



Global Information Assurance Certification Paper

Copyright SANS Institute
Author Retains Full Rights

This paper is taken from the GIAC directory of certified professionals. Reposting is not permitted without express written permission.

Interested in learning more?

Check out the list of upcoming events offering
"Network Monitoring and Threat Detection In-Depth (Security 503)"
at <http://www.giac.org/registration/gcia>

Tracking Online Counterfeiters

GIAC (GCIA) Gold Certification

Author: Emilio Casbas, ecasbas@gmail.com

Advisor: Adam Kliarsky

Accepted: March 16th, 2017

Abstract

The counterfeiting market makes-up a vast global business where the impact of fraudulent activity is hard to quantify. Counterfeiting is a global issue which has become more complex as black market activities moved to internet. The online counterfeiters create thousands of websites with different approaches as part of their strategy to lure unsuspected shoppers. This paper presents their most common tactics and its relation with the “Black market commoditization”. It will show its resilience against takedown efforts and it will provide some guidance about how to detect them. With the knowledge acquired, a new kind of threat intelligence feed could be generated. This information might be integrated into existing security technologies such as either proxies, Intrusion Detection Systems (IDSs) or Security Information and Event Management systems (SIEMs). The ultimate goal is to shed light on this increasing fraud vector so new detection capabilities can be deployed into existing services thus protecting users from unsafe sites.

1. Introduction

The last report (“Trade in Counterfeit and pirated goods, 2016”) published by the Organization for Economic Cooperation and Development (OECD) show that trade in counterfeit amounted up to 2.5% of world trade in 2013. Though the online environment in the context of counterfeiting of physical goods is nuanced, mainly due to its pace and industry-specific, more research is needed in this area to fully understand the issues involved. Fake luxury goods are sold all over the Internet. There is an underground economy where a large number of globally distributed criminals trade in data, knowledge, and services with the goal to defraud users and business (Kurt Thomas, 2015). Counterfeit goods are products that are manufactured or first sold with a trademark but without the authorization of the trademark owner. Counterfeiting is not a new phenomenon but with Internet the business is increasing rapidly. Online counterfeiters can easily set up websites and ship luxury counterfeit goods directly to the consumers. Using resources, infrastructure and services related to the malware ecosystem counterfeiters are able to divert consumers to rogue e-commerce websites.

This research is focused on the online marketplace that has resulted due to the booming e-commerce in global trade and the recent injunctions against market operators. For example, Cartier and other Claimants sought an injunction against the five main ISPs (Internet Service Providers) in the UK requiring them to block access to websites related with the online counterfeiting market (Cartier International AG & Ors v British Sky Broadcasting Ltd & Ors, 2014). Italian Police likewise had shut down a website related with the online counterfeiting (Reuters, 2016). And in a winning battle against cybersquatting by the Alexander Wang brand, that resulted in 459 counterfeiting websites being shut down (WWD, 2016). In the European Union (EU) customs seizures skyrocketed in 2015 (EU customs enforcement, 2015) and raised concerns about the online counterfeiting fraud being “*massively underestimated*.” This online fraud expands over 15% every year (Counting the cost of Counterfeiting, Netnames 2015). The online counterfeiting threat has traction in the Intellectual Property (IP) world whereas in InfoSec does not. “InfoSec also has a tendency to obsess over the technical sophistication of an attack instead of the impact it has on real people” (Stamos, 2016). Despite having

estimated revenues similar to well-known threats like the Zeus Trojan or even higher than the ClickFraud threat (Kurt Thomas & Danny Yuxing, 2015, page 7, table 1), there are still no existing intelligence sources or countermeasures to protect users against this online threat. Apparently this problem is not relevant in the security community due to observations as:

- Lack of counterfeit website trackers which exist for the Zeus malware.
- Inexistent proxy categories for the specific “counterfeit” threat. (Blue Coat Category descriptions, 2016)
- No related URL categories. (Zscaler URL categories, 2016)
- Lack of articles/research mentioning the scale of this online fraud.

The industries that are targeted by online counterfeiters range from apparel, software and music to automobile and airplane parts and toys. This research will focus on the brands which are more intensely targeted by the counterfeiters (OECD, 2016, Chapter 4, Figure 4.2) which do fit with the dataset collected by the research used to write this paper (*Appendix A contain some metrics about this research*).

1.1. An Easy and Free Approach to this Problem

One approach to solve this problem was developed with an online anti-counterfeiting website called Desenmascara.me (<http://desenmascara.me>).

Desenmascarama is Spanish and translates to “Unmask me” in English. This tool was developed as a proof of concept to promote security awareness among the web site owners (see *Appendix B* for some use case examples and to see signs of the tool in the web logs). The tool was presented in the Arsenal area of Black Hat Europe 2014 (Black Hat, 2014).

The tool was mentioned in some forums so the activity increased. This has led to an increase in user’s questions through the public contact form of the website. It turns out users were scanning websites to know the sites legitimacy and posted questions related to the result. This behavior triggered an analytics based research on the databases to discover a huge number of counterfeiting based websites. This finding was the turning

point of the goal of the online tool: to focus on the detection of counterfeit-based websites. The tool is available online and the only step required to use it is to enter a web address to scan and click a button. The process is described below:

Figure 1. Desenmascara.me main function



The goal of desenmascara.me is to show what is behind the scenes of a website in terms of metadata. Based on the steps of the figure 1:

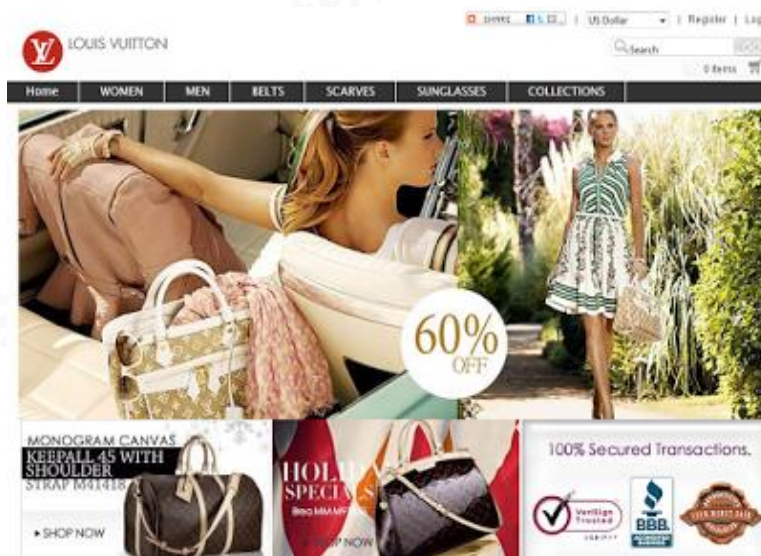
1. Any user can enter a website to scan (unmask)
2. Desenmascara.me will analyze all the metadata extracted (Hypertext Markup Language [HTML] code of the main site and Hyper Text Transfer Protocol [HTTP] headers).
3. Based on the findings it will show a report with 3 possible statuses:
 - **Site flagged as FAKE:** The website is a FAKE website related with the online counterfeiting.
 - **Site not flagged:** Any normal website. This should be the most prevalent result (score given based on the metadata analyzed).
 - **Site flagged as malware:** The website was blacklisted by Google SafeBrowsing¹. There are cases where a website is flagged as FAKE and as malware at the same time.

¹ **Google SafeBrowsing:** Google service that lets client applications check URLs against Google's constantly updated lists of unsafe web resources.

1.2. Counterfeit-related Websites

The counterfeit merchants no longer sell their wares on the nearest street corner of large metropolitan areas. Expensive-looking items heavily discounted to attract buyers. Potential customers didn't care about the risk of buying the black market items. The counterfeiters have expanded their market to the online marketplace, the difference is the customers are not as savvy and unaware they are purchasing from fraudulent sites. The focus of this paper is on counterfeit websites ('FAKE' websites). Figure 2 shows an image of a suspicious *Louis Vuitton* website which might appeal to the average uninformed internet consumer.

Figure 2. Fake Louis Vuitton website



In the next sections we will see different categories of FAKE websites and tactics that the online counterfeiters use to lure users. In order to better understand the magnitude of this online fraud in terms of brands affected, a longer overview of FAKE websites screenshots is available in the *Appendix C*.

1.3. The Online Counterfeiters and Their Relation with the Internet Underground Economy

The Internet underground economy is the culprit of the current well-known threats such as fake-antivirus, ransomware, Trojan banks and any kind of commodity crimeware available. "Crimeware represents malware infections within organizations that

are not associated with more specialized classification patterns” (Verizon DBR, 2015). This underground economy has well-defined roles with inter dependencies such as to buy and sell compromised web servers, exploit kits, spam distribution and wholesale access to stolen user credentials including usernames, passwords and credit card details among other sensitive personal data. The figure 3 represents just a single abuse monetization value chain (Kurt Thomas & Danny Yuxing, 2016). It starts by spamming to Twitter users through fake accounts to advertise FAKE sites, go through potential victims landing in those FAKE sites when clicking in the URL, and might end with an order placement including the victim’s credit card for payment (whose credentials might be stolen, Kaspersky 2016), ultimately the delivery is coordinated (in the best case) from the illegal manufacturers.

Figure 3. Full underground value chain



The above figure shows the full underground value chain required to make money from spamming FAKE websites related with the online counterfeiting. A research about the payment processors involved with online counterfeiting showed that 97% were handled by China’s largest three banks (McCoy, 2016). Moreover, the table in the figure 4 shows estimated revenues from some illegal products and services of this underground economy (Kurt Thomas & Danny Yuxing, 2016, page 7):

Figure 4. Estimated revenue by profit centers, highlighting the FAKE sites related.

Profit Center	Strategy	Estimated Revenue	Time Frame
Spamvertised products	Pharmaceuticals [97]	\$12-92 million	2007-2010
	Luxury knock-offs [152]	\$68 million	2013-2014
Scareware & Ransomware	Fake anti-virus [133]	\$130 million	2008-2010
	CryptoLocker [159]*	\$3 million	2013-2014
Clickfraud	ZeroAccess [115]	\$36 million	2013
	DNS Changer [149]*	\$14 million	2007-2011
Financial Scams	Pump and dump [150]*	\$120 million	2008-2013
	419 scammers [8]*	\$200 million	2006
Credit Card Theft	ATM withdrawal scam [118]*	\$45 million	1 day
	Zeus banking trojan [9]*	\$70 million	2009-2010
	Re-selling stolen cards [35]*	\$300 million	?-2013

Table 1: Estimated revenue from a multitude of profit strategies (irrespective of operating costs). These strategies span the spectrum of cybercrime: from selling illegal products to outright credit theft. We annotate all industry and government estimates of criminal revenue with an asterisk to emphasize an unknown collection methodology. We caution these values may be overestimates.

The profit centers are divided based on different kinds of fraud. While almost all these threats are included in the plethora amount of threat intelligence providers available nowadays, there is one threat which is being ignored by the security industry, and **the revenue for the bad actors is almost the same as the infamous Zeus banking Trojan**. For simplicity, in the table 1 we see the term by which we will refer to this threat in the current paper.

Table 1: The focus of this paper in short terms

Throughout the paper the word “FAKE” in all capital letters is used. This is a short term to indicate the site refers to *Luxury knock-offs* as highlighted in the Figure 4.

2. Fake Sites Classification

Once we have more background about the online counterfeiting threat, how it looks like and its relationship with the underground economy; let’s see how they do operate more in detail. The fraudsters use different methods to deploy the FAKE sites in an orchestrated way. While there might be many approaches to set up FAKE sites, the next four tactics are the most prevalent based on the data collected for this research. The next classifications will serve us for the purpose to know how an average Internet user might end up in a FAKE site related with the online counterfeiting and the peculiarities to be lured which we will use later to set up our lab to spot them.

Copycat Websites

Copycat websites are imitations of other well-known webs. They mimic either the legitimate agencies or vendors to misled users with the use of official-looking crown logos. These kinds of websites are the most prominent. It will show legitimate logos and pictures plus additional features to attract unsuspected users. These images show some suspicious signs:

Figure 5. Main part of the FAKE site



Figure 6. Brand included in the URL



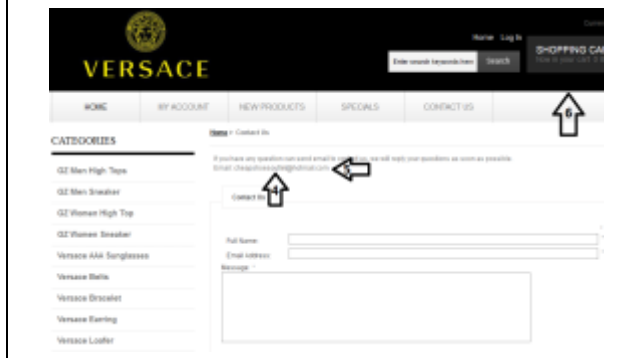
Figure 7. Official logo use



Figure 8. High quality picture of the good



Figure 9. Contact form of the FAKE site





There are some parts of the site highlighted with arrows and numbers in the previous pictures that are part of the seven main visible features of a copycat website to mislead users:

1. To use the targeted brand (Trademark) as part of the web domain along with a product optionally.
2. To use official logos (content and design elements that make it difficult for consumers to distinguish the FAKE site from the official).
3. To show high-quality pictures of the goods
4. To have generic inbox to contact them
5. To use free mail accounts in the contact form.
6. To use free software platforms with poor design.
7. To show big discounts and low prices.

