A Survey of Botnet Architecture and Batnet Detection Techniques

Cheng-Yi Liu[1], Chih-Hsiang Peng[2], and Iuon-Chang Lin[2,3]
(Corresponding author: Iuon-Chang Lin)

Department of Information Management,[1]
TransWorld University, Yunlin, Taiwan
Department of Management Information Systems,[2]
National Chung Hsing University, Taichung, Taiwan
Department of Photonics and Communication Engineering,[3]
Asia University, Taichung, Taiwan
(Email: iclin@dragon.nchu.edu.tw)

(Received Mar. 9, 2012; revised and accepted July 24, 2012)

Abstract

Botnet is an urgent problem that has impact on information security and reduces confidentiality, integrity and availability of certain service. Many large companies are one of the victims who are attacked by botmaster and it costs heavy economic losses. Some attacks happened due to the operation of botnet, such as DDoS and spam. To enhance security, many researches are concentrated on detecting and analyzing the architecture of botnet. How to detect the attacks from botnet and minimize the damage is an important issue. Many techniques, including intrusion detection system and honeypot, are designed to monitor packet data. Once identifying attacks, different solutions are used to block attacks on different levels. In this paper, the architecture of botnet, the methodology of detection and techniques are described.

Keywords: Botnet, DDoS, IDS

1 Introduction

With the rapid growth of technology, Internet has become a tool that can solve many problems in life [1, 2]. Although the usage of Internet is practical and it can also enhance overall efficiency, there exists several kinds of attacks in Internet and the number and type of attacks is still rising. Botnet is an urgent problem that has impact on information security and reduces confidentiality, integrity and availability of certain service [3]. In the research of information security, three standards, including confidentiality, integrity and availability, are discussed [4, 5]. Confidentiality is a standard to avoid unauthorized users to unfold the sensitive data intentionally or unconsciously. Integrity is the standard that data should be correct and in the integrity. And to availability, it is the standard that the system should be in normal operation when users need to access the system to get data or service. Because of some interest like financial gaining, attackers usually target large companies to implement attacks to achieve their goals. In the current network environment, botnet is still an urgent problem. In Taiwan, there are one-third of users which are infected with the botnet, and most of them are involuntary and without knowledge. The reason that the victims infect with the botnet is the bad usage of Internet, such as surfing unknown websites or downloading unauthorized software [6, 7, 8]. When the personal computers are injected with malicious software, the attacker can easily control them via the malicious software. Through the above model, number of infected user gains gradually day by day, and the scale of botnet will be extremely large when the infection has continued for a certain time period. Through the concatenation of network, the scale of attacks is too large that can paralyze a service in short time, and it is a typical example of distributed denial of service (DDoS). In 2002, attackers targeted some large companies, like Yahoo, eBay and Amazon, and distributed denial of service caused their services interrupted and large loss on finance [9]. In the past experience, any attackers could control botnet to launch attacks to the victim easily.

To enhance security, many researches are concentrated on detecting and analyzing the architecture of botnet. In recent years, researchers analyze botnet in different view and design many detection methodology and strategies to minimize the damage of attacks from botnet. In detection methodology, it is useful to detect botnet in statistics method and machine learning. Using the above two techniques, the pattern of attacks from botnet can be identified in different ways. The trained pattern can be the standard to determine the input packet data belongs to attacked data or normal data. In the analysis phase, there exists another issue about the selection of algorithm which is used to achieve feature selection, classification and prediction. By combining different algorithms, detection rate and overall performance will have variety changes. By applying different algorithms, a detection model can be trained and constructed. Because the packet data is different between simulation data and real data from the environment which exists numerous servers and personal computers, it is needed to set up in a real environment to test the performance with the trained detection model. After the detection phase, some techniques are applied to minimize the damage of attacks, and the techniques belong to hardware level and software level.

In this paper, several issues are discussed, such as the architecture of botnet, the attacks from botnet, detection methodology and strategy of detecting attacks from botnet. In the following sections, Literature review is the core section that contains several subsections and has the detailed description of botnet.

2 Literature Review

In this section, several issues about botnet are discussed. In the following article, Section 2.1 is the review of botnet, Section 2.2 is the attacks from botnet, Section 2.3 is the detection methodology and final Section 2.4 is the techniques that are related to detection. The purpose of this paper is to review current research articles, which are relevant to botnet, and get more reorganization about botnet.

2.1 Review of Botnet

Botnet is also called zombie network and it is closely related to IRC protocol, which is used to manage some functions like group chatting or recording chatting history [10, 11, 12, 13]. IRC is the abbreviation of Internet relay chat and it is programmed by Arkko Oikarinen in 1988. IRC is the first Internet chatting protocol in application layer and is frequently used in Internet chatting room like ICQ, MSN or other messenger software. The user of IRC has to use client software to connect with the IRC server and set unique username. The software that is frequently used is mIRC, ChatZilla or other plugin in the browser. At first, IRC is used to manage chatting server efficiently, and Greg Lindahl programmed an IRC bot which is called “game manager for the hunt the wumpus game” to enhance the efficiency of work. The concept of IRC server is then used by hacker to control the infected personal computers and implement the command from hacker. It is said that the hacker uses the convenience of IRC to develop botnet.

