Deep Learning Models for Intrusion Detection in Wi-Fi Networks: A Literature Survey

Achmad Eriza Aminanto and Muhamad Erza Aminanto

Abstract Recently, the number of devices that are connected to the Internet are increasing exponentially due to the rise of the Internet of Things (IoT) era. Despite many advancements of the IoT era, we have been exposed to cyber security threats. Moreover, in this Covid-19 pandemic situation, the trend of cyber crimes is also increasing sharply. In this paper, we discuss one of possible countermeasures to combat cyber threats, namely Intrusion Detection Systems (IDS). IDS usually leverage many different types of machine learning models to detect the unknown attacks. In order to avoid confusion for future researchers in this field, we examine several states of the art papers which leverage deep learning for IDS in Wi-Fi networks. For this purpose, we choose one common Wi-Fi networks dataset, called AWID dataset. By examining the recent studies, we are able to understand current problems of IDS in Wi-Fi networks and able to prepare the best machine learning model for the corresponding problem to achieve a safe environment with minimal risk of cyber threats.

Keywords Intrusion detection system · Deep learning · AWID dataset · Anomaly detection

1 Introduction

One of the biggest media in Japan, the Japan Times [1], has reported that four promising technologies these days are big data, blockchain, AI (artificial intelligence), and IoT (internet of things). Presented in [2] and [3], approximately 30 million hosts will be connected to the Internet, output about US$ 175 billion in 2020 all over the world. At the same time, based on [4], the South Korean Smart

A. E. Aminanto (✉) · M. E. Aminanto
School of Strategic and Global Studies, Universitas Indonesia, Depok, Indonesia
e-mail: eriza@sci.ui.ac.id

M. E. Aminanto
e-mail: erza.aminanto@ui.ac.id

L. Yola et al. (eds.), *Sustainable Architecture and Building Environment*,
Lecture Notes in Civil Engineering 161,
https://doi.org/10.1007/978-981-16-2329-5_14

Home Energy Management segment's revenue is estimated to US$ 600 Million by 2024. However, the challenges [5] are cyber-security concerns and privacy breach. In Indonesia, the trend of cyber crime continues to increase every year. This cyber crime is increasingly becoming a major concern for the police of the Republic of Indonesia, especially after the Covid-19 pandemic. More and more users surf in cyberspace, especially social media which is increasing significantly. However, technological sophistication is currently being misused as a medium to commit crimes, especially cyber crimes. Quoted from digital media antaranews [6], which compiles from the National Police Criminal Investigation Unit [7], the trend of cyber crime continues to increase and the majority of cybercrimes are dominated by fraud. National Police also held a content contest regarding cyber crime awareness [8]. This competition shows the seriousness of the Police in dealing with cyber crime and means that cyber crime is currently becoming a very important concern for the Police.

IoT environments will mostly consist of small, light battery devices, and interconnected via wireless medium. Unfortunately, they will not execute computationally-high tasks such as cryptography computation because it will drain the battery instantly [9]. The huge amount of sensitive data transmitted in the air is appealing for attackers, yielding the IoT environment a main target of cyber-crime. The rapid spread of IoT-enabled devices has caused wireless networks to expose passive and active attacks, and their number has increased dramatically [10]. The wide and rapid spread of computing devices using Wi-Fi networks has produced complex, large and high-dimensional data, which can cause confusion when capturing attack attributes and force us to strengthen system security measures. Intrusion Detection System (IDS) is one of the most common components in every network security infrastructure [11]. However, it is a difficult task to develop an IDS with autonomous machine learning functions in an IoT environment.

Machine-learning has been leveraged broadly in the IDS research topic. IDS use machine learning models as the classification algorithm because of their model-free properties and capability to be learned [12]. By using the state of the art of machine-learning models namely deep learning, we can achieve impactful benefits for improving current IDSs in Wi-Fi networks [13].

