Tracking Botnets

M.Freydefont and M.Papadaki

Network Research Group, University of Plymouth, Plymouth, United Kingdom
email:info@cscan.org

Abstract

Botnets are not only a threat for companies under the pressure of Distributed Denial of Service (DDoS) attacks, but also at the origin of massive information theft, targeting the banking credentials of home-users. It is widely accepted that nowadays, botnets are the most challenging threat available on the Web. This paper is an attempt to study the feasibility of a tracking system which would shut botnets down in an automated fashion. The study is realized with a review of botnets monitoring techniques as well as a trend analysis in bots specifications. The results show that it is not realistic to imagine such automated "botnet-killer" system. Instead, an end-point defense strategy should be applied, putting the accent on educating people and improving the usability of security products.

Keywords

Bot, monitoring, trend analysis, defense

1 Introduction

These last years, malicious activity on the Internet has moved from the hackers community, motivated by technological challenges, to well-structured criminal associations (Ilet, 2005). Distributed Denial of Service (DDoS) attacks, spamming, phishing, information theft ... all these frauds have merged and are now embodied by a single entity: *the botnet*. The latter has become the favourite tool of cyber-criminals and at the same time one of the most challenging threats available on the Internet (Abu Rajab *et al*, 2007). Their distributed nature makes them hard to eradicate. The wide range of services they offer to their controller, moreover, the opportunities to make easy money, contribute to the professionalization of the underground economy. As a result, bots are getting more and more sophisticated, so hard to eliminate.

This paper investigates existing work that has been done to monitor botnets as well as the new trends in botnets specifications. The aim is to recommend areas where the efforts should be focused and propose ways to defend against them.

2 Background

A *bot* is a piece of malware that can perform tasks requested by a remote user. A *botnet* is the name given to a group of computers infected by a bot that enables a third body to perform malicious and distributed activity. The victim hosts are also

sometimes called *zombies*, *slaves* or *drones* and the controller of the botnet, *master* or *herder* (Myers, 2006).

There are three aspects that define bots:

- *Communication:* how the bot interacts with the master and how the bots are linked together. A *Command and Control* (C&C) server is the host where all the slaves connect to wait orders from the herder.
- *Propagation:* the way the botnet gets bigger. This includes the reconnaissance and the contamination phases.
- *Services:* the actions a bot can undertake and that make it interesting for the cyber-criminal.

At the beginning, IP addresses of C&C servers were hard coded in bot's code (Schonewille et al, 2006). The herders soon realized the obvious limitations of such practice: when a C&C server is taken down, all the clients are simply lost. They fixed this problem replacing the IP addresses by dynamic domain names (Schonewille et al, 2006): a domain name associated with an IP address that can be changed. This provides a great flexibility in the sense that when a C&C server is shut down, the herder just has to relink the domain name with the IP address of the new server for displacing all the zombies to their new headquarters.

IRC is an old protocol for text-based conference (Kalt, 2000a). It allows users to connect to a server and join a chat-room called *channel*. Hackers have found an application of this protocol to botnets. The slaves and the master meet up in a channel hosted by a server (the C&C server) where they can receive the commands sent by the master. IRC is considered as the legacy protocol for botnets (Myers, 2006).

3 Monitoring botnets

Reseachers have studied different approaches to monitor botnets, both from the inside, infiltrating the botnet or from the outside, analyzing visible traffic.

3.1 Honeynets

The honeynets enable researchers to gather information about botnets (Honeynet Project, 2006). However, they have two main drawbacks:

- Honeynets only enable local observation, it is not possible to get a broad view of the entire botnet. In an IRC botnet for instance, all the members are not always visible, due to IRC server options, RFC 2811 (Kalt, 2000b).
- Honeynets doesn't not allow to choose which botnet to monitor as the researcher has to wait to capture malware first.

