Poster: Challenges of Accurately Measuring Churn in P2P Botnets

Leon Böck
Telecooperation Group
Technische Universität Darmstadt
Darmstadt, Germany
boeck@tk.tu-darmstadt.de

Shankar Karuppayah
Telecooperation Group
Technische Universität Darmstadt
Darmstadt, Germany
Universiti Sains Malaysia
Penang, Malaysia
karuppayah@tk.tu-darmstadt.de

Kory Fong
RBC Research Institute
Toronto, Canada
kory.fong@rbc.com

Max Mühlhäuser
Telecooperation Group
Technische Universität Darmstadt
Darmstadt, Germany
max@tk.tu-darmstadt.de

Emmanouil Vasilomanolakis
Center for Communication, Media and Information technologies (CMI)
Aalborg University
Aalborg, Denmark
emv@cmi.aau.dk

ABSTRACT

Peer-to-Peer (P2P) botnets are known to be highly resilient to takedown attempts. Such attempts are usually carried out by exploiting vulnerabilities in the bots communication protocol. However, a failed takedown attempt may alert botmasters and allow them to patch their vulnerabilities to thwart subsequent attempts. As a promising solution, takedowns could be evaluated in simulation environments before attempting them in the real world. To ensure such simulations are as realistic as possible, the churn behavior of botnets must be understood and measured accurately. This paper discusses potential pitfalls when measuring churn in live P2P botnets and proposes a botnet monitoring framework for uniform data collection and churn measurement for P2P botnets.

CCS CONCEPTS

• Security and privacy → Malware and its mitigation; **• Networks → Network dynamics**;

KEYWORDS

botnets; churn; monitoring; peer-to-peer

ACM Reference Format:
Leon Böck, Shankar Karuppayah, Kory Fong, Max Mühlhäuser, and Emmanouil Vasilomanolakis. 2019. Poster: Challenges of Accurately Measuring Churn in P2P Botnets. In *2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19), November 11–15, 2019, London, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3319535.3363281

ACM ISBN 978-1-4503-6747-9/19/11.
https://doi.org/10.1145/3319535.3363281

1 INTRODUCTION

Botnets are networks of malware infected machines, called bots, that can be remotely controlled by attackers. These so called botmasters abuse the bots for criminal activities, e.g., spam distribution and Distributed Denial of Service (DDoS) attacks. Countermeasures against such botnets greatly depend on the structure of the Command and Control (C2) channel implemented by the botnet. While centralized C2 channels are still highly popular, they suffer from the problem of having a single point of failure. To overcome this problem, botmasters opted for more resilient C2 channels. Among the most resilient and sophisticated C2 channels are fully distributed P2P botnets [8]. In a P2P botnet, each bot can be used to disseminate commands issued by the botmasters.

To successfully attack such a botnet, detailed information about its population and interconnectivity are required. However, even if sufficient information is present, takedown attempts are highly challenging and could still fail [9]. Such failed takedown attempts may alert botmasters and allow them to patch their botnet's vulnerabilities. A possible alternative to this problem is to use simulators such as the open source Botnet Simulation Framework (BSF)[1], as they allow to experiment, prototype and evaluate takedown approaches in a dynamic environment.

To facilitate realistic and accurate simulations in a simulator, two components are crucial: 1) the communication protocol and membership management, and 2) the churn behavior of the botnet, i.e., nodes joining and leaving the botnet. While the former can be precisely extracted by reverse engineering, the churn behavior has to be measured on the live botnet itself. Obtaining accurate measurements is crucial, as it greatly influences the botnet's topology and therefore its resilience and resistance to monitoring approaches [2].

The goal of this paper is to discuss the challenges of accurate churn measurements. To address these challenges, we propose Botnet Monitoring Framework (BMF) for accurate and uniform data collection across multiple P2P botnets.

[1] https://github.com/tklab-tud/BSF

The collected data can then be used to facilitate accurate simulations of P2P botnets.

