https://doi.org/10.31891/csit-2023-4-1
UDC 004.93

Maksym CHORNOBUK, Valeriy DUBROVIN, Larysa DEINEHA
National University «Zaporizhzhia Polytechnic»

CYBERSECURITY: RESEARCH ON METHODS FOR DETECTING DDOS ATTACKS

This article describes the problem of DDoS attacks, analyzing their nature and consequences. The paper covers common DDoS attack types, such as SYN flood, ICMP flood, UDP flood. Existing methods for detecting attacks from literature are reviewed, including machine learning approaches, including artificial neural networks, support vector machines and decision trees. The paper introduces a decision tree-based machine learning model for the detection of DDoS attacks. The model is trained and tested on a publicly available dataset. The dataset consists of 1,04,345 rows of data, where every row includes 23 features, such as source IP, destination IP, port number, number of bytes transferred from the switch port, etc. A similar set of characteristics can be obtained on a real network hardware using simple calculations, which makes it possible to approximate the model evaluation to real operating conditions. SYN flood, ICMP flood and UDP flood attack types are present in the data, as well as legitimate traffic. To avoid overfitting, only some columns were used, and columns such as IP addresses were discarded. The field "label" in each row of the dataset contains either 0 or 1 where 0 corresponds to legitimate traffic and 1 to malicious one. The problem of DDoS attack detection is therefore formally reduced to the task of binary classification of each row from the dataset. The constructed model achieves an average classification accuracy of 0.94 with a standard deviation at the level of 0.06 in detecting the above mentioned types of attacks. To objectively assess the effectiveness of the model and avoid distortion of the results, stratified 5-fold cross-validation was used. The developed model can be applied in the real world network hardware to filter malicious packets or as a tool for warning the administrator about an attack. This research advances cybersecurity by enhancing DDoS attack detection.

Keywords: machine learning, DDoS, decision tree, classification.

Максим ЧОРНОБУК, Валерій ДУБРОВІН, Лариса ДЕЙНЕГА
Національний університет «Запорізька політехніка»

КІБЕРБЕЗПЕКА: ДОСЛІДЖЕННЯ МЕТОДІВ ВИЯВЛЕННЯ DDOS-АТАК

У цій статті розглядається проблема DDoS-атак, аналізується їх природа та наслідки. Стаття охоплює поширені типи DDoS-атак, такі як SYN-flood, ICMP-flood, UDP-flood. Розглядаються існуючі методи виявлення атак з літератури, включаючи методи машинного навчання, такі як штучні нейронні мережі, метод опорних векторів та дерева прийняття рішень. У статті представлено модель машинного навчання на основі дерева прийняття рішень для автоматичного виявлення DDoS-атак. Модель навчена та протестована на загальнодоступному наборі даних. Набір даних складається з 104345 рядків даних, де кожен рядок містить 23 поля, такі як IP-адреса джерела, IP-адреса призначення, номер порту, кількість байтів, переданих із порту комутатора тощо. Подібний набір характеристик можна отримати на реальному мережевому обладнанні за допомогою простих розрахунків, що дає можливість наблизити оцінку моделі до реальних умов можливої експлуатації. Типи атак SYN-flood, ICMP-flood, UDP-flood присутні в даних, а також наявний легітимний трафік. Щоб уникнути ефекту перенавчання, використовувалися лише деякі поля, а такі поля, як IP-адреси, були відкинуті. Поле «label» в кожному рядку набору даних містить 0 або 1, де 0 відповідає легітимному трафіку, а 1 — зловмисному. Тому проблема виявлення DDoS-атаки формально зводиться до здійснення бінарної класифікації кожного рядка з набору даних. Побудована модель досягає середньої точності класифікації 0,94 зі стандартним відхиленням на рівні 0,06 при виявленні зазначених типів атак. Щоб об'єктивно оцінити ефективність моделі та уникнути спотворення результатів, була використана стратифікована 5-fold кросс-валідація. Розроблена модель може бути застосована в реальному мережевому обладнанні для фільтрації шкідливих пакетів або як інструмент для попередження адміністратора про атаку. Це дослідження покращує сферу кібербезпеки, розширюючи методи виявлення DDoS-атак.

