

A Survey of Botnet : Consequences, Defenses and Challenges

Yun-Ho Shin and Eul-Gyu Im

Department of Electronics and Computer Engineering, Hanyang University,
HaengDang-Dong, SungDong-Gu, Seoul, 133-791, Korea
{yhs, imeg}@hanyang.ac.kr

Abstract. As technology has been developed, the network of bot, botnet, has been huge matter in computer science society. Most botnet causes network security threats and they are based on C&C server such as IRC, HTTP common protocol [1] and recently botnet also constructs P2P connection and the bot's characteristics and activities are all different according to the structure of botnet. That is why the existed research is numerous, too, and it is beneficial to categorize and to classify defense mechanism of bot. The bot activities result in a lot of negative effects such as DDoS (Distributed Denial of Service) and Spamming. The mechanisms for bot detection and defenses can be categorized into C&C based bot detection and P2P based bot detection. Besides, these mechanisms recently can be combined algorithms from different studies. Several DDoS defense methods also will be introduced in this paper.

Keywords: botnet, bot detection, P2P bot, C&C bot

1 Introduction

As technology has been developed rapidly, global Internet threats are also increasing, especially, one of the threats is caused by botnet activities. It makes both attackers and defenders concentrate on them. The botnet activities will be stressed next section detail also botnet structure, too. The botnet activities conduct spamming, DDoS (Distributed Denial of Services) attacks and scanning [2]. A botmaster compromises vulnerable hosts as bots which mainly operates malicious operation. Then if vulnerable hosts are infected, in C&C structure, the botmaster can communicate with infected bots through C&C (command and control channel). Using this C&C, botmaster enables execution of distributed attacks remotely. Contrarily, a bot using P2P protocol can communicate with other bots without botmaster's authorization.

In this paper, the basis of botnet's knowledge is shown as section 2. It consists of several parts: negative consequences of botnet activities, the types of topology of botnet, e.g., IRC-based, HTTP-based and P2P botnet. In section 3, we will introduce several recent techniques for bot detection and defense. Our summary is provided with future challenges of botnet defense methods and future work.

2 Basic knowledge of Botnet

2.1 Basis of bot and botnet scenario

Malicious Software has two types of program fragments [3]. First type is a program which is dependent on a host program such as backdoors, logic bombs, Trojan horse and viruses. Another type is independent type, e.g., worms and bots which wait for attacker's attack command, e.g., zombie host [4]. The network bots connected together are also called botnet, which cause launching DDoS attacks or spreading spam mail [4, 5, 6] as mentioned before. To achieve their purpose, each bot can be installed with various methods to destroy victim host [14]. Because the number of bots is increasing dramatically, it is difficult to detect and prohibit the illegal results from botnet. That is, it is useful to perceive how bots are connected and how bots act. According to [7], propagation, communication and attack could be process of botnet system. Firstly, botmaster probe vulnerable hosts to compromise and makes them as bots, and then compromised hosts propagate themselves, finding vulnerable hosts and installing bot software. After propagation, bots organize C&C (command and control channel) to communicate each other and to make clear whether connected bot still is available to attack or not, especially on P2P botnet system. In other case, IRC, HTTP-based bots and botmaster communicate with C&C to launch attack instructions. After that, a botnet which is forced to attack by master can be on the way of attack phase. The types of attacks can be different by attacker's purpose.

2.2 The topologies on which botnet depends

Centralized C&C structure – The centralized architecture designed from original chat rooms can be categorized into IRC-based C&C and HTTP based C&C [1] shown as figure 1. A botmaster or an attacker commands to bot using central server, and all bots have connection with C&C server. C&C channel's primary aim is to make bot armed [20]. Most of these bots' act is keylogging the victim host and it can be serious threat by presenting availability that bots collect privacy information from victim host in [14]. Bailey *et al.* [7] asserts some weakness of C&C based botnet. In [7], since each end hop is located on the 'end point', the probability of exposure of a bot can be increased, and it is easy to notify central server and to compromise it. When central server is compromised, millions of nodes connected to the server can fail.

