Network Intrusion detection system using artificial intelligence

Clementine Mamogale - 216025117

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
clementinemamogale@gmail.com

Abstract. The widespread use of the internet and communication devices and the development of new technologies that can connect to one other has advanced, which increased the volume of data transfer over the internet. Simultaneously, this has opened many novel attacks which are beyond the human control capabilities, which also poses challenges to the current network securities. One tool that is used by the network securities to prevent possible network intrusion is an Intrusion Detection System (IDS), which works by monitoring the traffic over a network to ensure the availability, integrity, and confidentiality of the network. This tool used to be effective over the past years, but as the internet grows, the IDS has become less and less ineffective. Hence in this paper, I introduce a Network Intrusion detection system (NIDS) architecture that uses Artificial Intelligence (AI) as its backbone. This architecture will feature deep learning and machine learning techniques to develop a system that is adaptive and resilient to network intrusions, including zero-day intrusions. To demonstrate the effectiveness of this model, we use the old KDD+ and the new CICIDS 2017 datasets. The reason for these two datasets is because the KDD+ is widely used because it has well-known intrusions and can be used to create the baseline of the NIDS. Then the CICIDS is new and has new attacks that happened recently. Hence it can be used for training for zero-day attacks.

Keywords: Network Intrusion Detection System (NIDS), machine learning, deep learning, Network Security, Vector Machine, Naïve Bayes, Neural Networks.

1 Introduction

In this modern life, our daily activities are influenced by the widespread use of computer devices, which are connected to one another and share data and information. Using the information and communication technology (ICT) and its resilience, businesses and individuals were able to offer real-time global business continuity and a frontier solutions of interoperability [1]. The rise of this interoperability and data exchange among computers on the internet has opened vulnerabilities that are exploitable, and they can result in harmful effects on the end-users of the computers. Hence a

resistant and resilient intrusion detection system (IDS) is needed, an IDS which will be able to maintain and provide availability, integrity, and confidentiality of the system. After all, an IDS is recognized as the first layer of defense among the defensive mechanisms that address all the attack vectors [1].

A Network Intrusion Detection System works by scanning a network traffic as it comes and identify any violations if they exist based on customized detection levels that are preconfigured on the network and then report them, then block them from going through the network before they cause any damage to the data that is being protected by the network. IDS use the idea of behavioral features to differentiate between a legitimate and non-legitimate traffic. Hence the development of Intelligent Intrusion Detection system is needed, which will be able to learn about different behaviors of normal and abnormal network and use that learned behaviors to detect intrusions [2]. In this paper I propose a Network Intrusion Detection System (NIDS) that uses AI as its backbone to detect intrusion on a network in real time, automatic and fast, which aims to provide minimum false alerts. The KDD+ and CICIDS 2017 datasets will be used to demonstrate the effectiveness of the system and it will make use of the Naïve Bayes classification technique, Support Vector Machine and the Neural Networks which will work together to complement each other for maximum security of a network.

**********talk about the layout of the paper here**********

2 Literature review

The concept of intrusion on a network have been around for a very long time and having intrusion detection systems which worked by monitoring networks for any anomalous behaviors or misuse. It is only now that there is a rise of this intrusions on network, and many because of the rise of the internet and many things connecting and sharing data over networks [4].

James Anderson wrote a paper in 1980 for a government organization, which is accredited for introducing automatic IDS, a paper that was used to build the first IDS [4]. With the rise of the internet, some problems for that system arose, since it was using the signature based algorithm, it was a bit slow in terms of registering new signature from new threats and the system was only effective against known threats. Zero-day attacks could compromise it, easily [4].

Since then, other models have been developed in the industry to try and fix the problem that were present on the system that was built using the paper from James Anderson. Until the year 1999 when Todd Heberlein introduced a different kind of system called Network Security Monitor (NSM) which helped in monitoring and collecting data about network traffic [4]. This opened new ideas and there were more interests in the intrusion detection environment and new investments were made towards the market. This project featured some of the other projects that were made before it, such as the Distributed Intrusion Detection System which introduced the

idea of hybrid intrusion detection. As a results, the IDS field was revolutionized and brought to the commercial world [4].

Currently, by statistics. It shows that in the market of security vendors, IDS remains one of the top selling ones. Some of the commercial IDS that were introduced in the early years includes Network Flight Recorder (NFR) which uses libpcap, it was introduced in 1999, then 1998 APE which was later changed to Snort which is a packer sniffer was also developed. It uses libpcap also. And it remained the world's largest IDS that was used, with active users of over 300 000 [4]. However, due to fast grow of the internet and almost everything connecting to the internet. These methods become ineffective. at least, on their own. Hence the introduction of NIDS which uses AI which features more than one technique to get the work done.

2.1 Recent similar work

Ensemble Core Vector Machine (CVM) which is a better version of Support Vector Machine (SVM), which are a supervised learning machines. They are used for regression analysis and classifications [5]. The algorithms that are used in CVM and SVM are based on concept of minimum Enclosing Ball basis, which is an adopted function from statistic. The function helps in producing less false-positive rates and improvement in performance. Hence, they are considered the best prediction methods. However, unlike the SVM, the CVM has low overhead computation which makes it better than the SVM [2].

Internet of Things (IoT) is another field of study that has shown some interesting studies in IDS. IoT is a new and growing technological advancement, whereby the day-to-day "smart" devices, from different domains, connect and communicate to each other over the internet via data sharing [6]. These smart devices are embedded with sensors which generate high volume of data on the internet [5]. IoT brings lots of benefits to out everyday lives, but they also have vulnerabilities that can be exploited over the internet. These vulnerabilities on these devices are because most of them lack hardware security support, and their gateways also does not support such securities. IDS require high computational power which these smart devices gateways do not provide. They also have their own specifications and protocols that they follow. Implementing these complicated hardware and software security features on these smart devices, can cause them to misbehave or break. As a results, it remains a challenge to secure the environment of these smart devices while having them working perfect [6].

