Metric design for botnet mitigation

Santiago Aragón*, Daniel Meinsma*, Pallavii Jagannatha* Owais Ahmed*,
* University of Twente, Nederlands

***Abstract*—**

***Keywords*—**

I. What security issue does the data speak to?

It is impractical and very expensive to defend every possible vulnerability [1]. Therefore, decision-makers must plan effective security strategies about where and when to invest time and money [2]. To make decisions on security investments, security investment models are used and these models are build on security metrics [3]. Not all vulnerabilities will be attacked, so knowing the most likely attack vectors is key in planning effective security measurements [1]. For this knowledge, data about attacking behavior is needed. If we have this data, we can define metrics that measure different aspects of attacking behavior.

One way of getting data about attacking behavior is by using *honeypots*. Honeypots are systems made vulnerable intentionally to attract attackers and track their behavior [4]. An example of a honeypot is *Elastichoney* [3]. Elastichoney is an elasticsearch honeypot containing the RCE Remote code execution vulnerability -CVE- 2015-1427, which allowed attackers to execute java based code in its search bar [5]. The log files of Elastichoney contain data collected over two months and tracked about 8k attempts to attack from over 300 unique IP addresses [3].

When looking at the data collected by Elastichoney, a major security issue can be found, namely *Botnet Infection*. The attacking behavior of botnet infection found in the data can be summarised as follows:

- Using the code execution vulnerability, operation steps were performed on multiple different files: *wget* file, *chmod 777* on file location, execute file, removing file.
- Operations on the same file were requested by multiple different sources.

In this paper we will look at two parties who have great incentive to know about the attacking behavior of botnet infection, namely *Internet Service Providers (ISP's)* and *Enforcement agencies*.

Botnet infection is a major security issue for ISP's, because 27% of the overall unwanted traffic on the Internet can be attributed to botnet-related spreading activity [6]. And this is a great loss in productivity for ISP's.

Enforcement agencies want to counter botnet infection, because botnets are heavily used as a platform for various criminal business models like sending spam, committing click fraud, harvesting account credentials, launching denial-of-service attacks, installing scareware and phishing [7].

The rest of this paper will discuss metrics that measure different aspects of the attacking behavior of botnet infection and how these metrics can be used by ISP's and enforcement agencies in making effective security investments.

II. What would be the ideal metrics for security decision makers?

The security issue defined in section I have distinct effects depending on the issued party. We identify several parties that might be aimed by the security issue and we arrange them in two groups which are affected in similar ways. For each group, we identify how the botnet (BN) is negatively influencing their interests, as well as a set of ideal metrics which could help each party to derive better decisions to take the more effective counter measures to mitigate their problem while maximizing the return on security investment.

a) Internet Service Provider (ISP) & Botnet's victim: For this parties it is of main interest to avoid misuse of resources i.e. the ISP would like to block not legitim incoming traffic to its network while the botnet's victim would like to maintain availability of its resources.

b) Metrics:

- Attack prediction.
- BN identity detection.
- Resources (un)availability cost.

c) Regulator authority & Law enforcement agencies: The parties interested in prevent, prosecute and punish the ones behind a security issue, i.e., For the regulator authority is of main interest to transfer the risk to the ISP, to encourage the zombies machines to increase their security protection, while for a law enforcement agency is to have mechanisms to hunt down the attacker

d) Metrics:

- BN location detection

III. What are the metrics that exist in practice?

In the previous section we have seen which kind of metrics would be ideal for different parties facing the botnet issue. Before going to the metrics proposed for each party by this paper, first a list is presented of metrics that are already used by honeypots for botnet detection:

e) Network Telescope: This is a control metric.

A block of Ip addresses from the entire range of IPv4 addresses are unassigned to hosts. This network is called "darknet". These block of ip addresses are still advertised on the internet through Border gateway(BGP) protocol making it BGP reachable. If any host from anywhere in the world(on the internet) sends a packet to one of these addresses, this packet would travel all over the world, would reach the router that

advretises this routes, would be silently dropped (without any repsonses) but this would be logged. Network telescopes would be used to observe this internet traffic. By definition, this traffic is unsolicited since it does not have any hosts assigned to the addreses.Most of this unsolicted traffic would be malicious i.e traffic from malware, traffic from infected hosts that randomnly scan entire internet address space and so on [8].

Enforcement agencies use this to create metric out of samples of telescope data containing security event signatures. This metric would inform about possible network attacks, botnet activities and other misconfigurations.

f) Network Fingerprinting: This is a control metric.

With this method, one can create a metric that states which hosts communicates to which hosts. For example, one can track all the IP addresses , the honeypot communicates to. The honeypots will only communicate to the controller and in a peer to peer botnet , the honeypots will create multiple other members of the botnet [9].

Enforcement agencies use this to create metric which contains information of traffic logs that are automatically processed to extract a network fingerprint, the targets of any DNS requests, the destination IPaddresses, the contacted ports (and protocols), and whether or not default scanning behavior was detected. This would be used to differentiate between different types of botnets for example if the honeypot is part of a traditional command and control botnet [6].

g) IRC related features: This is a control metric.

