lastline

# LASTLINE WHITEPAPER

## Using Passive DNS Analysis to Automatically Detect Malicious Domains

6950 Hollister Avenue, Suite 101
Goleta, CA 93117
lastline.com | 805.456.7076

## Abstract

The domain name service (DNS) plays an important role in the operation of the Internet, providing a two-way mapping between domain names and their numerical identifiers. Given its fundamental role, it is not surprising that a wide variety of malicious activities involve the domain name service in one way or another. For example, bots resolve DNS names to locate their command and control servers, and spam mails contain URLs that link to domains that resolve to scam servers. Thus, it seems beneficial to monitor the use of the DNS system for signs that indicate that a certain name is used as part of a malicious operation.

In a pioneering scientific paper presented by the researchers *Leyla Bilge, Engin Kirda, Christopher Kruegel* and *Marco Balduzzi* at the Network and Distributed System Security Symposium (NDSS) in San Diego, in February 2011, the authors introduced EXPOSURE, a system that employs large-scale, passive DNS analysis techniques to detect domains that are involved in malicious activity. The authors used 15 features that they extract from the DNS traffic that allows them to characterize different properties of DNS names and the ways that they are queried. EXPOSURE was joint work between the University of California, Santa Barbara and Institute Eurecom. Lastline is a company based in Santa Barbara, California, and some of the people who were involved in the design of EXPOSURE are also involved in Lastline.

For a more detailed overview of the scientific prototype of EXPOSURE and a technical explanation of the different analysis steps in greater detail, the reader is referred to the official EXPOSURE publication: http://www.lastline.com/papers/ndss11_exposure.pdf. In this whitepaper, we give an executive summary of that scientific work and use excerpts from the paper.

Lastline has tech transferred the concepts behind EXPOSURE and has implemented a novel, improved production system to accurately and automatically identify malicious domains.

## Introduction

The Domain Name System (DNS) is a hierarchical naming system for computers, services, or any resource connected to the Internet. Clearly, as it helps Internet users locate resources such as web servers, mailing hosts, and other online services, DNS is one of the core and most important components of the Internet. Unfortunately, besides being used for obvious benign purposes, domain names are also popular for malicious use. For example, domain names are increasingly playing a role for the management of botnet command and

control servers, download sites where malicious code is hosted, and phishing pages that aim to steal sensitive information from unsuspecting victims.

In a typical Internet attack scenario, whenever an attacker manages to compromise and infect the computer of an end-user, this machine is silently transformed into a bot that listens and reacts to remote commands that are issued by the so-called botmaster. Such collections of compromised, remotely-controlled hosts are common on the Internet, and are often used to launch DoS attacks, steal sensitive user information, and send large numbers of spam messages with the aim of making a financial profit.

In another typical Internet attack scenario, attackers set up a phishing website and lure unsuspecting users into entering sensitive information such as online banking credentials and credit card numbers. The phishing website often has the look and feel of the targeted legitimate website (e.g., an online banking service) and a domain name that sounds similar.

One of the technical problems that attackers face when designing their malicious infrastructures is the question of how to implement a reliable and flexible server infrastructure, and command and control mechanism. Ironically, the attackers are faced with the same engineering challenges that global enterprises face that need to maintain a large, distributed and reliable service infrastructure for their customers. For example, in the case of botnets, that are arguably one of the most serious threats on the Internet today, the attackers need to efficiently manage remote hosts that may easily consists of thousands of compromised end-user machines. Obviously, if the IP address of the command and control server is hard-coded into the bot binary, there exists a single point of failure for the botnet. That is, from the point of view of the attacker, whenever this address is identified and is taken down, the botnet would be lost.

Analogously, in other common Internet attacks that target a large number of users, sophisticated hosting infrastructures are typically required that allow the attackers to conduct activities such as collecting the stolen information, distributing their malware, launching social engineering attempts, and hosting other malicious services such as phishing pages.

In order to better deal with the complexity of a large, distributed infrastructure, attackers have been increasingly making use of domain names. By using DNS, they acquire the flexibility to change the IP address of the malicious servers that they manage. Furthermore, they can hide their critical servers behind proxy services (e.g., using Fast- Flux) so that their malicious server is more difficult to identify and take down.

Using domain names gives attackers the flexibility of migrating their malicious servers with ease. That is, the malicious “services” that the attackers offer become more “fault-tolerant” with respect to the IP addresses where they are hosted.

The key insight behind EXPOSURE is that as malicious services are often as dependent on DNS services as benign services, being able to identify malicious domains as soon as they appear would significantly help mitigate many Internet threats that stem from botnets, phishing sites, malware hosting services, and the like. Also, the premise is that when looking at large volumes of data, DNS requests for benign and malicious domains should exhibit enough differences in behavior that they can automatically be distinguished.

In this whitepaper, we discuss the work by Bilge et al. EXPOSURE is a DNS analysis approach and a detection system. To effectively and efficiently detect domain names that are involved in malicious activity, the researchers use 15 features (9 of which are novel and had not been proposed before) that allows them to characterize different properties of DNS names and the ways that they are used (i.e., queried).

