ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

A Review on Machine Learning in Cyber Security: Application, Potential and Challenges

Chengwei Mo*

School of Computer, South China Normal University, 510631 Guangzhou, China

*Corresponding author. areswill@qq.com

Abstract

With the rapid development in the field of cloud computing, the Internet of Things, and big data, the massive amounts of data generated from network access points, networked devices, and network applications have brought huge difficulties and challenges to cyberspace security. Machine learning, as an important tool of artificial intelligence, has strengthened the combination of human and machine, which aims at mining and solving problems. It this context, it is necessary to take cybersecurity as the background and pay attention to in-depth discussion of machine learning and technical issues of cybersecurity. In this review work, it firstly elaborates the application of machine learning technology in cyberspace security research. Then it focuses on the solutions of machine learning in the field of cyberspace security, focusing on analyzing and summarizing the security features and commonly used machine learning algorithms in these solutions. Finally, it summarizes the existing problems, as well as the future development direction and challenges of machine learning technology in cyberspace security research.

Keywords

Machine Learning; Cyber Security; System Security; Network Security; Application Security.

1. Introduction

Cyberspace includes not only hardware and software such as the Internet, communication networks, various computing systems, various embedded processors and controllers, but also various data or information generated, processed, transmitted, and stored by these hardware and software, as well as the impact of human activities in it [1]. Nowadays, cyberspace development has been increasing rapidly due to the growth in the area of cloud computing, big data, Internet of Things, and software-based network. At the same time, countless network equipment and applications, and explosive network data, make the network environment increasingly complex and bring huge hidden dangers to cyber security [2]. The security of cyberspace not only affects the development of the national economy, but also affects social stability and national security. Therefore, cyberspace security (also named cyber security) has received extensive attention from the government, academia, and industry.

Facing the current situation of cyber security, traditional security technology appears inefficient and rigid. Manual analysis methods that rely on experience cannot discover and deal with 0-day vulnerabilities in real time [3]. Intrusion detection methods based on fixed rule matching cannot effectively cope with increasing network traffic, changing network environments, and evolving network technologies [4]. It is difficult to deal with high-dimensional data, poor performance, low self-adaptation and generalization capabilities, and unable to detect unknown network attacks [5]. As the cyber security situation continues to be severe, facing more complex security issues and explosive

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

growth of network data, this has put forward new requirements for cyber security research, and research can cope with massive data, diverse services, and rapidly evolving network environments to meet detection performance. The cyber security technology that requires self-adaptability and generalization has quickly received widespread attention from all walks of life.

In recent years, machine learning has been heated research topic and applied to various fields [6-9]. From the early Google Brain to the recent emergence of technologies such as Google AlphaGO [10] and unmanned driving [11], people have refreshed their understanding of machine learning and continuously opened up new areas of machine learning applications. As early as the 1980s, researchers applied machine learning to cyber security research, but due to the limitations of the conditions at the time, machine learning did not receive much attention from researchers [12]. With the application of communication, big data, cloud computing and other technology applications, and the continuous improvement of data search, storage and processing capabilities, machine learning uses a large amount of empirical data to improve the performance of the system itself by extracting useful information from massive amounts of big data, providing new ideas for solving current network security problems [13-16]. Therefore, applying machine learning to cyber security has very important research value and practical significance.

This work will sort out the current research status of cybernetwork security based on machine learning. At the same time, it will analyze the main problems existing in existing research and their reasons, and provide support for the subsequent research. The remaining of this work is organized as follows. Section 2 introduces the application procedure of machine learning in the field of cyber security. Section 3 extracts the related previous work on the application of machine learning in cyber security. Section 4 summarizes the whole article and analyzes the future prospects and challenges of applying machine learning in cyber security.

2. Application procedure of machine learning in cyber security

Generally, machine learning is considered as a set of algorithms that can use empirical data to improve the performance of the system itself. This work summarizes the general application process of machine learning from the perspective of applying machine learning technology to cyber security. As shown in Fig. 1, the general application procedure of machine learning in cyber security research mainly includes problem abstraction, data collection, and data preprocessing. And the six stages of security feature extraction, model construction, model validation and evaluation. In the entire application process, each stage cannot exist independently, and there is a certain correlation between each other.