A comparison of the DDoS attacks detected by incident reports (Peakflow SP statistic system) and honeynets (ShadowServer Foundation) showed that 13% of the attacks were detected by Peakflow against 2% for ShadowServer (Nazario, 2007). This demonstrates that botnet can not be tracked efficiently *only* using honeynets, a more global approach is required.

3.2 DNS Traffic analysis

In their paper entitled *"DNS as an IDS"* , Schonewille et al (2006) studied the potential of DNS servers for acting as detection systems. The hypothesis is the following: infected systems sometimes give information about themselves when making DNS queries. Information about the infection and the source may be extracted with analysis of those queries. The researchers of this study drew the conclusion that the DNS traffic analysis is limited is terms of botnet detection capacity, mainly due to the false positives raised. Also the analysis of the data is highly cpu-intensive and the cache on the client obscurates the real activity.

Another interesting study concerning DNS monitoring has been made, this time in the context of spamming botnets *"Revealing Botnet Membership Using DNSBL Counter-Intelligence"* (Dagon et al, 2006). DNSBL (DNS Black List) databases are normally used by regular mail servers (mail transport agent) to reject or flag spam messages (Wikipedia, 2007a). Nevertheless, they are also used by botmasters who perform lookups to check whether their spamming bots are blacklisted or not. Indeed, to rent a botnet for spamming purpose, the herder must insure that the bots are "clean"...(Dagon et al, 2006). The researchers used graphical analysis to distinguish legitimate lookups from reconnaissance queries. The origin and targets of suspicious queries are likely to be bots. *"With the ability to distinguish reconnaissance queries from legitimate queries, a DNSBL operator might be able to mitigate spam more effectively."* (Dagon et al., 2006).

3.3 Distributed Detection Systems and Algorithms

In their paper entitled "*A Distributed Host-based Worm Detection System"*, Cheetancheri et al (2006) present a distributed system for detecting large-scale worm attacks using only end-host detectors: End-host detectors monitor the traffic they can see and determine if there is an attack or not. But because of the limited view they have on the traffic, we cannot assume their detection quality is high. Therefore, information from many detectors is correlated and a Likelihood Ratio is then computed (probability that an attack actually occurs). In order to make the collaboration working, a complete protocol has been developed for exchanging the alert messages. It is a completely distributed approach with no single points of failure. Nevertheless, the system, while promising, is not finalized yet as it remains aspects to address. For instance, the system has only been tested using simulations and within a local area network; it does not take in account the worm traffic from outside (Cheetancheri et al, 2006). Moreover, this distributed system only relates to the early step of botnet construction (worms, whether they are mass-mailing or

propagates by exploiting vulnerabilities are the main vector for bot spreading). It is not designed to monitor existing botnets' activity.

Finally a team from the Portland State University in USA has developed an anomaly-based algorithm for detecting IRC-based botnet meshes (Binkley, Singh, 2006).Their system combines an IRC parsing component with a syn-scanner detection system. The algorithm is not signature-based and doesn't rely on any known port number or IRC command string. The system can clearly show the presence of botnets but there are more fuzzy cases where further analysis is necessary to determine whether the activity is actually suspicious or not. The technology employed here relies on attackers launching attacks (scans), therefore, there is no guarantee for every infected system to be detected. Also, one could argue that anomaly detection is "too late" (a host has already been exploited)

4 Trend Analysis

Bots' specifications evolve constantly and it is important to follow them to make sure for instance that the monitoring techniques developed do not become obsolete. This section presents the tendencies in terms of propagation, communication and services.

