



DEFENCE TECH

Terra, Cielo, Mare, Spazio, Spazio cibernetico.
PROTEGGIAMOLI

URL Obfuscation

Malware Lab Analysis Report

Summary

1. Our Malware Lab	03
2. Executive Summary	05
3. Analysis	08
3.1 The URL schema	09
3.2 Abusing Unicode to spoof URLs	11
3.3 Misleading TLDs: .zip Domains	12
3.4 IP address encoding formats	13
3.5 URL Shorteners	14
3.6 URL Doppelgangers and typosquatting	15
3.7 URL Redirects	16
4. Conclusions	17

This document is protected by copyright laws and contains material proprietary to the Defence Tech Holding S.p.A Società Benefit. It or any components may not be reproduced, republished, distributed, transmitted, displayed, broadcast or otherwise exploited in any manner without the express prior written permission of Defence Tech Holding S.p.A Società Benefit. The receipt or possession of this document does not convey any rights to reproduce, disclose, or distribute its contents, or to manufacture, use, or sell anything that it may describe, in whole or in part.

1

Our Malware Lab

1. Our Malware Lab

Defence Tech Malware Lab daily performs dissection of malware with the aim of timely understanding the technological evolutions of attacks, consolidating the knowledge of necessary to make more effective and faster the process of incidents responding, contributing to spreading information about emerging threats into the expert's community and among its clients.

Malware Lab analysts are continuously engaged in searching and experimenting new analysis tools, for increasing accuracy and scope of action with regard to

the proliferation of new evasion and anti-analysis techniques adopted by malwares.

The Malware Lab is also committed to the development of proprietary tools for malware analysis and supporting the management and response of incidents.

Besides malware analysis, Malware Lab ideated and implemented an automatic process of extraction of **Indicators of Compromise (IOC)** that is daily run on dozens of new malwares, intercepted in the wide for populating our Knowledge Base.



CORRADO AARON VISAGGIO

Group Chief Scientist Officer & Malware Lab Director

a.visaggio@defencetech.it



DEFENCE TECH

2

Executive Summary

2. Executive Summary

Uniform Resource Locator (URL) is commonly used to point to websites under the name of "links". By now checking the name of the website before opening a link is a common practice explained, in one way or another, to everyone who operates devices with sensitive information.

Behind URLs there is a proper structured standard defining how they should be

structured. In this report, we are going to illustrate the URL syntax and the way it can be -and is- abused by threat actors to attack average users.

Several obfuscation techniques can be used to obfuscate malicious URLs to make them look legitimate, and then these malicious links are used by attackers for Phishing or Smishing attacks:

- Phishing consists of fraudulent e-mails that try to trick the victim into opening a malicious link, masquerading as a trusted sender.
- Smishing, identical to Phishing, but it consists of taking advantage of text messages that often claim to come from the victim's bank or other trustworthy companies, asking for financial or personal data, or maybe to click on a malicious link.



Successful Phishing-based attacks originate in URL scheme manipulation, which is performed via known exploitations, by replacing characters of the original domain or through the abuse of Unicode characters inside the URL string.

Threat actors smartly use social engineering tactics in order to deceive naïve users into clicking a link which hides a malicious URL. Companies' employees are usually targeted by different social engineering techniques, based on Phishing, as:

