BotXrayer : Exposing Botnets by Visualizing DNS Traffic

Inhwan Kim, Hyunsang Choi and Heejo Lee
Div. of Computer & Communication Engineering, Korea University
[{neutrino37, realchs, heejo}@korea.ac.kr]

Abstract

Botnets pose a major problem to Internet security. They can cause various online crimes such as DDoS attacks, identity thefts and spam e-mails. While there have been many attempts to detect botnets, most of these studies have difficulties in detecting botnets due to their evasive techniques to resemble normal traffic. In this paper, we propose a visualization method, BotXrayer, to detect botnets. It displays DNS traffic on the plane of parallel coordinates using four carefully selected parameters that represent a botnet hierarchy and attack patterns efficiently. BotXrayer provides a view of graphs that helps humans recognize botnet patterns intuitively. Observing botnets frequently generate DNS traffic that forms unique patterns, we develop six botnet attack signatures. We adopt four logic operations (XOR, AND, OR, SUB) to find hidden botnet identities and to display distinct botnet graphs from noisy lines on the coordinates. Experiments with real traces in /16 networks show that the proposed mechanism can detect various botnets effectively. Furthermore, botnet activities, such as launching DRDoS, poisoning DNS cache entries and sending spams, were captured.

Keywords: Botnet, Visualization, DNS, Network Security

1. Introduction

Over the past few years, botnets have become a major Internet threat. Although it is imperative to protect PCs from being unwilling members of the botnets, effective defense solutions against botnets have not been adequately developed. Botnets have evasive techniques to prevent existence detection mechanisms. Thus, botnets resemble normal traffic.

One promising approach to overcome the limitations is visualizing complex situations in a simple and intuitive fashion [1]. Humans can easily recognize and identify graphical patterns from complex visual images [2]. Therefore, visual representation becomes essential in the Internet security field.

In this paper, we propose a visualization mechanism, BotXrayer, to detect botnets. The mechanism uses parallel coordinate that has many advantages, such as representing more than three values in a two dimensional space [3]. Four planes represent four distinct parameters in a DNS packet. Using DNS traffic, we obtain monitoring efficiency and accurate detection. The visualization yields colored lines that highlight the graphical pattern of botnets and their attacks. From the colored image patterns, we define six signatures that represent botnets and their malicious behaviors including DRDoS attack and DNS cache poisoning. To detect hidden botnets such as slow growing botnet or similar resemblance to normal traffic, we use logical filter operations such as XOR, AND, OR, and SUB. From the refined graph, we acquire a

This research was supported by National IT Industry Promotion Agency (NIPA) under the program of Software Engineering Technologies Development and Experts Education and the Ministry of Knowledge Economy, Korea, under the ITRC support program supervised by the NIPA (NIPA-2009-(C1090-0902-0016)) and the IT R&D program of MKE/IITA (2009-S-026-01, The Development of Active Detection and Response Technology against Botnet)

more definitive view of botnets and attacks.

Our study makes three main contributions. First, the visualization mechanism using the botnet character of DNS packets enables network administrators to uncover botnets and their activities intuitively. Second, using four X-raying filters, our approach provides a refined image that reduces noises and helps humans to recognize evasive botnet patterns. Third, the mechanism finds suspicious domain names for real world botnets, DNS cache poisoning and spamming using traffic taken from real networks.

2. Related Work

In this section, we review related works in botnet detection and visualization approaches related to botnet detection.

2.1 Botnet Detection Researches

Dagon represents a botnet detection and response approach [4] that analyzes the peculiarity of botnets rallying DNS traffic. However Dagon's approach is inefficient, since it generates many false alarms. Karasaridis [5] proposed an approach using IDS-driven dialog correlation based on a defined bot infection dialog model. However, it is IDS dependable. BotHunter [6] models the botnet infection life-cycle, as one sharing common steps. However, any malware not conforming to this model would seemingly go undetected using this approach. BotSniffer [7] is designed to detect botnets using either IRC or HTTP protocols. BotSniffer uses a spatial-temporal correlation detection method and relies on the assumption that all botnets, tend to communicate in a synchronized fashion.

These research approaches have weaknesses that are difficulty to detect botnet traffic, are evaded easily and have high overhead for realtime observing.

2.2 Botnet Visualization Researches

Visualization, where there are innovative approaches, is increasingly important in the field of network security to help analyze data using human intuition.

A few methods highlight botnet clues. IDS rainstorm [8] shows regular IDS alarm over two days which caused by botnets. DNS visualization [9] visualizes botnets in high rank DNS traffic. Krasser et al. [10] find IP scan patterns of botnets using parallel coordinates.

