Characters and Evaluation of P2P Botnet Node-based Detection

L. Wang
Langfang Teachers University
Langfang, 065000, China
liyagggwang@163.com

J. Shang
Langfang Campus of Nanjing Artillery Academy
Langfang, 065000, China

ABSTRACT: *In this paper, we proposed a novel node-based P2P detection. Comparing to traditional server-client bonnet on the Internet, the P2P (peer-to-peer) bonnet has capabilities to realize highly scalable, extensible and efficient distributed applications. The node-based P2P detection exploits the node profile generated from the novel as well as the degradation of the amount of traffic handled with sampling. It is expected to grow the detection rate. With these commonality features, flow-based techniques can institute rules for multiple bonnets detection, as well as for some unknown bot nets. The disadvantage of this method is that some legitimate applications may share the same flow features. This could be expected to result in a high false positive rate. Node-based detection extracts more general features of a bonnet. One node represents one bot machine. This technique detects bonnets from a macroscopic angle. We hope that it would help people find useful information quickly.*

Keywords: P2P Botnet Node, Characters, Evaluation, Detection.

Received: 4 Februar 2017, Revised 4 March 2017, Revised 11 March 2017

1. Introduction

Depending on our knowledge, there has been no research published regarding the application of the node-based bot detection [1]. The node-based bot detection is an effective and high-efficiency method in finding bots [2]. It stands at a higher level than both flow-based and packet-based detections. We expect that the node-based detection can result in better performance. Meanwhile, node-based detection has adaptability, since it is sensitive to new behaviors from bots implementing highly varied protocols. It is extremely important and necessary to design a system that can evaluate the performance of the detection online, instead of offline. It is equally important to train the detection system online, instead of an offline training process. Such a system is suitable for identifying new threats.

2. The characters of P2P botnet node-based detection

Botnet research in recent years is gradually beginning from anti-virus companies to academic and research institutions have done related research. The first research and response Botnet is anti-virus vendors. They depart from the malice of the bot program, which was seen as a tool by the back door, worms, Spyware and other malicious software technologies included in the scope of killing the virus. The famous major anti-virus vendors are several important bot program is written to the virus signature database. Symantec since 2004, in its semi-annual security trend analysis report, a separate chapter is given to the Botnet activity observations. Kaspersky also malware trend analysis report that the prevalence of bots is the 2004 the most significant changes in the field of virus.