We recognize that since a lot of previous studies used different machine learning models, there is confusion about how to properly adopt deep learning in IDS applications, particularly in Wi-Fi network attacks. Some studies only use deep learning methods in a partial sense, while other studies use shallow networks. The complexity of deep learning methods may be one of the reasons. In addition, deep learning methods require a lot of time to train properly. However, we found that some researchers have adopted deep learning methods throughout IDS. We compared the performance of several top-performance IDS in Wi-Fi networks. By comparing these IDSs, we expect other researchers who want to do research in this area can understand current situation and problems in the field of IDS in Wi-Fi networks.

2 IDS Background

Normally, we can separate IDS into misuse and anomaly detection [14]. The first model uses a rule to make use of actual portrayals to monitor the network. This model is also regularly referred to as a signature-based model. This model intends to recognize exact matching to the database. Despite the fact that this model is the most utilized everywhere [15], this model has evident drawbacks. The essential weak point is that it can't distinguish vague assaults on the grounds that it just thinks about the regarded traits of the assault. So as to preserve the exhibition of misuse detection, we have to refresh the assault signature without fail, which is problematic. Furthermore, attackers usually create an attack that does not consolidate previous assaults [16]. Such assaults make it extra tough to create fitting marks for the misuse detection. Then again, the focal point of inconsistency cognizance is to pick out irregular motion designs in the watched statistics [15]. Anomaly detection model for the most section control measurable examination and data mining techniques [17]. Since the arrangement model has the normal ability to do away with interruption examples and facts in the instruction stage, it can become aware of new types of assaults except in the past information.

Two normal strategies are typically utilized in IDS, to be particular clustering and classification undertakings. In the preliminary step, it is troublesome and costly to acquire infinite named community association information for regulated preparing. In modern years, clustering examinations have come to be a common anomaly discovery strategy [17]. Clustering itself is a solo information investigation technique that partitions a bunch of unlabeled records designs into numerous groups or corporations so the examples in the group are like one another, on the other hand not the same as the examples of exclusive clusters [17]. Simultaneously, classification is a method that acknowledges generous and vindictive offers based on the given information (commonly from clustering results). Clustering and classification can be done without difficulty utilizing AI strategies.

3 Dataset

For the purpose of this research, we choose one common dataset in Wi-Fi networks. This dataset released by Kolias et al., so-called Aegean Wi-Fi Intrusion Dataset (AWID) [18]. This dataset comprises two different types, the first type called "CLS" which has four labels, while another type called "ATK" which has 16 labels. Actually, the 4 labels in the CLS are big groups of 16 labels in the ATK. For example, simulated attacks in the CLS dataset consists of the Caffe-Latte, Hirte, Honeypot, and EvilTwin attack types listed in the ATK. The AWID dataset can also be divided into a full and partial dataset. In the partial one, there are 1,795,595 instances for training data, which comprises 1,633,190 normal instances and 162,385 attack instances. Meanwhile for testing data, they provided 575,643

Table 1 Class data distribution [18]

Type	Class	Training	Testing
Normal	Unbalanced	1.633.190	530.785
	Balanced	163.319	53.078
Attack	Impersonation	48.522	20.079
	Flooding	48.484	8.097
	Injection	65.379	16.682
	Total	162.385	44.858

instances in the partial dataset, consisting of 530,785 normal instances and 44,858 attack instances. This dataset contains 155 attributes such as frame length, radio tap present flag, radio tap data rate, wlan version, etc. Some data are presented in numerical and categorical form (Table 1).

4 State-of-the-Art IDS Using DL

In this section, we examine several state of the art publications which are the top performer in IDS for Wi-Fi networks. These publications can be regarded as the benchmark for next research in Wi-Fi networks attacks. At the end of this section, we summarize all research in the table.