4.1 Methodology

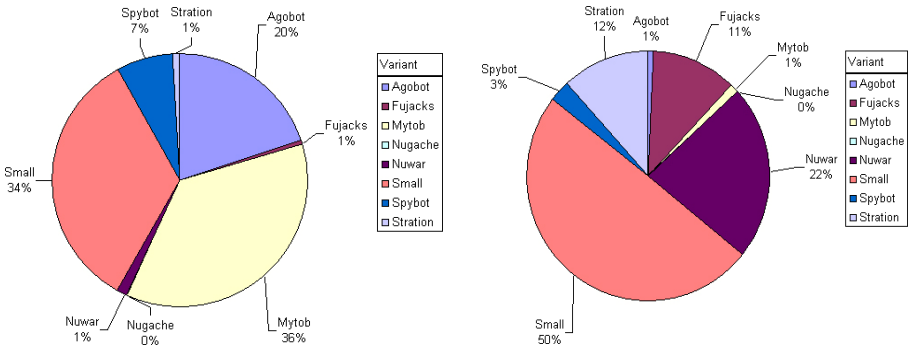
The antivirus vendor Trend Micro provides a comprehensive Virus Encyclopaedia (Trend Micro, 2007a), of malware variants, either caught by sensors or submitted by antivirus users. Trend Micro has been chosen because of the section *Statistics* available for any variant's description. It summarizes the number of infections made by the variants since it was firstly reported to Trend Micro. A program written in Java has been designed to crawl the website and collect information in an automatic fashion. This program receives as input a list of malware families and produces for each family, a file listing all the variants repertoried, their date of release and their total number of infections. The program also computes the number of days between the date of release of a variant and another specified date (6th of July, date chosen arbitrarily). The files generated by the Java Malware Crawler are used as sources of data imported in Excel sheets. This way, it is easy to sort the data following various criterions and create charts that give sense to the numbers collected. Seven families have been selected (the family names used come from the Trend Micro naming system):

- **PE_FUJACKS**: a recent family of profit-driven pieces of malware that have the particularity to propagate mainly infecting files.
- **TROJ_SMALL, WORM_NUWAR**: TROJ_SMALL.EDW aka Storm worm was the first outbreak of the year 2007. This variant creates botnets communicating with a peer-to-peer scheme. Moreover, it is an example of collaboration with another recent threat, WORM_NUWAR.
- **WORM_STRATION**: Stration is an HTTP-based botnet used mainly for spam.

- **WORM_NUGACHE**: Nugache is another example of peer-to-peer botnet but which also use encrypted communications.
- **WORM_AGOBOT:** probably the most popular malicious code since it first gave the possibility for hackers to assemble and thus to create their own variants, selecting modules through a user-friendly graphical interfaces.
- **WORM_SPYBOT, WORM_MYTOB:** other veteran families that have been on the front stage in the past years (i.e. often referenced in the literature).

4.2 Results

Each family has a total number of infection related which is equal to the sum of infections performed by all variants. The diagram shows the shares of infections amongst the families selected since 2001. The diagram on the right, on the other hand shows the infection shares only performed by variants released in 2007.



Since 2001 *In 2007*

Figure 1: Evolution of the infection shares amongst the selected families

Both statistics of infections and variants released (see table 1) demonstrate that the focus of the hackers' community seems to have moved from the legacy families Agobot, Mytob to new families such as Nuwar or Stration. The peak of interest in Small family is certainly explainable by the "success" of the variant Small.EDW, aka Storm worm, released in January 2007.

These recent families propagate by email, as for recent variants of Small family. Email propagation is not new but this channel is now prevalent. Talking only about emails is nonetheless passing by the real trend that stands behind: the social engineering attacks. Indeed, the choice of propagating over SMTP protocol is only relevant to bypass firewalls as well as IDS/IPS systems (of course when the mail has not been dropped by anti-spam counter-measures). But the human user still has to

fall in the social engineering scheme to trigger the infection. Unfortunately, the choice made by the malware creators seems to prove that it is more efficient to target the humans than software vulnerabilities for instance.