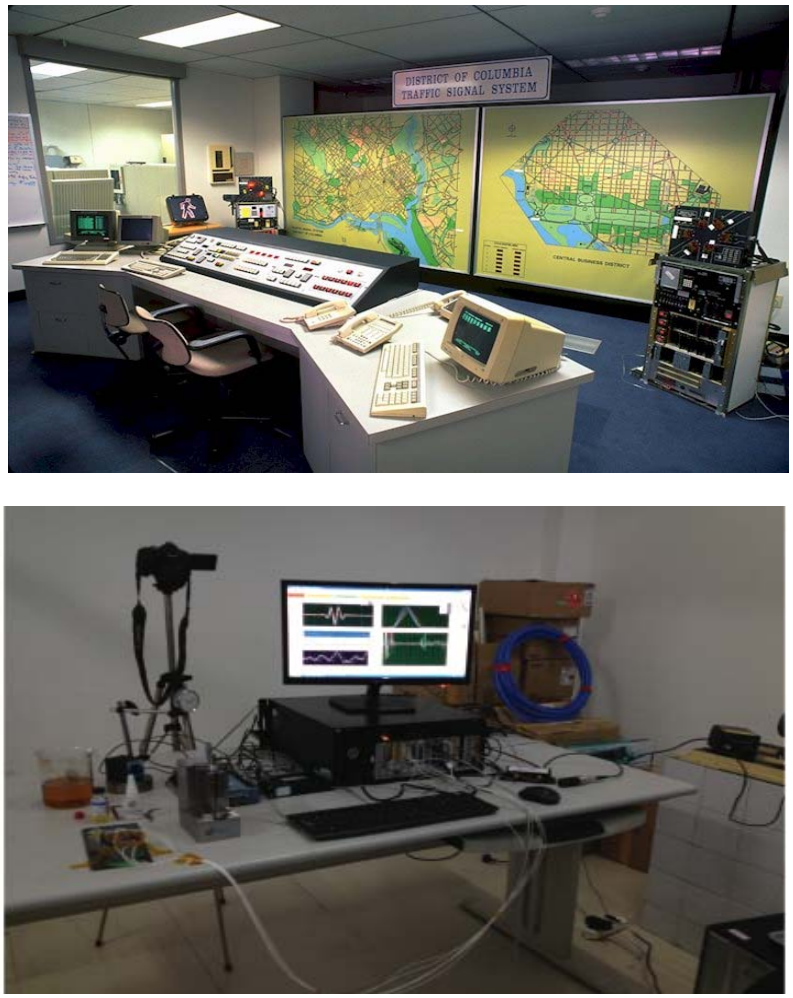

Figure 1. Actual Characters and Evaluation way

Academia began to place emphasis on the development of the Botnet. Some members of some of the Honeynet Project and Honeynet Research Alliance of the international use of the Honeynet analysis techniques Botnet Activities depth tracking and analysis, such as Azusa Pacific University, Bill McCarty, French Honeynet Project of Richard Clarke, University of Washington and the German Honeynet project. Germany in particular Honeynet Project in November 2004 to January 2005 and found nearly 100 Botnet tracked and released Botnet tracking technical report by deploying Win32 honeypots machine. Barnet is a major threat as a platform to attack the specified target launch DDos (Distributed Denial of Service) attacks, so DDos researchers also did research work on the Botnet. Organized by foreign DDosVax “Detecting Bots in Internet Relay Chat Systems” project, the analysis of the behavior characteristics based IRC protocol bot program in selected network traffic Optional correspondence, to detect the presence of Botnet. The methods of the organization set up by a Botnet experimental environment in for testing, the data obtained through statistical analysis, can effectively verify the results of the analysis on characteristics of Botnet traffic, but there are a few false positives.

China in 2005 started Botnet preliminary research. Peking University Institute of Computer Science and Technology in January 2005 began with the Honeynet project tracking Botnet of malware samples collected, using a sandbox, Honeynet both have their own advantages in its technical analysis, confirm whether it is the bot, and the bot Botnet control channel information to be connected to extract ultimately received more than 60,000 bot sample analysis reports, and more than 500 of which are still active Botnet tracking statistics View national distribution, size distribution information. These activities are described and the data. China's domestic online Botnet threat more seriously, need to attract Internet user attention. Malicious code research team began work in July 2005 for Botnet, through the large number of already know Botnet actual tracking and in-depth analysis, based on Botnet IRC protocol server side features are classified extraction form for Botnet judge rules on the server-side, so that the network can be the nature of the IRC server discrimination. Preliminary design and implement a Botnet automatic identification system, used in Chinese education and research computer network environment. As can be seen, from domestic to foreign countries since 2004 Botnet studies more and more attention to network security research personnel, research work has been greatly enhanced. But the work is not enough. In terms of detection and disposal Botnet there is much work to be done.

There are several key words in the concept of Botnet. "Bot" is the abbreviation of robot, is to implement the control function of the malicious code; "zombie computer" is the implantation of BOT computer; the control server (Control Server) "refers to the central server control and communication, based on IRC (Internet Relay Chat) control protocol in Botnet, that is to provide services IRC chat server. Barnet is a kind of malicious Internet behavior driven by the engine: DDoS attack is the use of service requests to exhaust system resources of the attack network, so that the network cannot handle the request of legitimate users. Duse attacks have many forms, but can see the most typical is the flow overflow, which can consume a large amount of bandwidth, but do not consume application resources. Duse attacks are not new. Over the past ten years, with the rise of the Botnet, it has gained rapid growth and widespread application. Barnet provides the required "firepower" bandwidth and computer infrastructure for DDoS attacks.

A Botnet first need is charged with a certain scale computer, and this scale is gradually with the introduction of some kind of diffusion or certain kinds of means of communication and the formation of bot program, there are several methods in this communication process: active vulnerabilities. The principle is to achieve by attacking the system loopholes access rights, and program execution bot code injection, will be to attack the system becomes infected with bots. The most basic way of infection belongs to this class is that the attacker manually using a series of hacking tools and scripts to attack, after obtaining permission to download bot program execution. An attacker will also bots and worms combine technology, so that the bot program can automatically spread, the famous bot sample, is realized automatically propagated bot program. E-mail viruses. bot program also itself, manifests itself by sending a large number of messages carrying virus in e-mail attachments and contains a link to download bot bot program executed in the message content and execution through a series of social engineering techniques to lure recipients click on the link or attachment, or by using a mail client vulnerabilities automatically, so that the receiver host become infected bots. Use instant messaging software to send links to execute bots friends list friends, and through social engineering techniques to trick their clicks, thus infected, MSN sexy chicken outbreak in early 2005, as used in this the approach. Attacker to provide Web services binding sites on the HTML page malicious script when visitors visit these sites will execute the malicious script, so bot program is downloaded to the host, and automatically executed. Disguised as useful software provided in the website, FTP server, P2P networks, trick users to download and run. Through the above numerous means of communication it can be seen in the spread formation Botnet the way viruses and worms and spyware complex functions very similar.

