International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34140-34143

A Survey on Command and Control Channel Based Botnet Detection Systems

[#]Beena J Stuvert, [*]Soniya B

Department of CSE, SCTCE, Pappanamcode, Trivandrum, Kerala
[#]beenajs13@gmail.com
[*]soniya.balram@gmail.com

Abstract
The internet has enabled access to widespread remote services in various environments. While internet offers a huge useful service which makes communication easier and faster than ever, it presents some threats too along the way. In recent years, the root cause of many security problems on the Internet are bots. Bots are software program installed in a vulnerable host that is capable of performing malicious actions. Botnets are arising as the primary source for various internet attacks such as DDoS attacks, spamming, phishing etc.
A botnet is a network of compromised computers under the control of bot code. The bot code can be installed in the victim machine by accessing infected sites or viral infection. The distinguishing characteristic of botnet from other classes of malware is its command and control (C&C) channels. Thus, bot detection based on the characteristics of the C&C communication is more effective. This paper is a survey of recent advances in botnet detection research based on C&C channels.

1 INTRODUCTION

In today's Internet, malicious software running on end-user machines is the major cause of many security problems. Malware are used to disrupt computer program operations or gather sensitive information or gain access to private computer system. Bots are one of the most efficient kinds of such malware. A software program installed in a vulnerable host that is capable of performing malicious actions is called bot. These bots can be installed on the victim machine by accessing infected sites or viral mechanisms. A network of compromised machine under the control of bot code is called botnet. These botnets are emerging as the most significant and growing threat to internet communication.
The botnets are controlled by botmaster or botherders. The distinguishing factor that characterizes other class of malware from botnets is its command and control channel. The botmasters can update or direct the bot through these command and control channels. Command and control channels use different communication protocols, and it is used by botherders to coordinates the activities of the bots and to give commands to their bots. Within the botnets, the bots are controlled by its C&C servers. Thus, the robustness and stability of the botnet is determined by its C&C architecture. The detection based on C&C is more advantages because packet inspection is not needed for detection. Thus detection is possible when encrypted traffic is used by the botherders.
The botnets C&C architecture can be classified into two such as centralized and decentralized. The centralized approach is similar to client-server model. All the bots acts as client and connect to centralized server. The command and control channel is responsible for sending commands to the bots. The centralized botnet can operate easily and it has quick reaction times. The centralized botnets are created easily as compared to decentralized, thus it becomes the most widespread botnets today.
The botnet life cycle includes mainly five phases such as initial infection, secondary infection, rallying, malicious activities and updating. In the rallying phase, the C&C channel is established and the Zombie is connected to the C&C server. The malicious activities are started at this point and the botmaster remotely control the bots through their C&C servers. Thus, the detection of the C&C channel can be performed at this phase.
Botnet detection can be mainly classified into two categories, one method is based on honeynet and other based on intrusion detection system. The information from bots can be collected using Honeynets. A honeynet [1] is a trap set to detect the unauthorized use of information systems. The features of the botnet characteristics can be analyzed and understand using the collected information. The second method is based on intrusion detection system. An intrusion detection system is an application to monitor the malicious activities of system services or report the policy violation to the management sites. IDS detection techniques are again classified into signature-based, anomaly-based and DNS based. Signature based detection is an IDS based detection system that applies behavior and signatures of known botnets [2]-[4]. The main idea is to extract features from packets of monitored traffic and maintain the features of existing bots in the database. Signature based detection is simple and it can performed easily by comparing every byte in the packets. In anomaly based detection techniques network traffic anomalies are used to detect botnets. Network traffic anomalies used to detect the presence of malicious bots in the network are traffic on unusual ports, high network latency, unusual system behavior and high traffic volume. Host based approach [5]-[7] and network based approach are the two categories of anomaly-based techniques. The detection method based on the DNS information generated by a botnet is DNS-based detection [8]-[9].
The botmasters use different techniques to evade the botnet detection. Thus, the existing botnet detection system becomes ineffective. The strength of the botnet depends on the communication with its C&C server, thus the common characteris-

tics of the botnet depend on the network activities of the C&C channels. Thus, the better way to detect botnet or individual bot in the botnet is based on the features of their C&C channel.

2 BOTNET DETECTION SYSTEMS BASED ON COMMAND AND CONTROL

A Command and Control system is set-up by the botherders to communicate with his bots indirectly because it does not want its identity to be published and want to cover the command sent. Within the botnet, the bots are controlled by botmasters using the C&C server. Different types of C&C servers exist such as centralized and peer to peer. The centralized C&C architecture provide simple, low-latency and efficient real time communication to the botherders. The centralized C&C use existing protocols such as IRC, HTTP for communication. P2P architecture is more robust than centralized architecture. It is more complex. The different botnet detection systems using the C&C are discussed below.