In contrast there are features which are hidden to the user and only visible with a deeper inspection with a tool such as a network sniffer. The tcpdump filter showed in the figure 13 will save a file with all the IPv4 HTTP packets to and from port 80 of the specified site and will print only packets that contain data.

Figure 12. Tcpdump command to capture HTTP traffic

```
tcpdump -w versace.pcap v 'tcp port 80 and (((ip[2:2] ((ip[0]&0xf)<<2))  
((tcp[12]&0xf0)>>2)) != 0)' and host www.versaceshoes.in.net
```

As the file with the packets captured is in binary format we need to open it with

the command showed in the figure 14:

Figure 13. Tcpdump command to read a pcap file

```
tcpdump -qns 0 -A -r versace.pcap
```

Now let's see some results to check those hidden features captured with Tcpdump. As highlighted in the figure 15 the website has signs of using Zencart platform, a free shopping cart system. As showed in the figures 15 and 16, the Zencart platform is using quite strange variables and paths for a legitimate company.

Figure 14. HTTP traffic captured with Tcpdump

```
23:31:08.283579 IP 192.168.1.8.53242 > 209.134.9.17.80: tcp 511
E..3>
@.@.s.....P.h...is not blacklisted by SafeBrowsing (Read more)
v.5..m..GET /images/versace160310/Versace%20Men%20Sweater/versace-men-sweater-1008.jpg HTTP/1.1
Host: www.versaceshoes.in.net
Accept-Encoding: gzip, deflate
Cookie: cookie_test=lease_accept_for_session zenid=ba1d080116a6a899e1ad5432454c038b; a2070_pag
Connection: keep-alive
Accept: */*
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/601.7.8 (KHTML, like Ge
Accept-Language: en-us
Referer: http://www.versaceshoes.in.net/
DNT: 1
```

Figure 15. Default path names

```
v.1..m..GET /includes/templates/mytemplate/css/style_gpe.css HTTP/1.1
Host: www.versaceshoes.in.net
Accept-Language: en-us
Accept-Encoding: gzip, deflate
Cookie: cookie_test=lease_accept_for_session; zenid=ba1d080116a6a899e1ad
Connection: keep-alive
Accept: text/css,*/*;q=0.1
If-Modified-Since: Tue, 23 Dec 2014 19:56:56 GMT
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_5) AppleWebKit/6
Referer: http://www.versaceshoes.in.net/
DNT: 1
Cache-Control: max-age=0
```

Additionally, in the Figure 17 we observe signs of Chinese tracking software which is common among the FAKE sites.

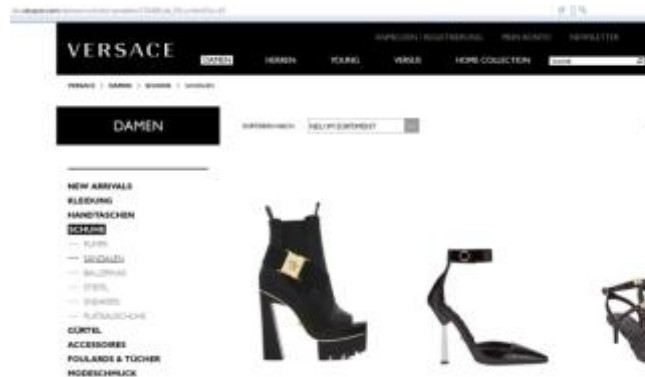
Figure 16. tcpdump result with Chinese tracking software signs

```
m...m..qGET /18822070.js HTTP/1.1
Host: js.users.51.la
Accept-Language: en-us
Accept-Encoding: gzip, deflate
Connection: keep-alive
If-None-Match: '5f2973285d3d11:0'
Accept: */*
If-Modified-Since: Fri, 01 Jul 2016 10:41:59 GMT
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_1
Referer: http://www.versaceshoes.in.net/
DNT: 1
Cache-Control: max-age=0
```

Now just for the sake of clarity let's see in the figure 18 how the official Versace website looks like:

Emilio Casbas, ecasbas@gmail.com

Figure 17. Main website of the Official Versace web site (German version)



The official version does share some of the common features of a copycat website (brand as part of the domain, official logos and high quality pictures), but as seen before, one thing they cannot copy but its hidden for the average user is the software the site is based on. Let's collect a packet capture to observe some slightly differences with the previous FAKE site:

Figure 18. capturing HTTP traffic related with versace.com site with tcpdump

```
tcpdump -w versace.official v 'tcp port 80 and (((ip[2:2]
((ip[0]&0xf)<<2)) ((tcp[12]&0xf0)>>2)) != 0)' and host versace.com
```

As before, we need to open the binary file with the below command as showed in the figure 20:

Figure 19. tcpdump command to read a pcap file

```
tcpdump -qns 0 -A -r versace.pcap
```

Now let's see just a result to show a key hidden feature such as signs of Demandware² software (Wikipedia 2017) as highlighted in the figure 21.

² Demandware is commercial software for retailers and brand manufacturers around the world

Figure 20. HTTP traffic captured with tcpdump

```

HTTP/1.1 302 Found
Date: Fri, 11 Nov 2016 08:08:28 GMT
Server: Apache
Set-Cookie: dwac_bcsvQ...
Set-Cookie: sid=qrQn...
Set-Cookie: dwpersonalisation_7254072e2668c23dc3bf6cca213a6657*; Expires=Thu, 01-Jan-1970 00:00:10 GMT; Path=/
Set-Cookie: dwresolutiondefined=true; Expires=Thu, 01-Jan-1970 00:00:10 GMT; Path=
Set-Cookie: dwgeop=DEde_DE; Domain=.versace.com; Expires=Sat, 11-Nov-2017 00:00:00; Path=
Set-Cookie: dwsid=nrAxc_cprJbEc5Kxi2eWXL-GyrP6YI0F230pdGmHCptJ6Ljs4KHBI0pA_fgbUbs...
Cache-Control: no-cache, no-store, must-revalidate
Pragma: no-cache
Expires: Thu, 01 Dec 1994 16:00:00 GMT
x-dw-request-base-id: ICBH-VglFHzIAaAK
Location: http://de.versace.com/on-demandware.store/Sites-DE-Site/de_DE/Home-Show
Vary: Accept-Encoding
Content-Encoding: gzip
Accept-Ranges: bytes
Content-Length: 806
Keep-Alive: timeout=5, max=100
Connection: Keep-Alive
Content-Type: text/html; charset=UTF-8

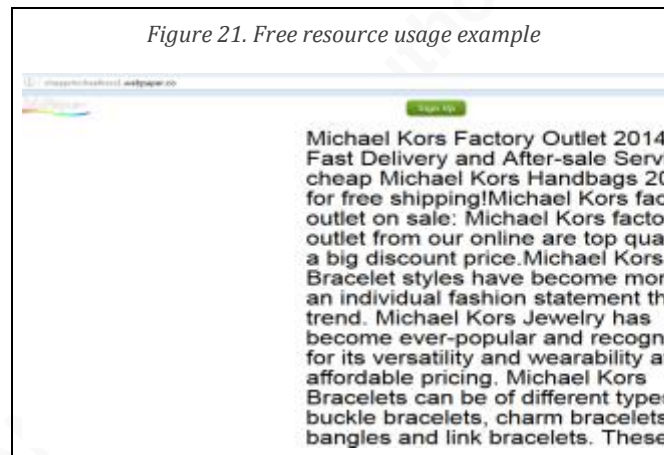
```

In this case the official Versace web is based on Demandware. Signs of this software are a strong indicator the site is a legitimate one, which has invested time and money to setup the web. But the drawback is that signs like these are hiding to the average user. The research Neural Signatures of User-centered Security (Ajaya Neupane & Nitesh Saxena, 2014) does show that real website detection is an easier task compared to fake website detection (with an accuracy of 46%). The global consumer shopping habits survey (MarkMonitor, 2015) revealed that 24% of consumers have bought products online that turned out to be fake. Therefore the brand names included in the website domain and the official brand logos along with pictures of the available products are hard to distinguish from official sites by an average Internet user. This strategy is uniform through the different FAKE sites classification with slightly differences as we will see in the next sections.

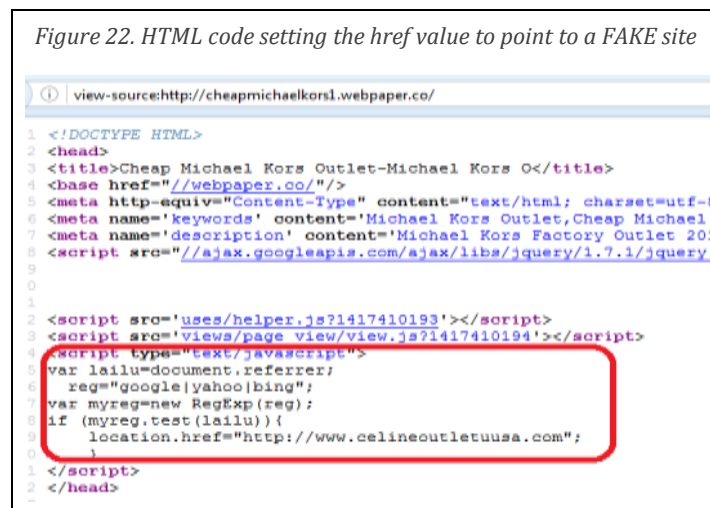
Free Hosting Based Sites

To have custom domains to store a FAKE website like the previous copycat websites it costs money. A cheaper approach is to leverage free online resources like blog post platforms and social networks. These kind of FAKE sites are the most easy to detect as usually the actors use free subdomains to create either FAKE shops or pages to support another related FAKE sites by using Black Hat Search Engine Optimization (SEO) tactics. Some free resources used by online counterfeiters include mainly among others those showed in the *Appendix D*.

Though most websites under this category are mainly used as pivotal websites to move traffic toward FAKE sites of the first category, which is the most prevalent. As these platforms are free, the counterfeiters (or the people of the underground economy focuses in this task) set up thousands of sites with SEO techniques so that the FAKE sites listings show up at or near the top of relevant search results and misdirect consumers searching for genuine products. As an example under this category we can see in the figure 22 a free hosting base site used as a pivot towards the FAKE site:



Initially it seems to be a static website with some advertisement and statements about a brand. But by looking what's behind it, in the source code as highlighted in the figure 23, we see its real intentions:



The website in the figure 22 will be seen like that only when the access is directly through the web address. Instead if an unsuspected user is using a search engine, as those highlighted in the figure 23 –looking for some Michael Kors product with big discount prizes- or similar, and it turns out a search result is the website above, the user clicking on it will end up in the fake website: *www.celineoutletusa.com*, as instructed by the code highlighted in the previous picture. This old technique of presenting different content to search engine spiders than presented to the user's browser is known as Cloaking, and it has been widely seen under this classification. Another free hosting based FAKE site example can be seen in the *Appendix F*.

Therefore under the free hosting based sites classification we observe 4 features:

1. Using the targeted brand as part of the web domain along with a product optionally.
2. Specific words to call the attention of a potential consumer to have access to the exclusive brand kind of: *outlet*, *sale* or *prizes*.
3. Bad grammar or redaction.
4. Used as pivots to forward traffic to FAKE websites using Black Hat SEO techniques.

Leveraging Compromised Websites

Compromised websites are websites that are hacked to include content from attack sites to either act as a hop in the malware distribution chain or to be replaced with a totally unrelated content ending up with defacement (Wikipedia, 2016). Though website defacement is commonly related to hacktivism, the fraudsters leverage their relationship with the underground actors to either compromise websites or to use existing compromised sites to store their FAKE ones. The compromised websites have an intrinsic value reflected in their existing links and search ranking. By using this technique the fraudsters drive user traffic to its FAKE site (compromised) without much effort. Additionally SEO techniques (Kirill, 2011) might be used in the FAKE front page to promote other FAKE sites for specific targeted keyword queries. As example we can see

the website: www.hcitalia.it in the *Appendix H*, which among many other legitimate websites was targeted in a massive campaign compromising websites to promote FAKE sites. These legitimate websites were compromised to either redirect or host links towards recent created FAKE sites targeting to 3 brands: Michael Kors, Oakley and Ray-Ban (Tecnologias Libres, 2015). This SEO technique is the most used within the compromised websites though sometimes the fraudsters as they may have total ownership of the website they might replace the legitimate content with the full FAKE shop as showed in the *Appendix L*. But while this technique can be seen, it is not widely used due to the low return of investment. The lifetime is low, usually until the owner is notified either by users of the website or by a third party provider, which ranges from 7 to 18 days (Elie, 2016). In contrast the lifetime of a FAKE website under a custom domain (copycat website) is on average 1.5 years based on the data collected in this research.