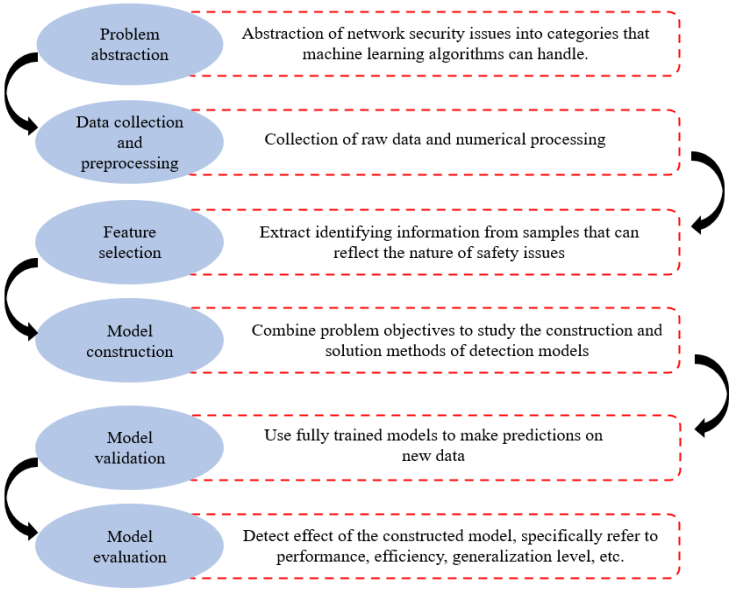

Fig. 1 Typical application procedure of machine learning in cyber security.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

The abstraction of security issues refers to the abstraction of cyber security issues into categories that can be handled by machine learning algorithms, such as classification, clustering, and dimensionality reduction [17]. Whether the problem mapping is appropriate or not is directly related to the success of machine learning technology in solving cyber security problems. Detection of inferior chips or hardware Trojan horses, detection of fake base stations, virtualization security, credit card fraud, etc. can all be abstracted as classification problems; device identity authentication, social network abnormal account detection, network intrusion detection, etc. can be abstracted as clustering problems; user identity authentication, forensic analysis, online public opinion, etc. can be abstracted as classification problems or as clustering problems. When it comes to the processing of high-dimensional data, it can be abstracted as a dimensionality reduction problem. For example, in device identity authentication and malicious webpage recognition, since the data dimension is too high, principal component analysis algorithms can be used to perform dimensionality reduction operations on the data [17].

Data collection and its preprocessing refer to the collection of raw data and numerical processing. Machine learning based cyber security research requires a large amount of security data containing effective information, and the data directly obtained in the actual network environment may contain many repetitive or missing content. To ensure the quality of the data and the effect of the model, it is necessary to use methods such as cleaning and normalization to process the data in advance. Data preprocessing uses statistical methods to analyze the data, and then cleans the abnormal or missing data in the data set, and finally use the normalization method to process the data. In addition, there is also the case of unbalanced data. Due to the quantitative difference between normal samples and abnormal samples, it is easy to have a greater impact on the effect of machine learning algorithms, such as bank card fraud detection [18], malicious traffic detection [19], mobile terminal Trojan detection. There are far more normal samples than malicious samples, and the detection effect of the directly constructed machine learning model is usually not ideal. Therefore, it is usually necessary to use oversampling and undersampling methods to construct a balanced data set to deal with the problem of unbalanced data. In addition to the collected data, there are some open-source datasets in cyber security field, as shown in Table 1.

Table 1. Open-source datasets in the field of cyber security

Number	Name of dataset	Description
1	DARPA Intrusion Detection Data Sets	Cyber intrusion detection data set from 1998-2000
2	UCI's Spambase	Spam email data set
3	Honeynet Project Challenges	Cyber-attack behavior data set
4	Internet Traffic Archive	Network packet data set, containing routing information
5	DMOZ Open Directory Project	URL address set
6	Phish Tank	URL address set of phishing website

Feature election refers to extracting identifying information from a sample that can reflect the nature of safety issues. Extracting effective features can greatly improve the quality and efficiency of detection, so it is necessary to study the methods of extracting effective features from massive data. Al-Rousan et al. [20] adopted the Markov method to model the BGP protocol, and then used the support vector machine algorithm to extract message features, and the detection accuracy of abnormal BGP reached 81.5%. Cheng et al. [21] extracted 33 traffic characteristics from the time characteristics of network traffic and historical traffic characteristics based on the time series analysis method, and constructed a network anomaly detection model based on LSTM. The recognition accuracy obtained was 10% higher than that of SVM [22], Bayes [23] and AdaBoost [24] method.