P2P Structure – This approach is more flexible than C&C structure although [19], and P2P botnet organizes several equal peers. According to [8], sometimes, attacker uses distributed P2P protocol to avoid topologies of centralized botnet structure due to weak points of C&C. Instead of connecting with one central server, P2P botnet has connection with each equal peer bot. Once the botmaster transmits attack commands to a bot, it spreads commands from master to all peer bot connected to itself and launch to command together. Furthermore, P2P channel includes not only botmaster's attack command, but also bots' chatting response to verify whether connected bot is alive or not. For this reason, response from P2P bot can be significantly important to measure degree of botnet consequences in many studies.

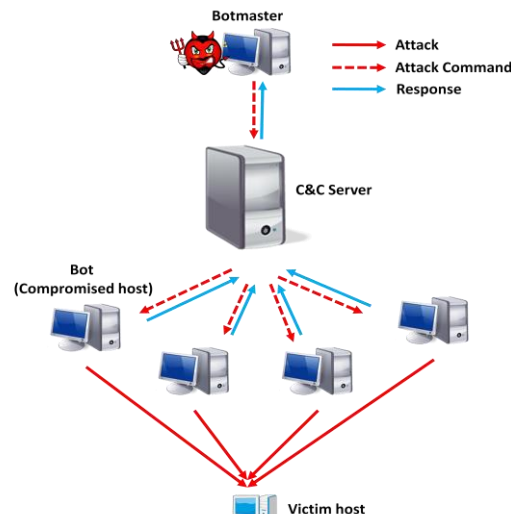


Fig. 1. IRC, HTTP-based centralized C&C botnet architecture

2.3 Consequences

In this section, the consequences caused from botnet can be categorized into DDoS attack and distributing spam mail. These negative effects will be considered briefly.

Spamming – Even botmaster cannot be aware of the number of bots because a number of bots are spread over the World. In 2008, the largest number of bots controlled by one botmaster was approximately more than 600,000, and it was bots not even calculated completely. These numerous bots are used in order to spread spam messages. Although recent e-mail servers have black list system to filter out spam mails, it does not work perfectly because botmaster can use a few other bots when botmaster's bots is filtered out by blacklist. This indicates that more recent vulnerable host have been compromised by bot agent recently than former days. Why spam is dangerous heads that since it contains malicious code or program, malware can cause other negative effects. Recently, when a user opens the spam mail, the log is updated and it is sent to botmaster so that he/she can be acknowledge that the email account to which the bot sent the spam message is activated and the activated mail can be used to compromise the victim. Our computer has been on the threats every day by this simple method.

DDoS(Denial of Service) – A DDoS attacks make destination servers waste their resources and make routers around the destination server jam since numerous bots send lots of data packages to destination host in [9]. Besides, due to the growth of a botnet organization, a kind of DDoS attacks is increasing and it can be divided into various features. First, it dependins on the searching method during the automation phase which comes before the DDoS attacks. The automation phase means the state

that attackers attempt to compromise vulnerable host and infect the hosts. In this phase, worms named *Nimda* or *Code Red II* use local subnet searching method to infect vulnerable hosts. Second, type of *IP spoofing* attacks is the measure to decide DDoS attacks. Direct attack and reflection attack are two types of *IP spoofing* attacks. Thirdly, kinds of exhausted resources, basically internal host resource and network data transmission bandwidth, also can be considered. *SYN flood* attack is one of the attacks caused by consumption of resources through following steps. The botmaster controls bots which send the packets with spoofed source IP address packets to host. The host tries to send SYN packet to the source, which the wrong IP address points. In the end, victim host is going to be shutdown because of the overhead of processing inaccurate source IP. Another attack method which exhausts targets' resources is to destroy host's availability to transmit data. In this attack, because of the amount of packet from wild botnet hops, victim host will lost its ability to transmit data or packets. Finally there is a method using vulnerability of system. If bots do not complete session with victim, resource can be exhausted.