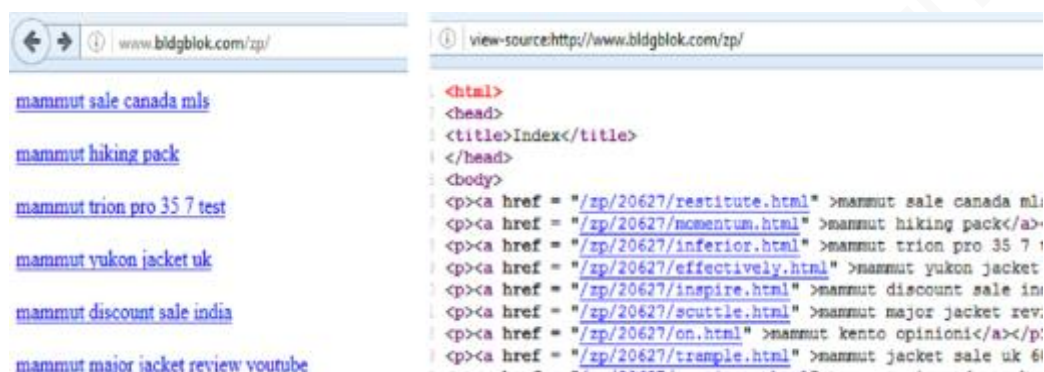
Leveraging Expiring Domains

Similar to the previous approach, leveraging expiring domains is not common but it is still used by the online counterfeiters due to its main benefit. The whole point of this technique is to leverage the existing links and traffic of the expiring domain. This will result in a FAKE site with existing potential users without much effort or SEO tactics. As example of this classification we can see the domain: www.bldgblok.com which was created on 2012-02-09 based on DomainTools³ by a Start-up company based on New York, by then it did show content about their goal and objectives as showed in Internet Archive⁴. The last twitter activity of this company was on 9 Jun 2014 (<https://twitter.com/bldgblok>), and then the domain was updated on 2016-02-10 and registered through a proxy to not disclose owner details, a usual tactic of the fraudsters. At the point of writing this paper, the website www.bldgblok.com did show a 403 Forbidden message, but the domain has still some suspicious directories with hundreds of links with the *mammut* string along with other words as the anchor text as showed in the figure 24.

³ <https://research.domaintools.com>

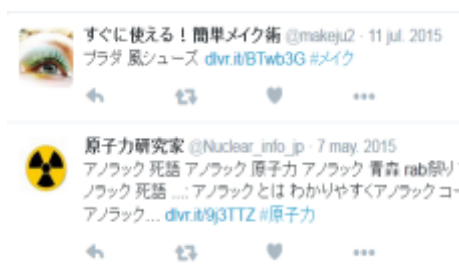
⁴ <https://web.archive.org/web/20121103155928/http://www.bldgblok.com/>

Figure 23. Suspicious directory with hundreds of links



Those links pointed to a vast amount of FAKE websites targeting to exclusive fashion brands as Gucci, Tiffany & Co., Moncler, Ray-Ban, and Nike. A full list of targeted brands and how the FAKE websites looked is available in *Appendix C*. Currently by clicking on any link of the figure 24 will led us to a website with a “Notice of Hacking⁵” message which is still available when writing this paper. This knock-off luxury campaign targeted many luxury brands at once, and as explained in section 1.3, it started spamming to twitter users through either FAKE or compromised twitter accounts.

Figure 24. Spam to promote the FAKE site bldgblok.com with shorted url through twitter



In *Appendix E* there is available an additional example of another FAKE site under the leveraging expiring domain classification.

3. Cat and Mouse Play

Most of the luxury and fashion brands targeted by this online threat are fighting against it – relentlessly. They hire specialized companies which in cooperation with

⁵ <http://servingnotice.com/Tvwvg5/intermediary.html?dec11go.com>

lawyers develop holistic strategies to take down the FAKE sites targeting them. But even with such actions, the FAKE sites owners utilize SEO tactics to propel new domain names to the top of search results after others are shut down. Even tactics so simple as to register consecutive domains as: www.christianlouboutinreplicas.com and www.christianlouboutinreplicas2.com are being used as a kind of high availability system of the FAKE sites (both FAKE sites are active at the time of writing this paper). When hundreds of websites are shutting down, hundreds more will pop it up again, and so on. As example of a brand taking action against the online counterfeiting, we can visit the domain: <http://www.rbioc.com>. This domain did store a FAKE Ray-Ban website. Currently there is a warning message of website shut down by a lawyer company and with full access to the Complaint text (Complaint Luxottica, 2016) which contains all the documentation needed to fulfill the Plaintiff against the owners of thousands of FAKE sites related to the same campaign. Within such documentation we can extract the full list of FAKE sites targeting to the brands affected. A short overview is showed in the *Appendix J*. Obviously, the defendants do not present themselves in the Court either because they conceal their identities to register the domains or they reside outside of the jurisdiction of the Court. The brands are constantly fighting this fraud and as result we can see take down messages by different brands with a special mention to Louis Vuitton (Louis Vuitton, 2016). This example along with the news referenced in the introduction section are enough to show the magnitude of this online threat which the brands are fighting behind the scenes. But the fraudsters are even using paid advertisements in the main social networks and search engines to promote their FAKE sites (*Appendix G*). While strict countermeasures are not in place to detect and flag these FAKE sites, the fraudsters will use all the tools and resources to promote this underground business.

4. Counterfeit Websites Detection

There are additional tactics used by the online counterfeiters but the four classifications explained here are the most common based on data gathered in this research. Let's quickly to recapitulate them:

- Copycat websites (the most common)

Emilio Casbas, ecasbas@gmail.com

- Free hosting based (to help promote the copycat with SEO tactics)
- Leveraging compromised sites (to drive traffic to the FAKE shop without much effort)
- Leveraging expiring domains (to make advantage of existing links and traffic)

At this stage we do know the most common tactics of the online counterfeiters, we do know how the FAKE sites look for the unsuspected users who might end up as victims of the online counterfeiting. Now let's see the issue from a security analyst's point of view instead of from a user perspective. For the sake of clarity let's see three potential scenarios where the FAKE websites detection could be useful:

1. An ISP is willing to block all the FAKE sites related with the online counterfeiting to protect their customers and to avoid feeding the counterfeiting dark business as some High Courts started to rule (TheGuardian, 2014)
2. A multinational company desires to implement a new categorization in its proxy environment to avoid their employees might end up victims of the online counterfeiting fraud or malware threats as already exposed sometimes are related.
3. A luxury brand being targeted by the online counterfeiting would like to collect all the FAKE sites infringing its trademark to take legal actions against them and to protect their potential consumers which could be misled.

The Manual Approach

It is out of the scope of this paper to show the process of gathering websites to analyze them and to show the full methodology to flag a website as FAKE, but we will see the key concepts based on information gathered by the previous classifications. The goal is to provide more flexibility to the analysts in order to let them chose the environment, tools and programming language that best suit its needs. Once the Fake website detection system has been set up and it is working, it would be up to the analyst the way to integrate it into the existing infrastructure as either analyzing traffic in real time or doing offline analysis, either focusing on the websites visited in the enterprise

(i.e.: proxy logs) or actively crawling the web to enrich the system continuously. Three potential scenarios are provided as example of usefulness of FAKE sites detection, but the possibilities are countless.

As we have learned in the *Fake sites classification* section, one common approach of the fraudsters to set up the FAKE websites is the element of the brand name along a name or adjective to call for the attention of the unsuspected user. They set up sites easily with open source software based CMS which can be easily installed with some clicks and customized afterwards with the prepared templates, whereas the web platforms used by the official websites usually are based on either commercial or quite customized software. Therefore this should be another key point to bear in mind. We should be able to analyze the HTTP headers and the HTML code of the website. With this in mind, to have the following items (given in Python examples), tools and APIs (Application Programming Interface) would be handy to set up our FAKES detection environment:

-Dictionary with brands of FAKE sites which we would like to detect (example given):

```
brands =  
[  
    'abercrombie', 'adidas', 'alexander wang', 'air jordan', 'armani', 'asics',  
    'balenciaga', 'barbour', 'belstaff', 'beats by dr dre', 'blahnik', 'burberry',  
    'lancel', 'lacoste', 'paul smith', 'prada', 'puma', 'reebok', 'ralph lauren', 'ray-ban',  
    'rayban', 'ray ban', 'rolex', 'roger vivier', 'real madrid',  
]
```

-Dictionary of decoys, words to use to call the attention of the unsuspected users (example given)

```
decoys =  
[  
    'sale', 'outlet', 'cheap', 'barat', 'oferta', 'online selling',  
]
```

-Dictionary with common CMSs used by the online counterfeiters with a height based on previous FAKE sites data analyzed or adjusted as needed. The less height the most likely a FAKE site (example given).

```
cms = { -5:['oscommerce', 'joomla', 'zencart'],  
        10:['prestashop', 'magnolia', 'percussion']  
}
```

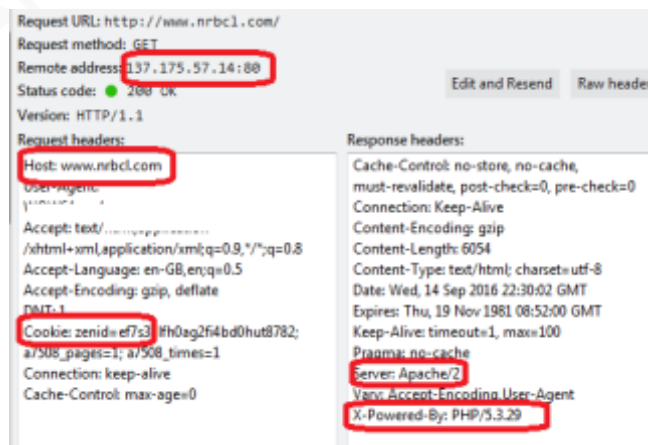
-Beautiful Soup python library to scrape the websites (Beautiful soup, 2016)

- Urllib2 python library for opening URL (urllib2, 2016)
- WhatWeb software to detect the technology behind the websites. (WhatWeb, 2016)
- Access to DNSDB -the passive DNS database of Farsight-. (DNSDB, Farshigt).

With the above arsenal in our hands, now let's take a look to the metadata of the FAKE websites to see what's behind them, and what intelligence we could leverage to correlate it with the previous items and tools. The following steps will show an analysis process example to have an idea of how to classify a website as FAKE:

1. We start analyzing the website: www.nrbcl.com
2. With the urllib2 library we will extract the HTTP headers to analyze them, but for simplicity, the HTTP headers are showed below with the developer tool included with Firefox:

Figure 25. HTTP headers of a website extracted with the Firefox developer tool



3. We have the 3 HTTP headers:
 - a. **Cookie:** with the value *zenid* showing that the website is based in ZenCart.
 - b. **Server:** It does show that is based on Apache server.
 - c. **X-Powered-By:** It does show that is not using a recent technology as such version -5.3.29- is end of life (PHP, 2014)

We save the IP address (Remote address) for later analysis.

4. Once we have the HTTP headers, we will continue by analyzing the HTML code with the beautiful soup library. Again for simplicity, we use the view page source function of Firefox to show the HTML code as in the figure 27:

Figure 26. HTML code extracted with the Firefox view page source option

```

1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-tr
2 <html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="es">
3 <head>
4 <title>Ray-Ban tiendas</title>
5 <meta http-equiv="content-type" content="text/html; charset=utf-8" />
6 <meta name="keywords" content="" />
7 <meta name="description" content="" />
8 <meta http-equiv="image:favicon" content="no" />
9 <link href="http://www.nrbel.com/favicon.ico" rel="shortcut icon" />
10 <base href="http://www.nrbel.com/" />
11 <link rel="canonical" href="http://www.nrbel.com/" />
12
13 <link rel="stylesheet" type="text/css" href="includes/templates/shop/css/styleSheet.css" />
14 <link rel="stylesheet" type="text/css" href="includes/templates/shop/css/styleSheet_productInfo.css" />
15 <script type="text/javascript" src="includes/templates/shop/javascript/jquery-min.js"></script>
16
17 </head>

```

5. With beautiful soup library we could extract the following HTML tags:
 - a. **Title:** which contains the brand Ray-Ban which we have in our dictionary of brands.
 - b. **Keywords:** Here is empty
 - c. **Description:** Here is empty (more on this later)
 - d. **HREFs:** which contain a pattern which is usually common among common FAKE sites (due to the templates used by the fraudsters)
6. We continue analyzing the HTML code and now we found the following section highlighted in the figure 28:

Figure 27. HTML code with red flags increasing the chance of a FAKE site

```

</style>
<div style="display:none;"><script language="javascript" type="text/javascript" src="http://js.users.51.la/18977308.js"></script>
</div>
</head><body id="indexHomebody">
<div id="headerWrapper">
<div id="nav_Currency_header">
<ul>
<li>
<form name="currencies form header" action="http://www.nrbel.com/" method="get"><input type="hidden" name="main page" value=""

```

The *display:none* style call our attention so we examine further the current line. What comes next is a typical sign of a FAKE website based on China actors. 51.la is a legitimate Chinese web counter site. As we have already seen in the packet dump from the figure 11, the online counterfeiters use this statistical tracking system to collect metrics about the profitability of their campaigns.

