*Computers, Materials & Continua*
DOI: 10.32604/cmc.2022.029969
***Article***

Tech Science Press

# Automatic Botnet Attack Identification Based on Machine Learning

**Peng Hui Li[1], Jie Xu[1,*], Zhong Yi Xu[1], Su Chen[1], Bo Wei Niu[2], Jie Yin[1], Xiao Feng Sun[1], Hao Liang Lan[1] and Lu Lu Chen[3]**

[1]Jiangsu Police Institute, Nanjing, 210000, China
[2]Public Security Department of Jiangsu Province, Nanjing, 210000, China
[3]The University of Adelaide, Adelaide, 5005, SA, Australia
*Corresponding Author: Jie Xu. Email: xujie_net@jspi.cn
Received: 15 March 2022; Accepted: 26 April 2022

**Abstract:** At present, the severe network security situation has put forward high requirements for network security defense technology. In order to automate botnet threat warning, this paper researches the types and characteristics of Botnet. Botnet has special characteristics in attributes such as packets, attack time interval, and packet size. In this paper, the attack data is annotated by means of string recognition and expert screening. The attack features are extracted from the labeled attack data, and then use K-means for cluster analysis. The clustering results show that the same attack data has its unique characteristics, and the automatic identification of network attacks is realized based on these characteristics. At the same time, based on the collection and attribute extraction of Botnet attack data, this paper uses RF, GBM, XGBOOST and other machine learning models to test the warning results, and automatically analyzes the attack by importing attack data. In the early warning analysis results, the accuracy rates of different models are obtained. Through the descriptive values of the three accuracy rates of Accuracy, Precision, and F1_Score, the early warning effect of each model can be comprehensively displayed. Among the five algorithms used in this paper, three have an accuracy rate of over 90%. The three models with the highest accuracy are used in the early warning model. The research shows that cyberattacks can be accurately predicted. When this technology is applied to the protection system, accurate early warning can be given before a network attack is launched.

**Keywords:** Honeypot; log; network attack; machine learning

## 1 Introduction

In the context of global security threats, network networks are happening all the time. In the network of oil, Iran has been threatened by the virus. In Europe and the United States, Zeng, the largest meat company by sales, paid $11 million to resolve a ransomware attack due to a pipeline in which the world entered a national emergency. The information of the incident also poses a huge threat

to the cybersecurity of national security. The network has produced a huge threat to the network of units and individuals [1]. The impact of attack methods on advancing cyberspace and careers.

In terms of attack methods, network attacks are divided into two categories, namely active attacks and passive attacks. Both attacks violate the basic attributes of information security. Active attack means that the attacker directly invades the target system while tampering or forging network information. Passive attack refers to an attack method in which the attacker collects information on the target system, or uses deception to obtain relevant information of the target system, instead of actively attacking the target system. In this article, from the perspective of analyzing TeaPot log data, the author divides the attack types into three types of attacks: Network Scanning, Botnet, and Hacker Attack based on the characteristics of the attack analyzed in the log. Network Scanning refers to the use of tools or other means to send data packets to the target system to obtain the target system information. Botnet refers to hackers who use their own programs to implant into the target machine, obtain certain control rights of the target machine, and form a command and controlling node to send forged data packets to attack a predetermined network. Hacker Attack refers to hackers using some tools and methods to attack the target system [2–5].

Botnet is the most characteristic of the three network attacks. It is a malicious software written by the attacker, which controls the network host and turns the network host into a broiler or puppet machine of the attacker [6]. Attackers can control these network hosts to launch detailed attacks on one or more targets according to their will. Botnet has the characteristics of high dissemination, high dissemination, and highly stolen secrets.

In network security protection, honeypots are used as an active defense tool to detect or defend against unauthorized operations or "traps" of Hacker Attack. Deploy some hosts as bait on the network. Such a "trap" program can induce an attacker to attack it. At the same time, the honeypot system has the functions of capturing and analyzing attack behaviors like the Fig. 1. After the defending party records the attacking party's attack data, it can more effectively understand the attacking methods and methods adopted by the attacking party. By inferring their attack intent and motivation, defenders have a clear picture of the security threats they face. This helps the defender to guard against the attacker's behavior when the attacker launches an attack, and at the same time, after the attacker conducts the attack, more effectively analyze the attack and defense attack behavior. At the same time, the collected attack data is also an attack method and tool to help network security experts analyze the attacker. At the same time, as a "fake target", the honeypot has a certain effect on delaying the attacker's attack on the real target [7–9].

In order to analyse data better and automatic identification, we have joined machine learning algorithms to this research. K-means is an unsupervised learning algorithm. The K-means algorithm is relatively easy to implement and has a good clustering effect. The K-means algorithm presets the number of clusters and continuously updates the centers of the cluster points. After several iterations, let the sum of squared distances from all data points to the center point stabilize. Finally, the clustering results are formed [10].