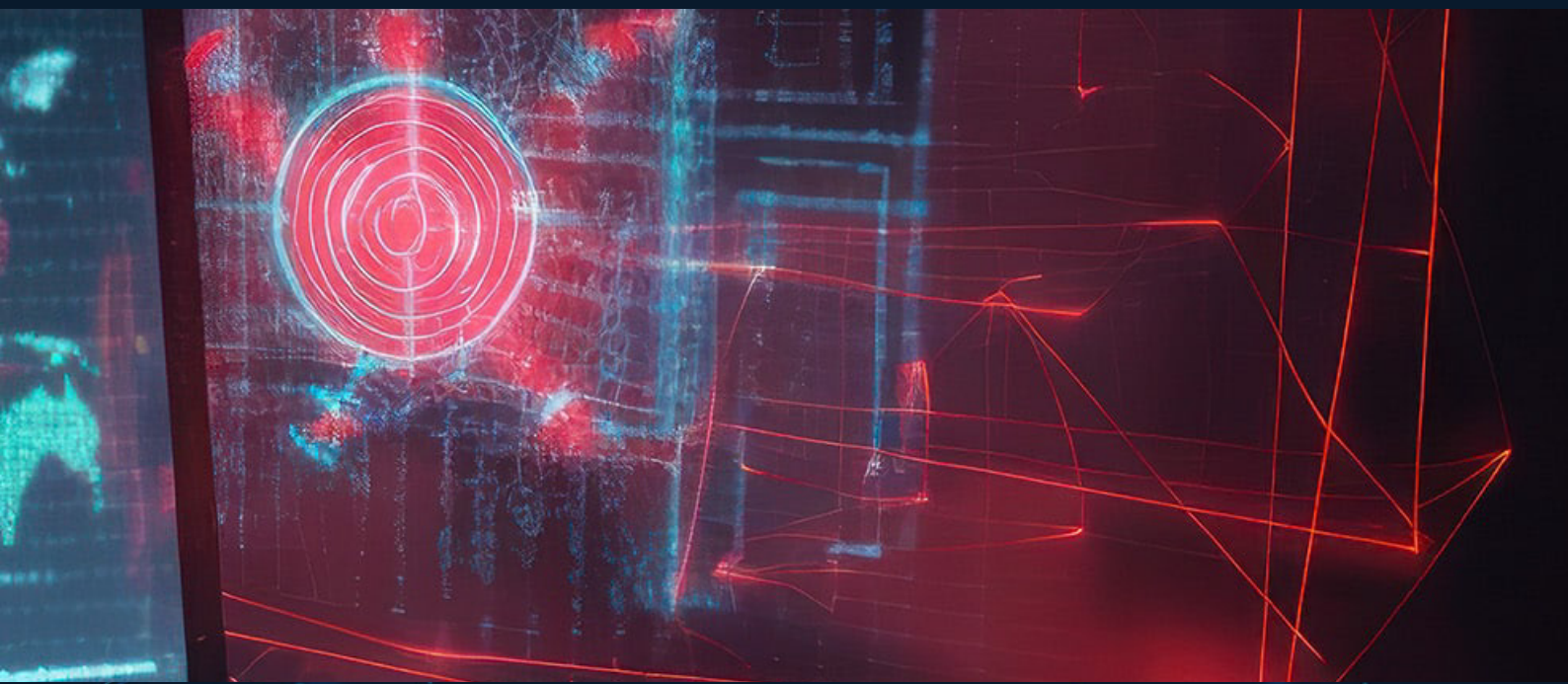
- Spear Phishing

Scam via communications, apparently by a trusted source, such as colleagues' faked e-mails, targeting a specific person, organization, or business, which leads to a fake web site. Although often intended to steal data for malicious purposes, cybercriminals may also want to install malware on a targeted user's computer.

- Whaling

It is identical to Spear Phishing, except for the importance of the target. Whaling is defined when the target is a high-level profile (e.g. company administrators). Moreover, Whaling is meant for the size of the targets or rather "catching a big fish".

Another different technique is: Typosquatting or URL Hijacking. This is based on spelling errors in the URL, or in specific cases compromised web sites which are used to carry out the attack.



3

Analysis

3. Analysis

Behind the scenes of every attack, which involves links concealing malicious URLs, there are different techniques.

We will break down some of the most complex obfuscation techniques by gradually explaining the various parts and standards that come into play when forging malicious but legitimate-looking URLs.

3.1 The URL schema

The syntax and the semantics of a URL are defined by the official Internet standard RFC1738¹. According to the standard, the structure of all URLs must be as in figure 1:

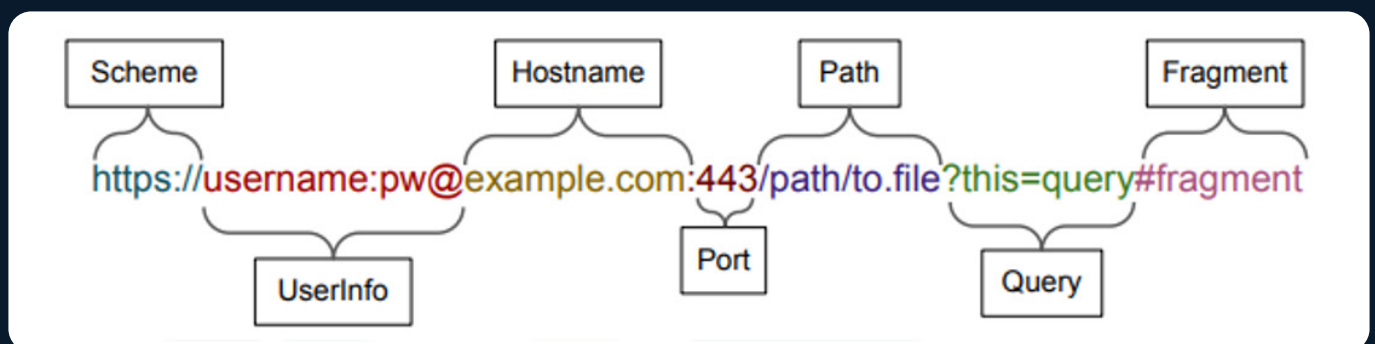


Figure 1. URL Syntactic Elements²

In practice correctly parsing URLs is surprisingly hard, this complexity has led multiple times to software bugs or even vulnerabilities in URL sanitization code, however in this report we will be discussing how this standard can be abused for phishing.