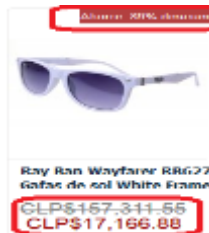
7. We continue analyzing the HTML code and now we came across the *alt* attribute with the value: *payments*, highlighted in the figure 29. This is a bad grammar mistake quite typical of the FAKE sites as we observed in the FAKE sites classification section.

Figure 28. HTML code with red flags increasing the chance of a FAKE site

```
<span class="footertexttitle">Comprea seguros</span>
<em>Aceptamos la mayoría de tarjetas de crédito</em>
</div>
</div>
<div id="siteinfoLegal" class="legalCopyright">
Copyright &copy; 2016 <a href="http://www.nxbcl.com/index.php/main_page=index">
</div>
```

8. As seen in the FAKE sites classification, either big discounts or low prices are the most common signs. In this case a discount of 89% is overwhelming as denoted in the figure 30:

Figure 29. CLP is the currency of Chile (1€ ≈ 740 CLPs)



9. In order to detect overwhelming discounts or lower prices, the *class* attribute name might be useful to look into as the highlighted examples in the figure 31:

Figure 30. Class attributes to show overwhelming discounts

```
<div class="indexPrice"><span class="normalprice">CLP$157,311.55 </span>&nbsp;<span class="pro
class="indexPrice"><span class="normalprice">CLP$157,311.55 </span>&nbsp;<span class="productSy
</h3><div class="indexPrice"><span class="normalprice">CLP$157,311.55 </span>&nbsp;<span class=
:span class="normalprice">CLP$157,311.55 </span>&nbsp;<span class="productSpecialPrice">CLP$17,:
```

10. In the step 5 we saw that the HTML tags: *keywords* and *description* were empty. This is a sign that the website is quite fresh, even not yet totally finished. By checking the domain creation details we confirm that indeed, the website was set

up yesterday (this line was written in September 15, 2016) as highlighted in the figure 32:

Figure 31. DomainTools information

Whois & Quick Stats	
Email	about@netel.com is associated with ~48,915,292 domains web@netel.com is associated with ~70 domains
Registrant Org	Ray White is associated with ~783 other domains
Registrar	GO Daddy COM, LLC
Registrar Status	clientDeleteProhibited, clientRenewProhibited, clientTransferProhibited, clientUpdateProhibited
Dates	Created on 2016-09-14 Expires on 2017-09-14 Updated on 2016-09-14
Name Servers	F1G3M62 DNSPOD.NET (has 2,813,818 domains) F1G3M62 DNSPOD.NET (has 2,813,818 domains)
IP Address	137.175.57.14 - 13 other sites hosted on this server

By only visiting a website to analyze the HTTP headers and the HTML code, we have already enough data to assess. As already seen there are many suspicious signs which point to a FAKE site. By correlating all the previous information and checking it against the profile, which previously we should have studied, of an official website of the brand, we come to the conclusion that the website analyzed (selling Ray-Ban sunglasses) can be categorized by the system as a FAKE site. This site fell under the “Copycat Website” fake site classification. A “free hosting based site” would have been easier to detect by following the same methodology. Now as an additional exercise, with the IP address of the previous FAKE website detected, we could investigate additional websites which would have a high probability to belong to the same campaign of the online counterfeiters. In order to do it, we will use passiveDNS. A passiveDNS search such the showed in the *Appendix I* will allow us to obtain historic data of the websites hosted in the same IP.

LIMITATIONS OF THIS APPROACH:

The steps devised in the manual approach are just some examples of how to spot FAKE sites with an accuracy of around 97% based on the results achieved. To have a

system like this detecting automatically FAKE websites have some pitfalls. By working on this research I noticed that the owners of some FAKE sites campaigns were denying access to the online tool desenmascara.me (which just works as a normal User Agent browsing the site). As in occasions all the FAKE sites hosted in the same IP block showed a 403 HTTP code as result (WAF or Block, 2015). The fraudsters take countermeasures to protect their FAKE websites for being scanned and flagged as such. Therefore additional measures as changing the public IP performing the scans and using different user-agents will override such countermeasures. Another solution is to have different machine instances scanning the websites and to switch between them when a block is detected. Cloud services available nowadays can automate these tasks quite easily.

API (the automatic approach)

In order to alleviate the pain to set up a system to detect FAKE sites, another alternative is to use an API, like the desenmascara.me API⁶. The results gathered by the online tool were accurate and reliable enough as to integrate them into the VirusTotal service. VirusTotal is the Google-owned virus and URL online scanning service. For example FAKE websites such as: www.rolexdaytona.in.net will be showed in VirusTotal as suspicious by desenmascara.me as showed in the *Appendix K*. For performance and privacy issues, not all the FAKE websites detected are synchronized with VirusTotal hence the best option to know if a website is flagged as FAKE is to query the desenmascara.me API. In order to do it we can proceed with the following API call:

GET [http://desenmascara.me/api/official/\\$URL](http://desenmascara.me/api/official/$URL)

The API takes a single parameter which is the URL to check against but without the HTTP (Hypertext Transfer Protocol) scheme. As an example, to check the website: www.bnkshops.com, the API call would be as follows:

<http://desenmascara.me/api/official/www.bnkshops.com>
And it will give us the result as follows:

```
{{"url": "http://www.bnkshops.com",
```

⁶ <http://desenmascara.me/api/howto>

```
"brand_affected": "vibram",  
"result": "FAKE",  
"current_status": "200",  
"last_check": "2016-09-12 22:04:08"}}
```

The API is read only at this time. If a website has not been previously analyzed by the desenmascara.me front end then the *result* field would be: “unknown”. The only way to check it out in live is going directly to the desenmascara.me public interface. The work to consume the API in real time, without the need to go to the website, is in progress.

Integration with Existing Security Tools

If we recall the figure 3 from the section “the online counterfeiters and their relation with the underground economy”, we know that IOCs (Indicators of Compromise) for most of the threads in such table are usually included in the vast amount of intelligence watchlists. Such watchlists either proprietary or open source are used to do some kind of: *correlation*, *validation* or *enrichment*. But currently there are no threat intelligence for the specific “luxury knock-offs” strategy of the profit center, *Spamvertised products*, whereas the revenue numbers are even higher than *ClickFraud* based on the aforementioned table. But now if the counterfeit website detection system has been successfully deployed in the environment, a new feed of FAKE sites could be gather to use it as best suit your needs (the appendix M contains an example of this potential feed). Otherwise you could use the desenmascara.me API or check websites against VirusTotal. At this point, one question that can arise is: can this feed of FAKE sites be referred as threat intelligence? Based on the Gartner definition (Gartner, 2013)

“...threat intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing menace...”

Threat intelligence would need come in a rich format as a processed indicator. In order to help us see various forms of security feeds, the DIKW Pyramid (DIKW, Wikipedia) comes handy here. The DIKW Pyramid refers to a class of models for representing structural relationships between its different layers represented as data, information, knowledge and wisdom. Based on the steps performed on “The manual approach” section to flag a website as FAKE, we can affirm that the feed of FAKE sites

we can collect by using this process would function in the knowledge layer of the DIKW pyramid. And the overall view of the metrics gathered over this layer will position us in the top of the DIKW pyramid; the wisdom layer. In the *Appendix A* you could take a look to some of the metrics about the data collected by the *desenmascara.me* project. This provides a new piece of information to leverage independently of the sector of your company. If some of the three different scenarios described at the beginning of this section would be a close match either within an ISP, a SME company, or a luxury or sports brand, you could then use the knowledge acquired here to fight against the increasing online counterfeiting threat.

5. Conclusion

The trade in counterfeit goods increased by more than 80% in five years (Trade in Counterfeit and Pirate Goods, 2016). The counterfeiters are improving their logistics networks, and leveraging the huge growth in online shopping as showed in this research. Lax domain registration process are making easy for the fraudsters to replicate brand's genuine websites very quickly making it easy to lure to the unsuspected online consumer.

The Internet is being used as “giant amplifier” for the sale of counterfeit goods. The online fraud presented in this paper will increase in the next years due to the growth of online and mobile shopping behavior. This affirmation is backed by the different studies referenced. Therefore, the goal of this paper is to shed light on the need to fight it. As showed it can be fight with easy and free solutions. The methodology presented here can be used as a starting point to deploy more advanced technologies to help the both parties affected by this fraud: the online users and the affected brands. SafeBrowsing (<https://www.google.com/transparencyreport/safebrowsing/?hl=en>) is a Google technology included by default in the major browsers to protect the users from unsafe sites such as: Malware sites and phishing sites. The ultimate ambitious goal of this research is to have widely-used technologies like the above to flag a new kind of unsafe sites: **FAKE websites related with the online counterfeiting.**

References

- Ajaya Neupane & Nitesh Saxena, (2014). Internet Society. Neural Signatures of User-Centered Security: An fMRI Study of Phishing, and malware warnings.
https://www.internetsociety.org/sites/default/files/11_5_0.pdf
- Beautiful Soup. Python (2016)
<https://www.crummy.com/software/BeautifulSoup/>
- Black Hat Europe (2014)
<https://www.blackhat.com/eu-14/arsenal.html#desenmascara-me>
- Blue Coat category descriptions, (2016).
<http://sitereview.bluecoat.com/categories.jsp>
- Cartier International AG & Ors v British Sky Broadcasting Ltd & Ors (2014).
<http://www.bailii.org/ew/cases/EWHC/Ch/2003/3354.html>
- Complaint Luxottica, (2016).
<http://gbcinternetenforcement.net/files/caseNo-16-cv-3989/COMPLAINT.PDF>
- Counting the cost of Counterfeiting, Netnames 2015
<https://www.netnames.com/assets/shared/whitepaper/pdf/NetNames-Counterfeiting-Report-A4-2015.pdf>
- Defacement. (n.d). In *Wikipedia, the free encyclopedia*. Retrieved on August 3, 2016
https://en.wikipedia.org/wiki/Website_defacement
- Demandware software (2017)
<https://en.wikipedia.org/wiki/Demandware>
- DIKW Wikipedia
https://en.wikipedia.org/wiki/DIKW_pyramid
- Elie Bursztein. (2016). Remediating Web Hijacking: Notification Effectiveness and Webmaster Comprehension.
<https://cdn.elie.net/publications/remediating-web-hijacking-notification-effectiveness-and-webmaster-comprehension-www-2016.pdf>
- EU customs enforcement of IPR, 2015
https://ec.europa.eu/taxation_customs/sites/taxation/files/2016_ipr_statistics.pdf
- Farsight DNSDB (2016), provided by Farsight Security, Inc.

Emilio Casbas, ecasbas@gmail.com

<https://www.dnsdb.info/>

Gartner 2013.

<https://www.gartner.com/doc/2487216/definition-threat-intelligence>

Kaspersky Lab Black Friday Threat Overview (2016)

<https://securelist.com/analysis/publications/76615/kaspersky-lab-black-friday-threat-overview-2016/>

Kirill Levchenko, Click Trajectories: End-to-End Analysis of the Spam Value Chain

<https://www.cs.uic.edu/~ckanich/papers/levchenko2011click.pdf>

Kurt Thomas, Danny Yuxing, David Wang, Elie Bursztein, Chris Grier, Thomas J., Christopher Kruegel, Damon McCoy, Stefan Safage, Giovanni Vigna. Framing (2015). Workshop on the economics of Information Security. Framing Dependencies Introduced by Underground Commoditization.

<http://research.google.com/pubs/pub43798.html>

Louis Vuitton (2016). La lucha de las marcas contra las falsificaciones online y la de Louis Vuitton.