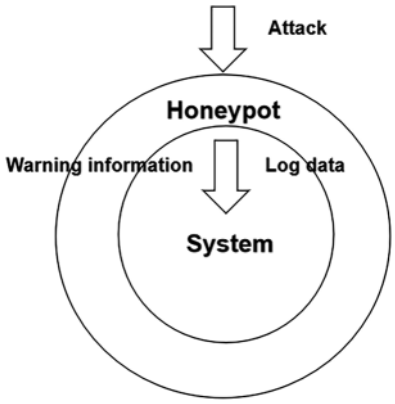


**Figure 1:** Honeypot system

## 2 TeaPot Log Introduction

### *2.1 Naming Traffic Log Files*

TeaPot will place the attack log files captured every day in the same folder, and name the placed folder as today's date, in the Fig. 2. for example: "November 7, 2021" will be named "20211107" and the format will be named "fake_data_local Port number" being "attacked_IP address" of remote attack. And the log file will be divided into two log files, log and asclog, for storage. The file with the suffix name is ".asclog", but the data stream format is changed to a string, and the content is the same as ".log". The Hex format log data is saved in the log file, which will be more conducive to importing the data into IDS (Intrusion Detection System) for attack detection, and the asclog file makes it easier to extract attack features.

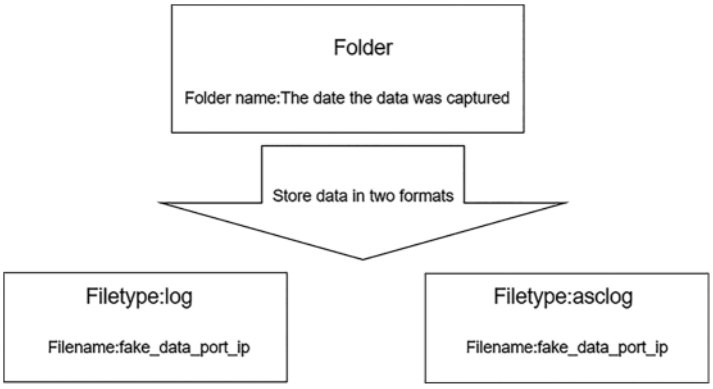


**Figure 2:** Two types of log files

### *2.2 Traffic Log Content*

TeaPot records the attack data interaction information in the log file. In the log file, it records the attack time, the attacker's IP address, the port being attacked, and detailed message interactions, and records sixteen input messages and normal messages. The text information, as shown in the first message in the Fig. 3 below, indicates that at "2021.10.23 04:16:04", the attacker's IP address is 89.248.168.226, attacked the port 21 of the machine.

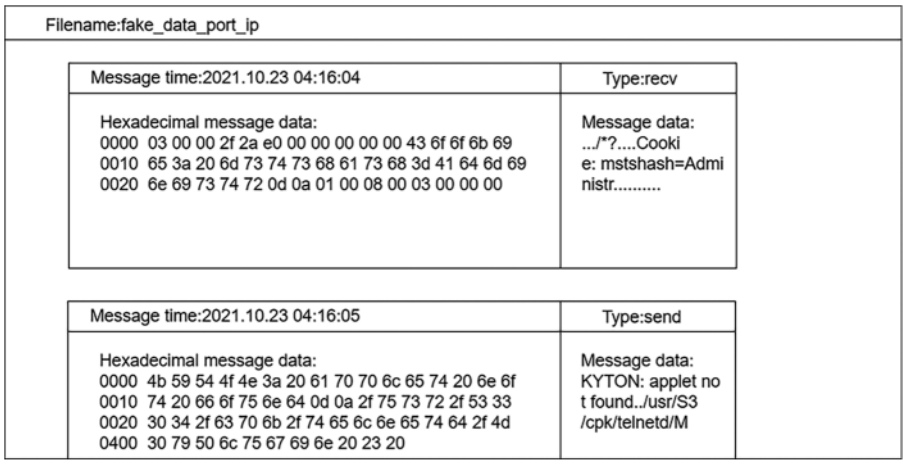


**Figure 3:** Log message format

### *2.3 Message Interaction Type Record*

TeaPot has the function of automatically bouncing the message to the attacker. Therefore, in the data packet, in order to easily see whether the message has been responded again by the attacker, the message is divided into "recv" and "send". The following is a detailed introduction:

recv: TeaPot will return all attack data to the third-party server. As shown in the Fig. 4 below, if an attack message from the target server is received, the target server will send a message to the server where TeaPot is located, and return to the port corresponding to the target server. In the message record, we record the message interaction type of target server sending to TeaPot as "recv".

send: The "recv" type is the message that the TeaPot server receives from the target server and returns to the target server. If the target server responds to the message that the TeaPot server bounces back to the target server at this time and due to the mechanism of the TeaPot server. The response message will be bounced back to the target server once again, TeaPot will record the interaction type of this bounced message as "send".

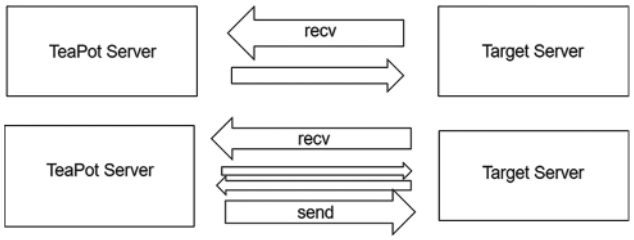


**Figure 4:** Log message type model

## 3 Data Attack Classification Based on TeaPot Logs

TeaPot's log will automatically record the plaintext data and ciphertext data of the message. As show in the Fig. 5 below. We separate the plaintext data in the log data from the attacker's IP, attacked port, attack time, attack duration, attack protocol, attack message and other characteristic information, divide the attack type into three categories: Network Scanning, Botnet, and Hacker

Attack [11]. These three types of data have relative characteristics with each other. For example, Network Scanning has a high probability of only one message in the messages recorded by TeaPot [12–14]. While Botnet and Hacker Attack will continue to attack the server Attacks will also result in multiple message interaction records in the TeaPot records. This article will focus on starting with Botnet, from cleaning data to analyzing data, and then analyzing the characteristics of Botnet. In addition, two types of attack related features, Network Scanning and Hacker Attack, will also be mentioned. For ciphertext data, we will extract its features for analysis, or try to decrypt the ciphertext data for analysis.