With the development of network technology, Internet is used more than past and more attacks appeared from Internet. Botnet is an urgent problem that needs to be investigated by researches and develop a completed methodology to detect it. As mentioned above, botnet is developed from IRC server. The concept of botnet is a series connection of infected personal computers and the infected ones could not resist the command from hacker and even do not know that they are infected with the botnet. Some users are infected by bad habits of using Internet, such as surfing unknown websites or downloading unauthorized software. Via the above actions, malicious software is injected into the personal computer and the infected user will wait for the command from hackers. Because the infected uses are volunteered to install or download malicious software, they always think they are away from infection. When the hacker has to launch attack, he will send command to the control and command server and the command and control server will sends the command to all infected personal computers which are called the bots. The attack will launch by the numerous bots to the certain victims.

In the classical botnet, a botnet has four characters, which includes bot master, control and command server, bot and the victim. Bot master is the unit that handles overall botnet, including managing, spreading and launching attacks. Managing botnet is the behavior that keeps the operation of botnet, spreading is the behavior that arranges malicious software or malicious websites that intent user to surf or download to install their malware, and launching is the behavior that starts attacking and sending command to the bots. The bot master often has the background of computer science and has knowledge in network. Owning abundant knowledge of network, it is easy to them to obtain benefits by using techniques of network. The bot master can obtain benefits from launching attacks, selling botnet service or other kinds of service that is related to botnet. Control and command server is the intermediate of communication between bot master and the bots. Control and command server is also called C&C server, and its functions are as its name: control and delivery command the botnet. If an attack is launching by a bot master, control and command server is needed to delivery command from bot master to the bots. In the classical botnet, control and command is a core part that will influence the operating. Through different kinds of architecture, control and command server has different tasks. Control and command server plays an important on delivering command and make attacks success. In the classical botnet, which is also called centralized botnet, the botnet can be broken down by detecting its control and command server. The bots are infected users that are controlled by bot master and they infected with botnet in different ways which are bad habits of using Internet. If the bots have no concept of monitoring the operation of their computers, it is difficult to find out they are the member of the botnet and controlled by bot master. Whenever the bot master launches the attack, the bots become the middlemen that actually attack the victim. According the model of attacks from botnet, the bot master just posts the command of attack, and all bots are awakened and execute the command from the bot master via the control and command server. As far as it goes, the bots do not know that they are infected and it makes the infection very serious in the current network environment. The last character in the botnet is the victim, which is the unit that receives attacks from botnet. The victim usually has some interest that attracts attacking from bot master. The victim can be a large company that makes money, a system that handle e-commerce and deal with a large number of orders, a military unit or a government website of certain country.

The relationship of the character in botnet is shown in Figure 1.

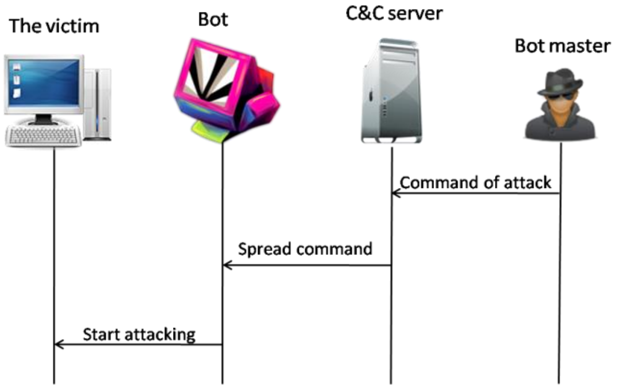

Figure 1: Attack model of botnet

The architecture of botnet can be classified into four types: star topology, multiple-server, hierarchical and random [14, 15]. Star topology, which is shown in Figure 2, is also called centralized botnet, which is most common and has the fastest speed of infection. Bot master launches attack by posting command to the control and command server, and the control and command server then broadcasts the command to all the bots. After the bots receive the command, the attack will start according to the attack pattern designed by the bot master. It is clear that the core part of this architecture is the control and command server, and it is very effective to break down botnet if the researcher or Internet service provider finds out the location of control and command server. By blocking the connection of the control and command server, the bots cannot receive the command from the bot master, and then the attack will not succeed. Compared with the star topology, the number of control and command server is modified. Due to the easiness of breaking down, multiple-server modifies the setting of control and command server. Two or more control and command servers are setting to post command, and each server is connected. When one of a server is detected and broke down, another one will replace it and keep the botnet operating normally. To the research or Internet service provider, it is more difficult to detect all servers than detecting only one server. To the bot master, the attack will keep going if one of the control and command servers exists, and it is safer with more servers. It also exists some disadvantages in the multiple-server architecture. To the bot master, it is harder to construct a multiple-server botnet because it is more complicated than star topology. In hierarchical botnet shown in Figure 4, there are several high level bots in it, and control and command server is not needed. High level bots are used as a C&C server, and its purpose is to make the real C&C server and bot master more hidden. If the bot master constructs the botnet by using hierarchical architecture, the botnet is not easy to be broken down due to the protection of the C&C server. The botnet just losses a part of bots if the high level bot is found. In the current network environment, the random botnet, which is also called P2P botnet, is still a serious problem that needs to be solved. The architecture of random botnet is shown in Figure 5. As shown in Figure 5, the random botnet does not have the control and command server [16]. Whenever one bot receives the command from the bot master, it will deliver the command to other bots which connect to it. Although the design of random botnet is very difficult, it has high safety because every bot is taken as a C&C server. In the centralized botnet, how to detect and break down C&C server is an important issue. In the P2P botnet, C&C server is very hard to detect. In the architecture of random topology botnet, if one of a bot is detected, it only has a little of effects that can break down the botnet because every bots are used as C&C servers.