¹ <https://www.rfc-editor.org/rfc/rfc1738>

² <https://cv.jeyrey.net/img?equivocal-urls>

Most components of the URL are well-known but there is one in particular that catches the eye: the UserInfo field. This field is interpreted by the browser as user credentials for the website and sent with the basic authentication header in the HTTP request.

This means that effectively all the text before the "@" character is ignored and the real domain name is the text that follows it, this easily lends itself to phishing by faking the Hostname of a URL. Figure 2 shows this effect in action.



`https://google.com@bing.com`

Figure 2. Typing the URL with the '@' operator

This URL will in fact open bing.com while attempting to login as the "google.com" user, in practice most browsers do not send authentication information anymore.

According to the standard the forward slash "/" character is reserved for path separators, this means placing it before "@" will lead to the browser parsing all the text before the first slash as a domain name. Figure 3 shows an example of this:



`https://google.com/search@bing.com`

Figure 3. Typing the URL with the slash character before the '@' sign

This will resolve to the expected domain which is google.com

This technique by itself may appear limited and easy to spot, however as we are about to see we can combine it with more text encoding quirks to produce believable phishing links.

3.2 Abusing Unicode to spoof URLs

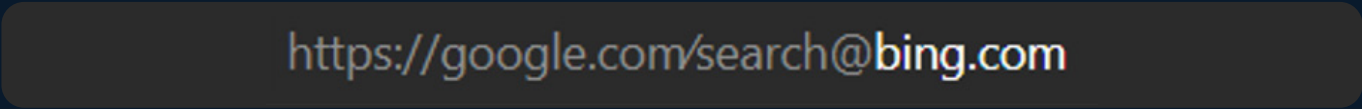
The first hurdle to overcome is the limitation of not being able to insert forward slashes in the UserInfo field.

A common trick is to use homographs that resemble a forward slash (or even indistinguishable when using certain fonts) but are actually a different Unicode codepoint, considering the following example:

- **Solidus**, Unicode Character `" / "` (U+002F), the only valid path separator
- **Fraction slash**, Unicode Character `" ⁄ "` (U+2044)
- **Division slash**, Unicode Character `" ⁀ "` (U+2215)

Only the first one is the real forward slash character that will interrupt the URL UserInfo parsing, the others are just look-alike characters with no special meaning.

Figure 4 shows a URL containing one of such characters as the username for a URL.



`https://google.com/search@bing.com`

Figure 4. Typing the URL with the Fraction Slash on MS Edge

There is a discussion in Chromium's bug tracker from 2016³ about mitigating this but it has been closed without any actual change due to compatibility concerns.

³ <https://bugs.chromium.org/p/chromium/issues/detail?id=584644>

3.3 Misleading TLDs: .zip Domains

Top Level Domains (TLDs) are the common well-known domain “extensions” such as .com, .net and so on. Traditionally these were meant to have specific meanings however recently the internet has seen the rise of gTLDs, where the g stands for generic, with a broader range of meanings not necessarily related to nationality or usage.

One gTLD recently introduced by Google is .zip, at the time of writing it has been a few weeks since such domains have become available for sale. The .zip TLD was met with criticism from cybersecuri-

ty researchers due to its name being associated with the well-known ZIP archive format.

By using this Top-Level Domain (TLD) and the URL UserInfo trick explained early, it is possible to forge very convincing phishing links, which can even deceive advanced users not paying attention.

The following is a very spot-on example by security researcher Bobby Rauch producing a fake Kubernetes download links which seemingly differs a little from the legitimate one:

Malicious: [hxxps://github.com/kubernetes/kubernetes/archive/refs/tags/@v1271.zip](https://github.com/kubernetes/kubernetes/archive/refs/tags/@v1271.zip)

Legitimate: <https://github.com/kubernetes/kubernetes/archive/refs/tags/v1.27.1.zip>

This example uses the same manipulation technique discussed earlier: the Unicode Fraction slash character allows faking a path and the “@” character is used to actually point to the domain “v1271.zip” (non-existent at the time of writing).