The remainder of this paper is structured as follows. Section 2 introduces related work on measuring churn. Section 3 discusses potential pitfalls and introduces BMF. Lastly, Section 4 summarizes our discussions and provides an outlook on future work.

2 RELATED WORK

In this section, we provide a brief overview on related work in measuring churn in P2P networks and botnets.

Stutzbach et al. analyzed the characteristics of churn in P2P filesharing networks [4]. They reasoned that crawling at high speed is essential for accurate churn measurements. Furthermore, they show that Weibull distributions are better suited to accurately fit churn measurements than exponential distributions.

Similarly to the previous work, Karuppayah [7] provided measurements and Weibull distribution fits for the Sality and ZeroAccess P2P botnets. Although, the churn behavior itself differs between filesharing and botnet P2P networks, the Weibull distribution is reported suitable to fit the churn behavior in P2P botnets.

Based on existing churn distributions, Böck et al. [2] presented an algorithm to replicate existing churn measurements accurately within their botnet simulation framework. This work also discussed the importance of accurate measurements to accurately replicate churn behaviors to investigate P2P botnets in simulators.

Many other works such as [5, 6, 8] have discussed the effects of churn in P2P botnets. They also highlighted the significance of obtaining accurate churn behavior of P2P botnets. Although none of them reported distribution fits for the data, they discussed on the impact of churn towards measurement accuracy and diurnal patterns. To the best of our knowledge, most of the work presented above utilized standalone crawlers, sometimes in combination with sensors to measure churn. As we will discuss in Section 3, such standalone monitoring approaches would fail to address many of the pitfalls of obtaining accurate churn measurements.

3 COMMON PITFALLS AND PROPOSED METHODOLOGY

In this section, we discuss potential pitfalls in measuring churn on live P2P botnets. Afterwards, we introduce an efficient and scalable botnet monitoring framework that aims to address most of the discussed pitfalls.

3.1 Pitfalls for Accurate Churn Measurements

To ensure that churn measurements are as accurate as possible, potential errors in collecting and interpreting the data must be considered. Stutzbach et al. presented a comprehensive list of potential pitfalls in measuring and fitting distributions to churn [4]. We summarize these pitfalls (1 – 7) and introduce three new pitfalls (8 – 10) we have identified for P2P botnets.

- **P1 Missing Data** Data must be complete for the period of distribution fitting. Otherwise, the missing data will adversely influence the churn measurements of the affected nodes.
- **P2 Biased Peer Selection** If a subset of peers is selected, they must be sampled in a non-biased fashion. Otherwise the measurements will be skewed towards the biased peer selection.
- **P3 Long Sessions** Given an interval τ and a minimum accuracy ϵ, $\frac{\tau}{\epsilon}$ short sessions, i.e., uninterrupted time-windows of availability, can be measured. However, only one session of length τ can be observed in the measurement period τ.
- **P4 False Negatives** False negatives within the measurements may occur due to network congestion or temporal failure of network connections. This may lead to bots appearing to be offline even when they are not.
- **P5 Brief Events** If the time between measurements is not granular enough, short-lived events may not be recorded. Consequently, sessions might be missed or multiple short sessions (mis)-interpreted as a single long session.
- **P6 NAT** Devices behind a Network Address Translation (NAT) device or firewall cannot be contacted directly over the Internet unless they first initiate the connection. Measuring such nodes may lead to skewed results due to frequently changing ports or shared IP address.
- **P7 Dynamic IP Addresses** Many Internet Service Providers (ISPs) uses Dynamic Host Configuration Protocol (DHCP) or Point-to-Point Protocol (PPP) to assign dynamic IP addresses. Reassignment of IPs will appear as separate leave-and-join events even though the bots remained online.
- **P8 Synchronization** When using multiple monitoring instances, it must be assured, that the clocks are synchronized. Otherwise aggregated results may lead to contradicting observations about the availability of bots.
- **P9 Non-persistent Identifiers** The lack of unique and persistent identifiers introduces measurement inaccuracies of the population, as new bots and re-joining bots can not be easily differentiated.
- **P10 Anti-monitoring Mechanisms** P2P botnets may deploy anti-monitoring mechanisms to hinder monitoring attempts. Such mechanisms must be considered and circumvented to obtain accurate churn measurements.