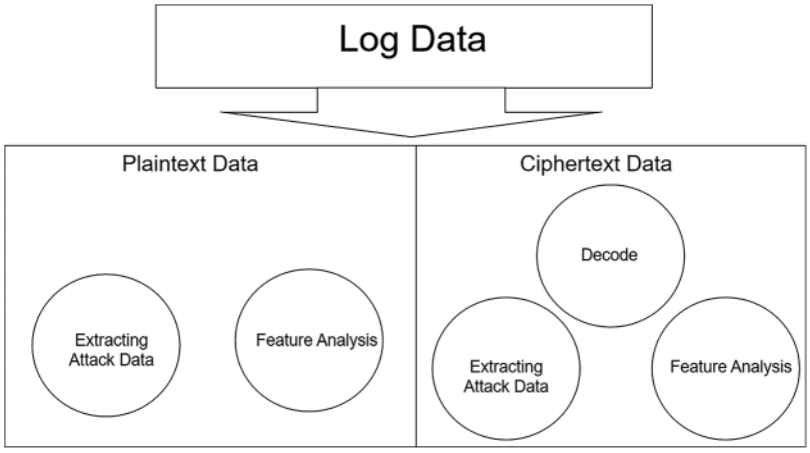


**Figure 5:** Plaintext and ciphertext data

## 4 Data Cleaning and Extraction

In the message information obtained by TeaPot, there are many encrypted messages. Start with the characteristics of plaintext data and ciphertext data, distinguish encrypted messages and cipher messages. After completing the message differentiation work, proceed to data clearing and data extraction:

1. Because the encrypted data has a higher degree of confusion, the encrypted data will rarely have 00 characters compared to the plaintext data. The number of 00 characters and the percentage of 00 characters are used to determine whether it is encrypted data.
2. In the message, there will be many invisible characters in the encrypted data. To judge by the visible characters, the plaintext data are generally visible characters.
3. Encrypted data will increase the entropy value of the data. The entropy value is used to determine whether it is encrypted message data.
4. Combine port numbers to analyze encrypted data and plaintext data, such as: 23 port telnet data is not encrypted, and 80 port web data is basically not encrypted
5. Combine the methods summarized above and combine machine learning methods to perform cluster analysis

Finally, we need to pay attention to that encrypted data is not useless data, encrypted data can also be analyzed. For example, we can start with the characteristics of encrypted data. If encrypted data appears continuously for a long time, it may be ciphertext blasting, attempted attacks, and it may

be a password attack. The specific attack type judgment needs to be combined with more features of the ciphertext data to make a comprehensive judgment.

## 5 Distinguish Botnets among the Three Types of Attacks

First, select the plaintext data for analysis. In order to analyze the Botnet attack data, it is necessary to analyze the Network Scanning type among the three main types of attacks. As show in Fig. 6, to classify by characteristics. Because Network Scanning has obvious characteristics, for example, only one message is sent in an attack and there is no code with attack characteristics. Botnet and Hacker Attack have long message exchanges with the target server during the attack, but Network Scanning may be a preliminary method of Hacker Attack, so we should pay attention to the distinction between this situation in data analysis. In addition, compared with Hacker Attack, Botnet and Hacker Attack have a very clear purpose of attack. Botnet is often based on DDoS and detection. Compared with Hacker Attack, Botnet will have more and more obvious features [15,16]. Below, several methods will be listed to distinguish between Botnet and Hacker Attack:

1. Distinguish between Botnet and Hacker Attack based on the characteristics of interactive messages. Botnet's attackers are often the victims. Therefore, the victim's server will attack our host by malicious attackers. If the message bounces back to it, then it will return the relevant message to us, but if it is a hacker's attack source, it often does not have this feature.
2. The number of attack ports and the type of protocol used can be used to distinguish whether the attack source is Botnet or a hacker's attack source. Because hackers' attack sources have a large number of ports that can be attacked, and there are many protocols used, it is difficult for Botnet to perform customized attacks on ports. From the perspective of the number of attacked ports, Botnet has obvious characteristics. Botnet usually only performs attack tests on a small number of ports, and the ports of these attacks are often fixed.