Our previous approach [11] described a visualization mechanism using DNS traffic. It can detect botnet evidence visually. However, there are unwanted noisy and false patterns. In this study, unlike the previous work, we develop a logical filter to acquire distinct images of botnets, and we define two more attack signatures related to botnets. In addition, we detect more abnormal behaviors such as DNS cache poisoning and spamming in real-life traffic.

3. BotXrayer

3.1 Benefits of Visualization in Botnet Detection

There are three main benefits in applying an information visualization approach to overcome heavy analysis for botnet detection.

First, visualization can easily deal with highly heterogeneous and noisy data from botnet traffic. The information of data acquired from a picture is much greater than having a human look at log files or textual data. Second, visualization can give us fresh insight into the analyzed data and allow us to deduce new hypotheses that are often lost in complex analysis. Even though an unknown botnet attack may have occurred, if an image pattern from the unknown attack is obtained, the attack can be quickly detected. Third, visualization is often much faster than other anomaly detection approaches. Looking at a picture enables a human to immediately realize what is really happening.

3.2 Selected Parameters

To distinguish between normal and abnormal traffic, the visualization mechanism uses DNS traffic from the botnets whose characteristics can be distinguished from legitimate DNS traffic [12]. We select four single-packet features from a DNS reply (answer) packet.

- **Host IP**: In a DNS reply packet, the destination IP address in IP header implies a DNS packet sender.
- **Target name**: A target name in an answer RR field is a queried domain name. The domain name can be a C&C name or an attack target.

– **Target IP**: IP addresses in an answer RR field represent an answer to the requested domain.
– **TTL**: TTL (Time To Live), in an answer RR field, means a time interval for caching before the source of the information should again be consulted. It is helpful in finding dynamic DNS(DDNS) [13] queries that often use fast flux service for botnets and are related to DNS cache poisoning attacks.

We also define four aggregate features calculated using the data from DNS.

– **Frequency**: Suspicious behavior, such as a DRDoS attack or DNS cache poisoning attack, frequently generates many DNS queries in a short time. We measure variance of DNS querying rates to measure the frequency.
– **Periodicity**: Queries generated by a program usually appear regularly. Therefore, measured periodicity can be used to find suspicious DNS queries generated by botnets.
– **Group size**: Botnets often send DNS queries in a coordinated fashion. We measure a number of the hosts that send similar DNS queries within a certain period to classify suspicious queries.
– **Abnormality**: If a DNS answer packet has a large TTL value or reserved IP address as an answer to a query or rapid change of DNS type (A,PTR,MX) occupation, the answer will be suspicious. We measure this abnormality by checking the aforementioned unusual queries.

3.3 Botnet Visualization Mechanism

In this section, we show how parallel coordinates can present botnets as a graph. Each coordinate is used to represent four different parameters: host IP, target name, target IP and TTL.We define the bottom of a coordinate to imply the minimum value (e.g., IP address coordinates, 0.0.0.0) and the top implies the maximum value (e.g., IP address coordinates, 255.255.255.255). We use a string hash function to point to the target name,. We use log scaled DNS TTL, because the value of TTL is not uniformly distributed. These four values enable the packet to be plotted as a connected line on parallel coordinates.

To amplify the abnormal pattern in a visualized image, we use colors to draw lines using suspicious rate estimation. Suspicious rate(S) is coupled with four aggreate features mentioned in the previous section; a frequency (F), a periodicity (P), a group size (G) and an abnormality (A). We derive an equation (1) as a measure for suspicious rate. Four aggregate features have a boolean value determined by their threshold (if they exceed the threshold, the value is 1. Otherwise, 0). α, β, γ and δ are the weights for each measure.

$$S = \alpha \cdot F + \beta \cdot P + \gamma \cdot G + \delta \cdot A \quad (0 \le S \le 1) \quad (1)$$