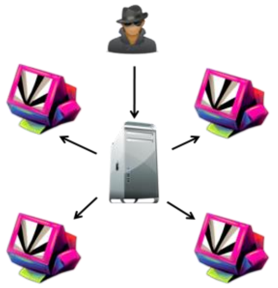

Figure 2: Star topology

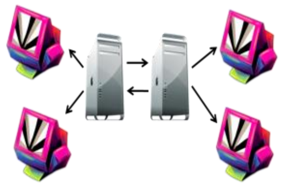

Figure 3: Multiple-server

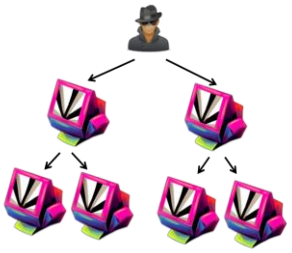

Figure 4: Hierarchical

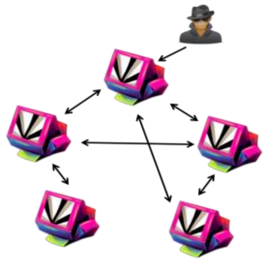

Figure 5: Random

2.2 Attacks from Botnet

There are several kinds of attack which is caused by botnet, and distributed denial of service, spam and personal information thieving are the most common among them. In this section, three kinds of attacks are described in detail.

2.2.1 Distributed Denial of Service

Distributed denial of service is one of the common attacks that are caused by botnet, and it usually makes a large quantity of financial loss [17, 18, 19]. The concept of distributed denial of service is to paralyze the operation of certain service that belongs to an organization like company, government or military. Some large companies are the targets that attract bot master to attack and receive benefit from the process of attacking. Some attacks are launched because of the economic benefits, and some are launched because of the self-satisfaction. By sending a large number of packets to the victim, the computation resource or bandwidth of the victim will be consumed, and this makes the victim’s service interruption. According to the method of paralyzation, distributed denial of service can be categorized into two types: bandwidth consumption and computation resource consumption. Two kinds of attack paralyze the victim in different ways. In bandwidth consumption, the attacker sends a great quantity of data to the victim and saturates the bandwidth. Via occupying the bandwidth, the victim has no more bandwidth to connect to the Internet. In the computation resource consumption, the attacker uses the vulnerability of the network protocol to launch attack.

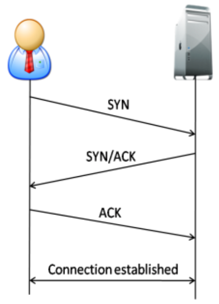

Figure 6: SYN flood

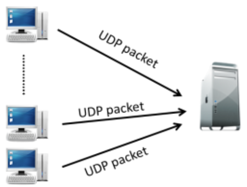

Figure 7: UDP flood

In the computation resource consumption, SYN flood is most common and easy to implement. SYN flood happened when the attacker used the vulnerability of TCP protocol [20]. In TCP protocol, three-way handshake is needed to ensure the integrity of data which is transmitted. The process of three-way handshake is that the user has to send a request to the server with a SYN packet, which is used to inform the server that a user needs to connect to it. Then the server will send a SYN/ACK packet back, which means the connection is approval. Finally, the user has to send an ACK packet to server and it means the start of the connection. Via three-way handshake in TCP protocol, data transmits more safely. The process is of three-way handshake is shown in Figure 6. SYN flood happened when the attacker sent a large number of SYN packet to the server to request a connection. With overmuch SYN packet, the server cannot handle too much request. If a legal user has to connect to the victim, the server will have too much request to deal with the legal user and it will be out of connection because it is full of SYN packet.