<http://blog.emiliocasbas.net/2015/06/la-lucha-de-las-marcas-contras-las.html>

MarkMonitor 2015.

https://info.markmonitor.com/ccc_barometer

McCoy, Damon. (2016). Bullet-Proof Credit Card Payment Processing

https://www.usenix.org/sites/default/files/conference/protected-files/enigma_slides_mccoy.pdf

OECD/EUIPO. (2016). *Trade in Counterfeit and Pirated Goods: Mapping the Economic Impact*, OECD Publishing, Paris. Chapter 4, Figure 4.2

<http://dx.doi.org/10.1787/888933345913>

PHP 5.3.29 Release announcement (2014)

http://de2.php.net/releases/5_3_29.php

Reuters, (2016). Italian police shut down fake Prada website.

<http://www.reuters.com/article/us-prada-fakes-website-idUSKCN0V024I>

Stamos, (2016). *Addressing security blindspots through culture*.

<https://www.facebook.com/notes/alex-stamos/addressing-security-blindspots-through-culture/10154390896047929>

Emilio Casbas, ecasbas@gmail.com

Tecnologias Libres –Massive campaign of FAKE sites (2015)

<http://blog.emiliocasbas.net/2015/11/massive-campagin-of-fake-sites.html>

TheGuardian (2014). Internet Service Providers must help crack down on fake goods, high court rules.

<https://www.theguardian.com/technology/2014/oct/20/internet-service-providers-fake-goods-high-court-rules>

Trade in Counterfeit and Pirated Goods. (2016). Mapping the Economic Impact, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264252653-en>

urllib2. Python library (2016)

<https://docs.python.org/2/library/urllib2.html>

Verizon Data Breach Report (2015).

<http://www.verizonenterprise.com/de/DBIR/2015/>

WAF or Block (2015).

<http://blog.emiliocasbas.net/2015/07/el-sitio-no-permite-desenmascaramiento.html>

WhatWeb (2016). Morningstar security.

<https://www.morningstarsecurity.com/research/whatweb>

WWD, (2016). Alexander Wang awarded \$90 million in anticounterfeit and cybersquatting case.

<http://wwd.com/business-news/legal/alexander-wang-awarded-90-million-in-anti-counterfeit-and-cybersquatting-case-10506074/>

Zscaler URL categories. (2016).

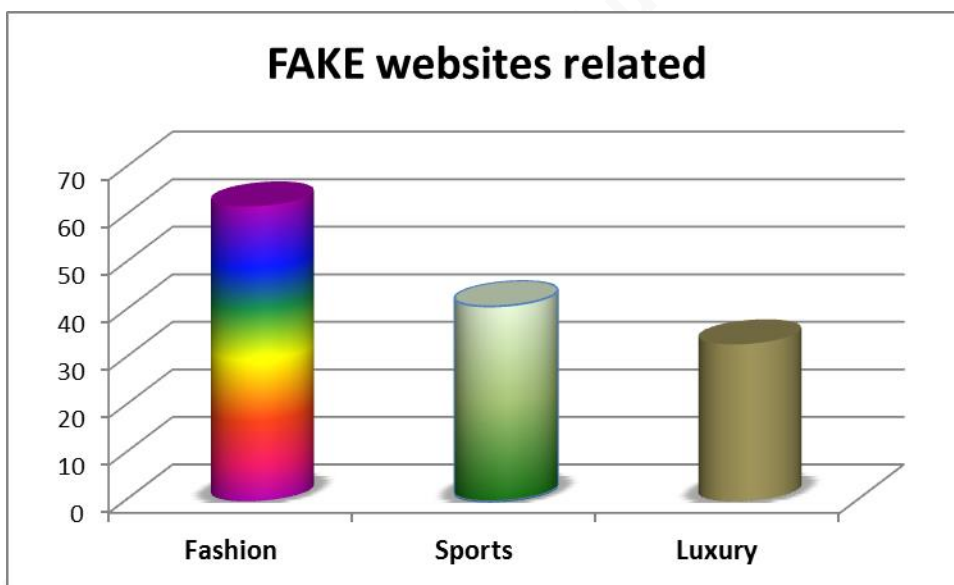
<https://www.zscaler.com/sitereview/help.html>

Appendix A: Desenmascara.me Metrics

Metrics about the data collected in this research in relation to the FAKE websites, categorized by industry, proportion of FAKE websites among the total websites analyzed, the Top brands more targeted by online counterfeiters and top TLDs of the FAKE websites.

This metric does show the industry targeted by online counterfeiters. Desenmascara.me detects 135 different brands which belong to the three main categories showed below.

Figure 32. Industry categorization of the FAKE websites.

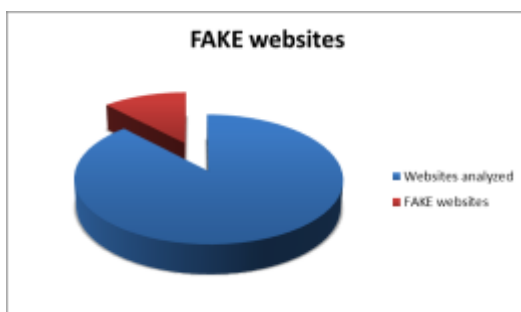


This metric does show graphically the proportion of FAKE websites among the sample of websites analyzed:

-Fake websites detected by desenmascara.me: **11399** (increasing daily)

-Total websites analyzed by desenmascara.me: **81724** (increasing daily)

Figure 33. Proportion of FAKE websites among the total of websites analyzed.



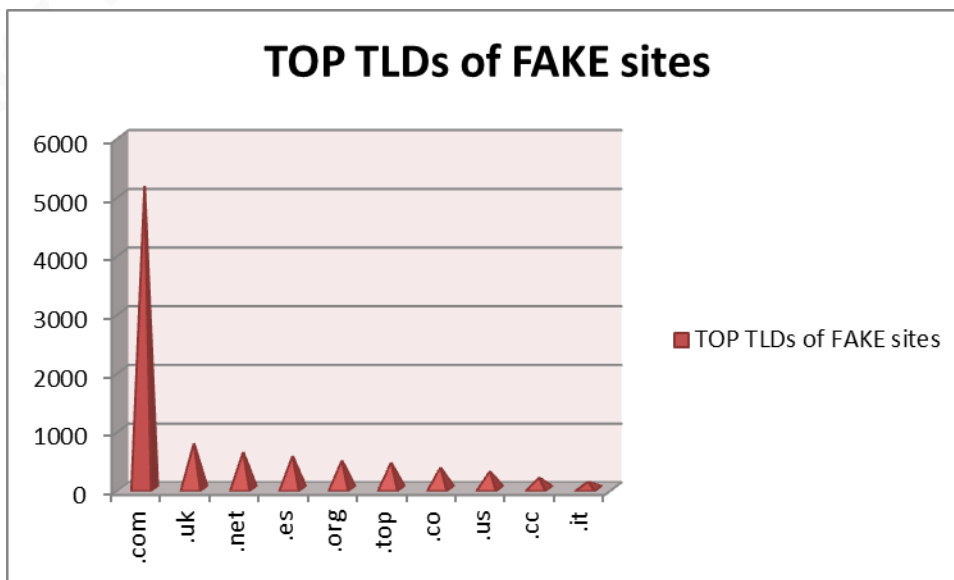
This metric does show the 10 brands more prevalent among the FAKE websites detected by desenmascara.me

Figure 34. Top FAKE sites by brand targeted



This metric does show the top 10 TLDs of the FAKE websites detected.

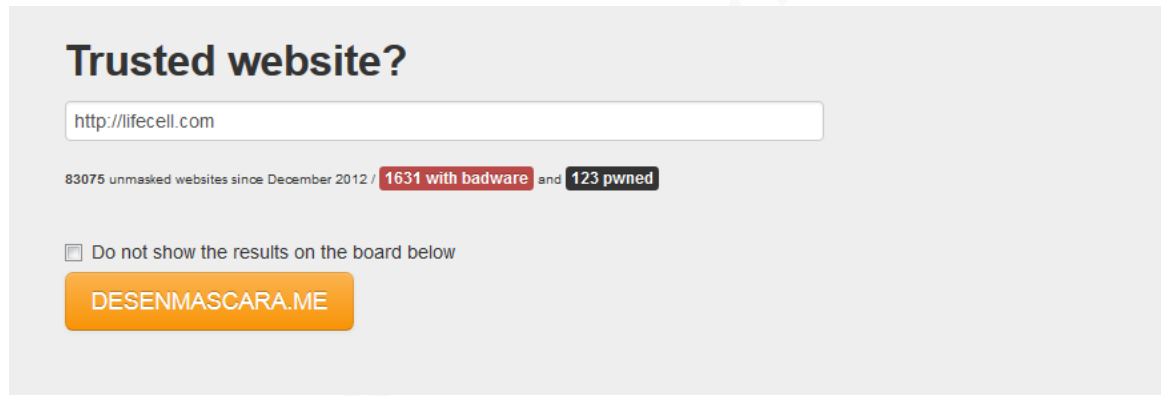
Figure 35. Top TLDs of FAKE sites



Appendix B: Using the desenmascara.me tool

The online tool desenmascara.me extracts all the metadata of websites submitted by any user. The goal was to make it as simple as possible, so anyone, regardless of its background, could use the tool. Therefore the only step required to run the tool is to type the web address of the website to analyze and to click the button as below:

Figure 36. Analyzing the <http://lifecell.com> website with desenmascara.me



The screenshot shows the 'Trusted website?' interface of the desenmascara.me tool. It features a text input field containing 'http://lifecell.com'. Below the input field, a status bar displays '83075 unmasked websites since December 2012 / 1631 with badware and 123 pwmed'. A checkbox labeled 'Do not show the results on the board below' is present. At the bottom, there is an orange button labeled 'DESENMASCARA.ME'.

After a few seconds a report like the below will appear:

Figure 37. Desenmascara.me report of the website lifecell.com

Web Site	http://lifecell.com
	☆☆ Let us know if the web site is OFFICIAL WEB NO OFFICIAL FAKE ☆☆
Awareness value:	-265 (with 20 or higher a website is considered somehow security awareness)
URL's MD5:	ce7b354d04af89a14309a373aec28b42
Unmasked on:	6 Sep 2015, 5:25 p.m.
Domain registered on:	292
Domain will expire in:	26-jul-2016 (324 days)
Web server:	Apache/2.2.23 (Unix) mod_ssl/2.2.23 OpenSSL/1.0.0-fips PHP/5.3.19 (vulnerability history)
Technology:	PHP/5.3.19 (vulnerability history)
Robots file:	Not found
HTTP methods:	Not found
Directory listing:	Not found
Third party content:	Not found
Electronic commerce:	Payment gateway or Paypal or own BBDO (Read more)
Private IPs:	No
Iframes:	Not found
Scripts:	11 (check the scripts)
Suspicious code:	Not found
Incrusted spam:	Not found
Location:	Not found
Google check:	Is not blacklisted by SafeBrowsing (Read more)
Metadata:	[http://lifecell.com [200] Apache/2.2.23] mod_ssl/2.2.23
Metadata:	* Cookies[fe_typo_user]
Metadata:	* Country[UNITED STATES][US]
Metadata:	* Google-Analytics [UA-9689002-1]
Metadata:	* HTML5, HTTPServer[Unix][Apache/2.2.23 (Unix) mod_ssl/2.2.23 OpenSSL/1.0.0-fips PHP/5.3.19]
Metadata:	* IP[166.78.77.8]
Metadata:	* JQuery[1.5.2]
Metadata:	* MetaGenerator TYPO3 4.6 CMS
Metadata:	* OpenSSL[1.0.0-fips]
Metadata:	* PHP[5.3.19]
Metadata:	* PoweredBy[TYPO3]
Metadata:	* Script[text/javascript]
Metadata:	* probably TYPO3, Title[LifeCell: Home]
Metadata:	* X-Powered-By[PHP/5.3.19/n]

To analyze the Metadata of a website and to provide a security awareness score based on it was the initial goal of the desenmascara.me tool. In the above case, the awareness score is quite poor: **-265**. Mainly due to the website showing many metadata which someone with knowledge could leverage it to attack the site. Based on the metadata found on this website with just 1 click, an attacker could extract the following information among other:

Brief explanation of the metadata found: The versions of Apache, mod_ssl and php showed have several security vulnerabilities which were patched around the third quarter of 2012. The jQuery version showed above is 1.5.2 which was released in the first quarter of 2011. The current version is 3.1.0. On top of that the metadata show signs of the TYPO3⁷ software version 4.6 which was released on the third quarter of 2011 which contains a critical security vulnerability of authentication bypass⁸.

The above information extracted through the metadata does show that the website owner is likely not following good practices of maintaining a website well patched and updated. Therefore the website would obtain a low score which when negative might be considered as a website prone to be compromised.