The model construction is the core step of the application of machine learning in network security. The construction and solution methods of the detection model are mainly studied in combination with the problem objectives. The principle of network anomaly detection is to create a normal behavior pattern. During each detection process, the detected pattern is compared with the normal pattern. If

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

the behavior is significantly different from the normal pattern, it is considered an abnormality. Based on the principle of decision tree and rough set, Li et al. [25] first selected 37 features of BGP in the public routing data set, and generated 3 feature sets of different sizes, and then used decision tree and ELM algorithm [20] to compare 3 features. A classifier is constructed from a feature set, and its accuracy in identifying BGP abnormal routes reaches 80.08%. Based on the work of Barford et al. [26], Kim et al. studied the IP packet data at the router exit through wavelet analysis, and found that when the characteristic value is greater than a given threshold, it can be regarded as abnormal traffic. Galeano et al. [27] used the ARMA model for anomaly flow detection. First, a regression model based on the data sequence was established, and then the residual was used to process the test set.

Model verification and evaluation, mainly to detect the detection effect of the model, specifically refer to performance, efficiency, generalization level, etc. For different detection targets, different indicators need to be used for evaluation: in the chip detection [28], malware detection [9] and other issues, the commonly used indicators [17] include accuracy, precision and recall ; in problems such as hardware Trojan detection, anomaly detection [29], network intrusion detection [30], the false positive rate [17] (FPR), false negative rate [17] (FNR) and f1-meaure (F1M) are often used [17] to measure the generalization ability of the model; in the authentication field, the false recognition rate [17] (FAR) and rejection rate [17] (FRR) are often used for model evaluation.

3. Practical application of machine learning in cyber security

Based on the above content, this section summarizes the network security research work based on machine learning in recent years according to the research scope, and organizes the following main contents.

3.1 System security

System-level security research, focusing on system security of computing unit, including system software security, chip security, and hardware security. In terms of system software security, security applications based on machine learning mainly include vulnerability analysis and mining [31-34], malicious code analysis [35-38], user authentication [39,40], and virtualization security [41]. In terms of system hardware security, machine learning is mainly used to solve identity authentication [42,43], side channel attacks [44], pseudo base station detection [45] and other issues. As for chip security, researchers utilize side channels, fingerprints, images and other information [17], combined with machine learning methods, to solve hardware Trojan horses [46,47], chip property rights protection [48] and other issues. According to the aforementioned work, this review article summarizes the relevant application of machine learning in cyber security at the system level as shown in Table 2.

Table 2. Application of machine learning in the system level of cyber security.

System security	Problems	Abstraction	Machine learning methods	Ref.
Chip security	Inferior chip detection	Classification	SOA, PCA, ANN	[48]
	Hardware Trojan horses	Classification, clustering, dimensionality reduction	KNN, ANN, SVM	[46,47]
Hardware security	Identity authentication	Clustering, dimensionality reduction	SOA, PNN, SVM	[42,43]
	Side channel attacks	Classification, clustering	PCA, SVM, Random Forest	[44]
	Pseudo base station detection	Classification	ANN, SVM	[45]
Software security	Vulnerability analysis and mining	Classification, clustering	PCA, RNN, SVM	[31-34]
	Malicious code analysis	Classification, clustering	SVM, Adaboost, KNN	[35-38]
	User authentication	Classification, clustering	SVM, KNN, DNN, LSTM	[39,40]
	Virtualization security	Classification,	SVM	[41]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

3.2 Network security

Network-level security is an important part of cyber security, including network infrastructure security, traffic detection and other related research. In terms of network infrastructure security, routing security and domain name security are the current research focus in the security field. In recent years, machine learning-based BGP routing detection [49,50] and DNS malicious domain detection [51] have achieved certain results. In terms of network traffic detection, it mainly extracts the characteristics of traffic or messages and uses machine learning to analyze the network security environment to detect potential dangers or malicious attacks that are occurring. Current research based on machine learning focuses on Botnet detection [52-54], network intrusion detection [55-57] and unknown traffic detection [58]. Based on previous study, this review article summarizes the relevant content of machine learning in the network layer security as shown in Table 3.