In the bandwidth consumption, the most common attacks are UDP flood, ICMP flood and HTTP flood [21]. It is similar to computation resource consumption, which paralyzes the victim by sending large number of packets. But there is still some difference between them because of the transmission protocol. UDP protocol (User Datagram Protocol) is not rigorous as TCP protocol which needs three-way handshake to ensure the establishment of connection. Compared with TCP protocol, the header is much simpler, which only has the port of source and destination, length, checksum and data. Despite the reliability of UDP protocol is not as high as TCP, it can transmit data in a massive and quick way. Thus the selection of protocol is important to enhance the security or efficiency of network. Because UDP protocol is connectionless oriented, which means that it does not check whether the data is received correctly, some attackers use the vulnerability of UDP protocol to launch distributed denial of service attack, which is called UDP flood. The attack model of UDP flood is shown in Figure 7. Different as SYN flood, UDP flood attacks the victim by sending a large number of packets to a random host which the port is open for certain services. If the number of members that belong to the botnet is larger, the scale of UDP flood will be larger. The characteristic of UDP flood, the bandwidth of the victim is huge due to the massive transmission of packet. After the bots receive the command from the bot master, the bots send the packet with high capacity to paralyze the bandwidth which the victim owns. If the victim encounters with UDP flood, the change of network traffic is very obvious and that makes the victim unable to connect to the Internet. ICMP flood is also a common distributed denial of service attack which belongs to bandwidth consumption. The purpose of the ICMP flood is to saturate the network by sending numeric ICMP packets. ICMP protocol (Internet Control Message Protocol) is used to reply the connection status between two hosts. Round-trip delay time is calculated by the number of successful reply and the time between them. It can check whether the destination host exists or not and compute the transmission time between source and destination host. The process of checking connection status which is also called ping is shown in Figure 8. Through “echo request” and “echo rely” sending by ICMP protocol, the reply time can check the connection status between two hosts. ICMP flood is similar to UDP flood, and it launches attack by sending numeric ICMP flood that makes the victim cannot afford processing it. If the victim receives too much ICMP flood, the bandwidth will be paralyzed. The effect is same as UDP flood, which is intercept the connection ability of the victim.

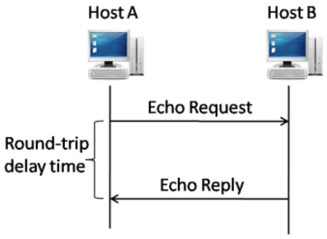

Figure 8: ICMP

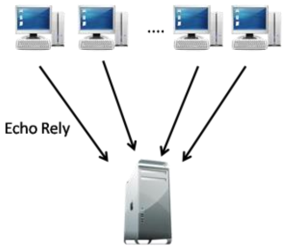

Figure 9: ICMP flood

2.2.2 Spam

Spam is also a serious problem which still exists in the current network environment. Due to the development of botnet, spam has become harder to solve because of its scale. Spam is also called junk mail, which attracts the click and links to the malicious website [22]. These kinds of website will make the victim infect with the botnet by downloading malicious software. Mail spam has some properties, like the content has malicious link or some attractive software. Mail spam usually sends by batch, which it means that spam sends numeric emails to the users in a way that is designed previously. Through the control with the command, the bots will send email to the receiver at certain time. The content usually has some links and attracts users to click. In spam, four characters are discussed, which are spammer, spam receiver and spam filtering service. Spammer is the unit that launches spam attack, which wants to gain some profit from it. The spam receiver is the victim who receives a larger number of spam mails. If the spam receiver cannot resist the attraction of the content in spam mail, some losses will happen, like personal information thieving and infecting with botnet. Spam filtering service is used to filter spam mail in some rules, and it is installed in client or mail server [23]. By Filtering mail with spam filtering service, spam mail cannot send to the end user easily. Some researchers and reports suggested that the mail, which the sender does not belong to friend list, should not be opened. According to the current network environment, if the sender is infected with botnet, which means the bot master can totally control it, the mail from the infected sender is not safe. Due to the behavioral model of spam from botnet, the spam mail is sent by command from bot master, and it makes the scale of spam grow quickly. Once the receiver clicks the links or downloads the file in it, it is possible that the user will post some private data to the server which bot master owns. If the receiver download the malicious software from the spam mail, it is possible that the infection happens. The receiver will become a member of botnet and wait for the command from the bot master.

2.2.3 Personal Information Thieving

Besides distributed denial of service and mail spam, one of the serious troubles from botnet is personal information thieving. Personal information is valuable, like name, email, address, account password and etc. This personal information is stored in the memory, and it is easy to be gotten by the bot master. As mentioned above in the behavioral model of botnet, some malicious software is injected into the victim’s computer and the bot master will control the victim via the control and command server. The purpose of personal information thieving is just information gathering, and gets financial profit from selling personal information. The personal information is thieved by two ways, which are uploading and recording. Uploading is that the infected user uploads some insensitive data to the bot master. Some information stores in cookie when the user saves their data intentionally or otherwise. Such file which stored in the memory usually has some insensitive data, like account and password. If the file that stores sensitive data is usurped, it is possible that the data inside can be extracted by information techniques. Besides uploading, recording is also a method to thieve personal information. Through the injected malicious software, it can record the input character when the infected user enters some sensitive data, like account, password, email or address. By recording illicitly by the software, the input data is gotten by the bot master and among them, “Spyeye” is one of a famous malicious software that thieving personal information. It is hard to be detected and monitor the victim to record the data from it. Some famous websites, like Google, Yahoo, eBay and Twitter, has stolen personal information by Spyeye. It monitors the victim by forcing downloading a configuration file and records the input illegally. If the personal information thieving occurs in the bank, it will be very severe because the stolen data is extremely sensitive. With more personal information that thieves from botnet, the bot master can have more financial benefit by selling to certain organizations.