In order to show an example of a website with low score which then become compromised⁹, the infamous Hacking Team website can be seen:

<http://hackingteam.com/> which had poor¹⁰ metadata like in the figure 39:

⁷ https://wiki.typo3.org/TYPO3_4.6

⁸ <https://typo3.org/teams/security/security-bulletins/typo3-core/typo3-core-sa-2015-001/>

⁹ <http://www.csoonline.com/article/2943968/data-breach/hacking-team-hacked-attackers-claim-400gb-in-dumped-data.html>

¹⁰ <http://desenmascara.me/consulta/5c629607f595904be96898e6d79f5194>

Figure 38. Compromised website with poor metadata

Metadata:	[http://hackingteam.com [200] Cookies[23150322952a0ca756c03222360ab078__cfduid.g
Metadata:	" Country[UNITED STATES][US
Metadata:	" Google-Analytics [UA-56018687-1
Metadata:	" HTTPServer[cloudflare-nginx
Metadata:	" HttpOnly[__cfduid
Metadata:	" IP[198.41.191.30
Metadata:	" maybe Joomla, MetaGenerator[Joomla! 1.5 - Open Source Content Management
Metadata:	" PHP[5.3.3
Metadata:	" Script[text/javascript
Metadata:	" Title[Hacking Team &git; Remote Control System
Metadata:	" UncommonHeaders[cf-cache-status,cf-ray
Metadata:	" X-Powered-By[PHP/5.3.3
Metadata:	" X-UA-Compatible[chrome=1
Metadata:	" cloudflare[In]

Metadata showing a vulnerable CMS like Joomla 1.5, one of the most used by the bad guys to compromise¹¹ websites, gives one of the lowest score despite of having signs or hardening or sitting behind cloudflare.

Another famous example is the Permanent Court of Arbitration website: <http://www.pca-cpa.org> apparently being victim of a watering hole attack¹² which showed ¹³metadata as below:

¹¹ <http://forum.joomla.org/viewtopic.php?t=246319>

¹² <https://www.threatconnect.com/china-hacks-the-peace-palace-all-your-eezs-are-belong-to-us/>

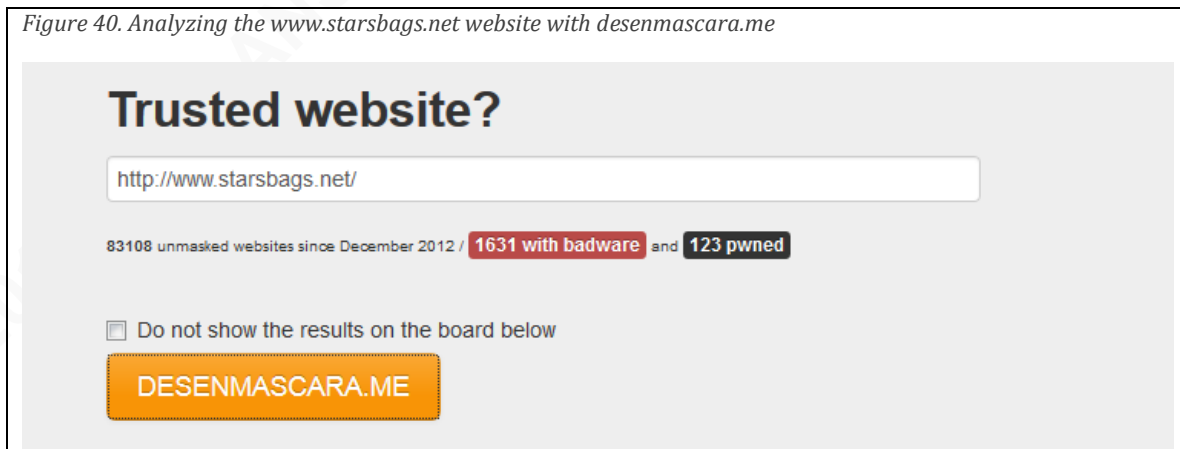
¹³ <http://desenmascara.me/consulta/0bba154ca7d0e73a61d770e5fc3affc7>

Figure 39. Metadata showing signs of non-update software

Metadata:	['http://www.pca-cpa.org [200] ASP.NET, HTTPServer [Microsoft-IIS/6.0]
Metadata:	' IP[194.26.24.152]
Metadata:	' Microsoft-IIS[6.0]
Metadata:	' Script[text/javascript]
Metadata:	' Title[PCA-CPA]
Metadata:	' X-Powered-By[ASP.NET]\n']

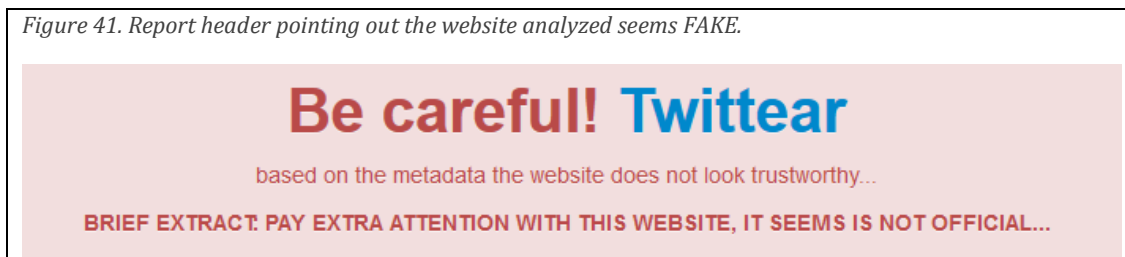
While this was the main purpose of the desenmascara.me tool: *to raise security awareness among web owners*, so they could better protect their websites. The tool is now focused in the detection of FAKE websites related with the online counterfeiting. For instance, let's see an example:

Figure 40. Analyzing the www.starsbags.net website with desenmascara.me



After a few seconds after clicking on the button we will see a report like below:

Figure 41. Report header pointing out the website analyzed seems FAKE.



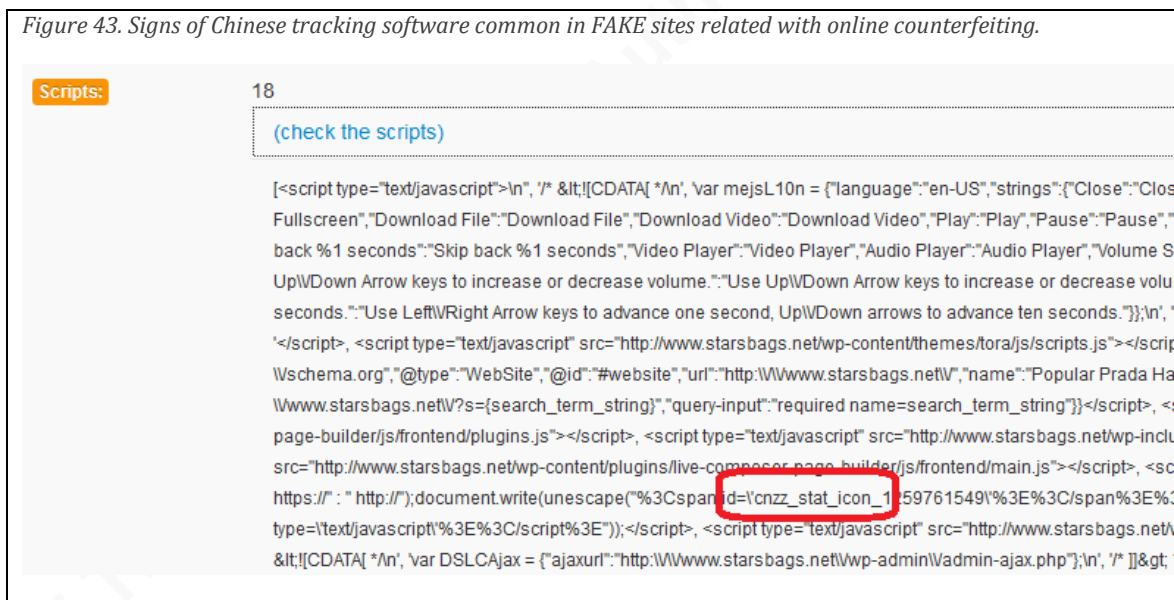
The full report is showed in the next page.

Figure 42. Desenmascara.me report of the www.starsbags.com FAKE website

Web Site	http://www.starsbags
	NO OFFICIAL (full URL available only in the first query. Contact with me if you need this information).
	Twitter fraude
Awareness value:	35 (with 20 or higher a website is considered somehow security awareness)
URL's MD5:	3cf602b66fd8d5de9c942b0c97e05584
Unmasked on:	24 Feb 2017, 8:31 a.m.
Domain registered on:	625
Domain will expire in:	08-jun-2017 (104 days)
Web server:	nginx
Technology:	PHP/5.4.45 (vulnerability history)
Robots file:	(check it out)
HTTP methods:	Not found
Directory listing:	Not found
Third party content:	Not found
Electronic commerce:	Payment gateway or Paypal or own BBDO (Read more)
Private IPs:	No
Iframes:	Not found
Scripts:	18 (check the scripts)
Suspicious code:	Not found
incrusted spam:	Not found
Location:	Not found
Google check:	Is not blacklisted by SafeBrowsing (Read more)
Metadata:	['http://www.st...
Metadata:	' Country[UNIT...
Metadata:	' HTML5, HTTPS...
Metadata:	' IP[198.55.30...
Metadata:	' JQuery, Open...
Metadata:	' PHP[5.4.45...
Metadata:	' PoweredBy[Wo...
Metadata:	' Script[appli...
Metadata:	' Title[Popula...
Metadata:	' UncommonHead...
Metadata:	' WordPress. X...

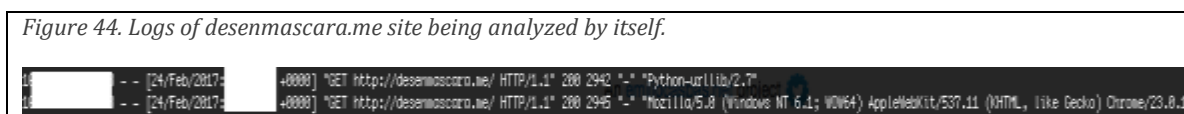
The website is flagged as “No official”, which in terms of an e-commerce shop of a famous brand means a FAKE website. Neither the full URL nor the metadata is fully showed into this “No official” classification. This was addressed as a caution to avoid users landing on the website through *desenmascara.me*. However when clicking on the “Check the scripts” link into the “Scripts” field, we will observe some signs of a FAKE site already discussed on this paper:

Figure 43. Signs of Chinese tracking software common in FAKE sites related with online counterfeiting.



As a side note, *desenmascara.me* performs an innocuous check on the website just as someone browsing the website. The two HTTP requests below will appear in the website’s logs of any web analyzed by *desenmascara.me*:

Figure 44. Logs of *desenmascara.me* site being analyzed by itself.



Below we can see an incomplete list of different kinds of metadata detected by desenmascara.me :

Figure 45. Web metadata detected by desenmascara.me

Server HTTP header metadata collected	
Server HTTP header	Description
AkamaiGHost	Web server using Akamai Global Hosting
AmazonS3	Web server using Amazon cloud
Apache/X.X	Web server using Apache technology
Microsoft-IIS/X	Web server using Microsoft IIS technology
PWS	Small Microsoft Web server for old Windows versions
nginx/X.X	Russian web server and reverse proxy
lighttpd/X.X	Web server optimized for speed-critical environments
OpenCms/X.X	Open source content management system written in Java
Netscape-Enterprise/X.X	Web server using old Netscape technology
Sun-ONE-Web-Server/X	Web server using iPlanet web server technology
Oracle-Application-Server-Xx	Web server using Oracle applications server
Lotus-Domino	Web server using IBM Lotus Domino technology
Sun-Java-System-Web-Server/X	Web server using Oracle iPlanet technology
Oracle-iPlanet-Web-Server/7.0	Web server using Oracle iPlanet technology
IBM_HTTP_Server/X.X	Web server using IBM technology (Apache based)
LiteSpeed/X.X	Web server using LiteSpeed technology (Apache based)
Alterian-CME/X.X	Web server using SDL ACM
Tengine	Web server using Tengine technology (nginx based)
eZ Publish	Web server using EZ technology
GSE	Web server using Google infrastructure (blogger)
gws	Web server using Google infrastructure (search pages)

sffe	Web server using Google infrastructure (static files)
tfe	Web server using Twitter infrastructure
YTS	Web server using Yahoo! infrastructure
cloudflare-nginx	Web server using CloudFlare infrastructure

Powered-by HTTP header metadata collected (
Powered-by HTTP header	Description
eBD/3.5.5	Web server using EBD technology
eWAY	Web server using eWay payment gateway
Express	Web server using nodejs with express
PHP/x.x	Web server using PHP technology
ASP.NET	Web server using Microsoft ASP technology
Servlet/X.X JSP/X.X	Web server using Tomcat application server
Plesklin	Web server using Parallels technology
(mod_rails/mod_rack)	Web server using Ruby on Rails technology
ARR/X.X	Web server using IIS with request routing technology
JSF/2.0	Web server using JavaServer Faces technology