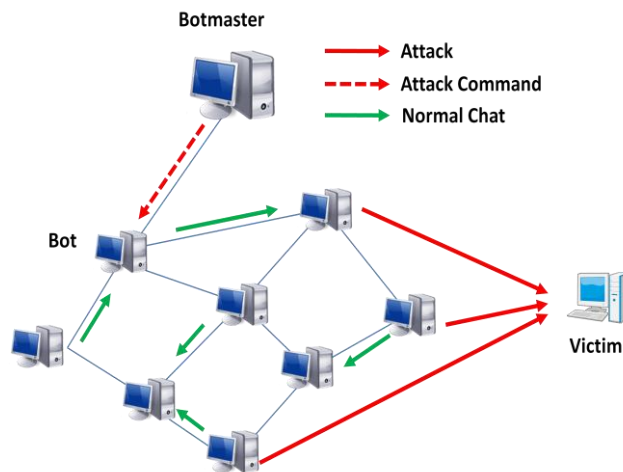


Fig. 2. P2P-based Botnet architecture

In a different way, traffic flooding attack sending a large amount of packets also can be measure to characterize DDoS attack. As mentioned before, DDoS attack can be characterized through many measurements. It can be the reason why DDoS attack is dealt with many studies around the world.

3 Defense Methods

In this section, a number of defense techniques, including C&C-based botnet detection approach and P2P-based detection approach and DDoS defense, will be introduced.

3.1 Botnet Detection

3.1.1 C&C-based bot defense

According to [10, 11], The Machine learning techniques are used to make sure and classify botnet traffic. The main thesis is that packet flows are consisted of IRC and non-IRC traffic and IRC traffic can also be divided into authentic traffic and bot flows, which bots use to communicate other bots through proposed 2 steps in [10]. For stage 1 and 2, Naïve Bayes, J45 and Bayesian networks is used to classify suspects. J45 is a decision tree in WEKA algorithm and, Naïve Bayes classifier is used to measure independence among the features of traffic, also Bayesian networks uses directed acyclic graph to capture the dependence. In this research, to classify IRC traffic, Naïve Bayes performed best in their test-bed and to be clear, more training sets approximately more than 10k flows were sufficient enough to get desired result. J48 and Bayesian networks were poor classifier in this proposed test-bed with IRC flows. In stage 2 to identify bot traffic, although Bayesian network classifier performed best, challenges existed absence of clear dataset. To solve this problem, they used regenerated virus set from Subseven Trojan. Another Machine learning algorithm, SVM, supports vector machine introduced by V. N. Vapnik in 1979, which was operated in [11], having beneficial performance than other Machine learning algorithms, which are Naïve Bayes and k-Nearest Neighbor(k-NN). Originally, the SVM algorithm has been used to classify pattern-matching, voice dictation and image recognition. In this proposed work, SVM could classify C&C session, using three categories such as session information vector, packet sequence vector and packet histogram vector. Session Information Vector generates total send/receive packet numbers, size and time from captured data. Packet size and packet interval time are generated by Packet Sequence Vector. Packet Histogram Vector was also dealt with data payload size. These vectors have good performances as a classifier excepting very high false-positive rate for detection of IRC session with Session Information Vector. Figure 3 indicates the result of this experiment.

While previous researches had tried to classify the traffic flows, following studies are dealing with botnet's behaviors in network. Mazzariello in [12] asserts that botnet performs malicious acts and it is very difficult to find a bot clearly. The IRC protocol is used to order and to obtain response back from a botmaster. The IRC channel is representatively used by humans to inter-communicate. The boundary between human activity and bots activity is clear in IRC environment because bots' operations or functions were not complicated and were small set of automated jobs. Furthermore, bots use limited dictionary and limited number of used terms. By analyzing pattern recognition techniques, it is easy to detect a botnet. According to [1], C&C servers and infected hosts were detected by the BotSniffer, and it could capture spatial-temporal correlation in traffic of network and it also can detect real-environment botnets with high precision and low false positive rate. Originally botnet C&C traffic is difficult to detect since it is similar to normal traffic and normal usage, and traffic volume is low. For these reasons, the activity of botnet is monitored by BotSniffer. The main types of response bot occurs are message response and activity response, which includes spamming, binary update and scanning [1]. In this paper, the BotSniffer has two engines, monitor engine to detect response and correlation engine

for grouping elements. The activity response collected by monitor engine is grouped by correlation engine for the analysis of group activity response analysis and group message response analysis. Same destination IP and port pair response are grouped, that is, same server client can be same group. In correlation engine, Response-Crowd-Density-Check algorithm and Response-Crowd-Homogeneity-Check algorithm are run to analyze group activity response and group message response respectively.