Since .zip domains are relatively new, there aren’t many documented attacks using them, but we expect that threat actors will be fast on picking them up for Phishing attacks, probably with an initial high rate of success.

3.4 IP address encoding formats

When thinking about IP addresses (more specifically IPv4 addresses) it's common to think about the decimal dotted notation that looks like "a.b.c.d" where each placeholder is a number from 0 to 255. And that is correct but not the full story: the IPv4 format in URIs was standardised⁴ only after being widely adopted; this leads to different operating systems interpreting seemingly invalid values as valid IPv4 addresses.

Mandiant⁵ investigated an obfuscation technique used by malware, following an example published on Twitter⁶:

```
hxxp://google.com@1157586937
```

Note that the destination which can be reached is after the "@" sign, where the Hostname is hidden by a simple format: 1157586937. As the previous examples, the destination of this URL is not Google, but the server represented by the Integer.

IPv4 addresses are represented in memory as unsigned 32-bit integers and indeed, most operating systems will parse these addresses from their numerical form; figure 5 shows a Linux system pinging the IP from the previous example. The same behaviour can be observed on Windows.

```
user@localhost:~$ ping 1157586937
PING 1157586937 (68.255.95.249): 56 data bytes
```

Figure 5. Pinging the integer which hides the IP address

⁴ <https://datatracker.ietf.org/doc/html/rfc3986#section-3.2.2>

⁵ <https://www.mandiant.com/resources/blog/url-obfuscation-schema-abuse>

⁶ https://twitter.com/ankit_anubhav/status/1592109955641126912

The integer IP format in the previous example acts as a second level of obfuscation concealing the real server from the average user.

Moreover, there are more uncommon formats that are correctly interpreted by browsers and can be abused by threat actors:

- Hexadecimal: 0xC0.0xA8.0x01.0x01
- Octal: 0300.0250.01.01
- Replacing the last two bytes with a 16-bit integer: 192.168.257

All the previous examples are correctly parsed by the “ping” utility and will connect to 192.168.1.1.

One more quirk is that, like IPv6, bytes containing zeroes in the middle can be

omitted allowing the use of addresses such as 20.2 which resolves to 20.0.0.2 and 20.1.2 which resolves to 20.1.0.2.

These different encodings can be combined producing examples more complicated to unpack visually; the following examples are valid addresses that resolve to 10.10.0.2:

- 10.655362
- 0xa.012.2
- 012.0xa0002
- 10.02400002

There are several open-source tools that can automatically obfuscate addresses, such as IPFuscator⁷, helping to hide and conceal malicious domains and addresses.

⁷ <https://github.com/vysecurity/IPFuscator>

3.5 URL Shorteners

A more common obfuscation technique is abusing URL shortening services like bit.ly, tinyurl.com, t.ly and similar.

These are legitimate services used to abbreviate long URLs typically to send them over channels with a limited character limit like SMS. For the same reason threat actors can abuse them to hide the real malicious URL behind the shortener domain.

In practice, most business won't rely on

public shortening services and will host their own, so an SMS link for an alleged warning from a bank or some kind of two factor authentication is likely to be a Phishing attempt and the users should be trained to spot these.

Most URL shortening services allow reporting scam URLs to block them but, for that to work, they need to be publicly known which means it may already be too late for a few victims.

3.6 URL Doppelgangers and typosquatting

It is a technique which replace one or more characters of a well-known domain with similar looking ones to deceive the user. An example could be:

www.paypal.com
www.paypal.com

The previous URLs seem identical because of the font "Calibri", but with a different font the difference is visible:

www.paypal.com
www.paypaI.com

It is reasonable to believe that this technique has a high rate of success when used for Phishing e-mails. So, by using a particular font, the user can be easily tricked by a doppelganger domain.

In the case of doppelganging, the attack vectors are more limited since non-ASCII characters are not allowed in domain names and instead are replaced with the Punycode⁸ encoding.

The examples shown in the previous section allowed faking URL paths because they used Unicode characters in the UserInfo part of the URL, which does not have such limitations.

Attempting to use Unicode characters to

create a fake path in a domain name such as "fakepath.com" will actually produce an ASCII domain name encoded as "xn--fakepath-h03d.com" which is correctly rendered by most browsers making it easier to spot.

A different but similar technique is typosquatting⁹, a technique where the attackers target users' typos or bad habits such as writing the URL directly in the browser, instead of searching through a search web engine.

A classic example is typing "google.com" instead of "google.com". A threat actor could emulate the legitimate web site with malicious intentions.

