Metric design for Botnet Infection problem

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

I. Introduction

It is impractical and very expensive to defend against every possible vulnerability [1]. Therefore, decision-makers must design and implement effective security strategies about where and when to invest time and money [2]. To make decisions on security investments, models are used and these models are build on security metrics [3]. Since the definition of these metrics are the basis of a good security model and therefore, they are a key element to produce valuable information for the decision-makers. For example, not all vulnerabilities will be attacked, thus knowing the most likely attack vectors is key in planning effective security measurements [1].

One way of gathering attacker information is by using *honeypots*. Honeypots are isolated and monitored systems that emulate to be vulnerable in order to attract attackers and track their behavior [4], i.e., attack attempt data could be used to define metrics that capture different aspects of attacker behavior. In particular we take a closer look to *Elastichoney* [3] project. It emulate to be a elasticsearch server suffering of a remote code execution (RCE) vulnerability identified as CVE- 2015-1427. This vulnerability allow attackers to execute Java based code by querying the server [5]. Furthermore, the Elastichoney project have gathered data over two months and tracked about 8k attempts to attack from over 300 unique IP addresses [3].

In Elastichoney logs the major security issue that can be found is *Botnet Infection* attempts, i.e. when a botnet machine is trying propagate the botnet by infecting another machine.

In this work we take a look to two different parties who have great incentive to know about the attacking behavior of botnet infection, namely *Internet Service Providers (ISP's)* and *Law enforcement agencies (LEA's)*.

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. On the other hand, LEA's would counter this security issue, 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].

In the rest of this paper we discuss ideal and state-of-the-art metrics that measure different aspects of botnet infection and how these metrics can be used by ISP's and enforcement agencies to make effective security investments. Finally, we propose, implement compare and evaluate our own metrics.

II. Ideal metrics

The security issue defined in section I have distinct effects depending on the issued party. For each party, we identify how the botnet infection (BNI) is negatively influencing their interests, as well as a set of ideal metrics which could help each party to be able to explain and react in a more informed fashion against the BNI issue, and therefore, derive better decisions to take the more effective counter measures to mitigate this problem while maximizing the return on security investment.

a) ISP: For this party 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.

- *BN propagation.* The growth and propagation trend are an ideal tool to quantify and predict the cost of malicious activity within the ISP network.
- *Resources (un)availability cost.* The ISP would like to know how much traffic of botnet infection activity is passing through his network. If is able to distinguish botnet activity from user activity he could measure how much resources are being misused and estimate the cost of having botnet inside his network and therefore justify the investment on security countermeasures.
- *Network infection rate.* The ISP could quantify his subnet infection rate prevention and reaction policies would be better targeted, i.e., if the subnet x.x.x.0 is heavy infected the ISP could target only this portion his network.

b) LEA: This party is interested in prevent, prosecute and punish the entity behind a security issue, i.e., to prevent BNI is of main interest to transfer the risk from the BN victims to the ISP and/or to encourage the zombies machines to increase their security protection, while for prosecution and punishment it is important to have mechanisms to understand and hunt down the BN.

- *BN geographical and digital location.* With geographical and digital location metrics the interested party could measure the infected population in a particular region either in the physical or digital world, i.e. how many infected machines are in the south of Chine, or how many infected machine does a ISP have.
- *BNI target machines.* By characterizing which kind of machines are the weakest link in a network, the LEA could launch prevention campaigns describing which kind of machines are more probable to become part of a botnet, i.e., operative system version, browser, etc.

III. State-of-the-art metrics

In the previous section we have seen which kind of metrics would be ideal for ISP's and LEA's in facing the botnet infection issue.

Current research mostly use honeypots to detect where botnet infection traffic is coming from and how it grows over time. In these cases, metrics set out the number of connections/attacks over time to the different IP addresses/longitudinal locations/domain names [6].

Other research has been done in differentiating between botnet types by looking at the connection and resources features of different botnet types. After classification, metrics set out the number of connections/attacks over time to the different botnet types [8], [9].