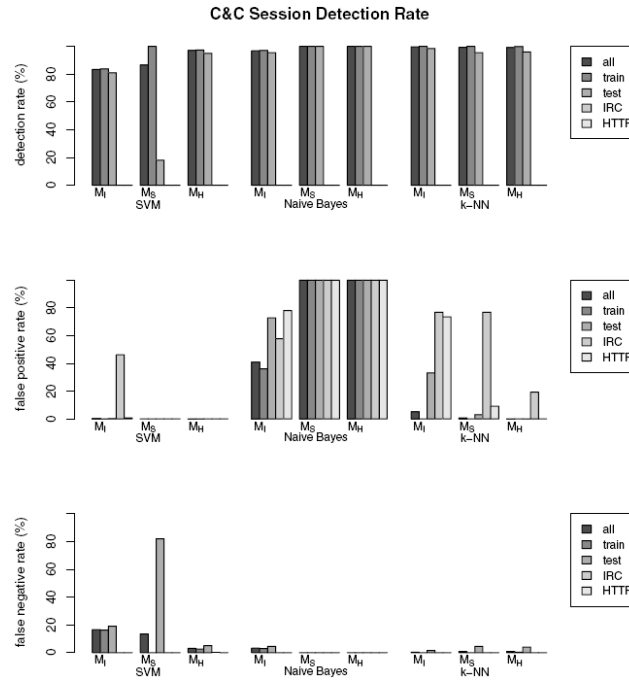


Fig. 3. Comparison of C&C Session detection Rate [11]

Likely, BotMiner in [13] detects behaviors of botnet traffic by notifying pattern between normal chat in C&C and malicious operations or command based on behavior and statistics. BotMiner has three modules, C-plane related with communication activities in traffic, A-plane related with malicious activities and Cross-cluster correlation to identify the hosts that share both activity patterns. C-plane monitor identifies who is talking to whom and A-plane monitor identifies who is doing what. Each plane monitor is connected with plane clustering and each clustering module uses logs gathered from each monitor. After this phase, cross-plane correlator finds patterns and detects the bot-traffic.

Botnet detection technology has been developed with immune-inspired algorithm in [14]. An ADCA (Dendritic Cell Algorithm) is the one of AIS (Artificial Immune System) and it can detect the compromised bot host according to Al-Hammadi et al.[14]. In human body, there is a DC (Dendritic Cell) that is the natural intrusion detection cell, and this cell is used to detect consequence to the host in this paper. DC has three signals and states and they are illustrated on figure 4. The DC exist in three

states according to signal, for example, if the combination of three signals' value is dangerous, the states will be changed to mature state. In IRC protocol, bot's keylogging and packet flooding behavior correlation values are used by DCA. The main thesis is that by running process or function call, the behavior of a bot could be changed and the function call can be antigen. For this reason, antigen, keylogging activities and time difference between sent time and received time could be signal, also these features are combined into a log file to calculate correlation. As a result, correlating the behavior by a single bot can be capable to detect malicious processes of a bot.

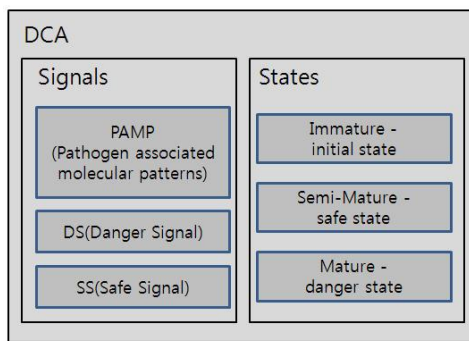


Fig. 4. DCA's three signals and states

3.1.2 P2P-based bot defense

Schoof and Koning [4] assert that bots based on P2P architecture could be detected by analyzing and monitoring bots' features. Using specific range of ports to communicate with other bots makes flows of network traffic monitored. Not only IP address lists used by the botnet, but also connection failures of bots in infection step can be significant evidence to identify the existence of botnet.

