at manipulating the training data to degrade model performance. Additionally, model evasion attacks can perturb input data in real time to evade detection by the AAD system. The dynamic nature of SDN-IoT environments and data drift further complicates detecting adversarial attacks [15]. Therefore, models must adapt to rapidly changing network conditions and evolving attack strategies [16]. Addressing these challenges requires innovative approaches that enhance the robustness and resilience of DL-based AAD systems in continual learning settings, thereby ensuring the security of DL models against adversarial threats [17].

The paper proposes an enhanced cross-validation strategy for continual learning environments to address the challenge of poisoning attack detection in DL-based AAD systems deployed in SDN-IoT networks. By integrating advanced cross-validation techniques with anomaly detection algorithms, the framework aims to maintain DL model robustness against poisoning attacks and enhance overall security in SDN-IoT networks [25]. Through empirical evaluations and comparative analysis, the research aims to provide insights into the efficacy of the proposed approach and contribute to the advancement of DL-based AAD for SDN-IoT environments in continual learning settings. This paper's contributions to the research community can be listed as follows:

- The paper presents an innovative method for detecting data poisoning attacks on DL models used in AAD systems deployed in SDN-IoT networks within continual learning settings.
- We validate the proposed method through extensive experiments using popular baseline datasets, experimenting on popular DL models for the anomaly detection, such as CNN and LSTM architectures.
- Furthermore, we suggest enhancing the detection approach by incorporating advanced techniques and opening new research directions for scholars.

The rest of the paper is structured as follows: Sect. 2 covering the background of the study followed by related work in Sect. 3. The Sect. 4 outlines the proposed poisoning attacks detection framework, while the Sect. 5 presents the experimental setup and findings. Discussions cover challenges, lessons learned, and future research directions in Sect. 6. Finally, Sect. 7 concludes the paper with the importance of cross-validation strategies for poisoning attack detection in DL-based AAD for SDN-IoT networks in continual learning settings.

2 Background

2.1 SDN-IoT Networks and Security

SDN and IoT have revolutionized network architectures, offering flexibility and scalability. However, the integration of SDN and IoT introduces unique security challenges. SDN-IoT networks are characterized by a dynamic and interconnected environment, amplifying the attack surface and making them vulnerable to cyber threats. Understanding the security implications of SDN-IoT networks is crucial for developing effective security mechanisms [6, 9, 13].

2.2 Deep Learning-Based Autonomous Anomaly Detection for SDN-IoT

DL techniques have shown promise in autonomously detecting anomalies in SDN-IoT networks by leveraging neural networks to analyze complex network data [8]. Various DL-based approaches, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), Long short-term memory (LSTM), autoencoders (AE), and restricted Boltzmann machines (RBMs) have been proposed for AAD in SDN-IoT environments. These methods can identify suspicious activities indicative of anomalies, providing a proactive and adaptive approach to network security [1, 18] (Fig. 1).

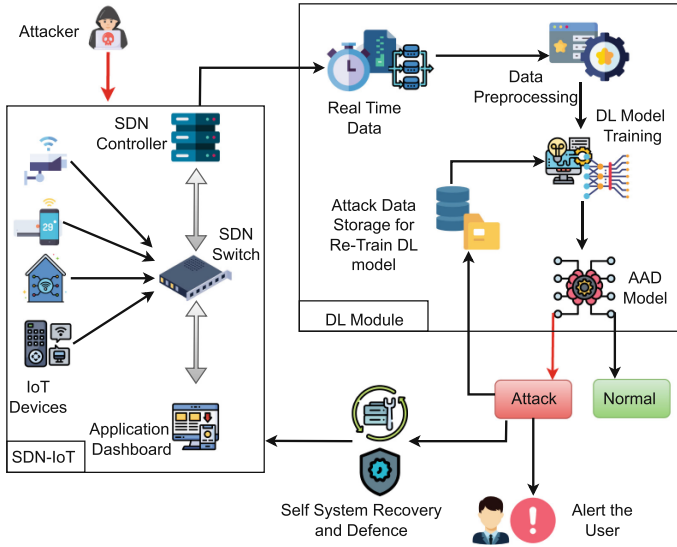


Fig. 1. DL-based AAD security system for SDN-IoT networks

The need for AAD over conventional Anomaly Detection (AD) methods in SDN-IoT networks arises from the challenges posed by the network’s dynamic nature and the diverse range of connected IoT devices. SDN-IoT networks exhibit dynamic behaviour, with devices joining and leaving the network dynamically and traffic patterns constantly changing. Conventional AD systems need help keeping pace with these dynamics and more adaptability to detect emerging threats effectively [23]. In contrast, AAD offers a robust and scalable solution by continuously learning and adapting to changes in network behaviour. By providing autonomous anomaly detection capabilities, AAD enhances the overall security of SDN-IoT networks, enabling organizations to detect and respond to security threats in real time, thus minimizing the risk of potential breaches and disruptions [7, 10, 21].