Adding phase, each host will be infected with a hidden attack on their bot program and added to the Botnet go, added a manner depending on the control and communication protocols varies. Botnet-based IRC protocol, the host infected with BOT programs will log on to the specified server and channel to go after the login succeeds, wait for the controller to send malicious commands in the channel.

In the control phase, the attacker sends a predefined control commands through a central server, so that the infected host to perform malicious acts, such as the launch DDos attacks, theft of sensitive information, the host, upgrade of other malicious programs. Figure 1 is observed in the control phase inward net spreads malicious programs Botnet behavior.

Different from other Internet malware, Botnet has its own unique characteristic, namely its control communication network [3]. Usually, the Botnet consists of a network of compromised computers controlled by a bot-master and has a large scale of the Internet. The disadvantage of C&C server is that it can be easily shut down or blocked by firewall once it has been aware by its victim. It depends on any computer on the system (P2P). Each computer can be act as a client or server to any other computer

in P2P network. If P2P bot program uses a fixed port, a bot can be identified by detecting specific features. However, most of the current bots change ports dynamically. Besides, some bots could use the normal ports such as port 80 to communicate, cheating the IDS. Thus, bot detection based on ports is infeasible.

The bot program may use length-fixed packets [4]. This feature can be utilized to detect bots. However, some normal applications may have the same packet length. Thus, bot detection which based on the packet length could cause problem. In addition, different bots usually have separate payloads. The unique sequence in a bot's payload can be extracted as the feature sequence of the bot. However, this is only beneficial to the bots that are known. It also has a lower generality. To resolve the problem, the concept of flow is introduced, which is the set of data packets with the same attributions [5]. The same attributions usually satisfy a tetrad property, which means they have the exact same source address and port, the same destination address and port. Some researchers thought it should be a quintuple, including the protocol. Nonetheless, the quintuple is not suitable because of the inability to identify an unknown flow protocol. Flow-based P2P bot detection is heuristic and intelligent. It has the capacity of detecting unknown bot nets.

The features extracted from flows are more general than those extracted from packets. Besides the usual properties of packets, the features of flows also include the number of the packets, the order of the packet arrives, the order of the interval between packets, the flow speed and the flow lasting time. Since processing a flow has less time than processing every packet in the flow. This makes that the detection based on flows have a higher efficiency than detection based on packets does [6]. With the chosen properties, data mining classification methods can be tailored to extract features and classify. Thus, this method can be used to detect alien bots.

Detection on P2P Botnet is difficult as it has no focal point (the C&C server) [7]. Any host connected to P2P Network can act as a C&C server. Once the postmaster obtains a list of hosts connected to a P2P network, he can control every host as he wishes. Although some computers are blocked by the firewall, once a bot is connected to at least one bot in another computer, it can receive any command indirectly from the postmaster through another computer.

The protection concept of detecting potential threat for the enormous scale of malicious software would be of strategic significance since such threats are serious and threatening. Bonnets, networks of malware-infected machines (bots) controlled by postmasters, usually carry out their nefarious tasks, such as sending spam, launching denial of service attacks, and even stealing personal data. Thus, how to detect bonnets and remove them has become an interesting and important problem in network security. Baptists also have a variety of types, including P2P bonnets, IRC bot nets, and HTTP bot nets, and so on. P2P bot is distributed [8]. If it can be detected in a reasonable time, network security can be improved significantly.

Signature-based P2P bot detection is traditional and deterministic [9]. It belongs to low level packet-based detection. Under a background of large data communication, detection based on packets generally has low process efficiency and a real bad time attribution, since it needs to process every packet in the flow. Furthermore, because the information contained in every packet is limited, an unknown bot cannot be recognised by this method.

The current methodology of signature-based detection mainly focuses on detecting a peculiar feature, for example, a specific port or a specific feature sequence in the payload. If the feature exists, the related source and destination addresses are stored into the bot dataset [10]. However, different bot program has changed communication protocols, different packet length, and different flow rate. Some of the bots even have their encryption mechanism to be protected. For a single packet, the features it can provide mainly include its source address and port, its destination address and port, the packet length and payload. As a result, it can be able to detect one bot at a time. 2 implementing green development ideas, and create new economic growth pole Eighteen major report proposed to focus on promoting green development, cycle development, low-carbon development "three major development", stressed the construction of ecological civilization in a prominent position, into economic development, promote intensive and efficient production space, living space livable moderate, efforts to build a beautiful China to achieve sustainable development of the Chinese nation. Promote green development, cycle development, low-carbon development is essentially asked us to promote the transformation of economic development, urban-friendly direction resource-saving and environment, and strive to achieve "low resource consumption, less environmental pollution, high value-added products, the production side. Formula Smart City has become the information element type intensive economic development." Smart City has become the decisive element in the information endogenous variables of economic development, not only to fully tap the potential of intelligent people and the social and material resource potential and achieve personal behavior and optimization of organizational decision-making, and can be infinitely replicated and repeated use, but not additional costs, not to cause