Family	2001	2002	2003	2004	2005	2006	2007
AGOBOT	0	3	68	644	423	38	26
FUJACKS	0	0	0	0	0	6	49
MYTOB	0	0	0	0	310	57	64
NUGACHE	0	0	0	0	0	2	2
NUWAR	0	0	0	0	0	7	716
SMALL	1	8	11	315	315	258	2931
SPYBOT	0	0	24	265	188	62	307
STRATION	0	0	0	0	0	148	293

Table 1: Number of variants released each year

Another recent strategy that has emerged is the collaboration with other pieces of malware, either from the same family or not. The best example of such strategy is the collaboration between Nuwar.CQ and Small.EDW (aka Storm worm).

A spam attack started at the beginning of January 2007. Recipients received emails with, as an attachment, a so-called video of the storms that hit the Europe in December 2006. Of course the attachment was not a real video but a Trojan TROJ_SMALL.EDW (Trend Micro, 2007b). So far, a classic scheme. What is less common however is that the Trojan Small.EDW downloads another mass-mailer worm, NUWAR.CQ which in its turn drops Small.EDW: this way they help each other to propagate.

The two pieces of malware have different goals and use two completely different topics as social engineering attacks. Small.EDW exploits the European storms while NUWAR.CQ used the incoming (at the time of the attack) Valentine's Day to fool the recipients (Trend Micro, 2007b): one can think he avoided the first trap but can still fall then into the second one!

A growing change concerns the topology and protocols for botnet communications. Specialists agree to say that the traditional IRC C&C servers are not trendy anymore. Instead, the shift has operated towards peer-to-peer architecture (Symantec, 2007a). Phatbot, Nugache or Small.EDW are examples of bots that has adopted peer-to-peer architecture. Small.EDW even use an open network, eDonkey (Dagon et al, 2007), that makes it harder to monitor since the activity is mixed with the rest of the users. Nevertheless, peer-to-peer botnets studied so far keep a failure point as they use static resources (hard coded list of hosts, cache servers, etc…) for the initial peer discovery. Their topology is not completely de-centralized yet and therefore, remains detectable and vulnerable.

In terms of services, DDoS attacks have decreased *"Symantec recorded an average of 5,213 DoS attacks per day, down from 6,110 in the first half of the year."* (talking

about year 2006, see Symantec, 2007a).The fact is DDoS have been studied and counter measures now exist like the Cisco Guard products (Cisco, 2007). Another possible explanation is that cyber-criminals try as much as possible to avoid direct contact with the victim. This is incompatible with the concept of extortion. It is also noteworthy to highlight that recent families amongst the ones that have been selected for this chapter (Stration, Nuwar, Fujacks), are not firstly designed to perform DDoS attacks (Trend Micro, 2007a).

On the other hand spamming and information theft are very active (Symantec, 2007a). Recent families such as NUWAR or STRATION make spamming their main goal:

- NUWAR broadcasts "pump-and-dump" spam to create artificial demand on financial stocks owned by the zombie network's creators (Trend Micro, 2007b).
- STRATION sends out pharmaceutical spam. It uses spam images in order to evade anti-spam rules (Trend Micro, 2007c).

Spam will certainly remain a privileged activity for the coming years.

5 Taking botnets down

An important question that motivated the authors to undertake this research was: is it possible to design a system that tracks and dismantles botnets in an automated fashion?

So far, it is does not seem possible, the challenges are far too numerous. Indeed such system should have a global view of the Internet (i.e. distributed and/or located at ISP level), detect accurately botnets, not affect legitimate traffic and act in agreement with the legal and ethical issues.

The monitoring techniques reviewed present good qualities but still suffers of limitations in terms of visibility (honeynets and distributed sensors), reliability (DNS as an IDS) or adaptability (the Portland University's botnet detector only works for IRC botnets). The use of fast-flux DNS (Lemos, 2007) makes even more complicated botnets takedown since the C&C servers are highly redundant. Finally, is the Cox Communication case (McKeay, 2007) well illustrates the problem of legal and ethical issues: this internet provider decided to re-route IRC traffic towards its own servers with the aim to uninstall the bots trying several commands. Whether the uninstallation attempt works or not, the bot is at least disconnected from its network. The idea would be nice if it did not also affect legitimate IRC traffic and thus the activity of some professionals using this protocol within their business. Moreover, as it is pointed out by Martin McKeay, we can wonder about the "intrusion" of the ISP in its customers system (McKeay, 2007).