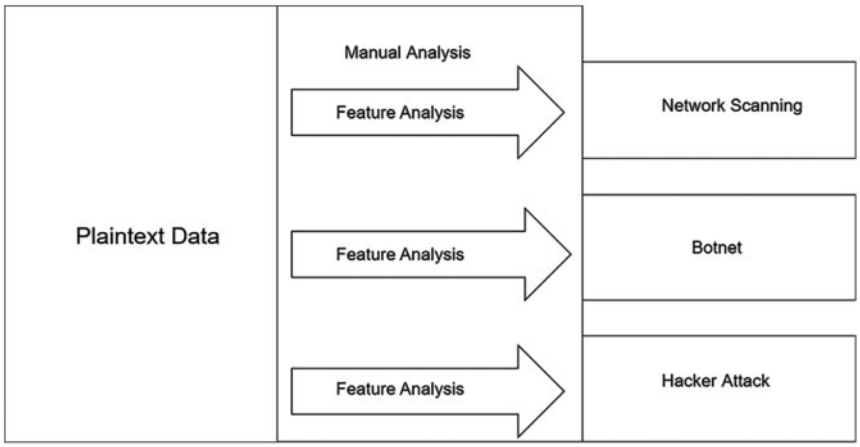


**Figure 6:** Attack type classification

## 6 Characteristic Analysis of Botnet

1. Interactive message type (T1): Find typical interactive Botnet interactive information in interactive messages. For example, there is "/MyPlugin" in the interaction data between TeaPot and the attacking server. Because TeaPot bounced the attack data to the attack server, causing

the attack server to have an effect of attacking itself. Botnet found that there was its own virus in the attacked server, so Botnet stopped the attack. Therefore, in the message, there will be special messages such as "/MyPlugin".

2. Download address type (T2): Find the download address of the Botnet virus in the message. The download address of the relevant Botnet domain name exists in the message. After finding the relevant download address and verifying it manually as the Botnet download address, these log files can also be marked as Botnet in batches. If the keyword is http://194.87.139.103/cleanfda/init.sh, the address is verified as Botnet's virus download address.
3. Action Statement Type (T3): It is judged as Botnet by looking for suspicious operation messages of Botnet. For example, in the ftp attack packet of port 21, the "STOR/Photo.scr" command appears, which is Botnet through the ftp protocol. Botnet copies the virus into the Photo directory. From this operation, it can be determined that the attack type is Botnet.
4. Characteristic character type (T4): In the message, look for characteristic characters. For example, in the attack message, the Mozi.m feature word appears, and "Mozi.m" is a Botnet. In normal interactive messages, the name of Botnet will not appear. In the interactive message of the characteristic character type, it can be judged as Botnet. The attack type is Botnet.
5. Specific port type (T5): By combining a specific port number and then according to the characteristic information of the port, the attack type is determined to be Botnet:

1. The attack packet of port 23 appears in the packet and bounces the shell, so it can be judged that the attack type is Botnet.
2. 5555 is the port that the Android adb service listens on by default. For attacks on port 5555, the attack packets are fixed. If device open port 5555, it is more likely to be Botnet. Port 5555 is open, and external attacks may be launched. Combined with manual analysis, certain tests are performed on suspicious IP addresses to determine the type of Botnet attack.
3. Port 6379 is the redis service. If the redis service is not configured properly, unauthorized access can be performed. The redis attack can be found more clearly in the log file ".asclog". In the message information, there is the redis version number, and in the interactive message, a rebound shell is found. Then it can be judged that the attack type is Botnet.

## 7 Experimental Data

On the public network, we use dozens of servers to capture network attacks, and we selected three servers in the early experimental data part. A TeaPot honeypot is built on the public network IP server to collect Botnet attacks from the public network. The log data is obtained by capturing attacks on the public network.

In the experimental part, 7252 pieces of attack data were found out of 24,918 pieces of data using 22 characteristic strings. And analyze the data. As shown in Tab. 1, 22 character strings are selected to be defined as attack characteristic strings, including five attack types. After sorting out the experimental data, as shown in Fig. 7, it is found that ports 6379, 8080, and 80 are vulnerable to attacks. Among the five strings, the characteristic string S4 appears the most times as shown in Tab. 2. We identify log data through feature strings, label and extract data. Attribute extraction is performed on the attack data and machine learning is performed. To build a model to early warning of attacks. And the accumulation of attack data is continuously increased to increase the efficiency of the model for identifying network attacks.

**Table 1:** Table character string

| Character string number | Character string | Instruction |
|---|---|---|
| S1 | T4 | Mozi.a |
| S2 | T2 | http://z.shavsl.com |
| S3 | T2 | http://crypto.htxreceive.top/ |
| S4 | T2 | http://oracle.zzhreceive.top/ |
| S5 | T2 | http://185.142.239.128/ |
| S6 | T2 | http://188.213.49.155/ |
| S7 | T2 | http://194.147.142.88/ |
| S8 | T1 | MyPlugin |
| S9 | T2 | http://103.209.103.16/ |
| S10 | T2 | http://199.19.226.117/ |
| S11 | T2 | http://205.185.121.185/ |
| S12 | T2 | http://86.105.195.120/ |
| S13 | T2 | http://104.192.82.138/ |
| S14 | T4 | Mozi.m |
| S15 | T2 | http://185.224.129.251/ |
| S16 | T2 | http://149.28.85.17/ |
| S17 | T2 | http://128.199.240.129/ |
| S18 | T3 | STOR/Photo.scr |
| S19 | T2 | http://194.87.139.103/ |
| S20 | T2 | http://45.133.203.192/ |
| S21 | T2 | http://85.239.33.9/ |
| S22 | T2 | http://34.66.229.152/ |