2.3 Poisoning Attacks on DL Models in Continual Learning Settings

Poisoning attacks cause a significant threat to the robustness of DL models, particularly in continual learning settings within SDN-IoT networks. These attacks can compromise the integrity and reliability of DL models over time by strategically injecting poisoned samples into the training data. By manipulating the model’s decision boundaries or inducing misclassifications of specific inputs, adversaries can degrade the performance and reliability of DL-based anomaly detection systems [15]. While essential for enabling DL models to adapt and improve their performance over time, continuous learning exposes them to vulnerabilities against poisoning attacks. As DL models continuously learn from incoming data streams, they become increasingly susceptible to malicious injec-

tions, which can alter their behaviour and compromise their effectiveness in detecting anomalies. Advanced detection mechanisms are essential to ensure the robustness of DL models in continual learning settings [20]. These mechanisms must be capable of detecting the poisoning attacks in DL-based AAD systems deployed in SDN-IoT networks (Fig. 2).

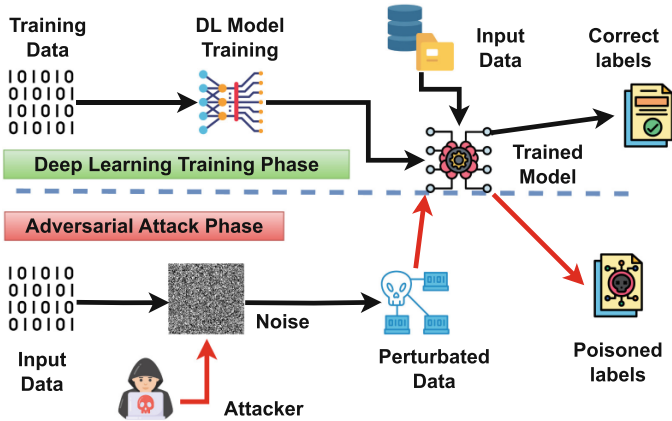


Fig. 2. Data level poisoning attack on DL model

2.4 Cross-Validation Strategies

The iterative process in continual learning introduces challenges such as model drift, catastrophic forgetting, and concept drift, which can compromise the performance of DL models over time [15]. Cross-validation strategies are one of the essential methods for maintaining the stability and reliability of DL models used in continual learning settings for DL-based AAD systems deployed in SDN-IoT networks. Cross-validation ensures the effectiveness and adaptability of DL-based anomaly detection systems in dynamically changing network environments [4]. These strategies help validate the performance of DL models on different subsets of data, ensuring that they generalize well and remain robust over time. In addition, techniques such as the Kolmogorov-Smirnov test are utilized to identify data drifts [26].

3 Related Works

DL-based AAD systems for SDN-IoT networks have seen significant advancements in detecting adversarial attacks and strengthening anomaly detection mechanisms. Numerous studies have explored the use of DL techniques, such as CNNs, RNNs, LSTM, and autoencoders, to identify abnormal patterns in

network traffic data, thus strengthening the security of SDN-IoT infrastructures [19,24]. For instance, Chaganti et al. proposed an LSTM-based approach achieving high accuracy in detecting network attacks [6]. Researchers have also developed detection mechanisms targeting adversarial attacks in DL models, including data poisoning and evasion techniques [11,22]. Alotaibi et al. discussed the challenges posed by adversarial attacks in ML/DL-based IDSs, emphasizing the need for robust DL models to enhance network security [2]. Bao et al. have proposed novel approaches to strengthen the robustness of DL models against increasingly sophisticated cyber threats [3].

Recent studies have focused on evaluating the performance of DL models using popular datasets, providing insights into the efficacy of various detection strategies [5,12]. Additionally, research efforts have extended to exploring cross-validation strategies tailored for continual learning settings to preserve DL model robustness against adversarial attacks, particularly those aimed at poisoning the training data [17]. However, recent works on continual learning have primarily focused on adversary-agnostic scenarios, overlooking the potential threat posed by data poisoning attacks [17].