Vaca et al. [19] proposed a WNIDS, which makes use of integrated learning. The advantages of integrated learning is that many basic learners are exploited to build a prediction model. By doing so, more accurate classification is expected. This paper also validated using AWID dataset. Wang et al. [20] examined various types of attacks in Wi-Fi networks using Deep Neural Networks (DNN) and Stacked Autoencoders (SAE). They achieved an impersonation detector in the AWID dataset.

Ran et al. [21] built a semi-supervised learning IDS model using ladder network. Their aim is to understand complex features with better discriminative ability for anomaly detection. The ladder network itself is constructed from Auto Encoders. Auto Encoder has the same number of neurons in both input and output. Ladder network used in this paper is composed of two sets of encoders, ordinary encoder and noise-encoder, and a decoder. They developed two steps of training, supervised and unsupervised steps [21]. AWID dataset is also used in this paper.

Parker et al. [22] proposed IDS for IoT infrastructures leveraging machine learning models. In this paper, they proposed two machine learning models, namely DEMISe and DETEReD. The first model combines feature extraction and mutual information. The Authors use two layers of stacked auto encoders concatenated with original attributes to be selected by mutual information method. This approach can be considered as filter based techniques since using mutual information for the feature selection step. THen the second machine learning model used is DETERed. The main difference is the wrapper based feature selection is used here.

Table 2 Publication comparisons of IDS that use AWID dataset

Paper	Dataset	Model	Classification type	Impersonation Acc (%)	Overall Acc (%)
Aminanto et al. [23]	AWID	FS+SAE	Binary	99.97	–
Parker et al. [22]	AWID	SAE+MI	Binary	98	–
Ran et al. [21]	AWID	DL-LN	Multi-class	89.32	98.54
Vaca et al. [19]	AWID	Ensemble	Multi-class	95.87	–
Wang et al. [20]	AWID	SAE+DNN	Multi-class	98.4	–

Aminanto et al. [23] developed an impersonation attack detector, D-FES using deep learning models. They used stacked auto encoder combined with other machine learning models. The stacked auto encoder is able to transform original attributes into different types of attributes that are more meaningful for the classification task. The extracted features are combined with original features to be selected by various feature selection techniques before classification. They successfully demonstrated the effectiveness of D-FES to detect impersonation attacks in AWID dataset.

From Table 2, we can see that Aminanto et al. [23] and Parker et al. [22] are focused to build the impersonation detector, while the rest three papers are multi class classification by distinguishing four labels provided in AWID. The best detection rate for impersonation class is Aminanto et al. [23] with 99.97%. While, the best overall detection rate is achieved by Ran et al. [21] with 98.54%. These numbers are benchmark data to be beaten for next research in IDS for Wi-Fi networks using AWID dataset.

5 Conclusions

We have examined several top performer IDS in Wi-Fi networks using state of the art machine learning models. As a summary, machine learning gives a significant impact in IDS research especially in Wi-Fi networks. Machine learning used as a feature engineering and classification step. In this paper we discuss many papers leveraging stacked auto encoder to extract more meaningful features from raw data. The output of this paper is the performance benchmark for next research in IDS for Wi-Fi networks, which are 99.97% in impersonation attack class and 98.54% of overall accuracy of four classes in AWID dataset. For future research, we believe that many ways will improve current work, such as decreasing the running time, making the computation decentralized to many devices, reducing false positive, and so on.