Dagon et al.[8] proposed a mechanism to identify botnet C&C servers using domain names that has high DDNS query rate. In this approach, a key metrics and various topological structures used to coordinate attack on botnet are identified. It is possible to consider the ability of different response techniques to degrade or disrupt botnets using the proposed performance metrics. The canonical DNS request rates and DNS density comparisons of botnets rallying DNS traffic are measured. However, the drawback of this approach is that the botmasters can evade the detection by using faked DNS queries, thus it generate false alarm due to misclassification of legitimate and popular domains.

Choi et al. [9] proposed an anomaly detection mechanism called BotGAD by monitoring group activities in DNS traffic. In this approach, unique features are defined to differentiate Botnet DNS queries from legitimate DNS queries. In the various stages of Botnet life cycle, the DNS traffic is appearing, thus it is possible to detect botnet based on the group activity of the botnet DNS traffic. BotGAD use the IP header information to detect botnet with encrypted channel and also detect unknown botnets in real time large network. Furthermore, the mitigation of C&C server can also detected using this approach. In large scale network traffic, huge processing time is required is the major drawback.

The work [10] uses a correlation algorithm to detect the presence of a botnet and identify C&C servers using passive traffic analysis. The method contains three stages, each of which uses flow characteristics: filtering traffic unlikely to be bot C&C, classifying traffic as likely to be IRC, and clustering related flows. Flows are then clustered and classified according to IRC-like traffic patterns in a five dimensional space considering packet inter-arrival times and packet sizes. Flow perturbation could be used to defeat each stage; the simplest approach targets the filtering of high bitrate flows by injecting packet- and flow-level noise.

Botsniffer [11] is an anomaly based detection method mainly used for detecting IRC and HTTP botnets in LAN without any prior knowledge of signatures. The detection approach is based on the fact that the bots in the same botnet respond to the botmaster's command and performs activities in similar fashion. To identify the crowd of hosts that exhibit similar response or activity pattern, several correlation and similarity analysis algorithms are used. Botsniffer mainly consist of two components such as monitor engine and correlation engine. The main function of monitor engine is to examine network traffic, generates connection record of suspicious command and control protocols, and detects activity and response behavior. The function of correlation engine is to analyses the spatial-temporal correlation and activity or message similarity observed by monitor engine. The main advantages of this system are it does not require prior knowledge of content signatures and it is able to detect encrypted C&C. But, this method has some disadvantages such as evasion by misusing the whitelist, evasion by encryption, evading protocol matcher, evasion by using very long response delay.

Botminer [12] is a botnet detection system that applies data mining techniques for detecting botnet command and control traffic. Botminer is an improved form of Botsniffer, which is independent of the command and control protocol, structure, and infection model of botnets. Botminer mainly consist of C-plane monitor, A-plane monitor and cross-plane correlator. The C-plane and A-plane monitors are used for capturing similar communication and similar malicious pattern respectively. Then Botminer clusters similar communication activities and clusters similar malicious activities. To identify the hosts that shares both similar communication patterns and similar malicious activity patterns, cross–cluster correlation is performed. Botminer can detect IRC-based, HTTP-based and P2P botnets and it also detect real-world botnets. However, it has some disadvantages such as it doesn't detect stealthier bots and can be evaded by bots using normal servers to hide the activity.

Disclosure [13] is a large scale, wide area botnet detection approach to detect botnet C&C servers by analysing the NetFlow. In this approach, the features such as flow size, client-access pattern and temporal behavior are used to distinguish C&C channel from bengin traffic. These features are effective in detecting current C&C channel and also it is relatively robust against the counter measures of the future botnets. An external reputation score is incorporated in this approach to reduce the false positive rate. It is the only systems that use NetFlow data and it does not assume a prior knowledge of the particular C&C protocols.

CoCoSpot [14] is a method to group similar botnet C&C channels. Recent botnets C&C protocols can be detected using traffic analysis features such as message length, the carrier protocol as well as the encoding scheme. Thus a fingerprint can be derived using this approach. The C&C candidates can be identified using the periodicity of the message. The main advantage of this approach is that, it is independent from payload byte signature which enables the detection of C&C protocols with encrypted message contents and the system is able to produce human-readable reports, thus analysis is easy. One disadvantage of this approach is the flow clustering of C&C, which is used to discover relationships between malware families based on the distance of their C&C protocol.

BotFinder [15] is a detection system that detects bots in the network traffic without performing packet content inspection. It is a system that detects malware infection in the network traffic by comparing the statistical features of the previously-observed bot activity. It works by automatically building mul-

ti-faceted models for C&C traffic of different malware families. The high level information about the network connection is required for this approach. The advantage of this system is, it identify bots in the network even the bots use encrypted C&C communication.

The work in [16] use anomaly based approach to detect the IRC based botnet. The main characteristics of IRC based botnet are the direction of ping-pong messages, homogenous response and group activity. Thus, to identify the anomalous host, the anomaly on the homogenous activity and group activity during the communication between C&C server and botmaster is used. It detects botnets without any prior knowledge of the botnets or their servers.