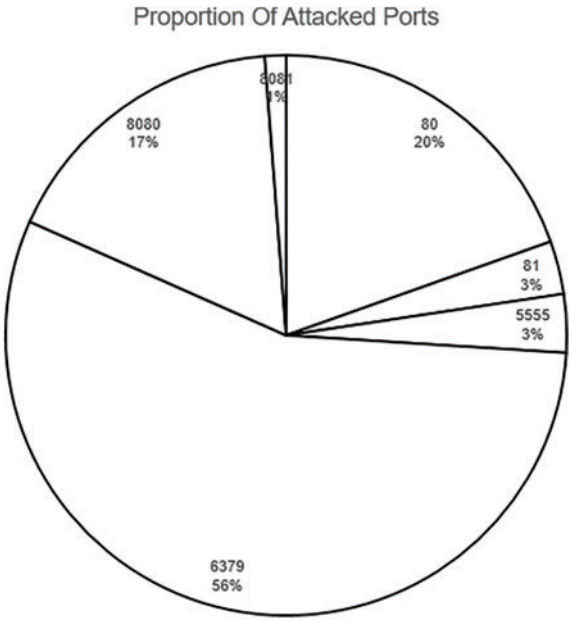


**Figure 7:** Proportion of attacked ports

**Table 2:** Attack statistics

| Character string number | Attack number | Proportion (%) |
|---|---|---|
| S1 | 295 | 4% |
| S2 | 126 | 1.70% |
| S3 | 136 | 3.20% |
| S4 | 3672 | 50.40% |
| S7 | 1 | 0.10% |
| S10 | 318 | 4.30% |
| S13 | 1173 | 16.10% |
| S18 | 217 | 2.90% |
| S19 | 841 | 11.50% |
| S21 | 157 | 2.10% |

## 8 Machine Learning

Attack types have been categorized above and perform feature extraction and data analysis on the attack data. By extracting the analyzed data, a library of feature strings has been constructed to identify the data. In order to be able to use machine learning methods to identify new log data. to determine whether it is attack data. For the existing data, K-means algorithm is used for cluster analysis. The existing more than 7000 pieces of attack data are extracted from the following Tab. 3 and then perform machine learning analysis on the extracted data. After the clustering results are generated by the K-means algorithm, the data are represented on the coordinate axis. In the data part of this experiment, K is set to 10, and the machine learning algorithm is used for automatic analysis to generate the results [17–19].

**Table 3:** Characteristics number

| Property number | Property name |
|---|---|
| 1 | Attacker port number |
| 2 | Attacked port |
| 3 | Packets number |
| 4 | Received packets number |
| 5 | Sent packets number |
| 6 | Average size of received packets |
| 7 | Average size of packets sent |
| 8 | Minimum value to send packets |
| 9 | Maximum received packets |
| 10 | Maximum number of packets sent |
| 11 | Variance of received packets |
| 12 | Variance of sent packets |

(Continued)

**Table 3:** Continued

| Property number | Property name |
|---|---|
| 13 | The time span of the first data and the last data |
| 14 | Average time interval |
| 15 | Maximum time interval |
| 16 | Minimum time interval |
| 17 | Median time interval |
| 18 | Variance of time interval |
| 19 | Attack type |
| 20 | Attack mode number |
| 21 | Log name |

As shown in Fig. 8, each point is represented as a piece of attack data. On the abscissa, it is represented as the number of the character string. For example, the number 1 represents the string number S1. The ordinate represents the cluster number of each attack data gathered after using the K-means algorithm [20–23].