Ключові слова: машинне навчання, DDoS, дерево рішень, класифікація.

Introduction

Every year, the importance of information and network technologies in human life, as well as in the economies of the countries of the world, is growing.

Along with the growing influence of information technologies, the risks associated with information security are also growing. One of the most important threats associated with information and network technologies are Distributed Denial of Service (DDoS) attacks. The essence of such attacks is the usage of huge arrays of resources in the network to generate malicious traffic against targeted network services. Such attacks have been popular for decades. A particular difficulty in the fight against them is their indistinguishability from legitimate traffic. In addition, there are a huge number of different types of attacks, which makes it even more difficult to detect them among legitimate traffic [1].

Yearly, the complexity and magnitude of DDoS attacks show consistent growth. Notably, in 2018, a single attack reached a terabit-per-second traffic size. A concurrent trend is the expansion of economic impact. According to [3], the average damage inflicted by attacks was below $10,000 in 2017, whereas in 2018, it increased significantly, averaging between $10,000 and $100,000.

A sufficiently large number of various methods for detecting DDoS attacks have been developed. Some methods are based on data about individual packets, while others signal an attack based on the capacity of packets arriving at the server over certain periods of time. However, DDoS attacks do not share specific distinguishing features, so no systems have yet been created that can accurately detect an attack of any unusual type.

Over time, an array of different methodologies have been developed to detect DDoS attacks. Certain strategies use features of individual network packets, while others base their detection on the volumetric properties of packets arriving on servers within specific temporal windows. However, DDoS attacks do not share specific distinguishing features, so no systems have yet been created that can accurately detect an attack of any unusual type [2].

In the ever-changing world of cyber threats, understanding DDoS attacks and creating better ways to spot them are really important to protect digital systems.

Related Works

One of the most common types of DDoS attacks is the SYN flood. This attack is based on some principles of the TCP protocol, which is a transport layer protocol and one of the main protocols of the Internet protocol suite.

TCP is a connection-oriented protocol, therefore it requires the connection between two nodes to be established before the actual data transfer can take place. One of these nodes is called “server” and the second one – “client”. While the server is listening for incoming connections, clients can establish such a connection by sending the SYN packet to the server. When the server receives a SYN packer, it starts a handshake procedure which consists in sending a SYN-ACK packet back to the client. Because computer RAM is finite, any server can only process a limited number of handshakes at a time.

SYN flood attack happens when malicious clients send SYN packets to the server without finalizing the handshake procedure. When server’s connection wait slots are exhausted, denial of service is happening, because legitimate clients are unable to establish new TCP connections to the server [2, 4].

Another well-known type of DDoS attack is the ICMP flood. It is based on the exploitation of Internet Control Message Protocol (ICMP), which is the network layer protocol that is used to send service messages between hosts on the Internet. ICMP includes the specification of the so-called “echo request”, which is a method normally used to determine the latency between two hosts. When a machine receives an echo request it’s obligated to send back the correct response.

Similar to the previous type of attack, one or more malicious hosts send legitimate echo request messages in such numbers that the attacked machine's resources run out. When legitimate users can no longer access the host, the denial of service is happening [2, 5].

UDP flood is another popular type of DDoS attack. As its name implies, it uses an important transport layer protocol called UDP. User Datagram Protocol (UDP) basically is a stateless protocol, so it does not require a connection to be established between hosts to send messages, nor does it check whether messages were successfully delivered.

Using such protocol features, attackers send random UDP messages to the target host. Processing such messages, the host spends resources, which are eventually exhausted. [2, 6].