And last, research has been done in training honeypots to automatically detect botnet infection traffic and metrics are used to show the percentage of correct detections/classifications for th different botnet types [10]. Some examples of this metrics are mentioned here:

c) Cluster quality : The parameters such as total number of flows, Net flow size, internal host count, concurrently active, start date , end date and length(days) of botnets are measured. The cluster quality is then analyzed by using accumulated botnet samples and traces [10].

Enforcement agencies use this metric to know about possible network attacks and botnet activities.

d) Detection rate : The parameters such as total number of flows, Net flow size, internal host count, concurrently active, start date , end date and length(days) of botnets are measured. The detection rate is then analyzed by using accumulated botnet samples and traces [10].

Enforcement agencies use this metric to know about possible network attacks and botnet activities.

e) Infection rate : The parameters such as total number of flows, Net flow size, internal host count, concurrently active, start date , end date and length(days) of botnets are measured. The detection rate is then analyzed by using accumulated botnet samples and traces [10].

Enforcement agencies use this metric to know about possible network attacks and botnet activities.

f) Connection ratio : This is the ratio of number of bots in the largest connected graph to number of remaining bots (peer to peer botnets).

Enforcement agencies use this metric to know how well a botnet survives a defence action.

g) Connection degree ratio : This is the ratio of average degree of the largest connected graph to average degree of original botnet(peer to peer botnets).

Enforcement agencies use this metric to know how densely the remaining botnet is connected together after bot removal.

h) Network diameter : Network diameter is the average geodesic length of the network. The dynamics of the network such as communication, information and epidemics are slow if ”l” is large. This metric serves to be relevant because with every message passed through botnet , probability of failure or disconnection exists. [11].

Enforcement agencies use this metric to know about efficiency of a botnet. A botnet is evaluated based on it communication efficiency depending on its role(used to forward command and control messages, update bot executable code, gather host-based information)

i) Botnet Robustness : Bots generally lose and gain new members over time. If the victim machines are performing tasks like storing files for download or sending spam messages from a queue, a higher degree of connection between bots provides fault tolerance and recovery [11].

Enforcement agencies use this metric to know about robustness of a botnet.

IV. Proposed metrics

In this section, we propose a number of metrics that can be used to extract valuable information from the *Ellastichoney* dataset. For each metric we define its input, output and the utility for the interested party mentioned in section II. For each ideal metric that we define in section II we name some metric implementations that help us to extract valuable information from the dataset in use.

A. BN propagation

1) BN growth: It measures the number new IPs where the BNI attempts are being performed. It is useful to see the aggregate growth of the BN, and to help to infer a BNI rate.

2) BNI attempts: It measures the number of attacks executed daily. We will use this data to analyze the BN activity. This metric could be useful to measure the effectiveness of mitigating controls.

B. Network infection rate

By summarizing the number of attacks that a particular IP perform and listing all the IP within a network, we can assign certain subnetworks an infection rate. I.e. Number infected IPs vs total IPs in a certain subnetwork.

C. BN geographical and digital location.

The input for this metric will include the number of attacks coming from various counties and also pinpointing the regional locations according to cities. This metric will help measure the scale of attacks coming from a particular geographical location. Furthermore, we can assess to block the IP ranges as per the geographical source of the attack coming from to mitigate the risk of potential impacts to critical infrastructure.