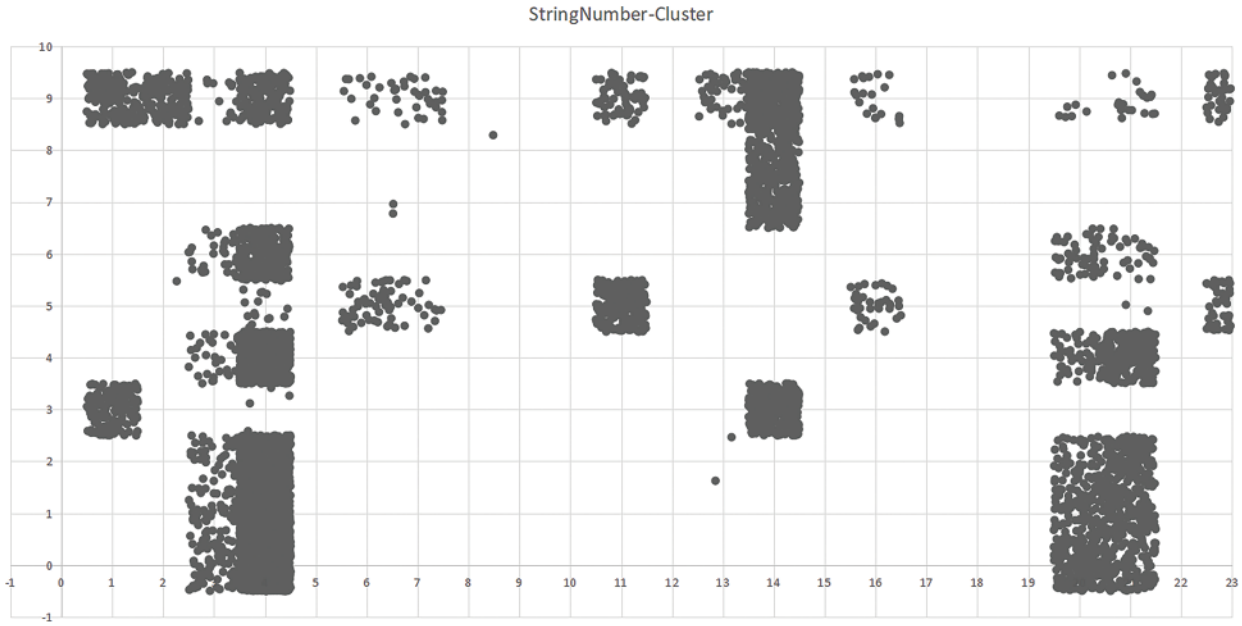


**Figure 8:** Machine learning analysis results of stringnumber-cluster

As shown in Fig. 9, each point is represented as a piece of attack data. The number of the attack type is represented on the abscissa. For example, the number 1 represents the attack type number T1. The ordinate represents the number of a cluster of attack data gathered [24–26]. It can be observed from both figures that the attack data are clustered into clusters, and the same feature string number or the same attack type tends to cluster into the same cluster [27–30]. The K-means algorithm can be used to identify the data well [31–35]. The following two figures represent. Through the K-means

algorithm, the same feature string data can be well grouped into a cluster [36–40]. Data of the same attack type can also be grouped into a cluster by this algorithm. There are prerequisites for automatic identification of attack data.

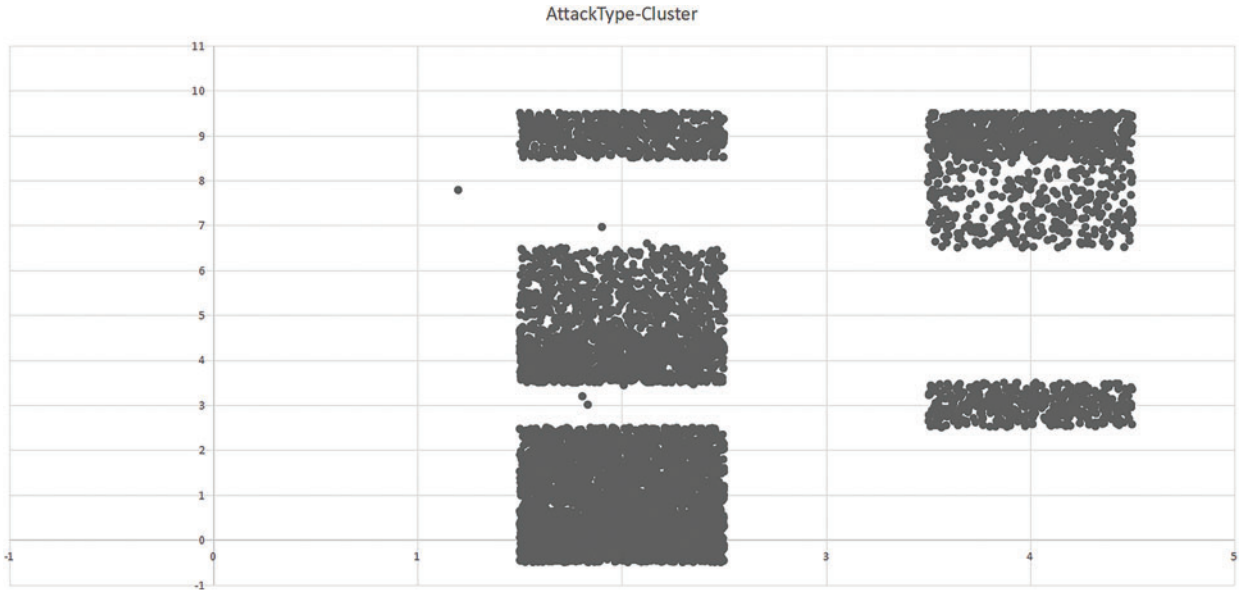


**Figure 9:** Machine learning analysis results of attacktype-cluster

As shown in Fig. 10, In TeaPot's log data, the Botnet attack data was identified by using the method mentioned above. The attributes of the extracted Botnet attack data were extracted, and more than 7,000 pieces of data that had been identified as Botnets were divided into two groups, one group accounted for 80%, and the other group accounted for 20%. In model building, 80% of the data is used as training data for the model. In order to understand which model is more suitable for early warning of network attacks, we have adopted models such as XBOOST, RF, GBM, ANN, and GLM. Train with an automatic learning framework, with model default parameters. And the data are modeled using different algorithms. Finally, another 20% of the real data is imported into the model as the test data set. As shown in Tab. 4. Accuracy, Precision, and F1_Score are finally generated by the model for the accuracy of early warning results for attack types. Through the results obtained above, we can analyze the effect of different models on early warning of network attack types. Finally, we found that the three models of RF, GBM, and XGBOOST are more suitable for Botnet attack warning. These three models have an early warning accuracy rate of more than 90% for attack information. Even in the XGBOOST algorithm, the accuracy rate for attack types exceeds 95%. Finally, the three models with the highest accuracy are used in the early warning model.