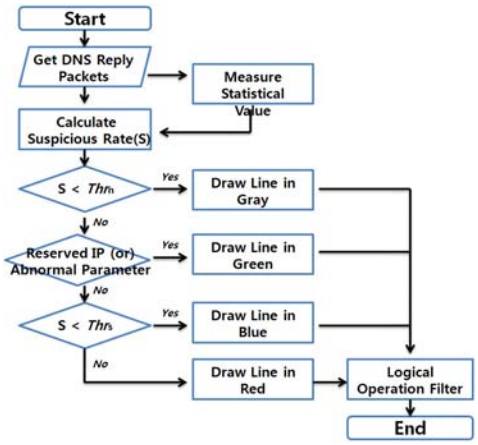


Fig. 1. Flowchart for line color selection

Using the suspicious rate, BotXrayer draws a suspicious DNS packet as red, green and blue colored lines. **Fig. 1** shows the flowchart to draw lines. BotXrayer determines the color of lines using two different thresholds, Thr_n, Thr_s. Each threshold is estimated from several experiments using real-life network traffic. If a DNS reply packet includes the loop-back IP address as a answer for the queried target name, or a large TTL value (more than a day), the packet is displayed as green. Traffic above the Thr_s threshold is drawn with a red line that denotes a suspicious object, otherwise a blue line is drawn. We can infer suspicious patterns intuitively using those colored lines.

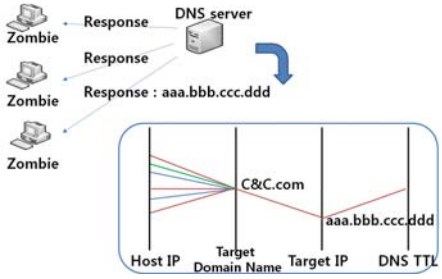


Fig. 2. Example of BotXrayer visualization

Fig. 2 depicts an example of the BotXrayer visualization. Suppose that some compromised hosts try to connect to a C&C server. At first, hosts send DNS queries to enquire an address of the C&C server. The hosts received apply packets and BotXrayer draws the domain feature that is a group of DNS packet lines. We can distinguish the image pattern of suspicious packets from the color of each line.

3.4 Graphical Signatures for Detecting Botnets

Table. 1. Graphical signatures of BotXrayer

	Pattern	Result
Botnet		Single Domain C&C
		Multi Domain C&C
		Fast Fluxed Domain C&C
Malicious Pattern		DRDoS Attack
		DNS Cache Poisoning
		Blacklisted Domain (Reserved IP)

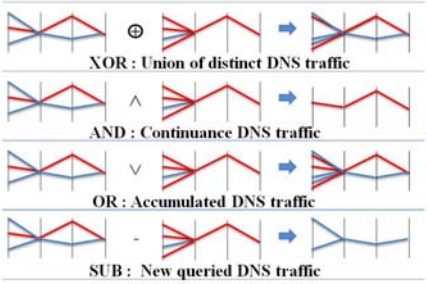
BotXrayer draws a graph on parallel coordinates and obains valuable patterns of graphs that we define as a graphical signature (**Table 1**). There are six visual signatures for botnets and malicious behaviors. A funnel-like pattern in **Table 1** indicates that infected hosts try to find the IP address of the C&C server or launch a DDoS attack using DNS. The pattern can also appear in normal communications. However, the measures used in estimating the suspicious rate can amplify abnormal patterns with colored lines. If a botnet has multiple C&C domain names, the visualized image generated by DNS of the botnet will be shown as a block-like pattern. Fast-flux, closely related to botnets forms a fish-like pattern, since multiple IP addresses are resolved to one domain name.

If a host generates DNS queries aggressively, the line is drawn as red. Several red lines indicates a DRDoS attack [14]. BotXrayer can display a graphical pattern of DNS cache poisoning attacks. Suppose that there is a domain name poisoned by Kaminsky's method [15]. In this case, the visualizer shows red lines pattern. If the attack succeeds, the IP address of the target domain will be changed.(The pattern can be distinguished using the OR filter that will be considereed in the following section). If the attacker sets a large TTL value to maintain an effect in DNS cache for a long time, the visualizer draws a large TTL value that looks like a tadpole pattern.

We observe a specific pattern that implies bots send DNS queries to know the IP address of disconnected C&C server that is blacklisted. This pattern is shown as a triangle with a connected line pattern as shown in **Table 1**.

3.5 X-raying Filter

Table. 2. Four logical operation filters



BotXrayer provides a filtering ability, termed X-raying. To reveal hidden botnets and to separate abnormal domain features. The X-raying filter amplifies a graph with more than two term images using logical operations, including XOR, AND, OR, and SUB as shown in **Table 2**. Regularly the visualization captures a current view image and saves previous term images in a history view. If a user selects two images in the history view with the XOR filter, the X-raying filter combines images. Only the remaining lines are shown after applying the XOR logical operation. When the remaining lines form an image, botnet DNS querying patterns, fast-flux pattern, DRDoS and DNS cache poisoning patterns will eventuate. Blacklisted C&C patterns isolated by the AND filter operation. Botnet hosts that have a centralized structure can be blocked to access the C&C server using DNS redirection mechanism. Once blocked, the bots continually send DNS queries for the blocked domain. Therefore, the domain features remain when applying the AND operation. The OR filter will merge images in order to monitor a long term view for slow attacks. Therefore a hidden botnet, like a slow growing botnet, can be revealed by the OR operation filter. The SUB filter is similar to the XOR filter. It shows newly appearing domain features. As a result, we can get clear view and image patterns of hidden attacks with the logical filters.