Table 3. Application of machine learning in the network level of cyber security.

Network security	Problems	Abstraction	Machine learning methods	Ref.
Network infrastructure security	BGP routing detection	Classification, clustering	SVM, ANN	[49,50]
	Malicious domain detection	Classification, clustering	Random forest	[51]
Network traffic detection	Botnet detection	Classification, clustering	SVM, Random forest	[52-54]
	Network intrusion detection	Classification, clustering	ANN, SVM, Bayesian	[55-57]
	unknown traffic detection	Classification, clustering	Logistic regression, Random Forest	[58]

3.3 Application security

Application-level security research mainly includes network application software security and social network security. In terms of network application software security research, the security issues of Email, Web and PDF are the current research focus. The current application software security research based on machine learning includes spam email detection [59], fake URL detection [60] and abnormal document detection [61,62]. In terms of social network security research, research related to machine learning includes abnormal social account detection [63,64], public opinion analysis [65], electronic forensics [66] and network fraud detection [67]. Based on previous work, this review article summarizes the relevant content of machine learning in application layer security as shown in Table 4.

Table 4. Application of machine learning in the application level of cyber security

Network security	Problems	Abstraction	Machine learning methods	Ref.
Application software security	Spam email detection	Classification, dimensionality reduction	SVM, ANN	[59]
	Malicious web detection	Classification, clustering, dimensionality reduction	Decision tree, Random forest, SVM	[60]
	Abnormal document detection	Classification	SVM, Decision tree	[61,62]
Social network security	Abnormal social account detection	Classification, clustering, dimensionality reduction	PCA, SVM, Random forest	[63,64]
	Network fraud detection	Classification	ANN, SVM, Decision tree	[65]
	Electronic forensics	Classification, dimensionality reduction	Logistic regression, ANN, SVM	[66]
	Public opinion	Classification, clustering	KNN, SVM	[67]

4. Conclusions and prospects

In the rapidly developing cyberspace, there are a large number of cyber security problems to be solved. It is this actual cyber security application requirement that has prompted researchers to apply classic machine learning algorithms to the field of network security. In recent years, cyber security research

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

based on machine learning technology have continuously appeared in various reports and documents. These researches have achieved good results in solving cyber security problems. Many machine learning algorithms have highlighted their solutions to cyber security maladies. However, the current technical solutions cannot fully meet the application requirements of cyber security. There are some problems that are difficult to solve and can be further researched. Using machine learning technology to solve cyber security problems is still a challenging task. While solving network security problems, machine learning itself also has certain difficulties. Therefore, how to choose a suitable machine learning algorithm to effectively solve cyber security problems requires further in-depth research. Subsequent research mainly focuses on three aspects: actual network data collection, unknown protocol feature extraction, and construction of adaptive incremental model. By collecting real security information from the actual network, extracting features from unknown protocol network data, adaptive and augmented detection model can be constructed to accommodate the characteristics of the actual network and detect the attacks in it, so as to better realize the application of machine learning-based cyber security research in the real environment.

References

[1] Z. Wang. The applications of deep learning on traffic identification[C]. Proceedings of the BlackHat USA, Las Vegas, 2015,1-10.

[2] M. Z. Gunduz, R. Das. Cyber-security on smart grid: Threats and potential solutions[J]. Computer networks, 2020, 169: 107094.

[3] B. Alhayani, H. J. Mohammed, I. Z. Chaloob, et al. Effectiveness of artificial intelligence techniques against cyber security risks apply of IT industry[J]. Materials Today: Proceedings, 2021.

[4] Z. Zhang, Q. Wen, W. Tang. Survey of mining protocol specifications[J]. Computer Engineering & Applications, 2013, 49(9):1-9.

[5] J. Caballero, H. Yin, Z. Liang, et al. Polyglot: automatic extraction of protocol message format using dynamic binary analysis[C]. Acm Conference on Computer & and Communications Security, Alexandria Virginia, 2007, 317-329.

