International Journal of Gender, Science and Technology UGC Care Group I Journal
..ISSN: 2040-0748 Vol-13 Issue-01 June 2024

A Convolutional Neural Network (CNN) Based System to detect network intrusions

Mrs. Anuradha A.[1], Ms. Sri Varsha Bh.[2]

[1] Associate Professor and Head in the department of IT at DVR & DR. HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District.

[2] MCA student in the Department of Computer Applications (DCA) at DVR & DR. HS MIC College of Technology, Kanchikacherla, NTR District.

ABSTRACT— The exponential increase of Internet users has made security a paramount concern in today's online environment. For the purpose of detecting and identifying intruders using data mining methods, several researchers have created intrusion detection systems in the past. The current methods, however, are unable to use data mining to get enough detection accuracy. To achieve this goal, we provide a novel intrusion detection system that can successfully identify and detect intruders in wireless networks, hence providing security for data exchange. By combining the current convolutional neural network with a novel feature selection technique based on conditional random fields and linear correlation coefficients, we can choose the most helpful features for classification. The suggested intrusion detection system achieves an overall detection accuracy of 98.88% and has been evaluated via tests. The suggested model's performance has been evaluated using tenfold cross validation.

INTRODUCTION

Changes have been more widespread and rapid as a result of advancements in PC and communication technologies compared to the past. Although new technologies have many positive effects on individuals, businesses, and governments, they also have some negative side effects. Consider the following: the availability of information, the security of stored data stages, the preservation of important data, and so on. Based on these considerations, tyranny based on fear in the digital realm is among the most pressing problems in the modern day. A number of groups, including criminal organizations, technically adept individuals, and digital

activists, have become so terrified of the internet that they threaten national security and open society. Therefore, Intrusion Detection Systems (IDS) were developed to proactively protect networks from cyberattacks. At present, 97.80% and 69.79% accuracy rates were achieved independently while learning the bolster support vector machine (SVM) computations to identify port sweep initiatives based on the new CICIDS2017 dataset. Alternatives to support vector machines (SVM) include convolutional neural networks (ANN), random forests, and 63.52 for CNN, 99.93 for Random Forest, and 99.11 for ANN.

Related works:

R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.

Port Scanning is one of the most popular techniques attackers use to discover services that they can exploit to break into systems. All systems that are connected to a LAN or the Internet via a modem run services that listen to well-known and not so well-known ports. By port scanning, the attacker can find the following information about the targeted systems: what services are running, what users own those services, whether anonymous logins are supported, and whether certain network services require authentication. Port scanning is accomplished by sending a message to each port, one at a time. The kind of response received indicates whether the port is used and can be probed for further weaknesses. Port scanners are important to network security technicians because they can reveal possible security vulnerabilities on the targeted system. Just as port scans can be ran against your systems, port scans can be detected and the amount of information about open services can be limited utilizing the proper tools. Every publicly available system has ports that are open and available for use. The object is to limit the exposure of open ports to authorized users and to deny access to the closed ports.

S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

Portscanning is a common activity of considerable importance. It is often used by computer attackers to characterize hosts or networks which they are considering hostile activity against. Thus it is useful for system administrators and other network defenders

to detect portscans as possible preliminaries to a more serious attack. It is also widely used by network defenders to understand and find vulnerabilities in their own networks. Thus it is of considerable interest to attackers to determine whether or not the defenders of a network are portscanning it regularly. However, defenders will not usually wish to hide their portscanning, while attackers will. For definiteness, in the remainder of this paper, we will speak of the attackers scanning the network, and the defenders trying to detect the scan. There are several legal/ethical debates about portscanning which break out regularly on Internet mailing lists and newsgroups. One concerns whether portscanning of remote networks without permission from the owners is itself a legal and ethical activity. This is presently a grey area in most jurisdictions. However, our experience from following up on unsolicited remote portscans we detect in practice is that almost all of them turn out to have come from compromised hosts and thus are very likely to be hostile. So we think it reasonable to consider a portscan as at least potentially hostile, and to report it to the administrators of the remote network from whence it came. However, this paper is focussed on the technical questions of how to detect portscans, which are independent of what significance one imbues them with, or how one chooses to respond to them. Also, we are focussed here on the problem of detecting a portscan via a network intrusion detection system (NIDS). We try to take into account some of the more obvious ways an attacker could use to avoid detection, but to remain with an approach that is practical to employ on busy networks. In the remainder of this section, we first define portscanning, give a variety of examples at some length, and discuss ways attackers can try to be stealthy. In the next section, we discuss a variety of prior work on portscan detection. Then we present the algorithms that we propose to use, and give some very preliminary data justifying our approach. Finally, we consider possible extensions to this work, along with other applications that might be considered. Throughout, we assume the reader is familiar with Internet protocols, with basic ideas about network intrusion detection and scanning, and with elementary probability theory, information theory, and linear algebra. There are two general purposes that an attacker might have in conducting a portscan: a primary one, and a secondary one. The primary purpose is that of gathering information about the reachability and status of certain combinations of IP address and port (either TCP or UDP). (We do not directly