4 Methodology: Poisoning Attack Detection Framework

The proposed poisoning attack detection framework within a continual learning setting employs an Enhanced Cross-Validation Strategy to ensure the robustness of DL models against data poisoning attacks. This strategy involves several vital steps to detect malicious data injections in data-level attacks. Specifically, the framework considers a task-incremental scenario where the model is continuously trained on new data while retaining knowledge from previous tasks. This setting allows the model to adapt to evolving network conditions and detect anomalies in real time.

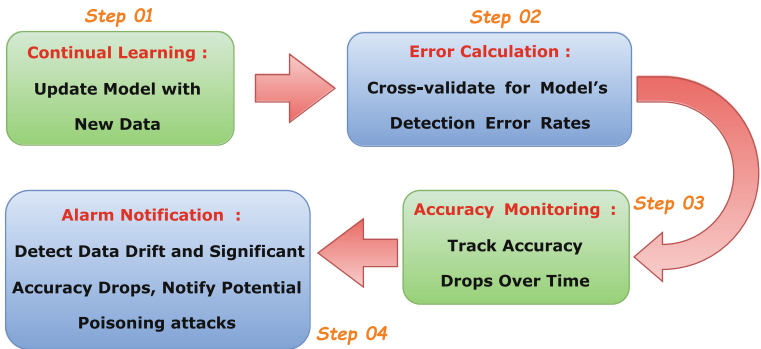


Fig. 3. Main steps of enhanced cross-validation strategy

The Enhanced Cross-Validation Strategy employed in this framework differs from traditional cross-validation approaches by focusing on identifying poten-

tial data poisoning attacks during model evaluation. While traditional cross-validation primarily assesses model performance on different subsets of data, the enhanced strategy incorporates additional checks for anomalies indicative of poisoning attacks. This strategy involves the utilization of specific metrics and techniques tailored for anomaly detection. In addition to detecting significant accuracy drops, it compares the model's detection error rates at different training stages to determine significant differences. If diffidence is identified, the framework employs the Kolmogorov-Smirnov test [26] to detect potential data drift or deviations in the training data distribution, which could indicate the presence of poisoned samples. This process is outlined in our previous study ¹ using a novel algorithm. In the current work, we enhanced it with additional steps. Main steps of the updated algorithm can listed as (see Fig. 3),

- **Continual learning:** Update the model with new data to ensure it remains relevant and accurate over time.
- **Error calculation:** Cross-validate for model's detection error rates.
- **Accuracy monitoring:** Track the model's accuracy over time to detect any drops that may indicate potential issues with its classification ability.
- **Alarm notification:** Detect data drift and significant accuracy drops, then notify potential poisoning attack or not.

5 Experiments

This section describes the experimental setup followed by the experiment results.

5.1 Experimental Setup

To validate our approach, we conducted experiments using three prominent datasets-CICIDS2017 ², InSDN ³, and CICIOT2023 ⁴ (refer to Table 1) that are chosen for their ability to represent diverse scenarios and challenges encountered in SDN-IoT environments. Specifically, we build a DL model to detect Distributed Denial of Service (DDoS) and malware threats using DL-based AAD in SDN-IoT networks. We used a modified CNN model and a simple LSTM model, which are likely supervised learning models. They are trained on non-poisoned data initially and continuously updated in a continual learning setup, indicating a supervised learning approach where the models learn from a labelled a examples. The overview of the AD model architectures we used in our study can be listed as follows:

¹ <https://github.com/tharinduyasarathna/Cross-validation-for-Detecting-Label-Poisoning-Attacks-A-Study-on-Random-Forest-Algorithm>.

² <https://www.unb.ca/cic/datasets/ids-2017.html>.

³ <https://aseados.ucd.ie/datasets/SDN/>.

⁴ <https://www.unb.ca/cic/datasets/iotdataset-2023.html>.

- Modified CNN Model: The CNN model consists of convolutional layers followed by activation functions (ReLU) and max-pooling layers to extract features from the input data. After flattening the output, two fully connected layers with dropout regularization are utilized to perform classification, followed by a softmax activation function to output probabilities for each class. The model has a total of approximately 4,083,586 trainable parameters.
- Simple LSTM Model: The LSTM model consists of a single LSTM layer followed by a fully connected layer with a softmax activation function for classification. The LSTM layer is used to capture temporal dependencies in sequential data. The model has significantly fewer parameters than the CNN model, with approximately 17,026 trainable parameters.

