Botnet Master Detection Using a Mashup-based Approach

Carlos Raniery P. dos Santos∗, Rafael Santos Bezerra∗, Joao Marcelo Ceron∗, Lisandro Zambenedetti Granville∗ and Liane M. R. Tarouco∗

∗Institute of Informatics - UFRGS
Rio Grande do Sul, Brazil
Email: {crpsantos, rsbezerra, jmceron, granville}@inf.ufrgs.br, liane@penta.ufrgs.br

Abstract—Botnets are considered by specialists, in both industry and academy, as one of the greatest threats to security on the Internet. These networks are composed by a large number of malware-infected hosts acting under a central command. They are usually employed to perform DDoS attacks or phishing scams. The behaviour of these botnets evolves due the adoption of new and sophisticated infection methods, changing of network protocols, and the employment of different command and control mechanisms. The security community, thus, is always dealing with such constant change. However, most botnet mitigation methods address just specific infection types or C&C protocols. We, therefore, propose a botnet mitigation approach based on the dynamic integration of pre-existing tools that can be employed together to achieve a more efficiently detection solution. To such end, we base our approach on a novel Web 2.0 technology called mashups to perform the information correlation. The proposal is extensible enough to allow even non-security information such as online mapping APIs to be integrated to create more sophisticated compositions, and displaying the results in a more meaningful way.

I. INTRODUCTION

The proliferation of personal computers and the decreasing costs of communication led the growing of computer networks, specially the Internet. Unfortunately, this growing comes associated with the increasing number of vulnerable hosts to infections. In the past, such infections were caused by viruses and worms that aimed to compromise and disable their targets. However this scenario has changed, virus creators no longer aim to just acquire recognition and attention from their community or the media. Now their main interest is making profit of their creations. Most of the infections are no longer committed to cause non functioning on the victims. Instead, infected hosts become bots, which can be remotely controlled by a human operator known as master or botmaster. Sets of such hosts are called botnets and are usually used to illegal purposes, such as large scale Distributed Denial of Service (DDoS) attacks, email spamming and phishing [1].

Botnets are considered the current largest security threat on the Internet [2]. One of the main reasons for this is the high number of infected computers. According to Vincent Cert [3], up to 25% of all computers on the Internet belong to botnets. Another reason lies on their effectively. Combining the efforts of hundreds or even thousands of machines widespread across the Internet to reach a single task such as DDoSing a server or spamming makes this approach very effective and hard to prevent. Due to such widespread nature, dismantling these networks is also a complex task. However, botnets conceptually have one single point of failure that can potentially be exploited, the botmaster. Identifying the botnet controller opens up many possible countermeasures, such as blocking communication, removing infected machines from the botnet, or even infiltrating this controller host in order to identify all participants of his botnet, as proposed by [4].

Most approaches to detect botmasters are based on detecting communication mechanisms used between the botmaster and the bots, which are known as command and control (C&C) channels. Centralized botnets have widely employed the Internet Relay Chat (IRC) protocol for many years [5]. Some efforts to extract information from the malware via IRC or disable its communication can be found, such as [1], [4] and [6]. However, more sophisticated botnets started to use other channels for communication. For instance, HTTP [7] is used to provide more stealthiness by mixing the C&C messages with regular Web traffic, also crossing firewalls more easily. Storm worm [8] employs Peer-to-Peer (P2P) communication to distribute the control of the botnet. Such approaches raise the difficulty of detecting botmasters, and nowadays there are just a reduced number of works that claim to address all of them [9].

One technique that can be used to discover information about the botmaster is the sandboxing. A sandbox is a controlled environment that allows safe execution of infected binaries and the observation of malware behavior. Such observation includes system calls, created and executed files and network activity. This technique has proven itself effective to detect the botnet controller [10]. However, its employment brings costs associated to the setup of the environment and the usage of commercial tools. Currently, there are some free, Web-based sandboxes, such as [11], [12] and [13], which employment can mitigate and even eliminate those costs, at the cost of a more limited functionality and database.

Given this scenario, we propose a mashup based approach to botmaster detection. Mashups are a recent Web 2.0 technology that proposes the agile and dynamic integration of external resources in a Web page, generating new functionality from such integration [14]. Our mashup based approach consists in integrating and correlating information from web available binary analysis tools (i.e., online sandboxes and anti-viruses).
We employ other free web based tools, such as geocoding services and map APIs, to better present the results obtained. Our main goal with this proposal is analyze if the usage and integration of free, web available tools is viable and effective. We also provide to the community a new tool to both detect and visualize botmasters.