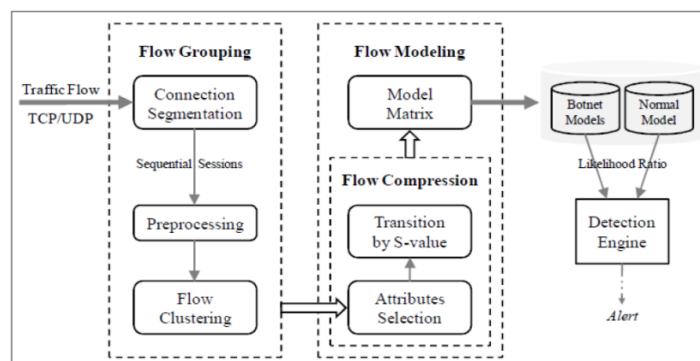


Fig. 5. Proposed framework in [8]

According to [8], because each bot in P2P architecture attempts to link to others, it

causes a number of traffics in order to find bot peer and to exchange their private. Furthermore, these generated traffics by bot have similar degree of fluctuating fixed form. This fixed form is useful to categorize TCP and UDP packet and to group the traffics which is used to be classified the activities of bots. Moreover Markov chain model is employed for making process model of transition using the fact that the P2P bot traffic flows could be different as depending on the status of bots. The entire mechanism includes Flow Grouping, Flow Compression and Flow Modeling. There are several phase to detect and to notify bots' existence. First, flow grouping is the process of clustering of TCP/UDP connection and measuring similarity amongst each flow. Second, flow compression phase calculates state value of each flow group from one phase. Third, matrix of transition is organized by the process of flow modeling. At the bottom of this framework, detection engine judges suspects of traffic from by likelihood ratio from model. Figure 5 is the framework to detect P2P bot.

3.2 DDoS(Distributed Denial of Service) defense

Since DDoS is also a consequence of botnet, in this part, DDoS defense mechanism is covered briefly. [15] suggests proactive detection of DDoS, using cluster analysis. The characteristics of packets entropy value of source IP/port, destination IP/port, packet type and the number of packets are calculated by algorithms and its values can be measured phase of attack. According to this paper, the attack phases consist of IPSweep, probing of live IP's to set up daemon or bot, breaking into the zombie host, installing DDoS software and launching the DDoS. In IPSweep period, small value of source IP address and increasing value of destination IP addresss are important measures, and in attack period, the value of source IP address is very small and the value of destination IP address are increasing dramatically. According to table 1 in [15], there are six clusters and cluster 5 indicates attack phase, displaying high value of the number of packets and entropy.

Table 1. Clustering analysis for DDoS deteciton in [15]

Variable	Cluster		
	1 Normal	2 Normal	3 Phase 1
Entropy of source IP	1.59	1.06	0.71
Entropy of source port	1.61	1.07	0.56
Entropy of destination IP	1.58	1.06	4.91
Entropy of destination port	1.50	1.07	0.55
Entropy of packet type	1.12	1.36	0.53
Number of packets	37.0	4.70	41.4
Occurrence rate of TCP SYN	0.02	0.44	0
Occurrence rate of UDP	0.00	0	0
Occurrence rate of ICMP	0.00	0	0.87
-			
Variable	4 Phase 2	5 Attack	6 Post-attack
Entropy of source IP	0.08	0.02	0.13
Entropy of source port	0.12	12.4	11.4
Entropy of destination IP	0.07	12.6	11.5
Entropy of destination port	0.12	12.6	11.5
Entropy of packet type	0.04	0.02	0.12
Number of packets	1.19	6225	2876
Occurrence rate of TCP SYN	0	0	0
Occurrence rate of UDP	0.99	0	0
Occurrence rate of ICMP	0	0	0

AMHI includes Address Matching and Hash Inspection in [16]. Address Matching has two rules, one is Localization Judgment to detect packet from out range of local subnet. If attacker forges IP address within the range of local subnet, the system cannot detect the packet. For this reason, another rule, MAC-IP Correlativity Judgment do not route the packets if source MAC address is not equal to the previously recorded one with statically allocated IP address is employed[16]. With these rules, Hash Inspection is also adoptable to detect DDoS packet. The degree of TCP, UDP, and ICMP flood through hash algorithm can be beneficial knowledge due to larger than 50% of similar degree among the data of packets from DDoS attack traffic.