[6] L. Lerman, S. F. Medeiros, N. Veshchikov, et al. Semi-supervised template attack[C]. Proceedings of International Workshop on Constructive Side-Channel Analysis and Secure Design, Berlin Heidelberg, 2013, 184-199.

[7] G. Hospodar, B. Gierlichs, E. D. Mulder, et al. Machine learning in side-channel analysis: a first study[J]. Journal of Cryptographic Engineering, 2011, 1(4):293-302.

[8] Y. Pang, X. Xue, A. S. Namin. Early identification of vulnerable software components via ensemble learning[C]. Proceedings of 2016 15th IEEE International Conference on Machine Learning and Applications, Orange, 2016, 476-481.

[9] A. Fairuz, N. Amalina. Evaluation of machine learning classifiers for mobile malware detection[J]. Soft Computing, 2016, 1(1):10-18.

[10] Google. AlphaGo [EB/OL]. https://deepmind.com/research/alphago, 2020.

[11] Baidu. Apollo [EB/OL]. http://apollo.auto/, 2020.

[12] D. E. Denning. An intrusion-detection model[M]. New York: IEEE Press, 1987, 143-149.

[13] L. Lippmann, P. Richard, B. Cunningham, et al. Improving intrusion detection performance using keyword selection and neural networks[J]. Computer Networks, 2000, 34(4): 597-603.

[14] Y. Li, J. Xia, S. Zhang, et al. An efficient intrusion detection system based on support vector machines and gradually features removal method[J]. Expert Systems with Applications, 2012, 39(1):424-430.

[15] J. Cannady. Artificial neural networks for misuse detection[C]. Proceedings of the 1998 National Information Systems Security Conference, Arlington, 1998, 443-456.

[16] W. Cynthia, F. Jerome, S. Radu, E. Thomas. Machine learning approach for IP-flow record anomaly detection[C]. Proceedings of IFIP Networking, Valencia, 2011, 28-39.

[17] L. Zhang, Y. Cui, J. Liu et al. Application of machine learning cyberspace security research [J]. Chinese Journal of Computers, 2018, 41(9): 1943-1975 (in Chinese).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

[18] P. K. Chan, S. J. Stolfo. Toward scalable learning with non-uniform class and cost distributions: a case study in credit card fraud detection[C]. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York, 1998, 164-168.

[19] R. Sommer, V. Paxson. Outside the closed world: on using machine learning for network intrusion detection[C]. Proceedings of the 2010 IEEE Symposium on Security and Privacy, Washington, 2010, 305-316.

[20] G. B. Huang, Q. Y. Zhu, C. K. Siew. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1-3):489-501.

[21] M. Cheng, Q. Xu, J. Lv, et al. MS-LSTM: a multi-scale LSTM model for BGP anomaly detection[C]. Proceedings of International Conference on Network Protocols, Singapore, 2016, 1-6.

[22] N. M. Al-Rousan, L. Trajkovic. Machine learning models for classification of BGP anomalies[C]. Proceedings of IEEE 13th International Conference on High Performance Switching and Routing, Belgrade, Serbia, 2012, 103-108.

[23] A. W. Moore, D. Zuev. Internet traffic classification using bayesian analysis techniques[J]. ACM Sigmetrics Performance Evaluation Review, 2005, 33(1): 50-60.

[24] W. Hu, S. Maybank. Adaboost-based algorithm for network intrusion detection[J]. IEEE Transactions on Systems Man&Cybernetics- Part B Cybernetics, 2008, 38(2): 577-583.

[25] Y. Li, H. J. Xing, Q. Hua, X. Z. Wang, P. Batta, S. Haeri, L. Trajkovic. Classification of BGP anomalies using decision trees and fuzzy rough sets[C]. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, San Diego, 2014, 1331-1336.

[26] P. Barford, J. Kline, D. Plonka, A. Ron. A signal analysis of network traffic anomalies[C]. Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurements, Marseille, 2002, 71-82.

[27] P. Galeano, D. Pea, R. S. Tsay. Outlier detection in multivariate time series by projection pursuit[J]. Journal of the American Statal Association, 2006, 101(474):654-669.

[28] D. Jap, W. He, S. Bhasin. Supervised and unsupervised machine learning for side-channel based Trojan detection[C]. Proceedings of the IEEE 27th International Conference on Application-specific Systems, Architectures and Processors, London, 2016, 17-24.