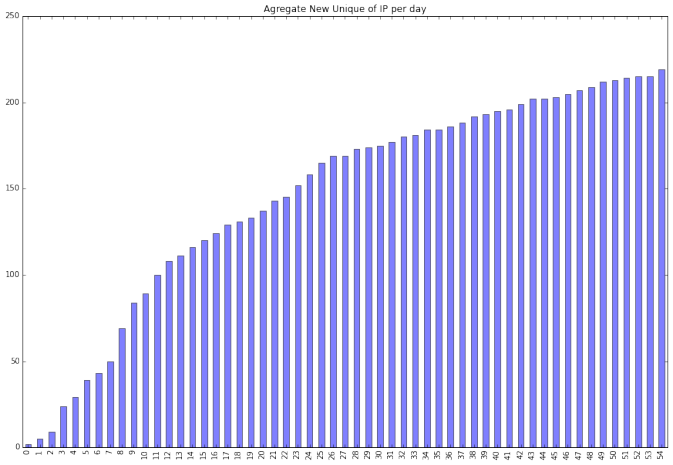
V. Results evaluation and conclusion

The evaluation and results of the metrics defined in section IV is presented in this section. We highlight the most useful metrics, however a previous and complete statistical analysis of the variables in the dataset have been performed. I.e., over 93% of the BNI attempts are coming from China, more than 90% of the requests were done via GET method, the most used OS is Windows 2000 and the most used browser is IE.

A. BN propagation

In figure 1 we can see the aggregate growth of unique IP addresses per day. As we can see, the number of IPs used to try propagate the BN(s) increments daily. Even when we know that ISPs use dynamic IP allocation and therefore, by itself it is not Personal-identifiable information (PII), this metric provide a rough idea of the BNI success and of the size of the problem that is being faced. Similarly, figure 2a presents the number of unique IP per day and figure 2b show the BNI attempts or BN activity in terms of number of requests per day. To compute the relation between this two metrics we have performed a Person's correlation resulting on a 0.5, suggesting that this two analysis are *strongly correlated* [12]. Consequently, the growth of the BN is strongly correlated with the increment of attacks, thus, or the BNI is succeeding and/or the BN activity is increasing and with this informations an interested party could motivate the investment on security countermeasures.

Fig. 1: BN growth



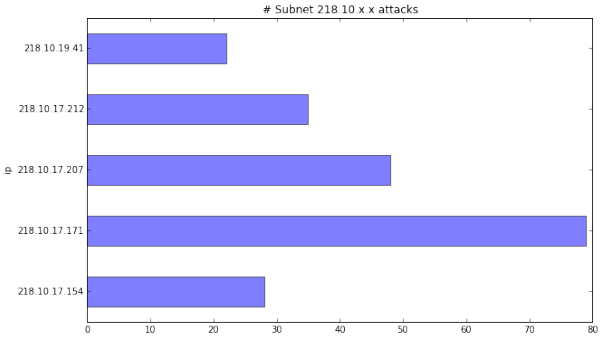
B. Network infection radio

An example of this kind of analysis is shown in figure 3, we list all the infected IPs in a subnetwork of an ISP, i.e. we have assumed that the ISP has class B IPs and one of his subnetwork is 218.10.0.0. Therefore, for each subnetwork within an ISP network we can see the activity of infected IPs and the infection radio of the subscribers in this network. Further analysis can be done by the ISP since he can use other data sources to get PII from the IPs and therefore compute this metric in a more accurate fashion.

C. BN geographical and digital location.

Geographical and digital location analysis is shown in figures 4 and 5. In figure 4 we show the geographical distribution of the BNI attempts where we can observe that the majority of attacks are coming from China, however this can not directly imply that its the most insecure region but the most active, since its also the most populated country in the world. Nevertheless, the impact for the ISP should not be underestimated since the business model of the ISP are *economics-of-scale*, thus they are really interested in wiping out unwanted traffic from their network. In figure 5a we show the top 10 most infected ISP and in figure 5b we show the top 10 most active BNI attempts. From this analysis we can observe that the the more ISP with more infected IPs are not necessary the most active BN.

Fig. 3: Subnet 218.10.0.0 attacks



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] 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.