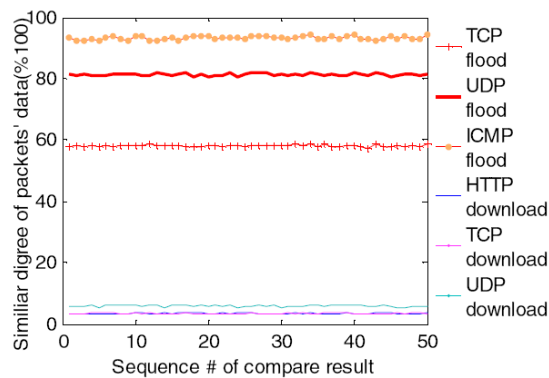


Fig. 6. Degree of hashed packets' data [16]

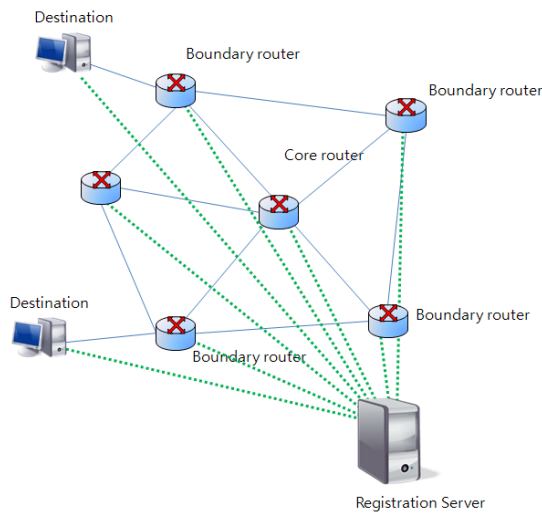


Fig. 7. Registered boundary objects

Figure 6 displays similar degree of hashed packets' data. Despite low degree of normal packets such as HTTP, TCP, UDP download packet, flood packets have high

degree of packets' data.

Unlike these approaches dealing with packet analysis, defending mechanism in [9] register routers in boundary and authenticate only registered routers in figure 7. Each boundary has one registration server in which all nodes in the region will be stored. To filter DDoS packet, router plays very important role as filter and notifier, including its several rules. As a result, the packages from other router must be discarded in authenticated boundary routers.

4 Challenges and Conclusion

Although botnets have been developed as more powerful threats on the Internet, our knowledge against botnet are not enough and it need that almost mechanisms should be changed and improved as botnet organization is rapidly developed. For example, in 2006, Zou and Cunningham [17] presented honeypot-aware advanced botnet although it is useful to capture valuable data of the gathered botnet activities such as attack commands and communication each other in [18]. That is why many of studies should be kept going continuously. Recently, computer science issues have not been dealt with only computer science technology. It means that in this case, many technologies of other fields can be applied to solve the problems such as Artificial Intelligent, Artificial Immune System and Machine Learning Skill.

In this survey, to detect C&C based bot, Machine learning technologies and bot behavior detections are mentioned, and SVM and Botsniffer are performed excellently. Moreover, DCA is also appropriate for detection of single bot based on the C&C.

This research is only one tip of the iceberg. The goal of our study is to understand how botnet works, affects, and is blocked. To make sure our main purpose, we analyzed and researched new mechanisms dealing with botnet defense technologies with this paper as just one step. After that, we will focus on the bot-net detection mechanism and will propose new bot-net detection mechanism step by step.