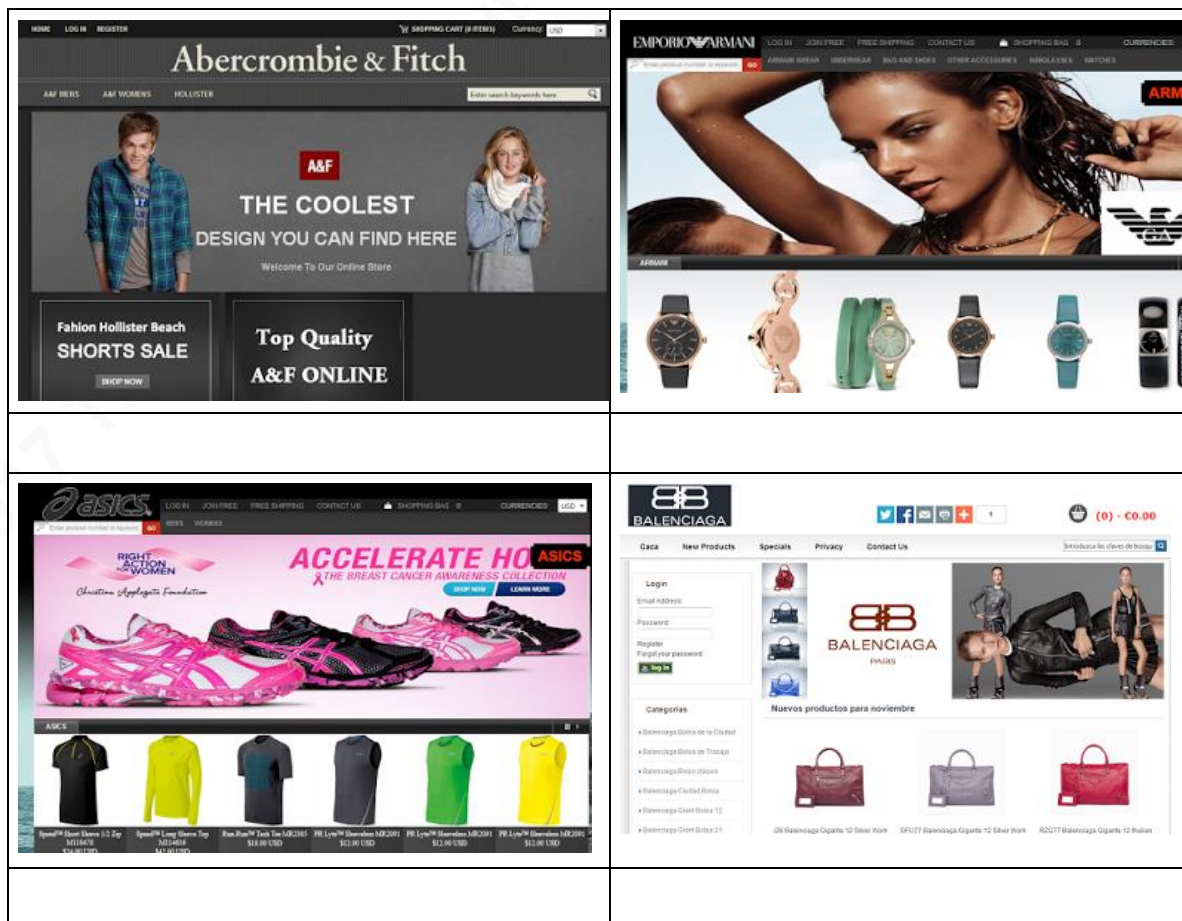
Uncommon HTTP headers collected (this headers aren't an HTTP standard)	
Custom HTTP headers	Description
access-control-allow-origin, access-control-allow-headers	Web server using HTTP access control (CORS)
x-generator	Web server running under Drupal
x-amz-	Web server running under Amazon services
x-cache-hits,x-timer,x-served-by, x-varnish, x-varnish-cache	Web server using Varnish cache technology
x-drupal-cache	Web server using Drupal technology
x-dynatrace	Web server using dynatrace technology
x-server-name	Web server using Websphere technology
strict-transport-security	opt-in security enhancement that is specified by a web application

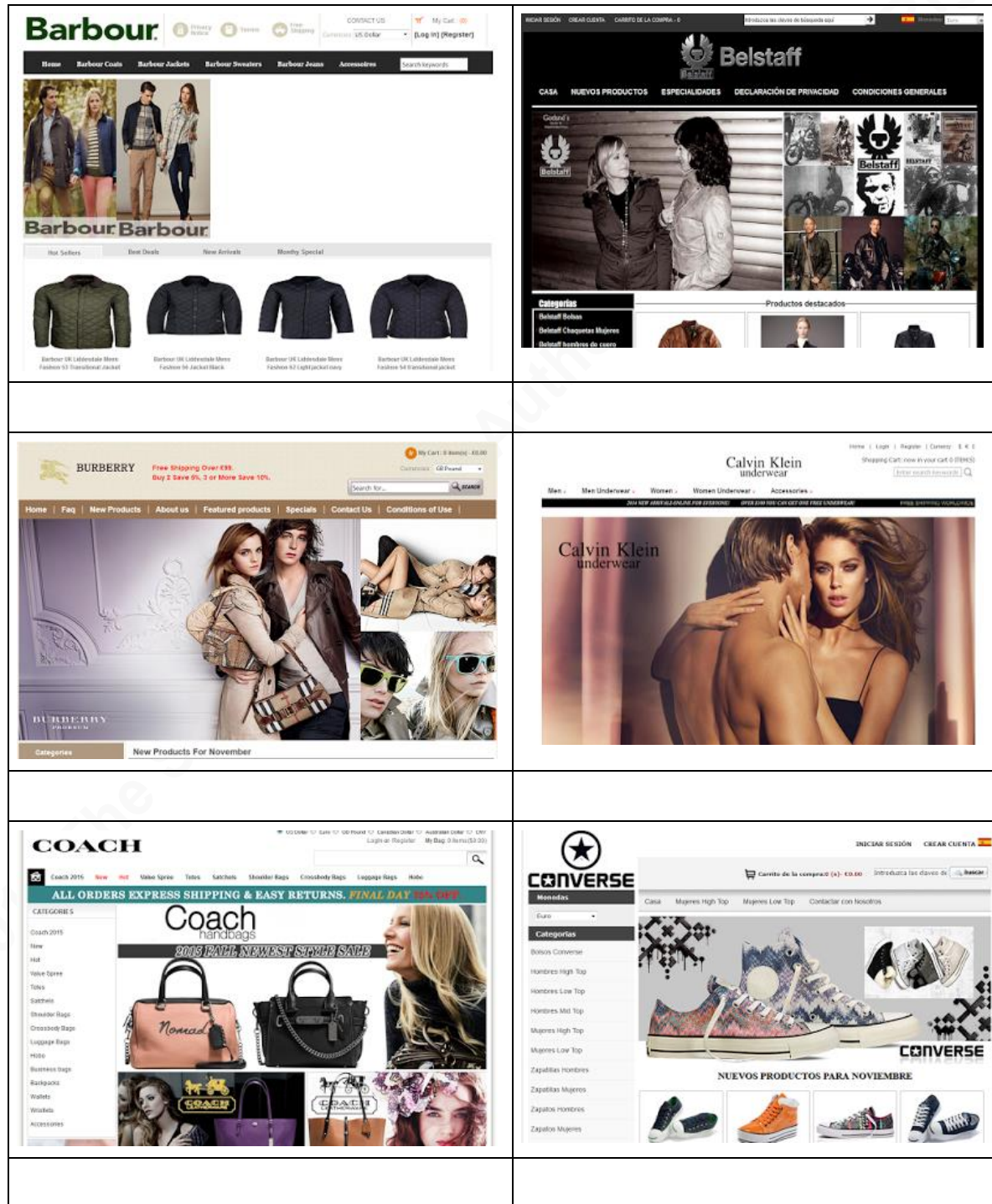
HTML metadata / HTTP headers collected which could allow fingerprinting	
HTML metadata	Description
moodle	Web server using Moodle technology
MetaGenerator[Infopark CMS Fiona	Web server using CMS Fiona technology
MetaGenerator[Sitefinity	Web server using SiteFinity technology
HTTPServer[BigIP / Cookies[BigIP	Web server using F5 technology
Cookies: PHPSESSID	Web server using PHP technology
Cookies: JSESSIONID	Web server using JSP technology
Cookies: ASPSESSION	Web server using ASP technology
Cookies: fe_typo_user	Web server using TYPO3 technology
Cookies: CFID	Web server using Coldfusion technology

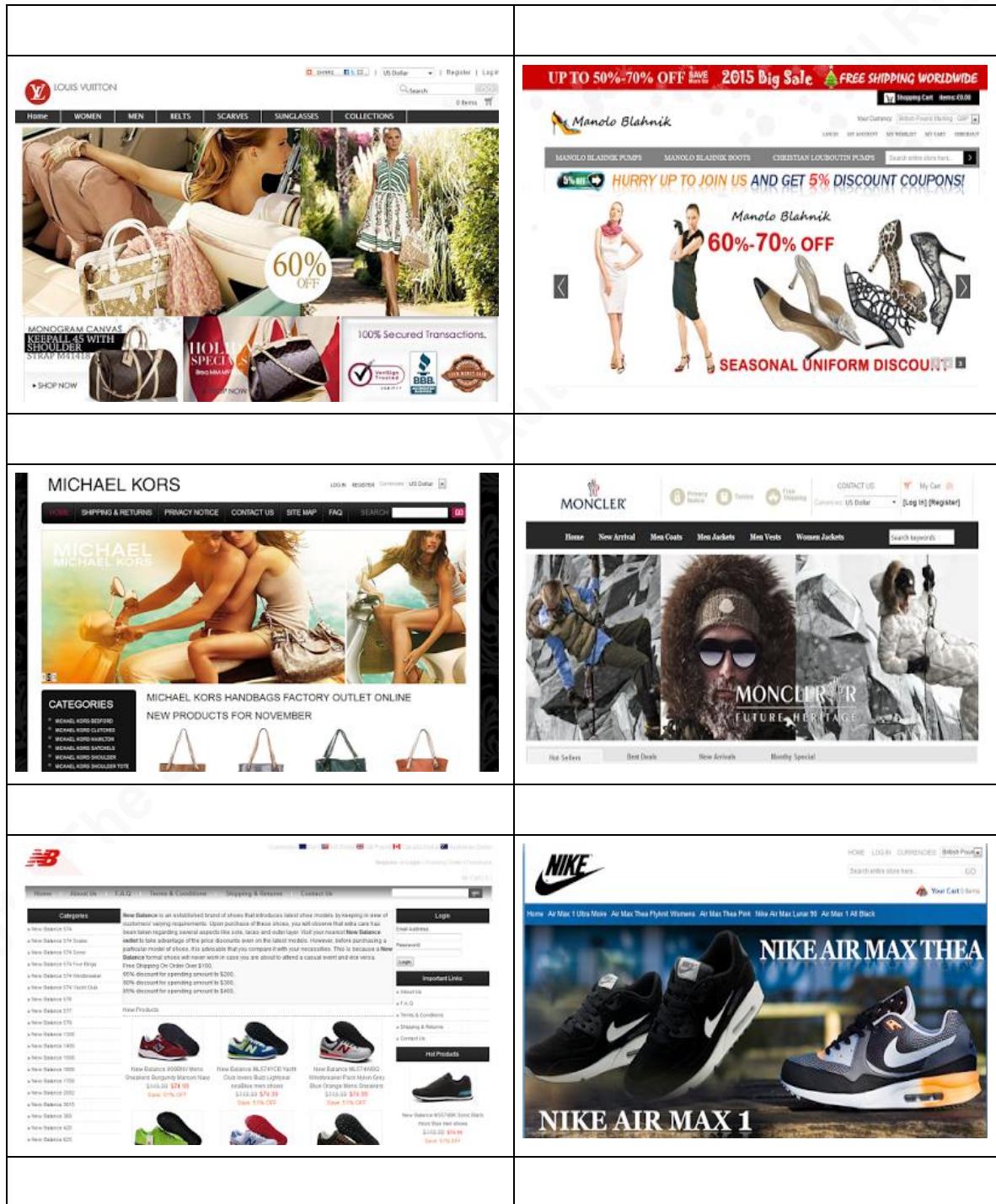
Cookies: CFTOKEN	Web server using Coldfusion technology
MetaGenerator[Square One, Meta-Author[Jeremy]	Web server using Square One CMS (light version of Joomla)
MetaGenerator[LFC]	Web server using LFS technology
MetaGenerator[Percussion]	Web server using Percussion CMS
RiOS[Web server using Riverbeed WAN optimization Riverbeed WAN optimization