2.3 Detection Methodology

In this section, detection methodology is described. In order to detect attacks from botnet, many researchers are concentrated on analyzing the characteristic of packet [24, 25, 26]. Via different methodology of analyzing attacks, attacks from botnet are detected and some standards are computed to evaluate the performance of the methodology. In the current issues that belong to detection, flow analysis is still an effective and flexible way. Flow analysis can be categorized into four steps: packet sniffing phase, data preprocessing phase, detection phase and evaluation phase. The concept of flow analysis is that for each packet flow in certain time zone, some messages are hidden and it can be used as a factor to detect attacks. For instance, the number of SYN packet is an important factor when SYN flood is needed to be detected. Compared with packet analysis, only one packet has no way to figure out whether the victim encounters SYN flood or not because SYN flood needs three-way handshake. If the detection is based on flow analysis, some attacks can be detected easily. In flow analysis, some methods are separated all packet data into several packet flows. The most common way to achieve segment is based on time fragment. For example, a day can be separated into twelve parts if the time segment is two hours. After setting the length of time fragment, several techniques are used to recalculate packet data in statistic methods.

Flow analysis is useful on analyzing logs, such as packet flow, DNS record, spam record and application logs. If flow analysis is used to detect attack from botnet, four core phases are commonly used to analyze packet data. The detail process about flow analysis is described in the following sections.

2.3.1 Packet Sniffing Phase

Packet sniffing is a way that can check the security of overall network environment. By analyzing the packet flow, some attacks can be detected and it can also achieve the detection of infected user from botnet. There are two models of packet sniffing, which includes active sniffing and passive sniffing. The first phase is packet sniffing phase, which the purpose is to collect packet data which contains attacked data and normal data. Packet sniffer is used to monitor packet data in the packet sniffing phase and some common sniffers are Wireshark, MRTG, Tripwire, Snort, snoop, tcpdump and etc. Except monitoring packet data, some packet sniffers can filter packet data by different rules, such as protocol, packet length, time and etc. Packet sniffer is a convenient tool that assists in researching by collecting packet data easily. Despite packet sniffing can be used on detecting attacks, it can also use on crime. By monitoring the packet flow, some information can be extracted, like the password which does not encrypted.

2.3.2 Data Pre-processing Phase

The second phase is data pre-processing phase, which the purpose is to recalculate data to be the form that can be used to detect attack. In this phase, feature selection and data calculation are two core components. In the flow analysis, a detection target is set before the packet sniffing phase. For different detection target, different features are used to be the factor of detection. For instance, the number of SYN packet is useful when the detection target is SYN flood, and it has high relationship with the detection result. However, it is useless for the detection of spam, because the attack model of spam does not rely on sending SYN packet continuously. In accordance with different kind of attack model, it is better that different set of features is designed. Despite the flexibility is low, it can have better performance on detecting different kinds of attack. It is difficult to design a pattern to detect all kinds of attack, and it has low performance averagely if only a set of features is used to be the factor of detection.

Feature selection is a way to enhance performance by selecting preferred set of features. It is useful when the scale of data is extremely large and dimension is high. Feature selection plays an important role on the flow analysis because the dimension of packet data is very large which contains more than twenty features. If too many features are used to be the factor of detection, some performances will decrease, like accuracy rate and computation time. Feature selection is strongly suggested before the detection no matter in statistic methods or machine learning methods. There are several algorithms about feature selection, and sequential forward selection is one of the famous algorithms among them, which is shown in Figure 10. It is proposed by Whitney in 1971 and it is a bottom up algorithm that is used to select preferred feature set. First, only one feature is considered and its performance is considered. The winner node will combine with other features to be a new feature set. This process will repeat until the preferred features set is found. Compared with sequential forward selection, the concept of sequential backward selection is similar. Through the process of feature selection is opposite, the purpose is still the same. Except sequential forward selection and sequential backward selection, it exists several algorithms that can achieve the goal of feature selection and statistic is also a suitable way to achieve feature selection, like correlation coefficient. Each feature has its influence on the category, and it can be easily computed. Feature selection is a process that evaluates which feature can have better influence on the detection and increase the performance of the system.

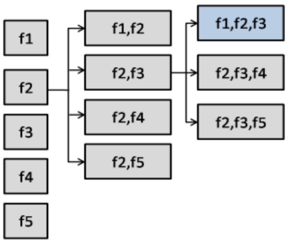

Figure 10: Sequential forward selection

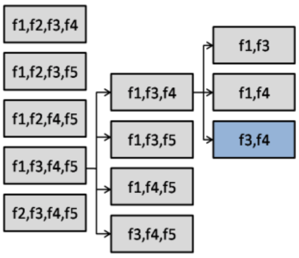

Figure 11: Sequential backward selection

After the feature selection, the feature set is decided by manual judgment or computing by algorithms. Packet data is recalculated into the form that conforms to the norm of features set. Some values cannot extract directly and it needs some computation, like average, standard deviation and the percentage. When the computation ends, the processed data will be the input of detection phase.

