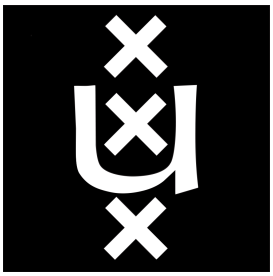
University of Amsterdam
System and Network Engineering
Security of Systems and Networks



Exhaustive Search on URL Shorteners

Alexandros Stavroulakis
Alexandros.Stavroulakis@os3.nl

Xavier Torrent Gorjón
Xavier.TorrentGorjon@os3.nl

Nikolaos Petros Triantafyllidis
Nikolaos.Triantafyllidis@os3.nl

December 9, 2014

Contents

Abstract

NOTE TO TEAM: This is just a first attempt on an abstract that can work as a guiding light. We'd better write the abstract after the report is finished. Which makes more sense. Peace. And love.

In this project we focus on URL shortening services, from a security point of view.
Our first aim is to determine the feasibility of an exhaustive mapping of all the short links to their respective long urls, estimating the cost in both time and computational resources. Secondly we try to discover the nature and the amount of sensitive (usernames, passwords, system configurations, user details, etc.) data that has been deposited to such services, and eventually pinpoint security holes that might have been leaked through them. Our final aim is to try and determine if there is some sort of mapping relationship between the long and short urls.
The research methodologies and software tools used for the project are described in detail. The results and interesting findings are presented and the appropriate discretion is applied where deemed necessary.

1 Introduction

URL shortening refers to the technique of taking any HTTP Uniform Resource Locator (URL) and producing a shortened version that links to the same Web resource, by issuing an HTTP redirect response. The purpose of this technique is to transform large (sometimes hundreds of characters long) and very descriptive URLs to something that is much sorter, easier to remember and be shared in an environment where typing space is limited (social media, mobile devices, instant messengers, etc.)

This technique has been around since the early 2000s but became really popular by the coming of Twitter, a social medium that only allowed a certain number of characters to be typed in each post (Tweet) of the user, and which started automatically shortening URLs more than 26 characters long. The first website to provide shortening services was tinyurl.com, with similar services including, among others, wp.me (by Wordpress), goo.gl (by Google) and bit.ly, with the last two being the most popular.

This report focuses on certain security issues that arise by the use of such services. The rest of this chapter is dedicated to the description of the problem we will be examining, presentation of previous work on this domain and mention of certain ethical implications that arise from our study. The second chapter is a description of the URL shortening methods in general and the two services that have been used in this study (goo.gl and bit.ly) in particular. The third chapter presents the research methods and software tools that we have designed, developed and used in this project. The fourth chapter demonstrates the results that have been produced by our research and the security implications that arise in terms of user privacy and system security. The next chapter is a discussion about suggestions and solutions that could help mediate the security problems of such services. The last chapter summarises the conclusions of the project and proposes ways to improve the current work.

1.1 Problem Description

URLs tend to carry a lot of information besides the location of the web resource. That can include, among others, file hierarchies, configuration parameters, IP addresses, port numbers and in more serious cases, usernames, clear text passwords, links for unauthenticated access to internal servers. Moreover, URLs can link to web content that would normally be inaccessible by non-authenticated users through permalinks and hotlinks. Namely that includes social network profiles of users, private pictures and videos, online documents etc. It is apparent that there is a lot of web content that can either be exploited for attacking the computer infrastructure of companies and organisations or can lead to user identification and leakage of their private data. Obviously, in all of the aforementioned cases this content has to be securely kept away from the public eye.

Hunting for URLs and web content that can be exploited is not something new.

1.2 Previous Work

Most of the previous work which has been done on URL shortening services is based on the use of short URLs and its correlation with SPAM and phishing techniques, whether that is to prevent spamming or to explain how these two are combined. One example of the latter is how the original URL is masked in a way that the receiver of such a malicious email will not be able to realize the fact that by clicking such a link, he or she will not be redirected to a legitimate website.

Another example was the investigation of specific countermeasures take from these particular services to defend against the manipulation of the shortened URLs for malicious purposes; also trying to determine and statistically analyze the extent of spamming given certain geographical locations in which the services were used.

As for the analysis of the URLs as independant links and their respective data, the primary focus was on short URLs collected by popular social media such as Twitter. And the statistics revolved around their popularity lifetimes and the expectancy of the amount of clicks these URLs would get.

1.3 Ethical Implications

2 Shortening Services

2.1 Goo.gl

2.2 Bit.ly

3 Research Methodologies

3.1 URL Crawling

3.1.1 Goo.gl

3.1.2 Bit.ly

3.2 Data Mining

3.2.1 RegEx

3.2.2 MongoDB Queries

4 Results

4.1 What can an attacker do?

4.2 What have we found?

4.3 User Privacy Implications

4.4 Sysem Security

4.5 Stats

5 Suggestions

After studying our research results and having witnessed what kind of options and possible offensive routes there are available for an attacker to choose from, we have some suggestions