A Hybrid Approach for detecting fast-flux botnets using HTTP request analysis and DNS based approach

Abstract

The emergence of botnets for our connected world has been a threat almost for the last two decades. The main target of these compromised computers is to manipulate target hosts for different malicious activities. The latest botnets evade detection systems using different methods and hide themselves in licit traffic. One of the evading techniques is the use of Fast Flux Service networks aimed at hiding the command and control channel which uses by an attacker to communicate the rest of the bots. The goal of this research is to find a hybrid approach that combines DNS based fast fluxing botnets detection methods with the HTTP request analysis.

Keywords Botnet, Detection, fast-flux botnets, machine learning, HTTP request analysis

1 Introduction

Due to an increasing growth of Internet usage, cybercrimes has been increasing at an Alarming rate and has become most profitable criminal activity. Botnet is an emerging threat to the cyber security and existence of Command and Control Server (c&c Server) makes it very dangerous attack as compare to all other malware attacks. Botnet is a collection of networked bots which is controlled by a Botmaster (Bot herder). In other words, the individual bots are software programs that run on a host computer allowing the botmaster to control host actions remotely. Bot-Master communicates with the remaining other bots using a command control channel. Botnets are used for different criminal activities including Distributed Denial of Service attack (DDOS), malware dissemination, phishing (stealing secret information like passwords, credit card numbers, and social security numbers), click fraud and email spams. The more modernized attacks include applications that installed on host machine that enquire the username and password for some type of embezzlement. The crime ranges from email spams to destroying targeted networked institutions like Stuxnet that target industrial control systems. They are the number one agents in cyber war between nations. The trend for detecting botnets uses different methods. As it will be discussed the section two, many researches are based on intrusion detection systems at the network traffic level on analyzing the traffic patterns or analyzing the incoming packets payload. The latest botnets use fast-fluxing service networks, cryptography and obfuscated malicious codes to evade detection mechanisms. There is a need to do HTTP request analysis for the encrypted botnets and also to detect fast-flux botnets. Fast-fluxing gives an attacker three major advantages. First, it offers simplicity, as the attacker can rely on only a few powerful backend servers, such as motherships. Second, fast-fluxing provides an extra layer of protection against tracking and discovery due to its use of blind proxies. Third, it extends the operational life spans of the motherships due to its extra layer of protection. This research will be proposing a hybrid detection mechanism using traffic analysis mechanisms and fast-fluxing detection techniques that provide a high detection rate[20].

1.1 problem statement

As mentioned earlier in section 1.1, there is a need to solve the evading techniques of the fluxing botnets that traffic payloads are encrypted. Most of the bots uses encrypted communication for their command and control channel.so one can came up with a mechanism that only analyze the transport layer data to back trace the bot channel. Most of the fast-fluxing detection mechanisms involve longer time to capture data and classify and also larger memory sizes for computation. These problems should be addressed in this research work.
To conclude, the research questions of this work are:

Q1. Can we combine HTTP request analysis with DNS query analysis to come up with new methodology?
Q2. Is the new proposed system time and storage efficient?
Q3. What is the detection rate of this hybrid system in comparison with other detection systems?

1.2 Objective

1.2.1 General Objective

The main objective of this research is to come up with efficient fast-flux botnet detection mechanism that evade previous detection methods using fast flux service network and whose message payloads are encrypted.

1.2.2 Specific Objective

- Ensure the new proposed system is time and storage efficient
- Compare and do an analysis the performance of the new approach with previous works

1.3 Scope

This researchs main focus is on detection botnets that use HTTP communication protocol whose IP packets are encrypted and botnets that use fast-flux service networks for DNS mapping for their botmaster address.

1.4 Contribution

After this research gets completed there will be a powerful botnet detection mechanism that combines data mining and correlation techniques. The detection rate will be higher than the previous detection methods.

1.5 Research Methodology

This research will follow different procedural steps.First deep Literature survey currently almost 90% has been done. Then we will experiment previous related works individually then we will design the hybrid algorithm.After that we will set an experiment to test the new approach and analyze the outcomes of the experiment. If didn't give the expected result make an optimization on the algorithm.Finally compare the new result with older detection methods.