The monitoring systems can prove useful to collect intelligence that will feed the security community and vendors but a lot of obstacles prevent actions to be undertaken afterwards to shut down botnets.

6 Conclusion

To defend against botnets, end-point defence strategy should be rather adopted: unfortunately, it is unrealistic to imagine cleaning the Internet of botnets given the state-of-the-art of monitoring technique as well as legal and ethical issues. Vendors offer products and services that can help to mitigate the threat, like Trend Micro's Botnet Identification Service (Trend Micro, 2007d) or Norton AntiBot (Symantec, 2007b). However, usability of security products in general should be improved again and again to foster their use by non-skilled people.

Educating/training users is essential: bots are first of all malware. As for any malicious pieces of code, the best way to be protected is not to execute them. The trend analysis showed that privileged propagation methods use social-engineering schemes. We could imagine a certification in "IT security awareness" that employees in a company, must pass. This certification would ensure that employees will not misbehave under social engineering attacks. Such certification could be required by the companies as a basic but should be light and quick to take. The simpler the certification is, the more chances there are that the certified people educate their family or friends afterwards. Communication through mass-medias can complete the population training.

Finally, working on new security architectures and/or protocols is certainly the key to make bots unusable. A fundamental difference between a bot and a human user is that the latter is...human. As a result, he is capable to pass very simple challenge while a program cannot, such as a CAPTCHA test (Wikipedia, 2007b). This difference should be exploited and integrated in new security architecture for operating systems.

7 References

Abu Rajab, M., Monrose, F., Terzis, A., Zarfoss, J. (2007) *My Botnet is Bigger than Yours (Maybe, Better than Yours)* [online] Available:http://66.102.9.104/search?q=cache:oeiQ7caR1E0J:www.usenix.org/events/hotbots07/tech/full_papers/rajab/rajab.pdf+prevalence+of+IRC+botnets&hl=en&ct=clnk&cd=2&gl=uk [Date accessed: Thursday 1th February 2007]

Binkley, J., Singh, S. (2006) *An Algorithm for Anomaly-based Botnet Detection* [online] Available: http://web.cecs.pdx.edu/~jrb/jrb.papers/sruti06/sruti06.pdf [Date accessed: 2nd March 2007]

Cheetancheri, S., Agosta, J.M., Dash, D., Levitt, K., Rowe, J., Schooler, E. (2006) *A Distributed Host-based Worm Detection System* [online] Available:http://delivery.acm.org/10.1145/1170000/1162668/p107-

cheetancheri.pdf?key1=1162668&key2=1411867711&coll=&dl=ACM&CFID=15151515&CFTOKEN=6184618[Date accessed: 27th April 2007]

Cisco (2007) *Cisco Guard DDoS Mitigation Appliance* [online] Available: http://www.cisco.com/en/US/products/ps5888/index.html [Date access: Wenesday 4th April 2007]

Dagon, D., Feamster, N., Ramachadran, A.(2006) *Revealing Botnet Membership Using DNSBL Counter-Intelligence* [online] Available: http://www-static.cc.gatech.edu/~feamster/papers/dnsbl.pdf [Date accessed: 17th March 2007]

Dagon, D., Grizzard, J., Nunnery, C., Kang, B., Sharma, V. (2007) *Peer-to-Peer Botnets: Overview and Case Study* [online] Available:http://www.usenix.org/events/hotbots07/tech/full_papers/grizzard/grizzard_html/ [Date accessed: Thursday 10th May 2007]

Honeynet Project (2006) *Know Your Ennemy: Honeynets* [online] Available: http://www.honeypot.org/papers/honeypot/index.html [Date accessed: Wednesday 10th January 2007]