Given the variety of types of DDoS attacks, as well as their extreme similarity to legitimate traffic, their detection is not a trivial task. All existing methods are somehow related to statistical analysis or machine learning.

The simplest methods are based on the analysis of some numerical characteristics of incoming traffic per unit of time. An example of such a method is [7], which describes the construction of a fuzzy estimator based on one value – mean packet inter arrival times. Nevertheless, despite the simplicity of the model, on certain types of attacks it has an efficiency of more than 80%.

More sophisticated detection techniques utilize popular machine learning techniques for classification problems, including artificial neural networks, support vector machines and decision trees [1].

An example of a more complex system based on machine learning technologies, in particular artificial neural networks, is [8]. The paper describes the model based on the particular artificial neural networks which classifies network packets into one of 4 types: DNS DDoS attack, CharGen DDoS attack, UDP DDoS attack and legitimate traffic. The classification is based on four packet parameters, which are packet arrival time, source IP address, destination IP address, used protocol and packet length. To test the model, public datasets were used, on which the model demonstrated an overall accuracy of 95.6%. However, the model has shown lower accuracy (82.1%) in the classification of UDP DDoS attacks.

Another example of a successful model based on neural networks is [9]. It is aimed to detect DDoS attacks in real time .The model was implemented in the Apache Spark cluster and tested on a public dataset. The average detection rate of the model is over 94%.

Proposed technique

Given the growing damage that DDoS attacks inflict on the network infrastructure, and subsequently the economy every year, it was decided to develop a model, which is able to detect such attacks. Taking into account the experience of well-known systems, as well as the variety of types of DDoS attacks, it was decided to use machine learning algorithms as the basis for the developed model.

The dataset used is publicly available [10]. An important feature of the dataset is the fact that it includes examples of the most common types of attacks: SYN flood attack, UDP flood attack, ICMP flood attack. It consists of 1,04,345 rows of data, where every row includes 23 features. There are extracted features such as source IP, destination IP, port number, number of bytes transferred from the switch port, etc. And there are calculated features such as the number of packets sent per second. A similar set of characteristics can be obtained on a real network hardware using simple calculations, which makes it possible to approximate the model evaluation to real operating conditions. All fields were converted to numerical values for the correct operation of the model. In particular, the Protocol field, which stored string values, was converted to integer values, where each protocol corresponds to a separate integer. All missing values were replaced with zeros. Source IP, destination IP and date and time fields were not used in order not to create information noise during training of the model.

The field “label” in each row contains either 0 or 1 where 0 corresponds to legitimate traffic and 1 to malicious one. The problem of DDoS attack detection is therefore formally reduced to the task of binary classification of each row from the dataset, that is, to building such a model that can precisely predict the value of the label based on other fields from the same row.

Decision trees were chosen as the basis for the developed model. A decision tree is a tree-like structure, where each internal node represents a test on an attribute, each branch represents an outcome of the test, and each leaf node belongs to one of the classes. Each sample starts from the root and, being subjected to tests, eventually comes to one of the leafs, which corresponds to the result of the classification of this sample.

Decision tree models have a long history of use in the field of classification problems. Over the decades, many algorithms have been developed to build such models: ID3, C4.5, and the latest – CART. All these algorithms are based on the "greedy" principle of building trees from top to bottom, but differ in details [11].

Popular Python library Scikit-learn [12] and its implementation of the CART algorithm was used to load and preprocess the dataset, to train the model and to evaluate it.

To objectively assess the accuracy of the constructed model the Scikit-learn [12] implementation of the stratified 5-fold cross-validation was used. This technique is used to effectively eliminate random information noise that can occur when splitting a learning dataset into training and test subsets.

Finally, the model was tested, the results of which are shown in Fig. for each of the folds.