In their approach, based on features that they have identified and a training set that contains known benign and malicious domains, they train a classifier for DNS names. Being able to passively monitor real-time DNS traffic allows them to identify malware domains that have not yet been revealed by pre-compiled blacklists. Furthermore, in contrast to active DNS monitoring techniques that probe for domains that are suspected to be malicious, the analysis is stealthy, and they do not need to trigger specific malicious activity in order to acquire information about the domain. The stealthy analysis that they are able to perform has the advantage that the adversaries, the cyber-criminals, have no means to block or hinder the analysis that is performed.

## Overview of the Approach

The goal of EXPOSURE is to detect malicious domains that are used as part of malicious operations on the Internet. To this end, the authors perform a passive analysis of the DNS traffic that they have at their disposal. Since the traffic they monitor is generated by real users, they assume that some of these users are infected with malicious content, and that some malware components will be running on their systems. These components are likely to contact the domains that are found to be malicious by various sources such as public malware domain lists and spam blacklists. Hence, by studying the DNS behavior of known malicious and benign domains, the goal was to identify distinguishable generic features that are able to define the maliciousness of a given domain.

***Extracting DNS Features for Detection***

Clearly, to be able to identify DNS features that allow to distinguish between benign and malicious domains, and that allow a classifier to work well in practice, large amounts of training data are required. As the offline dataset, the reseachers recorded the recursive DNS (i.e., RDNS) traffic from Security Information Exchange (SIE). They performed offline analysis on this data and used it to determine DNS features that can be used to distinguish malicious DNS features from benign ones. The part of the RDNS traffic they used as initial input to their system consisted of the DNS answers returned from the authoritative DNS servers to the RDNS servers. An RDNS answer consists of the name of the domain queried, the time the query is issued, the duration the answer is required to be cached (i.e., TTL) and the list of IP addresses that are associated with the queried domain. Note that the RDNS servers do not share the information of the DNS query source (i.e. the IP address of the user that issues the query) due to privacy concerns.

| Feature Set | # | Feature Name |
|---|---|---|
| Time-Based Features | 1 | Short life |
| | 2 | Daily similarity |
| | 3 | Repeating patterns |
| | 4 | Access ratio |
| DNS Answer-Based Features | 5 | Number of distinct IP addresses |
| | 6 | Number of distinct countries |
| | 7 | Number of domains share the IP with |
| | 8 | Reverse DNS query results |
| TTL Value-Based Features | 9 | Average TTL |
| | 10 | Standard Deviation of TTL |
| | 11 | Number of distinct TTL values |
| | 12 | Number of TTL change |
| | 13 | Percentage usage of specific TTL ranges |
| Domain Name-Based Features | 14 | % of numerical characters |
| | 15 | % of the length of the LMS |

Table 1: Features (LMS=Longest Meaningful Substring)

By studying large amounts of DNS data, the researchers defined 15 different features that they use in the detection of malicious domains. 6 of these features have been used in previous research, in particular in detecting malicious Fast-Flux services or in classifying malicious URLs.

## Feature Selection

To determine the DNS features that are indicative of malicious behavior, the researchers tracked and studied the DNS usage of several thousand well-known benign and malicious domains for a period of several months. After this analysis period, they identified 15 features that are able to characterize malicious DNS usage. Table 1 taken from the scientific publication gives an overview of the components of the DNS requests that they analyzed (i.e., feature sets) and the features that they identified.

The complete features that the researchers use in the detection and their rationale for selecting these features are explained in detail in the full scientific publication.

## Conclusion

The domain service (DNS) is a crucial component of the Internet. DNS provides a two-way mapping between domain names and their IP addresses. Just as DNS is a critical service for the functioning of benign Internet services, it has also started to play an important role for malicious activities. For example, bots resolve DNS names to locate their command and control servers, and spam mails contain URLs that link to domains that resolve to scam servers.

In this whitepaper, we discussed EXPOSURE, a public system that employs passive DNS analysis techniques to detect malicious domains. The key insight is that it is beneficial to monitor the use of the DNS system on a large-scale for signs that indicate that a certain name is used as part of a malicious operation. The experimental results show that the approach works well in practice, and that it is useful in automatically identifying a wide category of malicious domains such as botnet command and control servers, phishing sites, and scam hosts. Some of the researchers involved in the design of EXPOSURE are also involved in Lastline. Lastline has tech transferred some of the concepts behind EXPOSURE to build a production system that is able to automatically and efficiently perform automated passive DNS analysis to detect malicious domains.

## About Lastline

Lastline is a technology pioneer dedicated to stopping advanced malware, zero-day attacks, drive-by downloads and sophisticated Advanced Persistent Threats. Lastline's flexible Previct platform provides high-resolution analysis and protection; the required network security foundational layer capable of providing exacting security legacy APT, IPS, AV and next generation firewalls simply cannot see. The Santa Barbara based company is dedicated to providing the most accurate malware detection and defense available to our customers.