4. Evaluation

4.1 Experimental Setup

We acquired two different DNS traffic data sets to evaluate the implemented visualization system. First, we obtained DNS traces tapped from the gateway router of the 1Gb/s campus network that has a B class addresses, on May 19th, 2008. For the second, we got DNS traffic from the ISP network that has multi B class addresses, on July 3th, 2009. We created a prototype system of BotXrayer, shown in **Fig. 3**, to verify our mechanism.

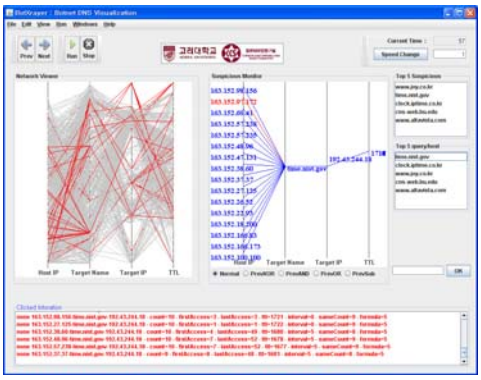


Fig. 3. Screenshot of a prototype implementation BotXrayer

4.2 Experimental Result

Test Using a Real-life Trace Using the obtained real-life network traffic, BotXrayer reports several images as suspicious patterns. We observed a funnel-like pattern as shown in **Fig. 4**. We found a target name '*time.nist.gov*' that hosts generated the largest DNS queries. The hosts infected by the storm botnet request DNS queries regularly to synchronize the clock time. The hosts create more than three queries per second. Moreover, the DNS traffic from the storm forms a suspicious pattern that can be distinguished from normal ones.

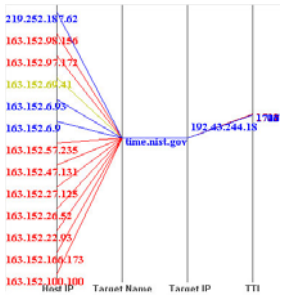


Fig. 4. The known domain feature of the botnet (Storm Botnet)

This is a multi domain botnet pattern that suspicious six hosts' queries of seven target names, as shown in **Fig. 5**. Each host creates one DNS query per second to the blacklisted domain. '*undernet.org*' provides an encrypted IRC chat service in which the botnet hides its command messages. Each domain like '*santaana.ca.su.undenet.org*' or '*mesa.az.us.undenet.org'*, is a different channel of Backdoor.IRC.zapchast .

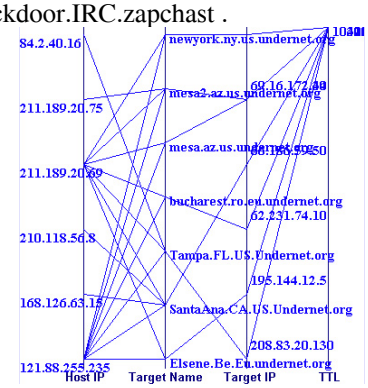


Fig. 5. The multi domain of the botnet (Backdoor.IRC.Zapchast)

Filtered Result To verify that BotXrayer can show a DNS cache poisoning attack using OR filter, we also experimented with DNS cache poisoning. A host performing the Kaminsky attack [15] causes many DNS packets to change an address of a target domain. After corrupting the DNS server, packets at the target site have relatively long DNS TTL values. Using the filter, the visualizer draws distinct views of changing the address, as depicted in **Fig. 6**.

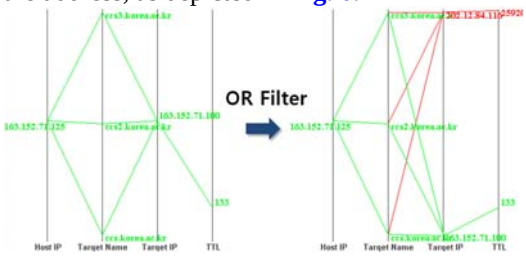


Fig. 6. Filtration effect on a DNS cache poisoning

Legitimate Site Fig. 7. Shows an example of the funnel-like pattern. These hosts create a DNS query per second that has zero second TTL. Therefore BotXrayer draws the pattern such as a single domain botnet. '*nasimg.nasmedia.co.kr*' is a banner advertisement service in a movie player. While the player is running, it creates DNS queries to obtain a new banner message regulary.