2.3.3 Detection Phase

In order to classify normal data and attacked data, some algorithms are used in detection phase, and the performance of the detection system is closely related to the design of the classification algorithms. Many researchers are concentrated on the improvement of algorithms or the combination of two or more algorithms. The only purpose is the increase of performance of the framework proposed by the researchers. In the detection phase, classification and prediction are the main parts and numeric algorithms are used. A concept of machine learning is also used, which contains supervised machine learning and unsupervised machine learning [27]. Machine learning is a science that can learn by itself and construct a pattern that can recognize unknown data. It is widely used in the domains which are image recognition, DDoS detection, biometric information, voice recognition and prediction of value. The difference between supervised and unsupervised machine learning is whether the correct answer exists or not. Some familiar supervised machine learning algorithms are artificial neural network, decision tree, k-nearest neighbor, support vector machine and etc. For unsupervised machine learning, some common algorithms are K-means self-organizing map and adaptive resonance theory. Two kinds of machine learning have different advantages. Supervised machine learning has higher accuracy rate than unsupervised machine learning. Unsupervised machine learning is much suitable than supervised machine learning when too many unknown patterns should be recognized because supervised machine learning just can identify known pattern that has been trained before.

2.3.4 Evaluation Phase

In this phase, some standards are set to evaluate the performance of the pattern which is trained in detection phase. In the research domain of intrusion detection system, performance is referred to four items, which are accuracy rate, false alarm rate, detection rate and reaction time. The first three standards can be calculated by confusion matrix and the last one can measure by experiments. Confusion matrix is commonly used to evaluate the performance in the domain of classification or prediction, and it is shown in Table 1. If the intrusion detection system belongs to anomaly detection, which means the data can just categorize into normal data and attacked data, confusion matrix is widely used to be the method of evaluation. In the confusion matrix, true positive, false negative, false positive and true negative can be gotten after the detection phase. According to Table 1, true positive (TP) means prediction result is normal when input data is normal, false negative (FN) means prediction result is attacked when input data is normal, false positive (FP) means prediction result is normal when input data is attacked, and false negative (FN) means prediction result is attacked when input data is attacked. Confusion matrix is used to calculate the number of different detection results, and accuracy rate, false alarm rate, detection rate can be computed from it. Accuracy rate is defined as the percentage of the correct classification or prediction, false alarm rate is the percentage of incorrect classification and detection rate is the percentage which the attacked data is correctly detected. These three items are the main standards that are useful on evaluating the performance of an intrusion detection system many researches will compare their performance by using confusion matrix. Except confusion matrix, reaction time is also an important parameter. Reaction can be defined as a time period between the generation of attack and the detection of attack. With shorter time, the intrusion detection system has better performance because the damage of the attack does not spread too widely. It is said that a trained pattern, which is used to detect attack, has better performance when the accuracy rate and detection rate are high and false alarm rate and reaction time are low.

Table 1: Confusion matrix

	Predicted	
	Normal	Attacked
Normal data	TP	FN
Attacked data	FP	TN

2.4 Detection Techniques

In this section, two techniques are discussed, which are firewall and intrusion detection system. In the research domain of intrusion detection system, intrusion detection system is commonly used to assist detection attacks from botnet. According to the detection methodology mention in Section 2.3, a framework of detection contains sniffing, analysis and report and it can rely on the operation of intrusion detection system. Among all kinds of detection technique, intrusion detection system plays an essential role on the detection and defending attacks from botnet. In the following sections, Section 2.4.1 is the description of firewall and intrusion detection system is discussed in Section 2.4.2.

2.4.1 Firewall

In order to enhance the security of a server or personal computer, firewall is commonly used to prevent attacks from botnet by separating normal user form attacks. It is a device that can prevent malicious attacks entering the protected unit by setting different network zone and the rules that control the access in and out flow. According to the type of firewall, it can be categorized into three types, which are network level firewall, application level firewall and proxy firewall. Through different level of protection, firewall can be used as a packet filter or a monitor of application. Despite some attacks or virus can be detected by firewall, it also exists some vulnerabilities. First, firewall may have vulnerabilities due to their protection rules that are set before, and it makes the firewall cannot detect all kinds of attack that has never seen. The attack model of attacks makes changes day by day and firewall becomes weaker. Another vulnerability is that some attacks happen when the attacks are generated in the local network because firewall can just filter the packet from WAN. The architecture of firewall is shown in Figure 12. As shown in Figure 12, firewall is installed between WAN and local area network and it is used to prevent attacks from WAN. Despite firewall is useful on detecting attacks form Internet, some vulnerabilities still exist and need to be solved.

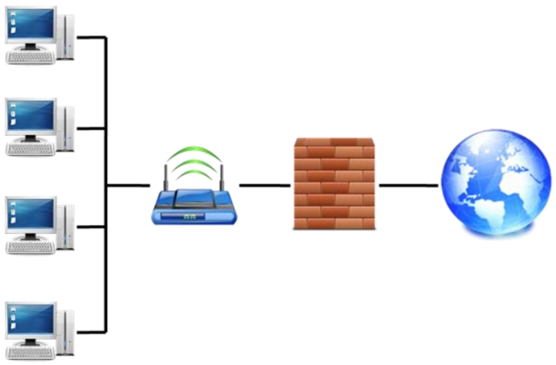

Figure 12: Firewall

2.4.1 Intrusion Detection System

Intrusion detection system or intrusion prevention system is used to improve information security further more. Intrusion detection system is installed between firewall and personal computer or server, and it is shown in Figure 13. When it detects the attacks, it will alarm the user that the attack is undergoing. For intrusion prevention system, it can solve the attacks in the certain method that is set previously when the victim encounters with the attacks. An intrusion detection system or an intrusion prevention system has three core parts, which are packet sniffing, detection and reporting, which the process is similar to the methodology that is used to detect attacks by the researchers. By monitoring and analyzing packet data, the detection is more cautious than firewall and it can reinforce the security.