Table 1. Dataset descriptions

Dataset	Total samples	Samples used	Features used	Labels used
CICIDS2017	2,829,385	450,000	78	DDoS, DoS, Bot, BENIGN
InSDN	343,889	342,275	77	DDoS, DoS, BOTNET, Normal
CICIoT2023	234,745	234,745	39	DDoS, DoS, Mirai, Benign

However, to simulate real-world scenarios, we introduced diverse perturbations to poisoning the training data in the $t = 4$ time step. By evaluating the model’s performance through cross-validation with the original training data, we could detect significant drops in accuracy and higher detection error, signalling the presence of a poisoning attack.

For our experiments, we followed a standardized training procedure to ensure the robustness of the models and preprocess the data to optimize the model’s learning capabilities. Before dividing the dataset into training and testing sets, we initially utilized half of the dataset to create subsets of data for retraining the model at each time step, ensuring coverage of all attack scenarios. Subsequently, the remaining half of the dataset was divided into a 70:30 ratio for training and testing purposes. These steps are crucial for maintaining continual learning settings in our method. During the continual learning process, at intermediate time steps, we introduced poisoned data samples into the training subset to simulate potential poisoning attacks. Our method then monitored the model’s ability to detect these poisoning attacks while ensuring it could distinguish between poisoning attacks and data drift using the Kolmogorov-Smirnov test. This test provided valuable insights into the model’s susceptibility to data poisoning attacks by comparing the cumulative distribution functions of the model’s outputs.

We employed a range of metrics to assess the performance and robustness of the DL model, including accuracy, precision, recall, F1-score, and ROC-AUC. Additionally, we measured detection error rates from the models, KS statistics and p-value generated from the KS test to determine the significance of any observed differences, helping us identify instances of potential data drift or data

poisoning attack. All experiments were conducted on a system featuring a 12th Gen Intel(R) Core(TM) i7-1265U CPU, featuring 10 cores and 32 GB of system memory operating within a Windows environment.

5.2 Experimental Results

The outcomes of the experiments are presented in Table 2. Due to computational constraints and technical limitations, we limited the number of continual learning iterations to five. In Table 2, the detection error for the CICIDS2017 dataset in the T_4 step is 0.05, the detection error for the InSDN dataset is 0.095, and the CICIoT2023 dataset is 0.7493 for CNN. Which means it is suspicious of data drift or data poisoning attacks. Then, the KS static for CICIDS2017, InSDN, and CICIoT2023 datasets in T_4 time step 0.05, 0.084, and 0.085, respectively. It means there is no possible data drift and alarm for the data poisoning attack. Furthermore, based on the results presented in Table 3, our proposed method demonstrates encouraging performance on the LSTM model as well. Our experiment results prove the effectiveness of the proposed cross-validation strategy.

Table 2. Experiment results for the poisoning attack on CNN for CICIDS2017, InSDN and CICIoT2023 datasets. $T_0 - T_3$ are non-poisoning continual learning steps that use the median value of metrics due to space limitation, and T_4 is the poisoning step used in continual learning settings.

Metrics	CICIDS2017		InSDN		CICIoT2023	
	$T_0 - T_3$	T_4	$T_0 - T_3$	T_4	$T_0 - T_3$	T_4
Accuracy	0.9848	0.9782	0.9996	0.9850	0.9967	0.7535
Recall	0.9848	0.9782	0.9996	0.9850	0.9967	0.7535
Precision	0.9856	0.9803	0.9996	0.9867	0.9967	0.8938
F1 score	0.9850	0.9788	0.9996	0.9853	0.9967	0.7788
ROC-AUC	0.9760	0.9757	0.9493	0.9407	0.9962	0.9182
Detection error	0.035	0.050	0.030	0.095	0.011	0.7493
KS-static	—	0.05	—	0.084	—	0.085

6 Discussion

6.1 Challenges and Areas for Improvement

Detecting adversarial attacks in DL models within continual learning settings presents several challenges and opportunities for enhancement. One key challenge lies in the dynamic nature of network environments, which complicates the detection of adversarial attacks as DL models adapt to new data. To address