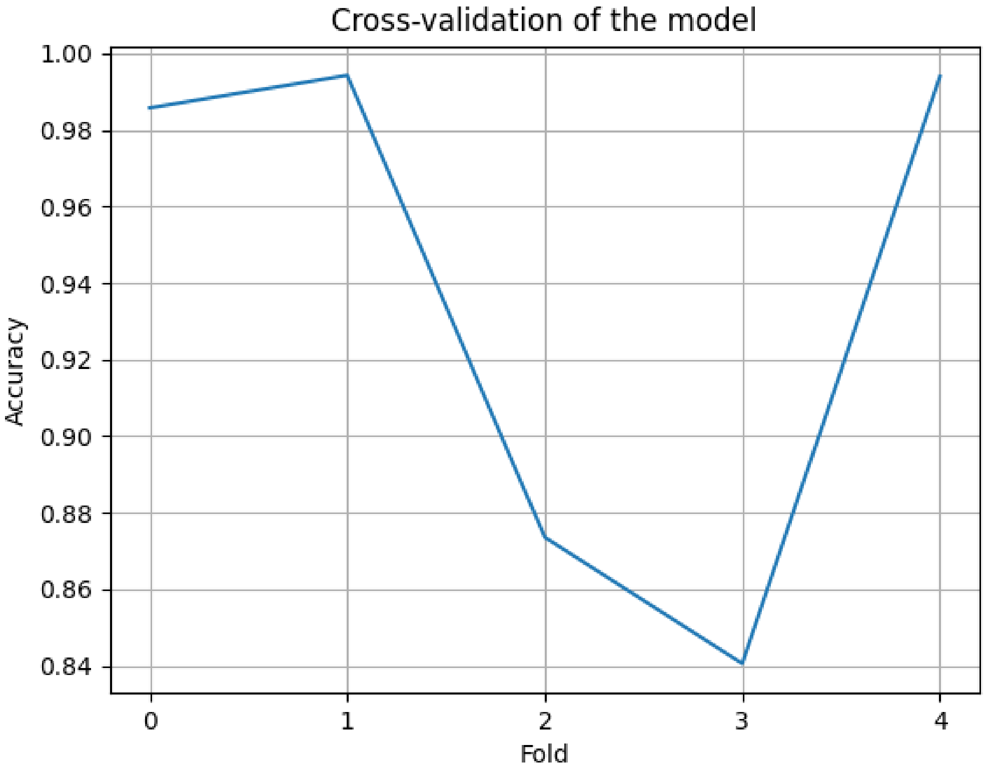

Fig. 1. The accuracy of the model during the 5-fold cross-validation

According to the test results, the average accuracy of the model was about 0.94, and the standard deviation was at the level of 0.06.

Test results point to high accuracy levels on the most popular types of attacks. It can be applied in the real world to filter malicious packets on network equipment, thereby significantly reducing the processing time for malicious packets and increasing the stability of the host during attacks by saving its resources. When used on critical infrastructure, where the accuracy of the model is insufficient, it can be used as a tool for detecting attacks for subsequent manual response to them by the administrators of the infrastructure.

Conclusions

In this paper, modern prospects for DDoS attacks and their economic impact were considered. The ever-increasing need to search for methods to protect against them, in particular their detection, was emphasized.

The main types of DDoS attacks, their nature and mechanisms are described. It is indicated that DDoS attacks are very similar to legitimate traffic, which complicates the task of detecting them.

Popular methods for detecting DDoS attacks from the literature, including methods based on machine learning algorithms, are considered.

A machine learning model based on decision trees has been developed that can effectively detect DDoS attacks. The model has been tested using a publicly available dataset [10]. The test results show a significant level of accuracy of the model, about 94%.

The constructed model is characterized by simplicity and high accuracy on the most popular types of attacks. It can be applied in the real world to filter malicious packets on network equipment. When used on critical infrastructure, where the accuracy of the model is insufficient, it can be used as a tool for detecting attacks for subsequent manual response to them.