Ilet, D. (2005) Official: Cybercrime is growing [online] Available: http://news.zdnet.co.uk/security/0,1000000189,39193449,00.htm [Date accessed: 22nd March 2007]

Kalt, C. (2000a) *RFC 2810 Internet Relay Chat: Architecture* [online] Available: http://www.irchelp.org/irchelp/rfc/ [Date accessed: Tuesday 16th January]

Kalt, C. (2000b) *RFC 2811 Internet Relay Chat: Channel Management* [online]Available: http://www.irchelp.org/irchelp/rfc/ [Date accessed: Tuesday 16th January]

Lemos, R. (2007) *Fast flux foils botnet takedown* [online] Available: http://www.theregister.co.uk/2007/07/11/fast_flux_botnet/ [Date accessed: 19th August 2007]

McKeay, M. (2007) *Should your ISP protect you from yourself ?* [online] Available: http://www.computerworld.com/blogs/node/5908?source=NLT_VVR&nlid=37 [Date accessed: Saturday 11th August 2007]

Myers, L. (2006) *AIM for Bot Coordination* [online] Available:http://www.mcafee.com/us/local_content/white_papers/threat_center/wp_vb2006_myers.pdf [Date accessed: 22nd March 2007]

Nazario, J. (2007) *Botnet Tracking: Tools, Techniques, and Lessons Learned* [online] Available: https://www.blackhat.com/presentations/bh-dc-07/Nazario/Paper/bh-dc-07-Nazario-WP.pdf [Date accessed: 25th April 2007]

Schonewille, A. ,Van Helmond, D-J. (2006) *The Domain Name Service as an IDS* [online] Available: http://staff.science.uva.nl/~delaat/snb-2005-2006/p12/report.pdf [Date accessed: 2nd March 2007]

Trend Micro (2007a) *Virus Encyclopedia* [online] Available: http://www.trendmicro.com/vinfo/virusencyclo/default.asp [Date accessed: Tuesday 17th July 2007]

Trend Micro (2007b) *TROJ_SMALL.EDW Storms into Inboxes, Teams Up with NUWAR to Create Unique Network* [online] Available:http://www.trendmicro.com/vinfo/secadvisories/default6.asp?VNAME=TROJ%5FSMALL%2EEDW+Storms+into+Inboxes%2C+Teams+Up+with+NUWAR+to+Create+Unique+Network&Page= [Date accessed: Monday 9th July 2007]

Trend Micro (2007c) *The STRATION Strategy* [online] Available:http://www.trendmicro.com/vinfo/secadvisories/default6.asp?VName=The+STRATION+Strategy [Date accessed: Wenesday 11th July 2007]

Trend Micro (2007d) *Botnet Identification Service* [online] Available: http://us.trendmicro.com/imperia/md/content/us/pdf/products/enterprise/botnetidentificationservice/ds03_bis_070725us.pdf [Date accessed: Thursday 9th August 2007]

Symantec (2007a) *Symantec Internet Security Threat Report Trends for July-December 06* [online] Available: http://eval.symantec.com/mktginfo/enterprise/white_papers/ent-whitepaper_internet_security_threat_report_xi_03_2007.en-us.pdf [Date accessed: Wenesday 4th April 2007]

Symantec (2007b) *Symantec Arms Consumers Against PC Hijackers with Norton AntiBot* [online] Available: http://www.symantec.com/about/news/release/article.jsp?prid=20070717_02 [Date accessed: Sunday 12th August 2007]

Wikipedia (2007a) *DNSBL* [online] Available:http://en.wikipedia.org/wiki/DNSBL [Date accessed: Tuesday 12th June 2007]

Wikipedia (2007b*) CAPTCHA* [online] Available: http://en.wikipedia.org/wiki/Captcha [Date accessed: Saturday 18th August 2007]