References

1. SHUSUKE MURAI. New Tokyo research center aims to boost Japan's fourth industrial revolution. Available on https://www.japantimes.co.jp/news/2018/07/05/business/tech/new-tokyo-research-center-aims-boost-japans-fourth-industrial-revolution/#.W2wwVigzbb0
2. Revenue of the internet of things in Japan from 2013 to 2020 (in billion U.S. dollars). Available on https://www.statista.com/statistics/512254/iot-revenue-japan/
3. Internet of Things (IoT) Connected devices installed base worldwide from 2015 to 2025 (in billions). Available on https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
4. iGateResearch. South Korea smart home market, volume, household penetration & key company analysis—forecast to 2024. Available on https://www.researchandmarkets.com/research/j22c7g/south_korea_smart?w=5
5. Laura W. South Korea smart home market 2017-2024: household penetration for smart home applications is expected to hit above 45%. Available on https://www.prnewswire.com/news-releases/south-korea-smart-home-market-2017-2024-household-penetration-for-smart-home-applications-is-expected-to-hit-above-45-300573827.html
6. Antara News Infographic. Cyber crime trend is on the rise. Available on https://www.antaranews.com/infografik/1571604/tren-kejahatan-siber-meningkat
7. Media Indonesia. Criminal investigation unit of POLRI: number of cyber crimes increasing since January. Available on https://mediaindonesia.com/read/detail/322327-bareskrim-polri-jumlah-kejahatan-siber-meningkat-sejak-januari
8. Welcoming Bhayangkara Anniversary. Bareskrim holds "Cyber Police Festival". Available on https://www.antaranews.com/berita/1565120/sambut-hut-bhayangkara-bareskrim-gelar-cyber-police-festival
9. Sforzin A et al (2016) RPiDS: Raspberry Pi IDS—a fruitful intrusion detection system for IoT. In: 2016 international IEEE conferences ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld). IEEE
10. Kolias C, Stavrou A, Voas J, Bojanova I, Kuhn R (2016) Learning internet-of-things security "hands-on". IEEE Secur Priv 14(1):37–46
11. Kolias C, Kambourakis G, Maragoudakis M (2011) Swarm intelligence in intrusion detection: a survey. Comput Secur 30(8):625–642
12. Sommer R, Paxson V (2010) Outside the closed world: on using machine learning for network intrusion detection. In: Proceedings of symposium on security and privacy, Berkeley, California. IEEE, pp 305–316
13. Anthes G (2013) Deep learning comes of age. Commun ACM 56(6):13–15
14. Saxe J, Berlin K (2015) Deep neural network based malware detection using two dimensional binary program features. In: 2015 10th international conference on malicious and unwanted software (MALWARE). IEEE, pp 11–20
15. Laskov P, Düssel P, Schäfer C, Rieck K (2005) Learning intrusion detection: supervised or unsupervised? In: International conference on image analysis and processing. Springer, Berlin, Heidelberg, pp 50–57
16. Zanero S, Savaresi SM (2004) Unsupervised learning techniques for an intrusion detection system. In: Proceedings of the 2004 ACM symposium on applied computing, pp 412–419
17. Tsang CH, Kwong S (2006) Ant colony clustering and feature extraction for anomaly intrusion detection. In: Swarm intelligence in data mining. Springer, Berlin, Heidelberg, pp 101–123
18. Kolias C, Kambourakis G, Stavrou A, Gritzalis S (2015) Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset. IEEE Commun Surv Tutorials 18(1):184–208

19. Vaca FD, Niyaz Q (2018) An ensemble learning based wi-fi network intrusion detection system (wnids). In: 2018 IEEE 17th international symposium on network computing and applications (NCA). IEEE, pp 1–5
20. Wang S, Li B, Yang M, Yan Z (2018) Intrusion detection for WiFi network: a deep learning approach. In: International wireless internet conference. Springer, Cham, pp 95–104
21. Ran J, Ji Y, Tang B (2019) A semi-supervised learning approach to IEEE 802.11 network anomaly detection. In: 2019 IEEE 89th vehicular technology conference (VTC2019-Spring). IEEE, pp 1–5
22. Parker LR, Yoo PD, Asyhari TA, Chermak L, Jhi Y, Taha K (2019) DEMISe: interpretable deep extraction and mutual information selection techniques for IoT intrusion detection. In: Proceedings of the 14th international conference on availability, reliability and security, pp 1–10
23. Aminanto ME, Choi R, Tanuwidjaja HC, Yoo PD, Kim K (2018) Deep abstraction and weighted feature selection for Wi-Fi impersonation detection. IEEE Trans Inf Forensics Secur 13(3):621–636