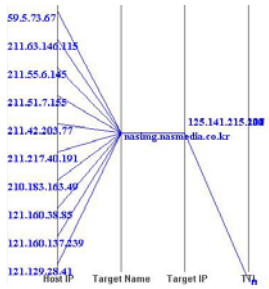


Fig. 7. Advertising site pattern

Spamming Pattern We acquired the unexpected attack pattern image shown in **Fig. 8**, during the evaluation using network traces. A deeper inspection, reveals a bots' DNSBL reconnaissance pattern. Botmasters frequently perform reconnaissance lookups to determine their bots' blacklist status [16]. DNSBL services respond whether or not the requested domain is blacklisted, using response IP address (127.0.0.1 for normal and 127.0.0.2 for spam). A host which has large DNS querying rates can be regarded as a bot's DNSBL reconnaissance. As mentioned in this section, BotXrayer shows the effectiveness for recognizing unknown and undefined malicious patterns visually.

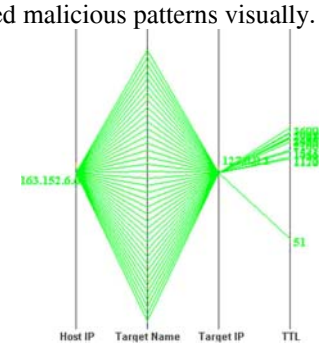


Fig. 8. Spamming botnet by DNSBL

5. Conclusions

In this paper, we described our development of a visualization mechanism using DNS traffic. The proposed visualization enables network administrators to discover botnets and malicious activities within normal traffic intuitively. Using the filter mechanism, we reveal hidden botnets and obtain distinct patterns. However, some false-positive patterns cause legitimate programs to look like botnets. It would be possible for a botnet to modify its behavior to bypass the mechanism. In future work, it is necessary to find more appropriate parameters to distinguish between legitimate programs and botnets.

References

[1] A. Karasaridis, B. Rexroad, D. Hoeflin, "Wide-scale botnet detection and characterization", in *USENIX Hotbot*, 2007

[2] M. Peck, "A brainy approach to image sorting", http://www.spectrum.ieee.org/apr08/6121

[3] A. Inselberg, "The plane with parallel coordinates", *The Visual Computer*, 1985

[4] D. Dagon, "Botnet detection and response", In *Proc. OARC Workshop*, 2005

[5] D. Keim, "Visual exploration of large data sets", *Communications of the ACM*, 2001

[6] G. Gu, P. Porras, V. Yegneswaran, M.Fong, W. Lee, "Bothunter: Detecting malware infection through IDS-driven dialog correlation", In *Proc. of USENIX Security Symposium*, 2007

[7] G. Gu, J. Zhang, W. Lee, "Botsniffer: Detecting botnet command and control channels in network traffic", In *Proc. of NDSS*, 2008

[8] K. Abdullah, C. Lee, G. Conti, J. Copeland, J. Stasko, "IDS rainstorm: Visualizing IDS alarms", In *Proc. of IEEE VizSEC*, 2005

[9] P. Ren, J. Kristoff, B. Gooch. "Visualizing DNS traffic", In *Proc. of VizSEC*, 2006

[10] S. Krasser, G. Conti, J. Grizzard, J. Gribschaw, H. Owen. "Real-time and forensic network data analysis using animated and coordinated visualization", In *IAW* 2005

[11] I. Kim, H. Choi, H. Lee, "Botnet visualization using DNS traffic", In *Proc. of WISA*, 2008

[12] H. Choi, H. Lee, H. Kim, "BotGAD: Detecting botnets by capturing group activities in network traffic", In *Proc. of COMSWARE*, 2009

[13] B. Zdrnja, N. Brownlee, D. Wessels, "Passive monitoring of dns anomalies", in *Proc. of DIMVA*, 2007.

[14] E. Courses, T. Surveys, "Motivation for behaviour-based DNS security: A taxonomy of DNS related internet threats", In *Proc. of SECUREWARE*, 2007

[15] D. Kaminsky, "Kaminsky on DNS rebinding attacks, hacking techniques", *Black Hat Briefings*, 2008

[16] A. Ramachandran, N. Feamster, D. Dagon, "Revealing botnet membership using DNSBL counter-intelligence". In *Proc. of SRUTI*, 2006