5 References

1. Gu, J. Z and W. Lee, "BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic," in Proceeding of 15 Annual Network and Distributed System Security, pp. 1-18, Jan, 2008.
2. Zhu, G. Lu and Y. Chen, "Botnet Research Survey," in Annual IEEE International Computer Software and Applications Conference(COMPSAC), pp. 967-972, 2008.
3. W. Stallings and L. Brown, "Cmputer Security Principles and Practice," Pearson Education. Published by 2008.
4. R. Schoof and R. Koning, "Detecting peer-to-peer botnets," <http://staff.science.uva.nl/~delaat/sne-2006-2007/p17/report.pdf>, pp. 1-7, Feb, 2007.
5. A. Karasaridis, B. Rexroad and D. Hoeflin, "Wide-scale Botnet Detection and Characterization," in USENIX Workshop on Hot Topics in Understanding Botnets, 2007.
6. N. Ianelli, A. Hackworth, "Botnet as a vehicle for online crime," CERT. Request for Comments (RFC) 1700, Dec, 2005.

7. M. Bailey, E. Cooke, F. Jahanian, Y. Xu and M. Karir, "A Survey of Botnet Technology and Defenses," in Proceeding of Cybersecurity Applications & Technology Conference For Homeland Security(CATCH), pp. 299-304, 2009.
8. S. K. Noh, J. H. Oh, J. S. Lee, B. N. Noh, and H. C. Jeong, "Detecting P2P Botnets using a Multi-Phased Flow Model," in proceeding of 2009 Third International Conference on Digital Society IEEE. pp. 247-254, 2009.
9. W. Zhang, S. Guo, K. Zheng and Y. Yang, "A Defending Mechanism against DDoS Based on Registration and Authentication," in Proceeding of The 9th International Conference for Young Computer Scientist(ICYCS.2008). pp. 2192-2197, 2008.
10. C. Livadas, B. Walsh, D. Lapsley and T. Strayer, "A Using Machine Learning Techniques to Identify Botnet Traffic," in Proceeding of The 2nd IEEE LCN Workshop on Network Security, 2008.
11. S. Kondo and N. Sato, "Botnet Traffic Detection Techniques by C&C Session Classification Using SVM," in proceeding of IWSEC, 2007.
12. C. Mazzariello, "IRC traffic analysis for botnet detection," in proceeding of The Fourth International Conference on Information Assurance and Security IEEE. (2008).
13. G. Gu, R. Perdisci, J. Zhang and W. Lee, "BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection," in proceeding of The 17th USENIX security Symposium, 2008.
14. Y. Al-Hammadi, U. Aickelin and J. Greensmith, "DCA for Bot Detection," in Proceeding of Congress on Evolution Computation(CEC) IEEE, 2008.
15. K. S. Lee, J. H. Kim, K. H. Kwon, Y. G. Han and S. H. Kim, "DoS attack detection method using cluster analysis," in proceeding of Expert Systems with Applications 34. pp. 1659-1665, 2008.
16. K. Bu and Z. Sun, "A Method Based on AMHI for DDoS Attacks Detection and Defense," in Proceeding of the 9th International Conference for Young Computer Scientists, Page 1571-1576, 2008.
17. C. C. Zou and R. C, "Honeypot-Aware Advanced Botnet Construction and Maintenance," in Proceeding of the 2006 International Conference on Dependable Systems and Networks (DSN'06) IEEE, 2006.
18. M. Vrable, J. Ma, J. Chen, D. Moore, E. Vandekieft, A. Snoeren, G. Voelker and S. Savage, "Scalability, fidelity and containment in the potemkin virtual honeyfarm," in Proceedings of the ACM Symposium on Operating System Principles (SOSP), Oct, 2005.
19. J. Yu, Z. Li, J. Hu, F. Liu and L. Zhou, "Using simulation to characterize topology of Peer to Peer Botnets," international Conference on Computer Modeling and Simulation (IEEE). pp. 78-83, 2009.
20. M. A. Rajab, J. Zarfoss, F. Monrose and A. Terzis, "A Multifaceted Approach to Understanding the Botnet Phenomenon," in Proceeding of IMC, pp. 41-52, 2006.