discuss ICMP scans in this paper, but the ideas can be extended to that case in an obvious way.) The secondary purpose is to flood intrusion detection systems with alerts, with the intention of distracting the network defenders or preventing them from doing their jobs. In this paper, we will mainly be concerned with detecting information gathering portscans, since detecting flood portscans is easy. However, the possibility of being maliciously flooded with information will be an important consideration in our algorithm design. We will use the term scan footprint for the set of port/IP combinations which the attacker is interested in characterizing. It is helpful to conceptually distinguish the footprint of the scan, from the script of the scan, which refers to the time sequence in which the attacker tries to explore the footprint. The footprint is independent of aspects of the script, such as how fast the scan is, whether it is randomized, etc. The footprint represents the attacker's information gathering requirements for her scan, and she designs a scan script that will meet those requirements, and perhaps other non-information-gathering requirements (such as not being detected by an NIDS). The most common type of portscan footprint at present is a horizontal scan. By this, we mean that an attacker has an exploit for a particular service, and is interested in finding any hosts that expose that service. Thus she scans the port of interest on all IP addresses in some range of interest. Also at present, this is mainly being done sequentially on TCP port 53 (DNS).

M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

Compared to the past security of networked systems has become a critical universal issue that influences individuals, enterprises and governments. The rate of attacks against networked systems has increased melodramatically, and the strategies used by the attackers are continuing to evolve. For example, the privacy of important information, security of stored data platforms, availability of knowledge etc. Depending on these problems, cyber terrorism is one of the most important issues in today's world. Cyber terror, which caused a lot of problems to individuals and institutions, has reached a level that could threaten public and country security by various groups such as criminal

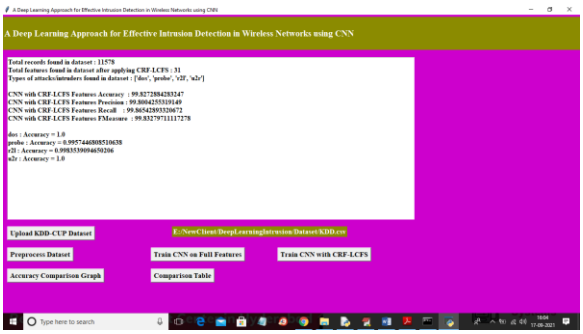
organizations, professional persons and cyber activists. Intrusion detection is one of the solutions against these attacks. A free and effective approach for designing Intrusion Detection Systems (IDS) is Machine Learning. In this study, deep learning and support vector machine (SVM) algorithms were used to detect port scan attempts based on the new CICIDS2017 dataset Introduction Network Intrusion Detection System (IDS) is a software-based application or a hardware device that is used to identify malicious behavior in the network [1,2]. Based on the detection technique, intrusion detection is classified into anomaly-based and signature-based. IDS developers employ various techniques for intrusion detection. Information security is the process of protecting information from unauthorized access, usage, disclosure, destruction, modification or damage. The terms ”Information security”, ”computer security” and ”information insurance” are often used interchangeably. These areas are related to each other and have common goals to provide availability, confidentiality, and integrity of information. Studies show that the first step of an attack is discovery [1].