⁸ <https://en.wikipedia.org/wiki/Punycode>

⁹ <https://support.microsoft.com/en-us/topic/what-is-typosquatting-54a18872-8459-4d47-b3e3-d84d9a362eb0>

3.7 URL Redirects

Many web sites provide the mechanism to redirect the user to another one, which could be hijacked by a threat actor to redirect users to a malicious page. An example of what this action looks like is:

<http://example.com/example.php?url=http://malicious.example.com>

The distracted user often does not see the second part of the URL after clicking on the link, especially when it's viewed from a small device (such as a smartphone), so he can be hoaxed to click on it.

It's important to note that this technique is made possible by a class of vulnerabilities known as "open redirects"¹⁰: these are often undesirable bugs and should be reported to the website administrator.



¹⁰ <https://cwe.mitre.org/data/definitions/601.html>

4

Conclusions

4. Conclusions

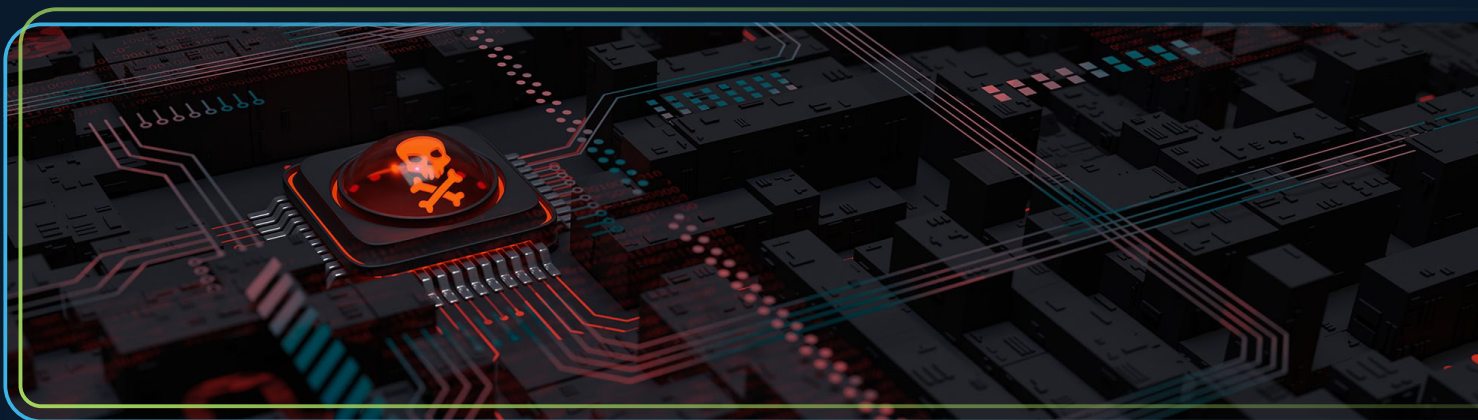
Phishing, be it via e-mail or SMS, is a major threat to companies and individuals alike; everyone has some basic awareness of the dangers of clicking links, that's why scammers employ a wide range of social engineering tactics to make the victim let his guard down. Most common techniques involve simulating an urgent matter that requires immediate action or spoofing the sender, so it looks like the message is from a legitimate person or business.

Obviously, our first recommendation towards end users is to always pay attention to everything that is received, especially for messages containing links. As we have shown distinguishing fake links is not always easy, in case of doubts,

there are online platform and tools, such as VirusTotal¹¹, that can help the user to check if the URL has been previously recognised as malicious.

Since many threat actors target companies or industries by attacking their employees, preventing phishing-based attacks is a critical matter for companies. Employees need to be periodically trained to understand basic cybersecurity awareness and to recognise suspicious senders and messages.

Every business needs to be aware on how to prevent, or to behave after, an attack; there is wide range literature and region-specific regulations on the matter, some guidance is also provided by NIST¹².



¹¹ <https://www.virustotal.com/gui/home/url>

¹² <https://www.nist.gov/itl/smallbusinesscyber/guidance-topic/phishing>



DEFENCE TECH

Terra, Cielo, Mare, Spazio, Spazio cibernetico.
PROTEGGIAMOLI

DONE IT
IT SECURITY

NEXT
INGEGNERIA DEI SISTEMI

FORAMIL
RADAR TECHNOLOGIES & DEFENCE SYSTEMS

INN·DESI
electronic systems

Defence Tech Holding S.p.A Società Benefit

Via Giacomo Peroni, 452 - 00131 Roma

tel. 06.45752720 - fax 06.45752721

info@defencetech.it - www.defencetech.it