References

1. D. G. DDoS detection and prevention based on artificial intelligence techniques. Scientific bulletin of naval academy. 2019. Vol. XXII, no. 1. P. 134–143. URL: https://doi.org/10.21279/1454-864x-19-i1-018 (date of access: 19.10.2023).
2. N. A. M., Zaboon K. H., Abdullah A. A. A review of the common ddos attack: types and protection approaches based on artificial intelligence. Fusion: practice and applications. 2021. P. 08–14. URL: https://doi.org/10.54216/fpa.070101 (date of access: 19.10.2023).
3. NETSCOUT's 14th annual worldwide infrastructure security report. NETSCOUT's 14th Annual Worldwide Infrastructure Security Report. URL: https://www.netscout.com/report/ (date of access: 19.10.2023).
4. Davidson J. An introduction to TCP/IP. New York, NY : Springer New York, 1988. URL: https://doi.org/10.1007/978-1-4612-4572-8 (date of access: 19.10.2023).
5. Seven deadliest network attacks. Elsevier, 2010. URL: https://doi.org/10.1016/c2009-0-61914-0 (date of access: 19.10.2023).
6. Data communication and networking concepts in user datagram protocol (UDP). International journal of recent technology and engineering. 2020. Vol. 8, no. 5. P. 2765–2768. URL: https://doi.org/10.35940/ijrte.d8758.018520 (date of access: 19.10.2023).
7. Real time DDoS detection using fuzzy estimators / S. N. Shiaeles et al. Computers & security. 2012. Vol. 31, no. 6. P. 782–790. URL: https://doi.org/10.1016/j.cose.2012.06.002 (date of access: 19.10.2023).
8. Artificial neuron network implementation in detection and classification of DDoS traffic / D. Perakovic et al. 2016 24th telecommunications forum (TELFOR), Belgrade, Serbia, 22–23 November 2016. 2016. URL: https://doi.org/10.1109/telfor.2016.7818791 (date of access: 19.10.2023).
9. Hsieh C.-J., Chan T.-Y. Detection DDoS attacks based on neural-network using Apache Spark. 2016 international conference on applied system innovation (ICASI), Okinawa, Japan, 26–30 May 2016. 2016. URL: https://doi.org/10.1109/icasi.2016.7539833 (date of access: 19.10.2023).
10. Ahuja N. DDOS attack SDN Dataset. Mendeley Data. URL: https://doi.org/10.17632/jxpfjc64kr.1 (date of access: 19.10.2023).
11. Han J. Data mining: concepts and techniques. 3rd ed. Burlington, MA : Elsevier, 2011. 703 p.
12. Scikit-learn. Scikit-learn. URL: https://scikit-learn.org (date of access: 19.10.2023).

Maksym Chornobuk **Максим Чорнобук**	student, Software Tools Department of National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine, e-mail: chornobuk.maksym@gmail.com https://orcid.org/0000-0003-3200-7306 Scopus Author ID: 58205318000	студент кафедри програмних засобів, Національний університет «Запорізька політехніка», Запоріжжя, Україна.
Valeriy Dubrovin **Валерій Дубровін**	PhD on Engineering, Professor, Software Tools Department of National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine, e-mail: vdubrovin@gmail.com https://orcid.org/0000-0002-0848-8202 Scopus Author ID: 7003406517, ResearcherID: L-2451-2018 https://scholar.google.com/citations?user=UZGC3S8AAAAJ	кандидат технічних наук, професор кафедри програмних засобів, Національний університет «Запорізька політехніка», Запоріжжя, Україна.
Larysa Deineha **Лариса Дейнега**	Senior Lecturer, Software Tools Department of National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine, e-mail: deynega.larisa@gmail.com https://orcid.org/0000-0003-0304-4327 Scopus Author ID: 57201676588, ResearcherID: K-3885-2018 https://scholar.google.com.ua/citations?user=DwbGeYUAAAAJ	старший викладач кафедри програмних засобів, Національний університет «Запорізька політехніка», Запоріжжя, Україна.