[29] M. Panda, M. R. Patra. Network intrusion detection using naive bayes[J]. International Journal of Computer Science and Network Security, 2007, 7(12): 258-263.

[30] N. B. Amor, S. Benferhat, Z. Elouedi. Naive bayes vs decision trees in intrusion detection systems[C]. Proceedings of ACM Symposium on Applied Computing, Nicosia, 2004, 420-424.

[31] F. Yamaguchi, F. Lindner, K. Rieck. Vulnerability extrapolation: assisted discovery of vulnerabilities using machine learning[C]. Proceedings of USENIX Workshop on Offensive Technologies, San Francisco, 2011, 13.

[32] R. Scandariato, J. Walden, A. Hovsepyan, et al. Predicting vulnerable software components via text mining[J]. IEEE Transactions on Software Engineering, 2014, 40(10):993-1006.

[33] D. Kim, J. Nam, J. Song, et al. Automatic patch generation learned from human-written patches[C]. International Conference on Software Engineering, San Francisco, 2013, 802-811.

[34] F. Yamaguchi, C. Wressnegger, H. Gascon, et al. Chucky: exposing missing checks in source code for vulnerability discovery[C]. Proceedings of the 2013 ACM conference on Computer & Communications Security, Berlin, 2013, 499-510.

[35] D. Arp, M. Spreitzenbarth, M. Hubner, et al. DREBIN: effective and explainable detection of android malware in your pocket[C]. Proceedings of the Network and Distributed System Security Symposium 2014, San Diego, 2014, 1-15.

[36] H. V. Nath, B. M. Mehtre. Static malware analysis using machine learning methods[C]. Proceedings of International Conference on Security in Computer Networks and Distributed Systems, Berlin Heidelberg, 2014, 440-450.

[37] N. Nissim, R. Moskovitch, L. Rokach, et al. Novel active learning methods for enhanced PC malware detection in windows OS[J]. Expert Systems with Applications, 2014, 41(13):5843-5857.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

[38] J. Wilhelm, T. Chiueh. A forced sampled execution approach to kernel rootkit identification[C]. Proceedings Heidelberg, of International Workshop on Recent Advances in Intrusion Detection, Berlin 2007, 219-235.

[39] Z. Nan. You are how you touch: user verification on smartphones via tapping behaviors[C]. Proceedings of 2014 IEEE 22nd International Conference on Network Protocols, North Carolina, 2014, 221-232.

[40] S. Asma, Z. Dema, S. Andraws, et al. Analysis of strong password using keystroke dynamics authentication in touch screen devices[C]. 2016 Cybersecurity and Cyberforensics Conference (CCC), Amman, 2016, 91-102.

[41] Y Zhang, A. Juels, M. K. Reiter, et al. Cross-VM side channels and their use to extract private keys[C]. Proceedings of the 2012 ACM conference on Computer and Communications Security, Raleigh, 2012, 305-316.

[42] O. Tekbas, N. Serinken, O. Ureten. An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions[J]. Canadian Journal of Electrical and Computer Engineering, 2004, 29(3):203-209.

[43] S. Dey, N. Roy, W. Xu, et al. AccelPrint: imperfections of accelerometers make smartphones trackable[C]. Proceedings of the Network and Distributed System Security Symposium 2014, San Diego, 2014, 1-16.

[44] G. Hospodar, B.Gierlichs, E. D. Minder, et al. Machine learning in side-channel analysis: a first study[J]. Journal of Cryptographic Engineering, 2011, 1(4):293-302.

[45] Z. Li, W. Wang, C. Wilson, et al. Fbs-radar: Uncovering fake base stations at scale in the wild[C]. Proceedings of the Network and Distributed System Security Symposium 2017, San Diego, 2017, 1-15.

[46] C. Bao, D. Forte, A. Srivastava. On application of one-class SVM to reverse engineering-based hardware trojan detection[C]. Proceedings of 15th International Symposium on Quality Electronic Design, Santa Clara, 2014, 47-54.

[47] T. Iwase, Y. Nozaki, M. Yoshikawa, et al. Detection technique for hardware trojans using machine learning in frequency domain[C]. Proceedings of 2015 IEEE 4th Global Conference on Consumer Electronics, Osaka City, 2015, 185-186.