Table 3. Experiment results for the poisoning attack on LSTM for CICIDS2017, InSDN and CICIOT2023 datasets. $T_0 - T_3$ are non-poisoning continual learning steps that use the median value of metrics due to space limitation, and T_4 is the poisoning step used in continual learning settings

Metrics	CICIDS2017		InSDN		CICIOT2023	
	$T_0 - T_3$	T_4	$T_0 - T_3$	T_4	$T_0 - T_3$	T_4
Accuracy	0.9510	0.8869	0.9762	0.8535	0.8625	0.8395
Recall	0.9510	0.8869	0.9762	0.8535	0.8625	0.8395
Precision	0.9498	0.8316	0.9755	0.8664	0.8567	0.8309
F1 score	0.9496	0.8517	0.9746	0.8434	0.8590	0.7946
ROC-AUC	0.8321	0.6353	0.7809	0.7237	0.8860	0.8180
Detection error	0.1358	0.3639	0.096	0.4148	0.2909	0.3555
KS-static	–	0.296	–	0.3270	–	0.2997

this, we focus on improving the detection mechanisms to identify data poisoning attacks effectively. An important aspect to consider is the difficulty in distinguishing between genuine network anomalies and adversarial perturbations. Adversarial attacks may mimic benign data, making them challenging to detect using traditional anomaly detection techniques. Enhancing our detection methods to discriminate between legitimate and adversarial samples accurately is crucial for effectively assessing the vulnerability of DL models to such attacks.

Additionally, scalability and efficiency are critical factors to consider in the context of adversarial attack detection frameworks. In large-scale SDN-IoT deployments, the computational overhead associated with detecting adversarial perturbations can be substantial. Therefore, optimizing our detection algorithms and computational resources is essential to ensure real-time detection and response to adversarial threats without compromising performance.

6.2 Future Research Directions

Future research directions in the domain of DL-based AAD systems for IoT-SDN networks can further improve the poisoning attacks detection method in continual learning settings:

- **Enhance Poisoning Attack Detection Approach:** Improve and enhance the detection of poisoning attacks by conducting experiments with a broader range of DL and ML techniques. Investigate new techniques to address remaining challenges, such as imbalanced data, by leveraging advanced approaches like Denoising Adversarial AutoEncoder-based techniques.
- **Testing with Other Network Attacks:** Extend the evaluation of our method to include other types of network attacks beyond DDoS and malware, such as Man-in-the-Middle (MitM) Attacks, Packet Spoofing attacks, and insider threats. Investigating the performance of the detection method

across a broader spectrum of network threats will provide a comprehensive understanding of its capabilities and limitations.

- **Integration of Advanced Techniques:** Investigate the integration of advanced techniques, such as graph-based anomaly detection, adversarial training, and explainable AI, to enhance the detection method accuracy. These techniques can improve the ability to detect sophisticated and stealthy attacks while providing insights into the reasoning behind the detection decisions.
- **Adaptation to Evolving Threat Landscape:** Develop adaptive detection mechanisms capable of dynamically adjusting to changes in the threat landscape of SDN-IoT networks. It includes identifying emerging attack patterns, anticipating novel attack strategies, and updating the detection model to maintain effectiveness against evolving threats.

7 Conclusion

In conclusion, the paper proposes an enhanced cross-validation strategy for poisoning attack detection in DL-based AAD systems deployed in SDN-IoT networks within continual learning settings. Empirical evaluations across popular datasets such as CICIDS2017, InSDN, and CICIoT2023 validate the method's efficacy across DL architectures like CNN and LSTM. Experiment results demonstrate the effectiveness of the proposed cross-validation strategy. Additionally, the paper identifies challenges and areas for improvement in current detection methods, suggesting future research directions to enhance the detection approach using advanced technologies. The paper contributes to advancing security mechanisms in DL-based AAD systems deployed in SDN-IoT networks under continual learning settings.