METHODOLOGY

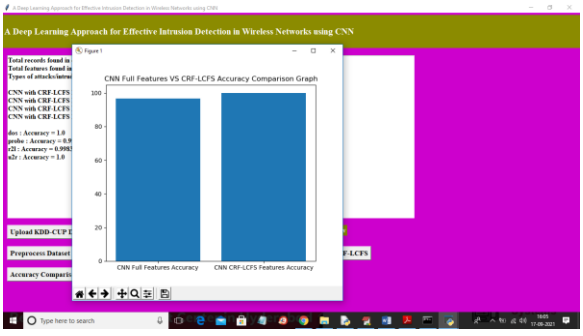
1. Upload KDD-CUP dataset: Using this module, uploading ‘KDD.csv’ dataset and we can see dataset loaded and some values are non-numeric and CNN will accept such values so we need to process them to assign numeric id to each non-numeric values.
2. Preprocess Dataset: Using this module, all values are converted to numeric format
3. Train CNN on Full Features: Using this module, training CNN on all features and CNN full features got 96% accuracy
4. Train CNN with CRF-LCFS: Using this module, training CNN with CRF selected features and got CNN CRF accuracy as 99% which his higher than CNN with full features
5. Accuracy Comparison Graph: Using this module, graph x-axis represents technique name and y-axis represents accuracy of that technique and in that graph CNN with CRF has got high accuracy
6. Comparison Table: Using this module, getting tabular format and getting predicted accuracy of each

attack using CNN with full features and CNN with CRF selected features and in both techniques CNN with CRF got better results.

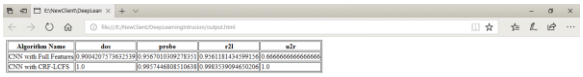
RESULT AND DISCUSSION



In above screen we can see selected features are 31 and we got CNN CRF accuracy as 99% which his higher than CNN with full features and now click on 'Accuracy Comparison Graph' button to get below graph



In above graph x-axis represents technique name and y-axis represents accuracy of that technique and in above graph we can see CNN with CRF has got high accuracy. Now close above graph and then click on 'Comparison Table' button to get below output



In above screen in tabular format we can see predicted accuracy of each attack using CNN with full features and CNN with CRF selected features and in both techniques CNN with CRF got better results.

CONCLUSION

Recent work has presented approximate support vector machine, artificial neural network, convolutional neural network, random forest, and deep learning calculations based on the state-of-the-art CICIDS2017 dataset. Compared to SVM, ANN, RF, and CNN, the results demonstrate that the deep learning calculation performed fundamentally better. In the future, we will use this dataset to conduct port sweeps and other sorts of attacks using artificial intelligence and deep learning calculations, Apache Hadoop, and Spark technologies in tandem. With the use of these computations, we are able to identify cyber attacks on networks. It works like this: if we go back far enough, we could find a lot of assaults. When

we find out when these attacks happened, we can record the characteristics that indicate at what values they occurred in datasets. Thus, we will anticipate the occurrence of cyber attacks by using these statistics. There are four algorithms that may make these predictions: support vector machines (SVMs), artificial neural networks (ANNs), random forests (RFs), and convolutional neural networks (CNNs). This research aims to determine which method has the highest accuracy rates, so it can detect cyber threats more effectively.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, "Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, "The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems," in Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca

algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.

[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in ICISSP, 2018, pp. 108–116.

[13] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141–149.

[14] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," IEEE Access, 2018.

[15] P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Security and Privacy, vol. 1, no. 4, p. e36, 2018.

[16] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[17] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.

AUTHOR PROFILES

Mrs Anuradha Anumolu completed her M.Tech (CSE) from Acharya Nagarjuna University. She has published more than10 papers in indexing Journals, currently working as an Associate Professor and Head in the department of IT at DVR & Dr HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District. Her areas of

interest are Data Structures, Data Mining, Cloud Computing, Artificial Intelligence.

Mrs. SRI VARSHA BHOGADI is an MCA student in the Department of Computer Applications at DVR & Dr. HS MIC College of Technology, Kanchikacherla, NTR District. She completed her B.Sc. in Mathematics, Physics, and Computer Science (MPCS) at S.V.L Degree College, affiliated with Krishna University. Her areas of interest include Machine Learning, Java, and Web Technologies.