With this method, one can create metric that differentiates between a member of a IRC type botnet and a non-infected member because these type of botnets send and receive signature commands over IRC channels [6].

Enforcement agencies use this to create metric which contains information of initial password to establish an IRC session with the server, the format of the nickname and username chosen by the bot, the particular modesset, and which IRC channels are automatically joined (with associated channel passwords).This is used to identify infected members(botnet) in the network [6].

Network fingerprinting and IRC related features provide enough information to join a botnet in the wild [6].

h) Longitudinal tracking: This is a incident metric.

With this method, one can create metric where you can visualise the number of attacks originating from a particular geographical location [6].

Enforcement agencies use this to create metrics containing information about geographical location to track the location of origination of a specific botnet.

i) DNS tracking: This is a incident metric. This method is almost the same as Longitudinal tracking, but instead of tracking the geographical location, this method tracks the domain names [6].

Enforcement agencies use this to create metrics containing information about domain names. This is used to probe the caches of a large number of DNS servers in order to infer the footprint of a particular botnet (total number of DNS servers giving cache hits) [6].

j) Botnet resources tracking: This is a incident metric.

In figure 1,you can see different resource aspects of botnet and each of these resources can be used to create metrics which differentiates between different types of botnets [9].

Enforcement agencies uses this to create metrics for each resource to characterize different botnets. It contains information about distinguishing characteristics. For example, peer-to-peer botnets would have network characteristics like distinctive communication graph, higher command latency and so on [9].

k) Signature tracking: This is a prevented losses/impact metrics.

Enforcement agencies uses this to create a metric that tracks the infection rate of a specific botnet by using the signature of a specific botnet, for example a trojan horse and its backdoor port as signature [10].

IV. A DEFINITION OF THE METRICS YOU CAN DESIGN FROM THE DATASET

Our main focus for the analytics will include identifying clear statistics on events from whom, from where and what attacks recorded in the JSON logs available. The following metric definitions will help us critically analyze the nature and characteristics of attacker behavior, the means of predicting the attack, the severity of the incident and resources needed to mitigate the impact of this vulnerability.

l) Attacker Behavior:: The inputs for this metric will include geographical locations / regions such as continent, country and city name. The different methods and types of various user agents used to exploit the vulnerability paying particular attentions to the application names. The majority of source IPs sending data traffic. The form and payload strings matching the actual exploit. The number of recurring attacks from particular IPs. The most important of all is the breakdown of the individual payload and its various instruction sets. Thus by analyzing these trends and rankings of different key data fields, one can predict the methodology and characteristic of the attackers behavior.

m)Attack Prediction / Probability:: The required data for this metric will include specific port numbers that are being used including the majority of source IPs sending data traffic via URLs with payloads attached. The method and nature of data transfer via URLs for direct path to vulnerabilities that are exploitable. The number of port scan attempts without including payloads and recurring attacks from particular IPs. These datasets will help to derive patterns for attack prediction and probability of successfully exploiting the vulnerability.

n) Scale of Incident:: The inputs for this metric will include the payload strings matching the actual exploit. The amount of valid and invalid vulnerabilities, malicious IPs and the percentage of false positives. The sum of individual attacks within a specific time interval. This metric will therefore help to determine the nature and level of attack strength and effort put in to gain access to critical infrastructure.

V. An evaluation of the the metrics you have defined. This should include graphical representations of the metrics (e.g., histograms, scatter plots, time series, bar charts).

References

[1] M. Rosenquist, "Prioritizing info security risks with threat agent risk assessment." Intel Information Technology, 2009.

[2] W. A. J. S. B. Sonnenreich, "Return on security investment (rosi) – a practical quantitative model." Journal of Research and Practice in Information Technology, 2006.

[3] R. Böhme, "Security metrics and security investment models," in *Advances in Information and Computer Security*. Springer Berlin Heidelberg, 2010, pp. 10–24.

[4] P. Wang and et al., "Honeypot detection in advanced botnet attacks," in *nternational Journal of Information and Computer Security 4.1*. USENIX, 2010, pp. 30–51.

[5] Common vulnerabilities and exposures:cve-2015-1427. [Online]. Available: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2015-1427

[6] M. Abu-Rajab and et al., "A multifaceted approach to understanding the botnet phenomenon," in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. ACM, 2006.

[7] R. Anderson and et al., "Measuring the cost of cybercrime," in *The economics of information security and privacy*. Springer Berlin Heidelberg, 2013, pp. 265–300.

[8] van Rijswijk-Deij. and et al., "Dnssec and its potential for ddos attacks: A comprehensive measurement study," in *Proceedings of the 2014 Conference on Internet Measurement Conference*. ACM, 2014.

[9] Grizzard, J. B., and et al., "Peer-to-peer botnets: Overview and case study," in *Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*, 2007.

[10] M. Andreolini and et al., "Honeyspam: Honeypots fighting spam at the source," in *Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop. USENIX Association*. USENIX, 2005.