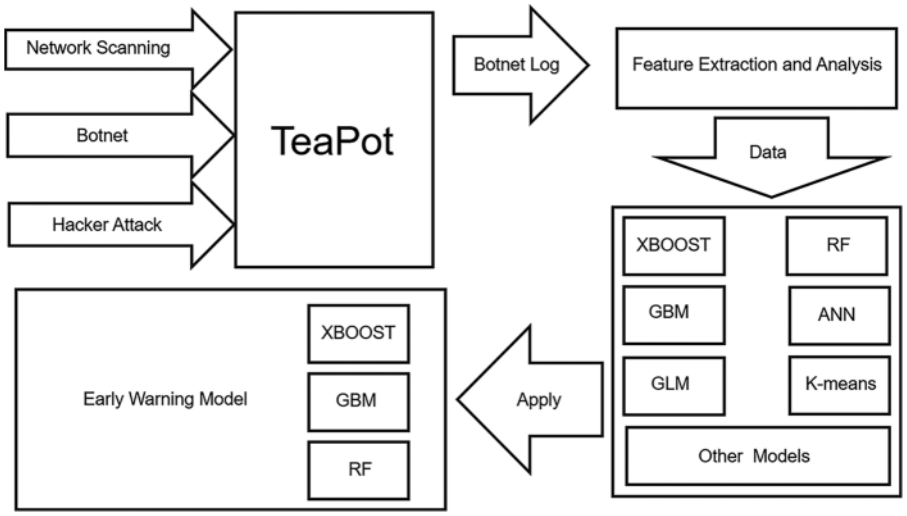


**Figure 10:** Model selection

**Table 4:** Model accuracy comparison

| No | Model name | Acccuracy | Precision | F1_Score |
|---|---|---|---|---|
| 0 | GLM | 0.6988284 | 0.523764 | 0.586902 |
| 1 | RF | 0.9207443 | 0.927358 | 0.911013 |
| 2 | GBM | 0.9352171 | 0.933697 | 0.930374 |
| 3 | XGBOOST | 0.9545141 | 0.953853 | 0.952518 |
| 4 | ANN | 0.7649897 | 0.648759 | 0.69469 |

## 9 Conclusion

Based on the analysis of TeaPot log data, this paper divides the attack types into Network Scanning, Botnet and Hacker Attack according to the characteristics of TeaPot log data. The attack characteristics, identification methods and some keywords of botnets are mainly analyzed. This paper analyzes the attack characteristics, which has far-reaching significance for enhancing and improving the honeypot function. At present, our team has collected tens of thousands of attack data by deploying the TeaPot system on the platform of dozens of public network servers, extracted millions of attack features from the attack data, and used machine learning methods to construct a model to determine the attack type. Automatically identifying attack types and providing targeted attack warning and prevention will greatly improve the security of the honeypot active defense system. In addition, this research plays an important role in using machine learning to automatically identify attack data and provide early warning of network attacks.

**Funding Statement:** The research of this paper is supported by the project of Jiangsu Provincial Department of Education (20KJB413002) and the science and technology research project of Jiangsu Provincial Public Security Department (2020KX007Z) and the Jiangsu Police Institute high level talent introduction research start-up fund" (JSPIGKZ, JSPI20GKZL404) and the 2021 doctor of

entrepreneurship and innovation in Jiangsu Province (JSSCBS20210599) and the Undergraduate Innovation and Entrepreneurship Training Program of Jiangsu Police College (No. 202110329053Y).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1] L. chen, X. S. Chen, J. F. Jiang, X. Y. Yin and G. L. Shao, "Research and practice of dynamic network security architecture of IAAS platform," *Tsinghua Science and Technology*, vol. 19, no. 5, pp. 496–507, 2014.

[2] I. H. Sarker, A. Kayes, S. Badsha, H. Alqahtan, P. Watters *et al.,* "Cybersecurity data science: An overview from machine learning perspective," *Journal of Big Data*, vol. 7, no. 1, pp. 1–29, 2020.

[3] S. Mahdavifar and A. A. Ghorbani, "Application of deep learning to cybersecurity: A survey," *Neurocomputing*, vol. 147, no. 2, pp. 149–176, 2019.

[4] Y. Xin, L. S. Kong, Z. Liu, Y. L. Chen, Y. M. Li *et al.,* "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.

[5] P. Dixit and S. Silakari, "Deep learning algorithms for cybersecurity applications: A technological and status review," *Computer Science Review*, vol. 39, no. 4, pp. 100317, 2021.

[6] W. Wang, Y. Y. Shang, Y. Z. He, Y. D. Li and J. Q. Liu, "BotMark: Automated botnet detection with hybrid analysis of flow-based and graph-based traffic behaviors," *Information Sciences*, vol. 511, no. 3, pp. 284–296, 2020.

[7] Y. B. Sun, Z. H. Tian, M. H. Li, S. Su, X. J. Du *et al.,* "Honeypot identification in softwarized industrial cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5542–5551, 2020.

[8] M. Baykara and R. Das, "A novel honeypot based security approach for real-time intrusion detection and prevention systems," *Journal of Information Security and Applications*, vol. 41, pp. 103–116, 2018.

[9] L. Y. Shi, Y. Li, T. X. Lin, T. Liu, B. Y. Shan *et al.,* "Dynamic distributed honeypot based on blockchain," *IEEE Access*, vol. 7, pp. 72234–72246, 2019.

[10] M. S. Yang and K. P. Sinaga, "A feature-reduction multi-view K-means clustering algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, 2019.