According to the method of detection, an intrusion detection system can be categorized into anomaly based detection and misuse based detection. Anomaly based detection is that once normal pattern is constructed, the data will be classified into attacked data if it does not belong to normal data. Anomaly based detection just has two types of result, which are normal and attacked data. Compared with anomaly based detection, misuse based detection detects attacks in different ways. Before misuse based detection, the detection target is already set. For instance, if the targets of detection system are SYN flood, UDP flood and HTTP Get flood, the detection system must have the attack pattern of these attacks. After the detection, the known packet flow can be classified into certain one category which has been set before. Different kinds of intrusion detection system have different advantages and disadvantages [28, 29]. For anomaly based detection, it can detect known attacks and it means it has better flexibility. Anomaly based detection often has high false alarm rate due to the changeful attack patterns. Compared with misuse based detection, its accuracy rate is oppositely low. For misuse based detection, the accuracy is high because the attack pattern is set in the system previously. Misuse detection has low flexibility because it can just detect the known attacks that are previously set in the detection system. It can be improved by updating the attack patterns continuously to make the intrusion detection system can recognize all kinds of attack. There is also another category of intrusion detection system, and it is classified by the method of collecting packet data. Three types are most common, which are network based, host based and application based. Network based IDS analyzes real time packet data, host based IDS analyzes log files and application based IDS analyzes application logs. Network based IDS belongs to real time intrusion detection system and host based IDS belongs to post analysis. Three kinds of IDS are used for different purpose. Network based IDS can detect attacks which is used to detect real time attack and for host based IDS, it is used to analyze the log file. Because host based IDS and application based IDS does not have time urgency, the researchers can analyze the packet data in a detailed way. After the analysis of host based IDS, the result can assist network based IDS or misuse based intrusion detection system to detect attacks.

3 Conclusions and Future Work

In this paper, several issues, which are closely related to botnet, are discussed. First, the architecture of botnet is described. There are four common architectures of botnet, which are star topology, multi-server, hierarchical and random botnet. DDoS and spam are two serious problems that are caused by botnet. DDoS can be categorized into bandwidth consumption and computation resource consumption, and several types of distributed denial of service, which attacks by sending different packets, are discussed. In order to detect attacks form botnet, a methodology is proposed and several researchers are concentrated on enhance their performances by modifying the analysis methodology. Finally, some techniques, that are helpful on detecting attacks, are discussed and intrusion detection system is described in detailed.

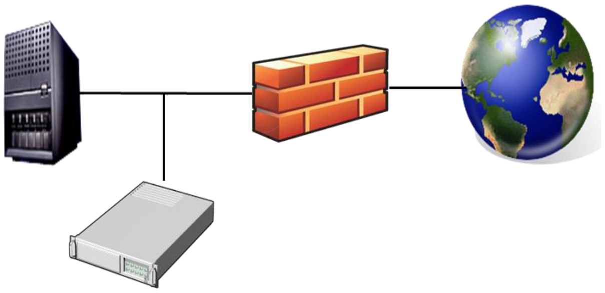

Figure 13: Intrusion detection system

To decrease the damage from botnet, the attacks need to be detected in an efficient way. In order to propose an intrusion detection system that has high performance, several issues are important, which are setting the target of attack, feature selection, algorithms of detection and the methods of evaluation. In the future work, feature selection and classification will be the main part of research. In several kinds of distributed denial of service, SYN flood and UDP flood are more serious and more common than other kinds of attack. To enhance the performance of the designed IDS framework, the target should not be too wide. If the detection target is set, the feature selection and the algorithm of classification or prediction will influence the performance of the detection phase. How to design a framework, which can collect packet and analyze it in an efficiency way, is an important issue that needs to be solved.

References

[1] Kotikalapudi, R. and Sriram, C., "Associating Internet Usage with Depressive Behavior Among College Students", IEEE Technology and Society Magazine, 2012, vol.31, pp.73-80

[2] Joerg, K., Andrea and G. K., "Consumer acceptance of the mobile Internet", MARKETING LETTERS, 2012, vol. 23, pp. 917-928

[3] Feily, M., "A Survey of Botnet and Botnet Detection", Third Internaiotnal Conference On Emerging Security Information, Systems and Technologies, 2009, pp. 268-273

[4] Baker, W.H., "Is Information Security Under Control?: Investigating Quality in Information Security Management", Security & Privacy, 2007, vol.5, pp. 36-44

[5] Hwang, S.Y. and Lee,C.H., "Reliable Web service selection in choreographed environments", Decision Support Systems, 2013, vol.54, pp. 4796-1476

[6] Plohmann, D. and Elmar, G. P., "Case Study of the Miner Botnet", International Conference on Cyber Conflict, 2012, pp. 1-16