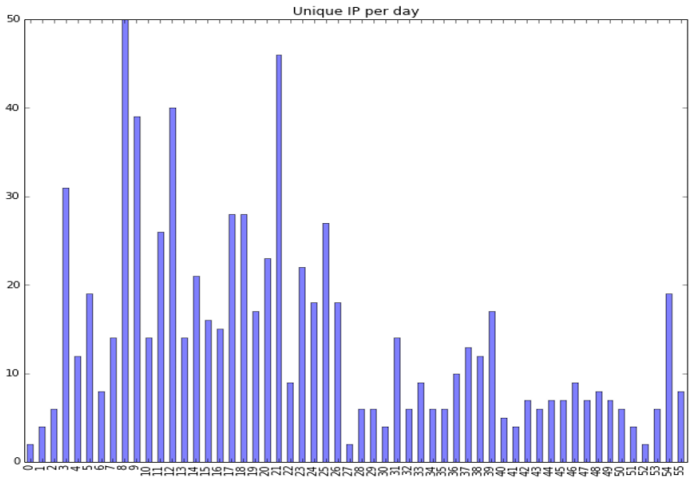
[9] M. Andreolini and et al., "Honeyspam: Honeypots fighting spam at the source," in *Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop. USENIX Association*. USENIX, 2005.

[10] F. Haltas, E. Uzun, N. Siseci, A. Posul, and B. Emre, "An automated bot detection system through honeypots for large-scale," in *Cyber Conflict (CyCon 2014), 2014 6th International Conference On*. IEEE, 2014, pp. 255–270.

[11]

[12] J. Wright. Statistics. [Online]. Available: http://faculty.quinnipiac.edu/libarts/polsci/Statistics.html

(a) Unique IP per day



(b) BNI attempts

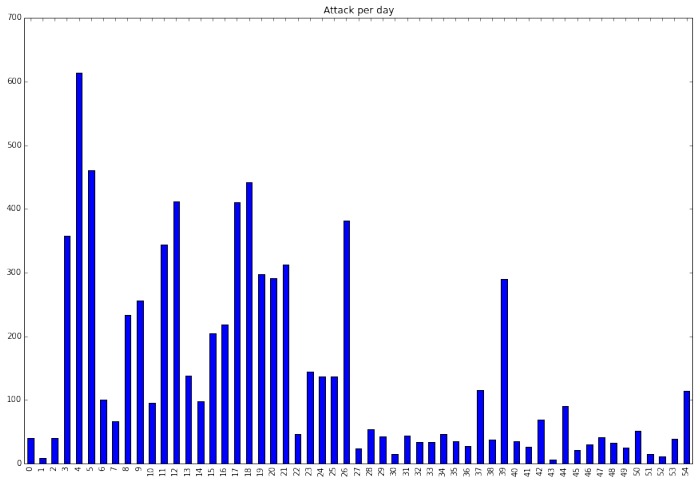
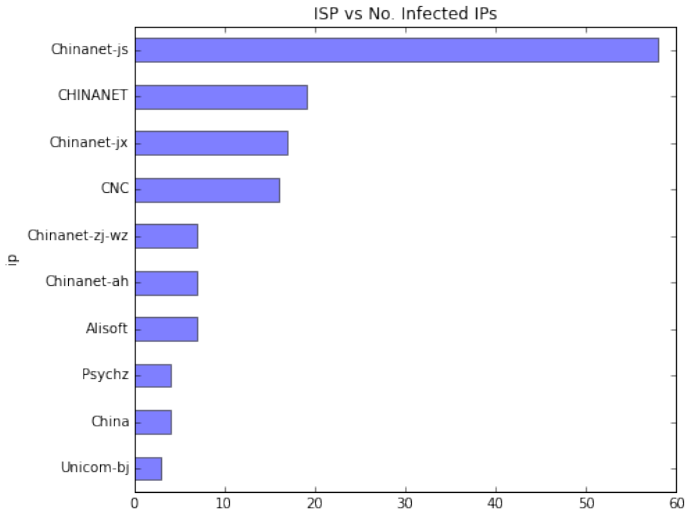


Fig. 4: BNI geolocation [13]

[13] ——. (2015) 60 days of watching hackers attack elasticsearch. [Online]. Available: http://jordan-wright.com/images/blog/elastichoney_elk/map.png

Fig. 5: ISP analysis

(a) Infected IPs per ISP



(b) BNI attempts per ISP