[11] S. S. Zhang, X. Y. Tang, Q. W. He, J. C. Liu and Z. L. Ying, "External correlates of adult digital problem-solving behavior: Log data analysis of a large-scale assessment," ArXiv Preprint ArXiv, vol. 2103, pp. 15036, 2021.

[12] T. A. Tuan, H. V. Long, L. H. Son, R. Kumar, I. Priyadarshini *et al.,* "Performance evaluation of Botnet DDoS attack detection using machine learning," *Evolutionary Intelligence*, vol. 13, no. 2, pp. 283–294, 2020.

[13] S. L. Tang, X. L. Huang, M. J. Chen, S. J. Sun and J. Yang, "Adversarial attack type I: Cheat classifiers by significant changes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 1100–1109, 2019.

[14] Y. Zhang, J. Niu, D. Guo, Y. L. Bao and L. X. Bao, "Unknown network attack detection based on open set recognition," *Procedia Computer Science*, vol. 174, no. 4, pp. 387–392, 2020.

[15] W. Feng and Y. Q. Wu, "DDoS attack real-time defense mechanism using deep Q-Learning network," *International Journal of Performability Engineering*, vol. 16, no. 9, pp. 1362–1373, 2020.

[16] Q. Yan, M. D. Wang, W. Y. Huang, X. P. Luo and F. R. Yu, "Automatically synthesizing DoS attack traces using generative adversarial networks," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 12, pp. 3387–3396, 2019.

[17] M. Ahmed, R. Seraj and S. M. S. Islam, "The K-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, pp. 1295, 2020.

[18] C. H. Yuan and H. T. Yang, "Research on K-value selection method of K-means clustering algorithm," *Multidisciplinary Digital Publishing Institute*, vol. 2, no. 2, pp. 226–235, 2019.

[19] S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan and T. C. Chang, "Two improved K-means algorithms," *Applied Soft Computing*, vol. 68, pp. 747–755, 2018.
[20] R. U. Khan, X. S. Zhang, R. Kumar, A. Sharif, N. A. Golilarz *et al.,* "An adaptive multi-layer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, pp. 2375, 2019.
[21] I. Rosenberg, A. Shabtai, Y. Elovici and L. Rokach, "Adversarial machine learning attacks and defense methods in the cyber security domain," *ACM Computing Surveys (CSUR)*, vol. 54, no. 5, pp. 1–36, 2021.
[22] A. Wang, W. Chang, S. Chen and A. Mohaisen, "Delving into internet DDoS attacks by botnets: Characterization and analysis," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2843–2855, 2018.
[23] V. Bontchev and V. Yosifova, "Analysis of the global attack landscape using data from a telnet honeypot," *Information & Security: An International Journal*, vol. 43, no. 2, pp. 264–282, 2019.
[24] S. Sun, Z. Cao, H. Zhu and J. Zhao, "A survey of optimization methods from a machine learning perspective," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3668–3681, 2019.
[25] C. Janiesch, P. Zschech and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021.
[26] R. Roscher, B. Bohn, M. F. Duarte and J. Garcke, "Explainable machine learning for scientific insights and discoveries," *IEEE Access*, vol. 6, pp. 42200–42216, 2020.
[27] J. Grimmer, M. E. Roberts and B. M. Stewart, "Machine learning for social science: An agnostic approach," *Annual Review of Political Science*, vol. 24, no. 1, pp. 395–419, 2021.
[28] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.
[29] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenbury, T. Verbelen *et al.,* "A survey on distributed machine learning," *ACM Computing Surveys (CSUR)*, vol. 53, no. 2, pp. 1–33, 2020.
[30] B. Schölkopf, "Causality for machine learning," Arxiv Preprint ArxIv, vol. 1911, pp. 10500, 2019.
[31] E. Erdemir and A. A. Altun, "A new metaheuristic approach to solving benchmark problems: Hybrid salp swarm jaya algorithm," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2923–2941, 2022.
[32] T. Jeslin and J. A. Linsely, "AGWO-CNN classification for computer-assisted diagnosis of brain tumors," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 171–182, 2022.
[33] J. Onshaunjit and J. Srinonchat, "Algorithmic scheme for concurrent detection and classification of printed circuit board defects," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 355–367, 2022.
[34] I. S. Kocher, "An experimental simulation of addressing auto-configuration issues for wireless sensor networks," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3821–3838, 2022.
[35] A. Berguiga and A. Harchay, "An IoT-based intrusion detection system approach for TCP syn attacks," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3839–3851, 2022.
[36] M. A. Samad and D. Choi, "Analysis and modeling of propagation in tunnel at 3.7 and 28 GHz," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3127–3143, 2022.
[37] A. M. Almars, "Attention-based bi-LSTM model for arabic depression classification," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3091–3106, 2022.
[38] C. Cheng and D. Lin, "Based on compressed sensing of orthogonal matching pursuit algorithm image recovery," *Journal of Internet of Things*, vol. 2, no. 1, pp. 37–45, 2020.
[39] M. A. Haq, "CDLSTM: A novel model for climate change forecasting," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 236–2381, 2022.
[40] I. Sood and V. Sharma, "Computational intelligent techniques to detect DDOS attacks : A survey," *Journal of Cyber Security*, vol. 3, no. 2, pp. 89–106, 2021.