[48] K. Xiao, D. Forte, M. Tehranipoor. Circuit timing signature (cts) for detection of counterfeit integrated circuits[M]. Switzerland: Springer, 2016, 211-239.

[49] T. Qiu, L. Ji, D. Pei, J. Wang, J. J. Xu, H. Ballani. Locating prefix hijackers using lock[C]. Proceedings of 18th Conference on USENIX Security Symposium, Montreal, 2009, 135一150

[50] M. Cheng, Q. Xu, J. Lv, et al. MS-LSTM: a multi-scale LSTM model for BGP anomaly detection[C]. Proceedings of International Conference on Network Protocols, Singapore, 2016, 1-6.

[51] B. Leyla, S. Sevil. Exposure: a passive dns analysis service to detect and report malicious domains[J]. ACM Transactions on Information and System Security (TISSEC), 2014, 2014(14):1-28.

[52] F. Tegeler, X. M. Fu, G. Vigna, et al. BotFinder: finding hots in network traffic without deep packet inspection[C]. Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies, New York, 2012, 349-360.

[53] X. Hu, M. Knysz, K. G. Shin. Rb-Seeker: auto-detection of redirection botnets[C]. Proceedings of Annual Network & Distributed System Security Symposium, San Diego, 2009, 1-17.

[54] G. F. Gu, J. J. Zhang, W. K. Lee. BotSniffer: detecting botnet command and control channels in network traffic[C]. Proceedings of Annual Network & Distributed System Security Symposium, San Diego, 2008, 1-18.

[55] A. Bivens, C. Palagiri, R. Smith, B.Szymanski, M. Embrechts. Network-based intrusion detection using neural networks[J]. Intelligent Engineering Systems through Artificial Neural Networks, 2002, 12(1): 579-584.

[56] L. Lippmann, P. Richard, B. Cunningham, et al. Improving intrusion detection performance using keyword selection and neural networks[J]. Computer Networks, 2000, 34(4): 597-603.

[57] Y. Li, J. Xia, S. Zhang, et al. An efficient intrusion detection system based on support vector machines and gradually features removal method[J]. Expert Systems with Applications, 2012, 39(1):424-430.

[58] M. Conti, L. V. Mancini, R. Spolaor, et al. Analyzing android encrypted network traffic to identify user actions[J]. IEEE Transactions on Information Forensics and Security, 2015, 11(1):114-125.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202110_7(10).0054

[59] C. H. Wu. Behavior-based seam detection using a hybrid method of rule-based techniques and neural networks[J]. Expert Systems with Applications, 2009, 36(3): 4321-4330.

[60] S. Lee, J. Kim. WarningBird: a near real-time detection system for suspicious urls in twitter stream[J]. IEEE Transactions on Dependable and Secure Computing, 2013, 10(3): 183-195.

[61] N. Srndic, P. Laskov. Detection of malicious pdf files based on hierarchical document structure[C]. Proceedings of the 20th Annual Network & Distributed System Security Symposium 2013, San Diego, 2013, 1-16.

[62] W. Xu, Y. Qi, D. Evans. Automatically evading classifiers[C]. Proceedings of the 20th Network and Distributed System Security Symposium 2016, San Diego, 2016, 1-15.

[63] G. Stringhini, C. Kruegel, G. Vigna. Detecting spammers on social networks[C]. Proceedings of ACM Computer Security Applications Conference, Austin, 2010, 1-9.

[64] B. Viswanath, M. A. Bashir, M. Crovela, et al. Towards detecting anomalous user behavior in online social networks[C]. Proceedings of the 23rd USENIX Security Symposium, San Diego, 2014, 223-238.

[65] S. Bhattacharyya, S. Jha, K. Tharakunnel, et al. Data mining for credit card fraud: a comparative study[J]. Decision Support Systems, 2011, 50(3): 602-613.

[66] M. N. Khan, C. R. Chatwin, R. C. Young. Extracting evidence from filesystem activity using Bayesian networks[J]. International Journal of Forensic Computer Science, 2007, 1(1):50-63.

[67] H. Liu. Internet public opinion hotspot detection and analysis based on K means and SVM algorithm[C]. Proceedings of 2010 International Conference of Information Science and Management Engineering, Shanxi, 2010, 257-261.