1.6 Paper Organization

- Section 1 discusses Introduction
- Section 2 contains Literature review
- Section 3 discusses the proposed Methodology
- Section 4 and 5 show the Time line Budget respectively

2 Literature Review

2.1 Botnet Life-cycle

several literatures discuss about the life cycles of Botnet development and categorized them in four main stages. These are Initial injection, Secondary injection, Connection phase, Command and Control Server and upgrading and maintenance phase [2,5,6]. Bot master follows this cycle to create, infect and control the Bots.lets look at them briefly.

2.1.1 Initial Injection

In first step, an attacker searches for vulnerable hosts and infect them using various exploitation techniques like sending spam emails, phishing, creating backdoors etc. First task of botmaster is to make the infected machines as a part of Botnet and it must know the address of command and control (c&c) server. In order to inject there are several ways of analyzing vulnerability of a host by setting vulnerability matrix in which it searches unprotected computers over the INTERNET. The address of the c&c server is hard coded in case of centralized architectures.

2.1.2 Secondary Injection

In this phase infected machine downloads the script that searches for malware binary and installs it.This code is called shell code. After installation the target converted to one of the zombie army. This activity uses HTTP, IRC and P2P protocols.

2.1.3 Connection

The main phase of Botnet Life cycle is establishing a connection where infected machines are connecting with command and control (c&c) server to receive commands from botmaster. In this phase, bots receive commands from bot-master and send reply to command and control (c&c) server. In this phase, a bot-master restart all the bots to check activeness of the bots and to make sure that bots are able to receive commands.

2.1.4 command and updating

Actual working of Botnet starts from command and control server. Attacker controls zombie army or bots from common c&c server. All infected machines need to know the address of a c&c server while downloading the script, so that they can easily communicate with the server.In this stage the server might change frequently in the case of P2P protocol for longer survivability. This process of changing role is called server mitigation[2].

2.2 command control Architecture

Based on the command and control architecture Botnets can be divided into centralized,decentralized and hybrid [6,12,17].

2.2.1 Centralized

The simplest model for creating Botnet is Centralized Model. This structure is having a single central Command and Control server for communication and creating all other Bot. All Bots are directly connected with a single c&c server. This central point needs a very high bandwidth. From one c&c server Botmaster sends and receives messages to all other bots in network. Many previous Bots like AgoBot, SDBot and RBot were created with centralized model[6] .The Biggest weakness of this architecture is single Command and Control Server which makes it more vulnerable. If single command and control server has been detected all other network can be easily detected.But the main advantage of this architecture is its simplicity.

2.2.2 Decentralized

The biggest weakness of a centralized model is single command and control server which makes the Botnet very easy to detect. To overcome this vulnerability bot owner developed decentralized Botnets. In this model, Botnet is not control by single command and control (c&c) server. Bot master can use any bot to act as a c&c server. Decentralized model can create large number of bots in a single Botnet and it is very difficult to trace.

2.2.3 Hybrid

Hybrid Model is a combination of both centralized and decentralized model. As each design of Botnet have some benefits and drawbacks. For creating a strong Botnet architecture, Botmaster take advantages of hybrid model. In hybrid model, encryption key is used to hide Botnet traffic within normal traffic. This model use random vulnerable port and sends encrypted messages from any bot in the network[6].

2.3 communication protocol

To communicate bots to c&c server there are different protocols that are related to the architectures.[6] categorized them as three.[12] categorized them as four.lets see of them.

2.3.1 IRC

IRC term stand for Internet Relay Chat works with real time internet text messages. Botmaster use IRC channels to create and communicate with all Botnet. Due to its simple architecture and high flexibility, IRC Botnet is mostly used by many Botnet owners at the Botnet beginning era. This Botnet is very effective in creating and reusing the bots. One limitation with IRC based network is its common usage of IRC protocol for communication which gets easily traced simply by seeing IRC traffic as part of normal traffic[6].