There is a huge gap between security capability and requirements, which puts these smart devices at risks of cyber-attacks such as DoS and spoofing attacks. Many records of IoT devices being attacked are recorded, and it shows how bad they can affect these devices, such as damaging the hardware, or interrupting the system. The Marian malware is one example that was recorded to be the most famous attack that happened in 2017 which affected 300 000 IoT devices, which further turned the devices into zombies to create DDoS attack [6].

For security purposes, IDS for smart devices are placed at the gateways to monitor the network for incoming and outgoing traffic and checks for malicious activities or not. This is still not enough since it does not check the internal interconnected traffic of the devices, which might cause internal attacks [6]. A better solution that was proposed for IoT was the use of Deep Neural Network (DNN) which will learn and create profiles for abnormal and normal networks from already existing datasets and based on that, it can find abnormalities in a network easy, and it get deployed live on the IoT. This method worked so well because it could identify even non-linear complex features which are used in real life attacks [6].

3 First Section

4 First Section

4.1 A Subsection Sample

Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation etc. does not have an indent, either.

Subsequent paragraphs, however, are indented.

Sample Heading (Third Level). Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

Sample Heading (Forth Level). The contribution should contain no more than four levels of headings. The following Table 1 gives a summary of all heading levels.

Table 1. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Run-in Heading in Bold.** Text follows	10 point, bold
4th-level heading	*Lowest Level Heading.* Text follows	10 point, italic

Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

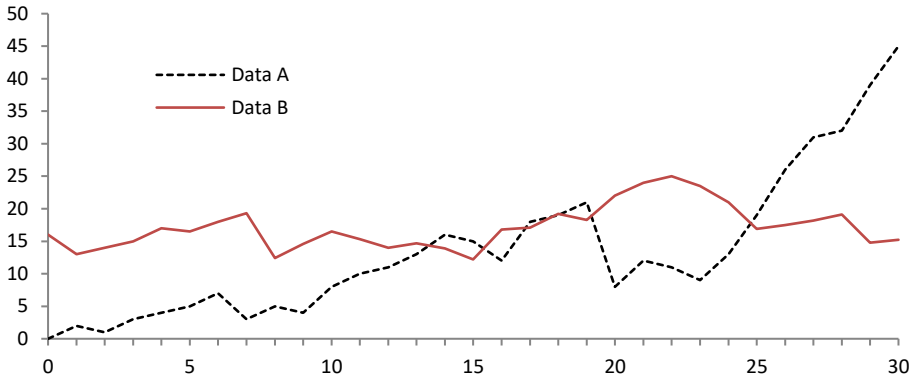

Fig. 1. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4], as well as a URL [5].

References

1) Author, F.: Article title. Journal 2(5), 99–110 (2016).
2) Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
3) Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
4) Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
5) LNCS Homepage, http://www.springer.com/lncs, last accessed 2016/11/21.

Klngfdnkddnkjnkngfd0

1. https://reader.elsevier.com/reader/sd/pii/S1877050921011078?token=B9BD6FFE3C86E2DD74A456A01956CF6DA067E92A474645A6F1360558C7F127895C1D0E990B5C4E33630BC1CF08F91064&originRegion=eu-west-1&originCreation=20220612193324
2. https://reader.elsevier.com/reader/sd/pii/S1877050918321136?token=71A22DCD7E8606EF7E519D3A1E7939A5D7DCC70FB252A9610845A124066DE2BA46EEC85A07DD74A4E29468BC9D812487&originRegion=eu-west-1&originCreation=20220612193328

3. https://www.researchgate.net/journal/Transactions-on-Emerging-Telecommunications-Technologies-2161-3915/publication/344726867_Network_intrusion_detection_system_A_systematic_study_of_machine_learning_and_deep_learning_approaches/links/61a604c86864311d938a92c5/Network-intrusion-detection-system-A-systematic-study-of-machine-learning-and-deep-learning-approach-es.pdf?_sg%5B0%5D=3jDD1L58VEnesSzxdTvYvwp0_61rq6ICOLKl2tjjxvTed3YAWsWX1Q2dcaSsqUPCQlbRSAcqxceJPx-oDt5DdSg.JQJoBClZHuZcttTiLuMOW9UQnisaHqaJSDDs2oMBM9v1hKbmtPj0EtMlE5h3jY_ZGyqDu7T6anRV87RsntUh0w&_sg%5B1%5D=uQS6nkX6_GaKkhJD6vUMzkZKM7ySW_iU1XkoxouJsHKLna-CYA4zsKXdCEG3H_Y9BkeIFaegXckRbCvEEqs_9a0Sdvs6hd82ofvwu1AYxRSlH.JQJoBClZHuZcttTi-LuMOW9UQnisaHqaJSDDs2oMBM9v1hKbmtPj0EtMlE5h3jY_ZGyqDu7T6anRV87RsntUh0w&_iepl=
4. https://reader.elsevier.com/reader/sd/pii/S1877705812021613?token=B75AFECDC757F2A8D54A0A4081F752ED705134BDCEB05B6A7237305EE9E84A908F48746A94C60ED84B82A7FCE295F847&originRegion=eu-west-1&originCreation=20220613230743
5. https://mdpi-res.com/d_attachment/mathematics/mathematics-10-00530/article_deploy/mathematics-10-00530-v2.pdf
6. https://iopscience.iop.org/article/10.1088/1742-6596/1518/1/012040
7.