3 CONCLUSION

Botnets has becomes the most prominent threats on the internet and it provides the key platform for many cyber-crimes such as distributed denial of service, sending spam, stealing personal information, computing resources, identity theft etc. These botnets are controlled and managed by botherders. They use various approaches to retain their bots secure. Therefore botnet detection is a big challenge in internet security. There exist various botnet detection techniques such as signature based system, anomaly based system, DNS based systems. These detection methods are developed based on the prior knowledge of the known bots, their statistical communication pattern, and based on payload inspection. These methods have both advantages and disadvantages. However, it becomes ineffective due to certain reasons such as changes in statistical patterns, encrypted traffic, randomizing bot communication etc. The botnet detection based on C&C is more effective compared to other techniques. In C&C based system, the features of the communication between C&C server and the bots are used for detection. So the inspection of packet payload is not needed. Thus the detection is possible at both encrypted traffic and normal traffic. Also, the C&C based methods detect the unknown bots without any prior knowledge. Thus, the better way to detect botnet or individual bot in the botnet is based on the features of their C&C channel.

The detection becomes a challenge when the botnets change their C&C architecture. The main challenge in botnet detection is the difficulty to testing the detection approaches with real world datasets.

References

[1] F. Pouget, M. Dacier, "Honeypot-based Forensics", Asia Pacific Information technology SecurityConference, 2004.

[2] J. Goebel, T. Holz, Rishi: identify bot contaminated hosts by IRC nickname evaluation, in: Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets, USENIX Association, Berkeley, CA, USA, 2007, p. 8.

[3] P. Wurzinger, L. Bilge, T. Holz, J. Goebel, C. Kruegel, E. Kirda, Automatically generating models for botnet detection, in: M. Backes, P. Ning (Eds.), Computer Security – ESORICS 2009, Lecture Notes in Computer Science, vol. 5789, Springer, Berlin/Heidelberg,2009, pp. 232–249. 10.1007/978-3-642-04444-1 15.

[4] Y. Kugisaki, Y. Kasahara, Y. Hori, K. Sakurai, Bot detection based on traffic analysis, in: The 2007 International Conference on Intelligent Pervasive Computing, IPC, 2007, pp. 303–306.

[5] M. Szymczyk, Detecting botnets in computer networks using multiagenttechnology, in: Fourth International Conference on Dependability of Computer Systems, DepCos-RELCOMEX'09, 2009, pp. 192–201.

[6] E. Stinson, J.C Mitchell, Characterizing bots' remote control behavior, in: Proceedings of the 4th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, DIMVA'07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 89–108.

[7] M. Masud, T. Al-khateeb, L. Khan, B. Thuraisingham, K. Hamlen, Flow-based identification of botnet traffic by mining multiple log files, in: First International Conference on Distributed Framework and Applications, DFmA 2008, 2008, pp. 200–206

[8] D. Dagon, C. Zou, W. Lee, Modeling botnet propagation using time zones, in: Proceedings of the 13th Network and Distributed System Security Symposium NDSS.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989

[9] H. Choi, H. Lee, H. Kim, BotGAD: detecting botnets by capturing group activities in network traffic, in: Proceedings of the Fourth International ICST Conference on COMmunication System software and middlewaRE, COMSWARE '09, ACM, New York, NY, USA, 2009, pp. 21–28.

[10] W. Strayer, R. Walsh, C. Livadas, D. Lapsley, Detecting botnets withtight command and control, in: Proceedings 2006 31st IEEE Conference on Local Computer Networks, pp. 195–202.

[11] G. Gu, J. Zhang, W. Lee, BotSniffer – detecting botnet command and control channels in network traffic, in: 15th Annual Network & Distributed System Security Symposium, The Internet Society (ISOC), San Diego, 2008.

[12] G. Gu, R. Perdisci, J. Zhang, W. Lee, BotMiner: clustering analysis of network traffic for protocol- and structure-independent botnet detection, in: Proceedings of the 17th Conference on Security Symposium, USENIX Association, Berkeley, CA, USA, 2008, pp.139–154.

[13] W. R. E. K. Leyla Bilge, DavideBalzarotti, "Disclosure: Detecting botnet commandand control servers through large-scale netflow analysis," *ACM*, 2012.

[14] ChristainJ.Dietrich, ChristainRossow, Norbert Pohlmann, CoCospot: Clustering and recognizing botnet command and control using traffic analysis, computer networks, Elsevier 2012.

[15] Florian Tegeler, Xiaoming Fu, Giovanni Vigna, Christopher Kruegel, BotFinder: Finding bots in network traffic without deep packet inspection, ACM 2012.

[16] Chia-Mei Chen, Hsiao-Chung Lin, Detecting botnet by anomalous traffic, in journal of information security and applications, Elsevier 2014.