Poster: Measuring the Lifecycles of Malicious Domains

Kevin Warrick, Roberto Perdisci, and Kang Li
Department of Computer Science, University of Georgia, Athens, GA
{*warrick, perdisci, kangli*}*@cs.uga.edu*

***Abstract*—DNS is an intermediate for a great breadth of computer security threats such as the propagation of malware, the command and control of botnets, and spam and phishing campaigns. To track the evolution of malware domains, we implemented Digger, a tool for reliable distributed active probing of malicious domain names. In this paper we describe the architecture of Digger, and we present preliminary results from on-going experiments we are conducting to track the lifetime of malicious domains. Studying the lifecycles of malicious domain names will provide insight into the many classes of criminal networks that depend on DNS, and inspire the development of new, more effective countermeasures.**

I. Introduction

Recent abuses of DNS known as fast-flux make criminal networks much more resistant to traditional countermeasures such as blacklisting. Fast-flux is characterized by domain name records with low (time-to-live) TTL that resolve to many different IP addresses – hundreds and sometimes thousands each day. Remediating such attacks requires an assessment of (1) the longevity of malicious domains, (2) the number of IP addresses these domains resolve to, (3) how often domains change addresses, and (4) any lifecycle characteristics that may correlate malicious domains within the same network. Besides fast-flux networks, many other types of criminal networks abuse the DNS and leverage domain names to provide agility to their malicious network infrastructures.

Passive DNS traffic monitoring [2] has been used in the past to identify and study the behavior of malicious domains. Unfortunately, passive DNS monitoring has the drawback of limited visibility. It is very hard to cover large scale, Internet-wide domain name behaviors. There may exist malicious domains that are never queried by the contributing networks, and therefore no information about these domains would be accessible through small scale passive DNS monitoring.

In order to collect information about the "behavior" of malicious domains that often do not appear in the passively monitored network traffic, we have developed a tool called Digger, which performs distributed active probing of domain names in an efficient and reliable way. The infrastructure of Digger is shown in Figure 1. Digger takes as input a list of domain names to be monitored, and actively queries a set of recursive DNS (RDNS) servers to resolve these domains.

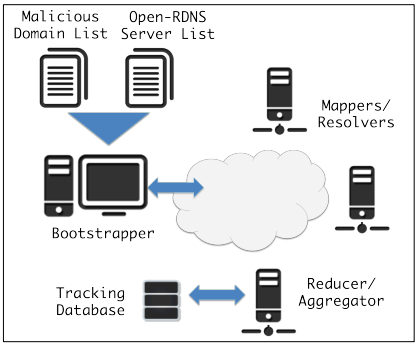

Figure 1. Digger's architecture.

The list of RDNS servers can include local servers as well as remote, public servers such as the ones provided by Google Public DNS [3], OpenDNS [4], and other similar services.

II. Approach

Digger is a distributed DNS lookup utility running on Hadoop MapReduce [1], [5]. Mappers act as querying queues and Reducers as response aggregators. Domains are randomly assigned a mix of public and local RDNS servers through which they will be queried. Hadoop nodes are run as virtual machines (VM) on commodity hardware. Effectively, Hadoop enables us to easily distribute pre-configured VM nodes to collaborators around the world and elastically expand the Digger network. To provide an Internet-wide view of DNS responses, Digger utilizes geographically distinct nodes and a diverse list of open RDNS servers. The need for a diverse list of distributed RDNS servers is motivated by a few observations. First, malicious domain names may resolve into different IP addresses, depending on where the DNS query originated from. Secondly, an attacker may control the authoritative nameserver (ANS) and may customize the ANS software to detect active probing attempts and return misleading responses in an attempt to pollute the tracking data. By using a large and diverse set of RDNS servers, it would be hard for an attacker to identify probing attempts or blacklist all of the available resolvers.

Digger collects information both on the resolved IP addresses for the monitored domains and on the name and IP addresses of the ANS servers that have authority over

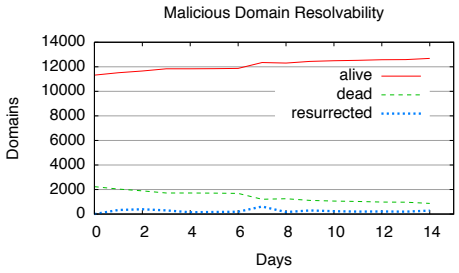

Figure 2. Domain resolvability.