3.2 Botnet Monitoring Framework (BMF)

In this subsection, we introduce BMF. This framework which is depicted in Figure 1, aims at improving the efficiency of future botnet monitoring activities with increased accuracy. Specifically, in the context of obtaining churn measurements, BMF attempts to address most of the pitfalls explained in Section 3.1.

BMF consists of three major components: 1) a NoSQL database, 2) (modular) monitoring modules and 3) a coordinator. The sheer amount of obtainable data through monitoring activities does not scale with relational databases. Moreover, the obtainable type of information varies across different botnets. Hence, NoSQL databases, e.g., MongoDB, are best suited to store such information.

The most commonly used monitoring mechanisms are honeypots, crawlers and sensors. However, as pointed out by many researchers [7, 8], each of the methods has their own advantages and disadvantages. Hence, the information from all of these mechanisms are often combined during analysis, e.g., churn measurement analysis. BMF is designed to enable a modular monitoring system that integrates existing monitoring mechanisms as well the possibility to include newer ones in the future. These monitoring modules are controlled by a Coordinator which has an overall view of the

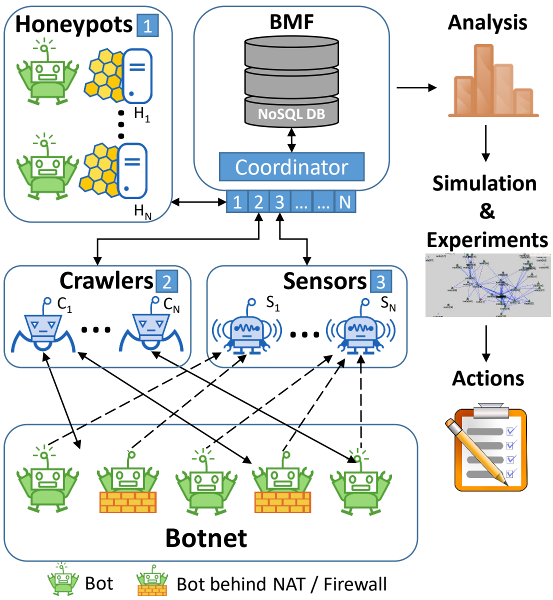

Figure 1: Illustration of a BMF setup

entire monitoring process and can synchronize the logged data from different monitoring nodes (**P8**).

Since information of all monitoring mechanisms needs to be considered as a whole, it only makes sense to introduce a modular monitoring system. For instance, bots that are executed in a honeypot or sandbox, would continuously communicate with other active bots. From that, BMF can keep track of recently active peers in the botnet. Therefore, once a crawler and a sensor are implemented, they can be bootstrapped into the botnet using the list of active peers, instead of stale bootstrap entries. Moreover, crawlers could also be leveraged to bootstrap or popularize the deployed sensors in the same manner (**P6,P8**).

Recent P2P botnets and the state of the art in this domain have seen many new anti-monitoring mechanisms that could hamper monitoring; hence affecting the quality of the monitored data [1, 3]. For instance, one of the most common restriction mechanism is the blacklisting of IP addresses. BMF can circumvent such mechanisms by coordinating the crawlers to yield better discovery without triggering the anti-monitoring mechanisms (**P10**). Moreover, the ability to coordinate the monitoring modules also enables the possibility to assign monitoring nodes from different networks to probe a single bot. Consequently, this can reduce the false positives from network congestion or temporal network failures (**P4**). In addition, the coordinator could also distribute the monitoring activities.

The design of BMF is also intended for long-term botnet monitoring (**P2,P3**). Moreover, the granularity of the monitoring frequency can be adjusted as needed (**P5**). The data generated from such long-term granular monitoring could be useful for many purposes; ranging from understanding to taking them down.