References

1. Ahmed N et al (2022) Network threat detection using machine/deep learning in SDN-based platforms: a comprehensive analysis of state-of-the-art solutions, discussion, challenges, and future research direction. *Sensors* 22(20):7896
2. Alotaibi A, Rassam MA (2023) Adversarial machine learning attacks against intrusion detection systems: a survey on strategies and defense. *Future Internet* 15(2):62
3. Bao Z et al (2021) Threat of adversarial attacks on dl-based IoT device identification. *IEEE Internet Things J* 9(11):9012–9024
4. Bates S et al (2023) Cross-validation: what does it estimate and how well does it do it? *J Am Stat Assoc* 1–12
5. Chaabouni N, Mosbah M, Zemhari A, Sauvignac C, Faruki P (2019) Network intrusion detection for IoT security based on learning techniques. *IEEE Commun Surv Tutor* 21(3):2671–2701
6. Chaganti R, Suliman W, Ravi V, Dua A (2023) Deep learning approach for SDN-enabled intrusion detection system in IoT networks. *Information* 14(1):41
7. Cook AA, Misirlı G, Fan Z (2019) Anomaly detection for IoT time-series data: a survey. *IEEE Internet Things J* 7(7):6481–6494

8. Erhan L, Ndubuaku M, Di Mauro M, Song W, Chen M, Fortino G, Bagdasar O, Liotta A (2021) Smart anomaly detection in sensor systems: a multi-perspective review. *Inf Fusion* 67:64–79
9. Haji SH et al (2021) Comparison of software defined networking with traditional networking. *Asian J Res Comput Sci* 9(2):1–18
10. Hariharan A, Gupta A, Pal T (2020) Camlpad: cybersecurity autonomous machine learning platform for anomaly detection. In: *Advances in information and communication: proceedings of the 2020 future of information and communication conference (FICC)*, vol 2. Springer, pp 705–720
11. Hathaliya JJ, Tanwar S, Sharma P (2022) Adversarial learning techniques for security and privacy preservation: a comprehensive review. *Secur Privacy* 5(3):e209
12. Jazaeri SS, Jabbehdari S, Asghari P, Haj Seyyed Javadi H (2021) Edge computing in SDN-IoT networks: a systematic review of issues, challenges and solutions. *Clust Comput* 1–42
13. Khalid M, Hameed S, Qadir A, Shah SA, Draheim D (2023) Towards SDN-based smart contract solution for IoT access control. *Comput Commun* 198:1–31
14. Khan H et al (2022) Adversarially robust continual learning. In: *2022 International joint conference on neural networks (IJCNN)*. IEEE, pp 1–8
15. Korycki L, Krawczyk B (2023) Adversarial concept drift detection under poisoning attacks for robust data stream mining. *Mach Learn* 112(10):4013–4048
16. Kozal J et al (2023) Defending network ids against adversarial examples with continual learning. In: *2023 IEEE international conference on data mining workshops (ICDMW)*. IEEE, pp 60–69
17. Li H, Ditzler G (2022) Targeted data poisoning attacks against continual learning neural networks. In: *2022 International joint conference on neural networks (IJCNN)*. IEEE, pp 1–8
18. Li L, Xie T, Li B (2023) Sok: certified robustness for deep neural networks. In: *2023 IEEE symposium on security and privacy (SP)*. IEEE, pp 1289–1310
19. Matheu SN, Robles Enciso A, Molina Zarca A, Garcia-Carrillo D, Hernández-Ramos JL, Bernal Bernabe J, Skarmeta AF (2020) Security architecture for defining and enforcing security profiles in DLT/SDN-based IoT systems. *Sensors* 20(7):1882
20. Mundt M et al (2023) A wholistic view of continual learning with deep neural networks: forgotten lessons and the bridge to active and open world learning. *Neural Netw* 160:306–336
21. Qiu H et al (2020) Adversarial attacks against network intrusion detection in IoT systems. *IEEE Internet Things J* 8(13):10327–10335
22. Ramirez MA, Kim SK, Hamadi HA, Damiani E, Byon YJ, Kim TY, Cho CS, Yeun CY (2022) Poisoning attacks and defenses on artificial intelligence: a survey. *arXiv preprint [arXiv:2202.10276](https://arxiv.org/abs/2202.10276)*
23. Ren K, Zheng T, Qin Z, Liu X (2020) Adversarial attacks and defenses in deep learning. *Engineering* 6(3):346–360
24. Said Elsayed M, Le-Khac NA, Dev S, Jurcut AD (2020) Network anomaly detection using LSTM based autoencoder. In: *Proceedings of the 16th ACM symposium on QoS and security for wireless and mobile networks*, pp 37–45
25. Wang Y, Mianjy P, Arora R (2021) Robust learning for data poisoning attacks. In: *International conference on machine learning*. PMLR, pp 10859–10869
26. Wang Z, Wang W (2020) Concept drift detection based on Kolmogorov–Smirnov test. In: *Artificial intelligence in China: proceedings of the international conference on artificial intelligence in China*. Springer, pp 273–280