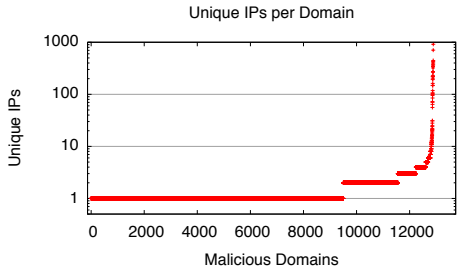

Figure 3. Resolved IP distribution.

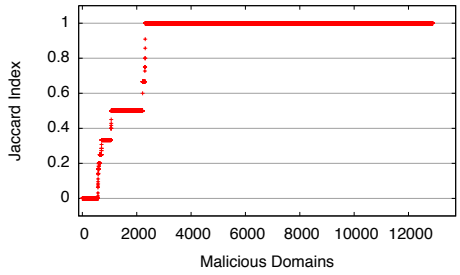

Figure 4. Jaccard index for set of IPs resolved from different resolvers.

the domains. This allows us to track the volume of domain names hosted at a particular set of ANS servers, which may be useful in particular for domain reputation systems because it enumerates which ANS servers have authority over large numbers of malicious domains.

III. Preliminary Results

In this section we discuss results obtained from a preliminary 14-day run of the Digger cluster. We used two primary malicious domain blacklists from malwaredomains.com and malwaredomainlist.com. As we anticipated, by the time we started our experiments on the blacklisted domains many were already dead, and only 11,818 of the initial 27,310 domains aggregated in the bootstrap phase were marked as active by Digger.

Figure 2 shows the resolvability of all actively queried domains per day including successfully queried (alive), unsuccessfully queried (dead), and successfully queried but previously unresolvable (resurrected) domains. The number of live domains climbs from the 11,818 domains initially aggregated to the 13,554 domains actively monitored at the end of the fortnight. Contrary to our intuition, such a trend indicates that many of the domains are long-lived and more domains are being introduced than are dying. Interestingly, the number of resurrected domains gravitates around 200 everyday revealing a number of domains that are intermittently inactive, which could potentially be an evasion mechanism or a correlating characteristic of instability – we are still investigating this anomaly.

Figure 3 shows the number of unique IPs each domain resolved to. Over 70% of queried domains resolved to a single address; however, 25 domains resolved to over 100 unique IP addresses.

Finally, Figure 4 shows the Jaccard index of the sets of IPs each malicious domain resolved to using sixteen different RDNS servers. Over 78% of domains have an index equal to 1, meaning complete consistency across resolvers. The remaining 22% of domains resolve to different IPs when queried from different RDNS servers. This underlines the importance of a diverse set of RDNS servers showing that a number of domains resolved differently across servers.

IV. On-Going Work

DNS is a dependency for the ever-growing criminal networks that pose some of the most serious security risks today. Digger is a late-breaking tool we have developed to reliably probe malicious domain names in a distributed way. We plan to use data collected using Digger for on-going research of the many classes of security threats that depend on DNS. Our preliminary results affirm our design choices and promise substantial data. As our understanding of domain lifecycles matures, we anticipate the ability to perform more complex assessments such as computing the reputation of ANS, for example. Ultimately, we hope that the research made possible by Digger will help improve how the community responds to malicious domains and suggest more effective remediation techniques.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. OCI-1127195. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] J. Dean and S. Ghemawat. *MapReduce: Simplied Data Processing on Large Clusters*. 6th Symposium on Operating System Design and Implementation(OSDI), 2004.
[2] F. Weimer. *Passive DNS replication*. In Proceedings of FIRST Conference on Computer Security Incident Handling, 2005.
[3] Google Public DNS Service. https://developers.google.com/speed/public-dns/
[4] OpenDNS Public DNS Service. http://www.opendns.com/supp
[5] Hadoop MapReduce. http://hadoop.apache.org/mapreduce/