Finally, the collected data can be analyzed for purposes such as churn analysis to produce a churn model. Considering that long-term monitoring data is available, duration of the analysis can be chosen such that no interruptions that could affect the resulting analysis exists (**P1**). The churn model could then be used in simulators such as BSF to conduct realistic churn simulations to allow defenders to study the impact of a potential botnet takedown.

4 SUMMARY AND FUTURE WORK

Within this paper we discussed the necessity for accurate churn measurements of P2P botnets to evaluate P2P botnet takedowns and enable realistic simulations. Therefore, we outlined ten pitfalls in measuring churn in live P2P botnets. Moreover, we propose BMF, a framework that addresses these pitfalls and allows uniform measurements of churn in various P2P botnets. As part of ongoing work, we also intend to address pitfalls **P7** and **P9**, which are not yet addressed. We observed these two pitfalls in an early measurement of the Hide'n Seek (HnS) botnet. After 148.16 hours of measurement, a group of bots disconnected simultaneously from a particular ISP. We presume, that this was caused by an IP address re-assignment. However, due to the lack of a unique identifier, we were unable to re-identify the bots under their new IPs. Our goal is to address these issues in the future by fingerprinting bots based on their shared information and response behavior, e.g. TTL and round-trip times.

For future work, we intend to implement BMF and conduct long term measurements for P2P botnets such as Sality, ZeroAccess, HnS and Hajime. Based on these measurements, we want to address three major goals: 1) enable more realistic simulations of P2P botnets, 2) compare churn across different botnets and analyze the difference between traditional and Internet of Things (IoT)-based botnets, and 3) make BMF and the collected dataset freely available to foster collaboration and advances in the fight against botnets.

ACKNOWLEDGMENTS

This work was supported by the Royal Bank of Canada (Grant Agreement 2761478.4) and German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP.

REFERENCES

[1] Dennis Andriesse, Christian Rossow, and Herbert Bos. 2015. Reliable Recon in Adversarial Peer-to-Peer Botnets. In *Proceedings of the 2015 ACM Internet Measurement Conference, IMC 2015, Tokyo, Japan, October 28-30, 2015*. 129–140. https://doi.org/10.1145/2815675.2815682

[2] Leon Böck, Emmanouil Vasilomanolakis, Max Mühlhäuser, and Shankar Karuppayah. 2018. Next generation P2P Botnets: monitoring under adverse conditions. In *International Symposium on Research in Attacks, Intrusions, and Defenses*. Springer, 511–531.

[3] Leon Böck, Emmanouil Vasilomanolakis, Jan Helge Wolf, and Max Mühlhäuser. 2019. Autonomously detecting sensors in fully distributed botnets. *Computers & Security* 83 (2019), 1–13. https://doi.org/10.1016/j.cose.2019.01.004

[4] Stutzbach Daniel and Rejaie Reza. 2006. Understanding churn in peer-to-peer networks. In *Internet Measurement Conference*. ACM, 189–202.

[5] Steffen Haas, Shankar Karuppayah, Selvakumar Manickam, Max Mühlhäuser, and Mathias Fischer. 2016. On the resilience of P2P-based botnet graphs. In *Communications and Network Security (CNS)*. IEEE, 225–233.

[6] Stephen Herwig, Katura Harvey, George Hughey, Richard Roberts, and Dave Levin. 2019. Measurement and Analysis of Hajime, a Peer-to-peer IoT Botnet. In *Network and Distributed System Security Symposium*.

[7] Shankar Karuppayah. 2016. *Advanced monitoring in P2P botnets*. Ph.D. Dissertation. Technische Universität Darmstadt.

[8] Christian Rossow, Dennis Andriesse, Tillmann Werner, Brett Stone-Gross, Daniel Plohmann, Christian J Dietrich, and Herbert Bos. 2013. Sok: P2pwned-modeling and evaluating the resilience of peer-to-peer botnets. In *IEEE Symposium on Security and Privacy*. 97–111.

[9] Tillmann Werner. 2013. Peer-to-Peer Poisoning Attack against the Kelihos.C Botnet. (2013). https://www.crowdstrike.com/blog/peer-peer-poisoning-attack-against-kelihosc-botnet/