environmental pollution, promote green economic growth. Meanwhile, the construction of smart city based on things, 3S (GIS, GPS, RS) and cloud computing as an important core technology, and their applications will continue to grow a new generation of information technology industries and creative industries, software and information services and other emerging industries new technology development groups, as well as intelligence, biology, nanotechnology and other gathering is conducive to the development of strategic new industries, create new economic growth point. Among cloud computing. Example, market size of cloud computing the next few years will reach an annual growth rate of about ninety percent.

3. Evaluation indexes

In order to evaluate the performance of a botnet detection technique, we need to introduce a quantitative measurement. In our detection technique, we basically classify the network traffic data into normal or anomalous/suspicious groups. Any deviation from the normal traffic pattern is considered as suspicious. Hence we need to define true positive (TP), true negative (TN), false positive (FP) and false negative (FN) to determine true positive rate (TPR) and false positive rate (FPR). The Table 4 defines TP, FP, TN and FN.

Now, the true positive rate (TPR) which is also known as sensitivity and the false positive rate (FPR) can be calculated using the following equations.

$$DR = TPR = \frac{TP}{TP + FN} \quad (1)$$

$$FPR = \frac{FP}{FP + TN} \quad (2)$$

$$\Pr ecision = \frac{TP}{TP + FP} \quad (3)$$

The true positive rate (TPR) evaluates the performance of a botnet detection technique in terms of the probability of a suspicious data reported correctly as anomalous. In other words it evaluates how well the model detected anomalous packets. On the other hand the false positive rate (FPR) evaluates the performance of botnet detection technique in terms of the probability of a normal traffic reported as suspicious generating false alarms.

Some related research on detection performance uses precision as the performance measurement. However, there is no research on the correlation between the detection rate (DR) and the precision. It can be seen from the following proof that the trend of FPR (i.e. DR) can be reached by precision.

Besides, both the detection rate and the precision have the equivalent importance in the detection system. Thus, we proposed a combination of the two measurements, called Comprehensive Evaluation Index (CEI), which has a strategic significance for the evaluation of detection performance.

$$CEI = DR * 50\% + \Pr ecision * 50\% \quad (4)$$

4. Conclusions

This paper analyzed the generality and detection rate of different detection methods. In summary, flow-based detection generalizes the commonality features of flows via the analysis of many known botnets. With these commonality features, flow-based techniques can institute rules for multiple botnets detection, as well as for some unknown botnets. The disadvantage of this method is that some legal applications may share the same flow features. This could result in a high false positive rate.The node-based detection extracts more general features of a botnet. One node represents one bot machine. This technique detects botnets from a macroscopic angle.

References

[1] Crotti, M. (2006). A Statistical Approach to IP-level Classification of Network Traffic. *Istanbul*. 22 170-176

[2] Stone-Gross, B. (2011). Analysis of A Botnet Takeover. *IEEE Security & Privacy*. 9 64-72.

[3] Romana, D. A. L. (2007). Detection of Bot Worm-Infected PC Terminals. *Information-an International Interdisciplinary Journal*. 10. 673-686.

[4] Braun, Lothar., Munz, G., Carle, Georg. (2010). Packet Sampling for Worm and Botnet Detection in TCP Connections. Network *Operations and Management Symposium*. 33 264-271.

[5] Wang, S. (2010). Method of Choosing Optimal Characters for Network Intrusion Detection System. Computer Engineering. 36. 140-144.

[6] Wang, X.-L. (2011). Research of Automatically Generating Signatures for Botnets. *Journal of Beijing University of Posts and Telecommunications*. 34 109-112.

[7] Gu, G., Zhang, J., Lee, W. (2008). BotSniffer: Detecting botnet command and control channels in network traffic. 4 (8) 45-49.

[8] Karasaridis, A., Rexroad, B., Hoeflin, D A. (2007). Wide-Scale Botnet Detection and Characterization. HotBots. 7. 7-7.

[9] Wang, P., Sparks, S., Zou, C C. (2010). An Advanced Hybrid Peer-to-Peer Botnet. *IEEE Transactions on Dependable and Secure Computing*. 7 (2) 113.

[10] Ramachandran, A., Feamster, N., Dagon, D. (2006). Revealing Botnet Membership Using DNSBL Counter-Intelligence. SRUTI. 6 49-54.