The remainder of this paper is organized as follows. In section 2, we depict our mashup-based solution. A qualitative evaluation of our proposal is presented in Section 3. In the Section 4 we lay out the obtained conclusions and present the future work.

II. NEW APPROACH

In this section, we will discuss our botnet mitigation solution. It consists on a botmaster detection tool that integrates and correlates information from online binary analysis, displaying the botmasters found on a map based on the geolocation of their IPs. It also provides an interface for binary submission and an online database that stores all the information about the binaries analyzed.

A. Created Mashup

We developed a mashup for botmaster and C&C Channel detection. This mashup, presented in Figure 1, integrates and correlates information from Web available binary analysis tools, such as sandboxes and anti-viruses. The obtained results are geolocated in a map to show the obtained results. Besides the components Geolocation and Map, already presented, we also developed two others: one to access sandbox and anti-virus online tools, and another to access an IP to ASN translation service (e.g., Whois).

- **Analysis**: an Adaptation component that receives the binary file uploaded by the network administrator, and forwards it to online analysis tools. Its output is a set of information regarding the infection kind, C&C Sever IP address, and time of the analyze;
- **Whois**: Adaptation component that retrieves the ASN based on an IP address. It was used to discover from which ISP the C&C Server is based. The information is retrieved from a Whois Web page, which receives just the IP address.

The mashup for botmaster detection receives as input infected binaries, which can be submitted by users or gathered in honeypots set by network administrators. Such binaries are then submitted to three online binary analysis tools. CWsandbox and Norman, will execute the binary and analyze its behavior, and Virus Total, will scan the binary for virus signatures. The information gathered from these three tools is then correlated. Such correlation aims to discover information about a possible C&C channel, and if successfully, try to identify the C&C Server.

With the C&C information, we employ our **Geolocation** and **Whois** components in order to discover both location of the C&C Server, and its responsible ISP. This information is then aggregated and sent to the **Map** component, which generates an interactive map-based Web page. This Web page allows the visualization of C&C Servers spread around the world through markers representing incidences of binary vectors of that botnet, as show in Figure 1. Furthermore, the information gathered is also available in a data feed of detected botnet vectors and C&C Servers using different data formats (i.e. XML and JSON). This information is also timestamped, allowing the creation of temporal representations of the infections.

The composition of results from many sandbox and anti-virus online tools is an alternative to the identification and mitigation of botnets based on a centralized approach. It replaces the usage of complex and paid tools with the integration of free, simpler ones. The adaptability obtained with such mashup-based approach is important due to the heterogeneity of technologies employed in C&C channels. Since each sandbox tool is able to analyze and detect a specific set of technologies, the usage of multiple tools allows the mashup to detect different kinds of infections. Furthermore, it allows constant self-atualization, through the addition of new tools to detect new kinds of infections. The support of new analysis mechanisms is possible by the creation of wrappers, which is a straightforward process.

Finally, mashups are built with reusability in mind. In this case, this characteristic allows the created mashup to be used as a building block for more sophisticated botnet mitigation mashups. For instance, it is possible to develop a mashup able to identify P2P botnets. Further possibilities include the implementation and addition of firewall filters to automatically isolate the C&C channel from the network potential hosts, and automatically detect infected hosts inside the network through flow analysis.
III. Evaluation and Verification

In this section, we evaluate our mashup for botnet mitigation approach. This evaluation is divided in two aspects. Firstly, we evaluate our mashup based approach qualitatively, doing an analysis of its advantages and limitations. We then make a comparative evaluation, where we compare our approach with the current botnet mitigation and information gathering techniques presented in Section 3.

A. Qualitative evaluation

In order to qualitatively evaluate our approach, we determined some characteristics that will be discussed in details following. Such characteristics, presented bellow, will be the focus of this subsection:

- **Flexibility**: relates to the capability of adaptation to new scenarios, such as dealing with new botnets or new C&C technologies;
- **Extensibility**: the possibility of reuse by different tools or different approaches;
- **Implementation effort**: the effort needed to implement the theoretical approach;
- **Usage effort**: the effort needed to use the botnet mitigation mashups once they are implemented;
- **Reliability**: how reliable are the botnet mitigation mashups, specially under critical scenarios.

The flexibility of our mashup is based on the capability of integrating new botnet information and mitigation tools. This is special interesting for botnet mitigation since such networks are always evolving. New protocols for C&C channels, infection methods, and obfuscation techniques are always being used and obligating the employment of new techniques and tools to deal with them. Wrappers can be created to these new tools [15], enabling them be integrated in our botnet mitigation mashup. Additionally, non-exclusive security tools can be employed to increase the effectiveness of the mashup approach. For instance, a module to analyse Netflow flows and find C&C channels can be developed, and a firewall module can dynamically create firewall rules to stop such communication.