2.3.2 HTTP

Hyper Text Transfer Protocols are also very popular for creating Botnet. It can easily hide Botnet traffic with Normal Traffic. With the HTTP protocol Botmaster can easily bypass the firewall and hide malicious traffic with normal http traffic. HTTP based bots often download the instruction from web based Command and Control Server which is more accurate then IRC based Botnet [6]. Http based model is used by many famous Botnet as like Bobox, Click Bot, Rustock [6].Many Researchers has purposed some web based models to Detect HTTP Based Bot from network traffic[10].[19] also proposed HTTp bsed request analysis by correlating request uniformity and the time gap between them.

2.3.3 P2P

Peer to peer model provide decentralized model where all bots are connected with each other. The main aim of peer to peer protocol is to hide the command and control server. Botmaster uses different bots to issue commands every time. Bots are dependent on previous or other connected bots without having Single c&c server. Nearly 70% Botnets are created using P2Pprotocols. Some advanced Botnets like Phatbot, Nugache used P2P based bots for creating a Botnet[6]. P2P based model creates a very complex architecture which is very hard to trace. Botmaster can use any peer to communicate with other bots. It act as client server model where each node can work as client or server. For creating a new bot it just need an address of any bot connected in Botnet for communication. Major benefit is detection of single bot doesnt mean detection of whole Botnet.

2.3.4 DNS Based

DNS is a fundamental protocol in the Internet. Namely, DNS protocol is not only a fundamental protocol to resolve domain names to IP addresses but also is used by many Internet applications such as email exchanging.The most recent Botnets uses fluxing networks for evading detection. The fluxing network works that a set of IP address change frequently corresponding domain name[4].

3 Detection Techniques

There have been different Botnet detection techniques over the litratures.most of the litratures classify them as honeypot based and Intrusion Detection techniques.The second one further divided into four parts as signature based, anomaly based, DNS based and data mining based.[6]

3.1 Honeypot Based

Honeynet project[15] which were conducted in 2007 was one of the pioneering informal studies of the Botnet problem. However, efforts are in progress to quantify the Botnet problem, detect the presence of Botnets, and design defenses against attacks by Botnets. But [2,6,12] discuss that honeynets are mostly useful to underrstand Botnet technology and characteristics, but do not necessarily detect bot infection.

3.2 Intrusion Detection systems

This is an approach based on the identification of traffic patterns.sometimes it is called as Network based approach.there are different categories of IDs

3.2.1 Signature Based

Knowledge of useful signatures and behavior of existing Botnets is useful for Botnet detection. For example, Snort [9] is an open source intrusion detection system (IDS) that monitors network traffic to find signs of intrusion. Like most IDS systems, Snort is configured with a set of rules or signatures to log traffic which is deemed suspicious. However, signature-based detection techniques can be used for detection of known Botnets. Thus, this solution is not useful for unknown bots[1,2,6,9,12,16,17].But [1] proposed a framework which is called BnNeSSy(Botnet security system) which is based on the traffic analysis using artificial neural network model. it has a mechanism of detection Botnets with unknown signatures because of its adaptive nature for training unknown illicit traffic.it has the capability of modifying the tracebility of unknown bots.

3.2.2 Anomaly Based

Anomaly-based detection techniques attempt to detect Botnets based on several network traffic anomalies such as high network latency, high volumes of traffic, traffic on unusual ports, and unusual system behavior that could indicate presence of malicious bots in the network[2,3,6,12,13,17].This technique is used to detect unknown bots or which are called zero day bots.[3] proposed a novel anomaly based detection frame work by using Genetic Neural Network(GNN).

3.2.3 DNS based

DNS-based detection techniques are based on particular DNS information generated by a Botnet. DNS-based detection techniques are similar to anomaly detection techniques as similar anomaly detection algorithms are applied on DNS traffic. As mentioned in the life cycle, bots typically initiate connection with c&c server to get commands. In order to access the c&c server bots perform DNS queries to locate the respective c&c server that is typically hosted by a DDNS provider. Thus, it is possible to detect Botnet DNS traffic by DNS monitoring and detect DNS traffic anomalies[2,6,12,17].[18] proposes a novel DNS based Botnet communication detection method.[20] currently work on this problem using Evolving fuzzy neural network (EFuNN) for detecting Botnets in fast flux networks.DNS sink-hole is another method to catch the concern of many bots and then trace the c&c server.

