Proceedings of the 11th graduate workshop SPRING of the special interest group Security - Intrusion Detection and Response (SIDAR) of the German Informatics Society (GI) (SPRING 2016), Darmstadt, Germany, June 02-03, 2016

Monitoring of the DNS Infrastructure for Proactive Botnet Detection

Christian Dietz[*][†], Anna Sperotto[†], Gabi Dreo[*] and Aiko Pras[†]

[*] Universität der Bundeswehr München
85577 Neubiberg, Germany
{Christian.Dietz, Gabi.Dreo}@unibw-muenchen.de

[†] University of Twente
7522 NB Enschede, Niederlande
{C.Dietz, A.Sperotto, A.Pras}@utwente.nl

1 Introduction

Botnets enable many cyber-criminal activities, such as DDoS attacks, banking fraud and cyber-espionage. Botmasters use various techniques to create, maintain and hide their complex C&C infrastructures. First, they use P2P techniques and domain fast-flux to increase the resilience against take-down actions. Second, botnets encrypt their communication payload to prevent signature based detection. However, botnets often use the domain name system (DNS), e.g., to find peers and register malicious domains. Since, botmasters manage a large distributed overlay network, but have limited personal resources, they tend to automate domain registration, e.g. using domain name generation algorithms (DGAs). Such automatically generated domains share similarities and appear to be registered in close temporal distance. Such characteristics can be used for bot detection, while their deployment is still in preparation. Hence, the goal of this research is early detection of botnets to facilitate proactive mitigation strategies. Using such a proactive approach prevents botnets from evolving their full size and attack power. As many end users are unable to detect and clean infected machines, we favour a provider-based approach, involving ISPs and DNS registrars. This approach benefits from its overview of the network that allows to discover behavioural similarities of different connected systems. The benefit of tackling distributed large-scale attacks at provider level has been discussed and demonstrated in previous studies by others. Further, initiatives to incentive ISPs centred botnet mitigation are already ongoing. Previous research already addressed the domain registration behaviour of spammers and demonstrated DGA based malware detection. In contrast, our approach includes the detection of malicious DNS registration behaviour, which we currently analyse for the .com, .net and .org top level domains. These domains represent half of the registered Internet domains. By combining DNS registration behaviour analysis with passive monitoring of DNS requests and IP flows, we are able to tackle botnets throughout their whole life-cycle.

2 Research Problem & Questions

The goal of this research is to enable early botnet detection in provider environments. Therefore, our approach is based on large-scale DNS registration behaviour analysis, as this will allow to discover botnet activity in the (pre-)deployment phase of its life-cycle. Thus, our novel approach can prevent the botnet from becoming deployed and actively used. Furthermore, the proposed approach takes into account the dynamics of botnet malware and the Internet infrastructure, high data rates, incompleteness of data and encrypted bot communication. In order to tackle the early botnet detection problem, we ask the following questions: (i) How do botnets interact with the domain name system? (ii) Can domain registration characteristics be used for botnet detection, and if yes, how?

3 Approach

The goal of this research is to allow faster botnet detection and mitigation. Current approaches are usually limited to detect bots after they already became active or while they are used in attacks. Our approach targets botnet detection in the pre-deployment phase. Therefore, our approach is based on two components: (1) passive monitoring of communication characteristics and (2) DNS registration behaviour analysis. DNS registration analysis allows to detect the preparatory actions of deployment of the C&C infrastructure and the bots. Therefore, our approach allows botnet early detection and consequently facilitates proactive botnet mitigation. In addition, our approach allows botnet detection in the subsequent phases of the bot life-cycle (preparation, infection, peer discovery, malware update, command propagation and attack) by using passive DNS and flow monitoring solutions. Figure 1 provides an overview of our novel approach.

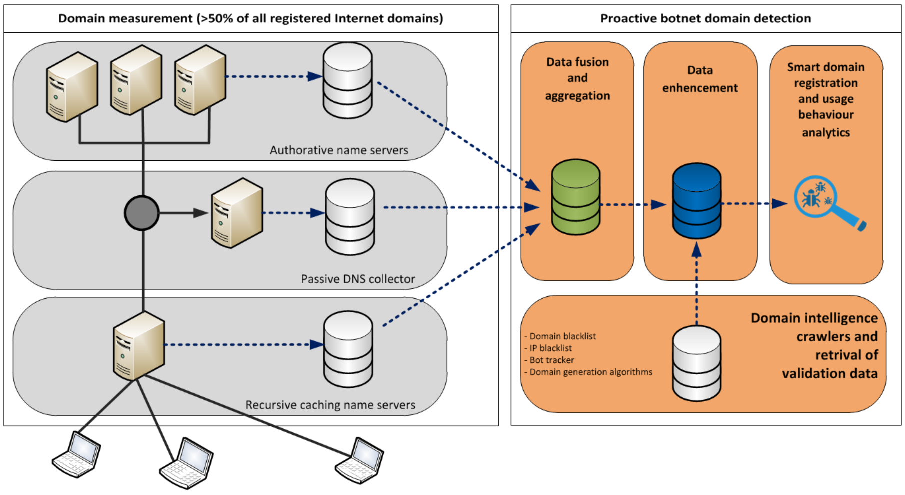
Domain measurement (>50% of all registered Internet domains)
Authorative name servers
Passive DNS collector
Recursive caching name servers
Proactive botnet domain detection
Data fusion and aggregation
Data enhencement
Smart domain registration and usage behaviour analytics
- Domain blacklist
- IP blacklist
- Bot tracker
- Domain generation algorithms
Domain intelligence crawlers and retrival of validation data

Figure 1: Components of the passive measurement and smart analytics infrastucture.

Research question (i) aims to get insight into the deployment and management of botnets. Therefore, we collect DNS registration data on a daily basis for the *.com*, *.net* and *.net* domains, representing half of the domains registered on the Internet. Second, we query different botnet tracking services and use DGAs to find botnet related records in the domain registration dataset. Research question (ii) aims to extract characteristics of botnets in their deployment phase to allow an early detection and mitigation. To answer this question, we use registration databases of top level domain registrars. Currently, our study involves the *.com, .net, and .org* top level domains. We will validate our novel approach based on simulations and real-live environments. Further, we compile different datasets. First, we crawl the registration database of multiple top level domains, different botnet domain and IP blocklists with time stamps. This allows us to measure the temporal difference between botnet deployment and detection. Second, we passively capture IP flow data and DNS requests in multiple provider networks to evaluate (a) how accurate our approach can detect the large-scale similarities between distributed bots and (b) determine the temporal delay between malicious domain registration and the first activity. This evaluation also uses IP address and DNS blocklists that our crawlers collect on a regular basis.

Acknowledgments This work is supported by the EU FP7 Flamingo Project (ICT-318488).