The flexibility is obtained due the modularity of our mashup system. We proposed and implemented modules (i.e., wrappers) to access different online sandboxes and anti-viruses systems. Such modules was composed with pre-existing modules (i.e., geolocation and mapping) created originally to other context (i.e., BGP peering) [15]. The result of the composition (i.e., the botnet mitigation mashup presented in Section 4), can be reused and extended by other users of the mashup system, but also, by users whom don’t use the system, for example, by accessing the report available in different data formats (i.e., JSON, CSV, XML). This report includes information the following information: type of infection, botnet name, ASN of the infected host, and geographic and chronologic information.

The reduction of the implementation effort is obtained by the existence of wrappers to access the external systems. In fact, using a mashup-based approach reduces this effort to the creation of independent wrappers by software developers. The integration of those wrappers is responsibility of the network security administrator, who knows well about security but is not necessarily a skilled programmer. About the usage effort, the mashup approach gives the end user the possibility to easily manipulate the components and alter the composition, which adds another layer of extensibility, as previously discussed.

We are aware of the reliability limitation of our approach. This happens because the mashup depend on free online and external tools, which can fail. Problems such as offline services, overloaded services, discontinued services and API changes may arise with the usage of such online tools. Redundancy can be, and was, used to mitigate this issue. Similar services are available online, such as the sandboxes employed in our implementation. Furthermore, once the information about botnet infection is gathered, it is stored in the mashup system. In case of failure of the external services, binary analysis is compromised, but existing reports remain functional and useful. Thus, reliability of our approach is adaptable to the administrator needs, but it demands an adequate choice of components.

IV. Conclusions

In this paper, we proposed a mashup-based botnet mitigation approach. In our proposal, we cover the integration of various botnet mitigation tools and techniques with external resources in botnet mitigation mashups. Based on this proposal, a mashup prototype was developed with information about botnets detected using free online sandboxing and anti-viruses tools with geocoding, Whois and mapping online services.

Due to the possibility of integration of various techniques, including ones yet to come, to cover different kinds of botnets, we reached the conclusion that our mashup-based approach is adequate to the ever changing scenario of botnet mitigation. However, this approach does not aim to substitute other botnet mitigation efforts, but to leverage and correlate them to optimize the obtained results and create new ones from the integration. This is further reinforced by the comparison with other approaches performed in our evaluation section, where botnet mitigation mashups showed superior criteria coverage compared with current Vertical Correlation and Horizontal Correlation approaches. It is our understanding that the fluid and dynamic nature of botnets as a security threat demands extensible and flexible solutions, as is ours.

Having implemented our botnet mitigation mashups manually and aided by a mashup system, we have come to the conclusion that its usage reduces the effort needed to materialize our approach and further enhances its flexibility and extensibility. With the aid of a mashup system with adequate wrappers, the integration can be performed by a network administrator and/or a security expert, and the results better reflects their needs. Changing and extending this mashups becomes similarly easy. Furthermore, creating wrappers to the resources needed (IPSs, IDSs, anti-viruses, sandboxes) is a less demanding process than implementing the mashups manually.
Our implementation, while providing good information and coverage, used freely available components (i.e., online sandboxes, anti-viruses, geolocation tools, mapping APIs), remaining completely free. However, reliability and maintainability suffer due to the lack of guarantee of service of these resources. In more critical implementations, we recommend the usage of more reliable tools.

Our proposal sees botnet mitigation mashups as extensible tools, and our implementation reflects this view. To this end, the C&C information map code is available to be extended and the information used to create the map is available in a report. We delineated cases where this information can be further extended. For instance, the integration with a firewall to automatically create rules to block communication with botmasters and/or the integration with flow analysers to identify infected hosts within the network are both cases to be explored.

The search for online resources to be integrated and how these resources can be leveraged to mitigate botnets is also a research opportunity. Plenty of tools, be them security related (such as computer diagnosis tools) or not (such as RSS feeds) are available online to be mashed up, and they bring a host of possible new mashups. On the other hand, the evaluation of the integrability of current security tools such as IDSs, IPSs, firewalls, anti-viruses in a mashup based approach is also a research opportunity which we intend to chase. Evaluate if this tools offer, for instance, web interfaces, web service interfaces and computer readable reports and how those can be used in a mashup based approach.

REFERENCES