3.2.4 Mining Based

Data mining technique capture the high volume of network traffic and find the malicious traffic from it. With the help of existing techniques abnormal traffic can be differentiated. But to detect c&c server and recognize its pattern used for Botnet is very difficult. As encrypted c&c Server hide itself with normal traffic so to detect this kind of Botnet attacks Data mining techniques are used with data Classification and Clustering [2,6,12,17].[10] proposes detection and characterization of Botnets using passive analysis based HTTP protocol.This method the advantage of security as it depends on the TCP header only.[7] also derived an algorithm and a platform to monitor group of hosts that perform at least one malicious activity in one step and then find the hosts that show similar communication patters among them.

3.2.5 Machine Learning based

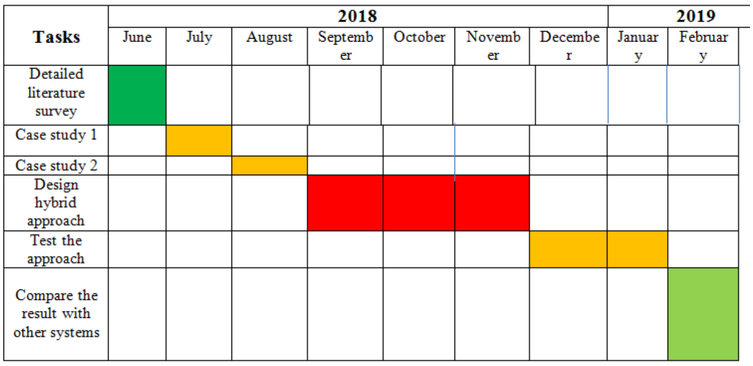
Different Machine Learning techniques such as Classifiers, Decision Trees etc. are also used to detect the Botnet as they are more effective to detect chat Bots but cannot work well to Detects the Command and Control Server.[13] uses a novel flow based detection system that relies on supervised machine learning for identifying P2P based Botnets.[14] uses a machine learnig approach for detecting IRC based Botnets. But this approach is unable to detect when the payloads in the network packet is encrypted.

4 Proposed Methodology

My detection method is based on both DNS based and Machine learning based. First network traffic data and DNS name service data in fast flux networks collected passively for a period of time and stored in database. The data collected from different network is then filtered on basis of protocol used which is in our case HTTP. The selected HTTP requests are organized and correlated to identify the HTTP Bots web activities in addition to group similarity correlation to look for any evidence of HTTP Botnet[19]. Then the results that show bot like characteristics are finely traced with their fluxing rates. The pattern in their command and control requests is analyzed using supervised machine learning techniques. Using fast flux hunter [20] we can trace the botmaster. Knowing individual bots doesnt mean knowing the whole botnet but knowing the botmaster is the key in detecting decentralized botnet archtectures.

5 Time Line

I have planned to complete this research with in 10 months. The schedule will look like as in the table below. The last month will be document preparation and submission

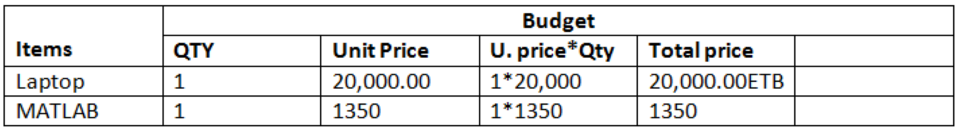


Tasks	2018							2019	
	June	July	August	September	October	November	December	January	February
Detailed literature survey	🟩								
Case study 1		🟨							
Case study 2			🟨						
Design hybrid approach				🟥	🟥	🟥			
Test the approach							🟨	🟨	
Compare the result with other systems									🟩

Figure 1:

6 Budget

The implementation will be using software like Wireshark-win 64-1.12.6 and WinPcap(both are open source softwares) to capture online traffic data, as well as MATLAB version 7.14 . I will use intel core i7 processor that has NVIDA GPU that operates on Windows 10 (64 bit),16 GB of memory (RAM) .