[7] Sun, W., "The Botnet Defense and Control", 2011 International Conference on Information Technology, Computer Engineering and Management Sciences, 2011, vol.4, pp. 339-342

[8] Zang, L., "A Survey on Latest Botnet Attack and Defense", 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, 2011, pp. 53-60

[9] Zargar, S., "A Survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks", Communications Surveys & Tutorials, pp. 1-24

[10] Ma, X. and Guan, X.,"A Novel IRC Botnet Detection Method Based on Packet Size Sequence", 2010 IEEE International Conference on Communications, 2010, pp. 1-5

[11] Wang, Z. and Li, F., "The Research of Detecting IRe Botnet Based on K- means Algorithms", 2010 Second International Conference on Communication Systems, Networks and Applications, 2010, vol. 1, pp. 208-210

[12] Wang, Z., "The Detection of IRC Botnet Based on Abnormal Behavior", 2010 Second International Conference on Multimedia and Information Technology, 2010, vol.2, pp. 146-149

[13] Mazzariello, C., "IRC Traffic Analysis for Botnet Detection", Fourth International Conference on Information Assurance and Security, 2008, pp. 318-323

[14] Wang, Y., Jin, Z. and Zhang, W., "Analysis of Botnet Attack and Defense Technology", 2011 International Conference on Computer Science and Service System, 2011, pp. 3021-3023

[15] Zhu, Z., Lu, G., Chen, Y., Roberts, P. and Han, K., "Botnet Research Survey", 32nd Annual IEEE International Conference on Computer Software and Applications, 2008, pp. 967-972

[16] Dittrich, D., "P2P as botnet command and control: A deeper insight", 3rd International Conference on Malicious and Unwanted Software, 2008, pp. 41-48

[17] Wei, W., Chen, F., Xia, Y. and Jin, G., "A Rank Correlation Based Detection against Distributed Reflection DoS Attacks", Communications Letters, 2013, vol.17, pp. 173-175

[18] Zhang, C. W., Cai, C. P., Chen, W. F., Luo, X. and Yin, J., "Flow level detection and filtering of low-rate DDoS", COMPUTER NETWORKS, 2012, vol.56, pp. 3417-3431

[19] Yu, S., Zhou, W., Doss, R. amd Jia, W., "Traceback of DDoS Attacks Using Entropy Variations", IEEE Transactions on Parallel and Distributed Systems, 2011, vol. 22, pp. 412-425

[20] Haris, S.H.C., "TCP SYN flood detection based on payload analysis", IEEE Student Conference on Research and Development, 2010, pp. 149-153

[21] Lau, F., Rubin, S. H., Smith, M. H. and Trajkovic, L., "Distributed denial of service attacks", 2000 IEEE International Conference on Systems, Man, and Cybernetics, 2000, vol. 3, pp. 2275-2280

[22] Dhinakaran, C. and Lee, J. K., "An Empirical Study of Spam and Spam Vulnerable email Accounts", Future Generation Communication and Networking, 2007, vol. 1, pp. 408-413

[23] Xia, H., Fu, Y., Zhou, J. and Xia, Q., "Intelligent spam filtering for massive short message stream", The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 2013, vol. 32, pp. 586-596

[24] Rahmani, H., Sahli, N. and Kamoun, F., "DDoS flooding attack detection scheme based on F-divergence", Computer Communications, vol.35, pp. 1380-1391

[25] Casas, P., Mazel, J. and Owezarski, P., "Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge", Computer Communications, 2012, vol.35, pp. 772-783

[26] Li, M. H. and Li, M., "An Adaptive Approach for Defending against DDoS Attacks", MATHEMATICAL PROBLEMS IN ENGINEERING, 2010, vol. 2010, pp. 1-15

[27] Stevanovic, D., Vlajic, N. and An, A. J., "Detection of malicious and non-malicious website visitors using unsupervised neural network learning", APPLIED SOFT COMPUTING, vol. 13, pp. 698-708

[28] Modi, C., Patel, D., Borisaniya, B., Patel, H., Pater, A. and Rajarajan, M., "A survey of intrusion detection techniques in Cloud", Journal of Network and Computer Applications, 2013, vol.36, pp.42-57

[29] Zaman, S. and Karray, F., "Lightweight IDS Based on Features Selection and IDS Classification Scheme", International Conference on Computational Science and Engineering, 2009, vol. 3, pp. 365-370

Cheng-Yi Liu is a lecturer in the Department of Information Management at TransWorld University. He received a master’s degree in Business Administration from the National Changhua University of Education in 2001. His current research interests include soft-computing, neural network, and information security.

Chih-Hsiang Peng born in Hsinchu County, Taiwan, in 1988. He received the B.M. degree from National Chung Hsing University (NCHU), Taichung, in 2011 in management information systems. He is currently pursuing the M.S. degree with the Department of Management Information Systems. His research is concentrated on the topic of network security, especially in the intrusion detection system.

Iuon-Chang Lin received the Ph.D. in Computer Science and Information Engineering in March 2004 from National Chung Cheng University, Chiayi, Taiwan. He is currently a professor of the Department of Management Information Systems, National Chung Hsing University, Taichung, Taiwan. His current research interests include electronic commerce, information security, cryptography, and cloud computing.