	Budget				
Items	QTY	Unit Price	U. price*Qty	Total price	
Laptop	1	20,000.00	1*20,000	20,000.00ETB	
MATLAB	1	1350	1*1350	1350	

Figure 2:

References

[1] Nogueira, A., Salvador, P. and Blessa, F., 2010, June. A Botnet detection system based on neural networks. In Digital Telecommunications (ICDT), 2010 Fifth International Conference on (pp. 57-62). IEEE.

[2] Feily, M., Shahrestani, A. and Ramadass, S., 2009, June. A survey of Botnet and Botnet detection. In Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09. Third International Conference on (pp. 268-273). IEEE.

[3] Yin, C., Awlla, A.H., Wang, J. and Yin, Z., 2015, October. A novel framework towards Botnet detection. In Computer and Computing Science (COMCOMS), 2015 3rd International Conference on (pp. 9-12). IEEE.

[4] Zhang, L., Yu, S., Wu, D. and Watters, P., 2011, November. A survey on latest Botnet attack and defense. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on (pp. 53-60). IEEE.

[5] Asha, S., Harsha, T. and Soniya, B., 2016, May. Analysis on Botnet detection techniques. In Research Advances in Integrated Navigation Systems (RAINS), International Conference on (pp. 1-4). IEEE.

[6] Kaur, N. and Singh, M., 2016, August. Botnet and Botnet detection techniques in cyber realm. In Inventive Computation Technologies (ICICT), International Conference on (Vol. 3, pp. 1-7). IEEE.

[7] Zeidanloo, H.R., Manaf, A.B., Vahdani, P., Tabatabaei, F. and Zamani, M., 2010, June. Botnet detection based on traffic monitoring. In Networking and Information Technology (ICNIT), 2010 International Conference on (pp. 97-101). IEEE.

[8] B. Saha and A, Gairola, Botnet: An overview, CERT-In White PaperCIWP-2005-05, 2005.

[9] Lee, N.Y. and Chiang, H.J., 2010, December. The research of Botnet detection and prevention. In Computer Symposium (ICS), 2010 International (pp. 119-124). IEEE.

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

[11] Bailey, M., Cooke, E., Jahanian, F., Xu, Y. and Karir, M., 2009, March. A survey of Botnet technology and defenses. In Conference For Homeland Security, 2009. Cybersecurity Applications and Technology (pp. 299-304). IEEE.

[12] Amini, P., Araghizadeh, M.A. and Azmi, R., 2015, September. A survey on Botnet: Classification, detection and defense. In Electronics Symposium (IES), 2015 International (pp. 233-238). IEEE.

[13] Stevanovic, M. and Pedersen, J.M., 2014, February. An efficient flow-based Botnet detection using supervised machine learning. In Computing, Networking and Communications (ICNC), 2014 International Conference on(pp. 797-801). IEEE.

[14] Carl, L., 2006. Using machine learning techniques to identify Botnet traffic. In Local Computer Networks, Proceedings 2006 31st IEEE Conference on. IEEE.

[15] Honeynet Project and Research Alliance. Know your enemy: Tracking Botnets, March 2005. See http://www. honeynet.org/papers/bots/.

[16] Zeidanloo, H.R., Shooshtari, M.J.Z., Amoli, P.V., Safari, M. and Zamani, M., 2010, July. A taxonomy of Botnet detection techniques. In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on (Vol. 2, pp. 158-162). IEEE.

[17] Raghava, N.S., Sahgal, D. and Chandna, S., 2012, May. Classification of Botnet detection based on Botnet architechture. In Communication Systems and Network Technologies (CSNT), 2012 International Conference on (pp. 569-572). IEEE.

[18] Ichise, H., Jin, Y. and Iida, K., 2015, July. Detection method of DNS-based Botnet communication using obtained NS record history. In Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual (Vol. 3, pp. 676-677). IEEE.

[19] Eslahi, M., Abidin, W.Z. and Naseri, M.V., 2017, November. Correlation-based HTTP Botnet detection using network communication histogram analysis. In Application, Information and Network Security (AINS), 2017 IEEE Conference on (pp. 7-12). IEEE.

[20] Almomani, A., 2018. Fast-flux hunter: a system for filtering online fast-flux Botnet. Neural Computing and Applications, 29(7), pp.483-493.