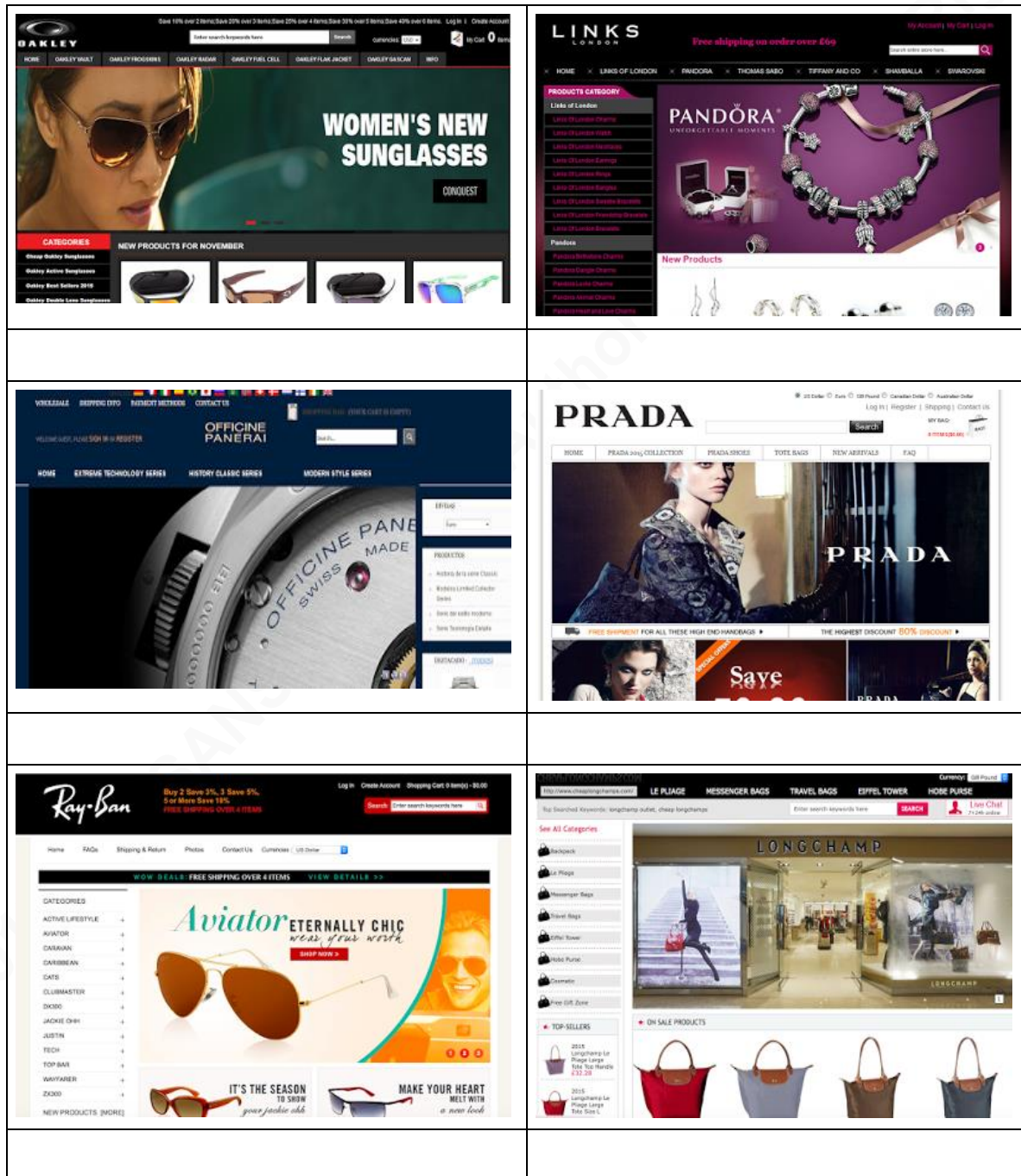
Appendix C: Massive Campaign of FAKE Websites

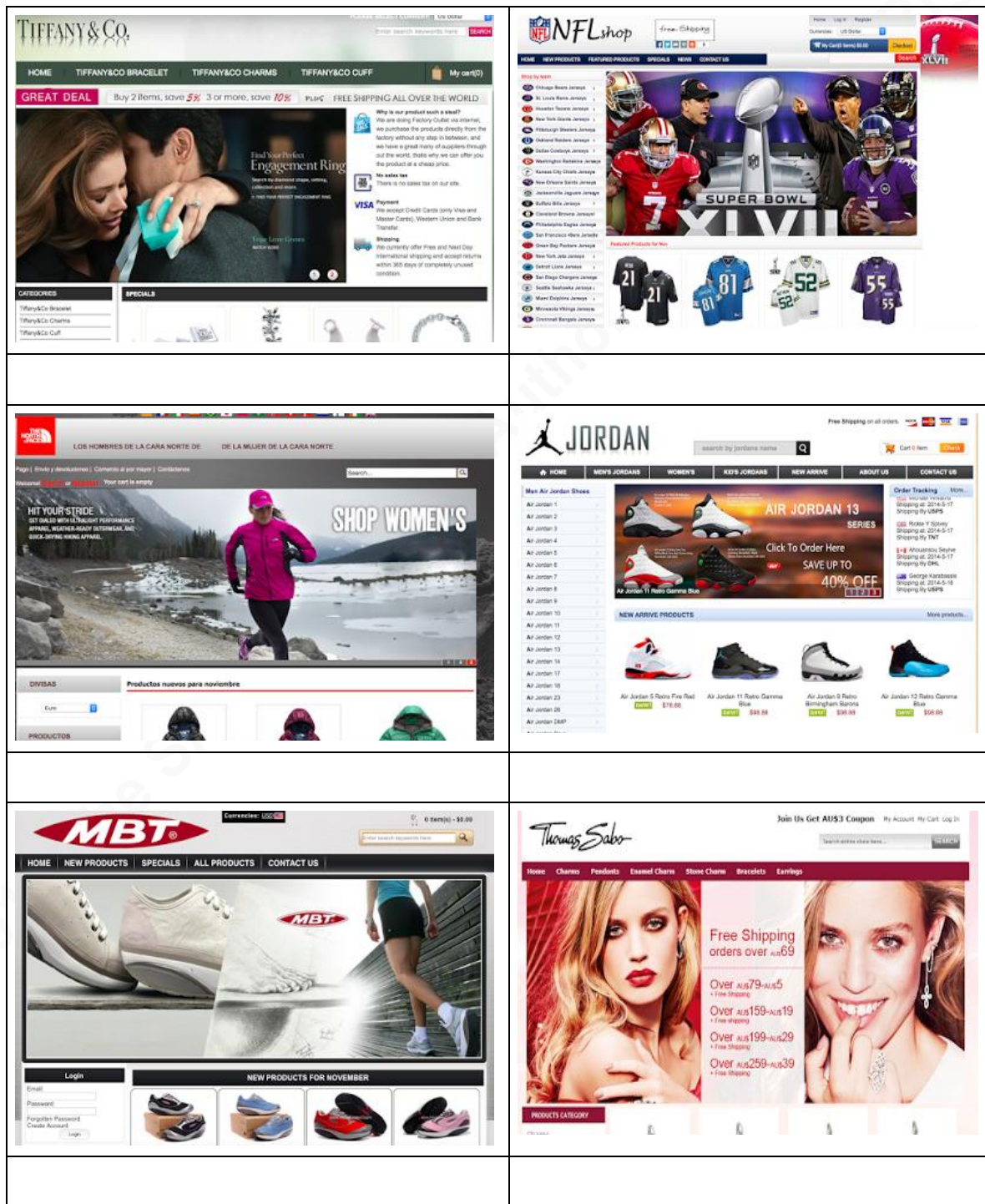
The below list of FAKE websites screenshots was taken during a research of a massive campaign of FAKE websites due to the proximity of the Black Friday and Christmas seasons.

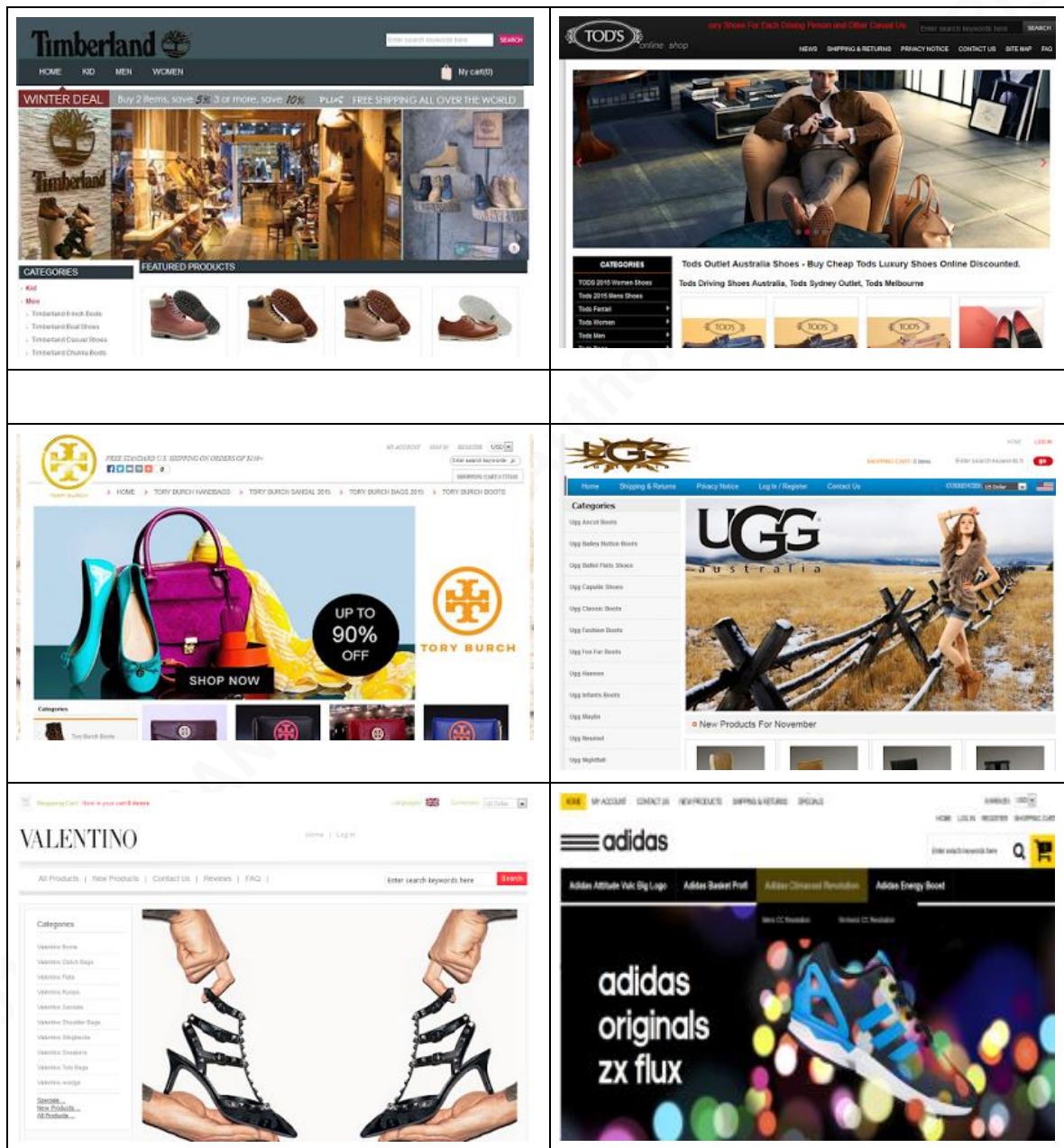












Appendix D: Free Hosting Resources Used by Online Counterfeiters

List of some free resources used by online counterfeiters spotted while doing this research. Though this list is not quite extensive does show the most common free resources abused to support the online counterfeiting fraud.

Title	URL
Free website builder	http://webpaper.co
The EASIEST Way To Create Landing Pages	http://1minutesite.es
Community-powered social commerce website	http://polyvore.com
Easy and free blog platform	http://blogspot.it/com
Drag & Drop Site Builder	http://weebly.com
Make a free website with our free website builder.	http://yolasite.com
Build a free website with our easy to use, free website builder	http://tripod.com
Choose Your Own Domain Name and Create a Unique Design	http://wordpress.com
Award-winning free website builder tool that lets you create a website in minutes	http://iconosites.com
Website builder it's easy and virtually free.	http://ucoz.com
Website builder, websites for free, eshops for free!	http://webgarden.com
No coding skills needed. Choose a design, begin customizing and be online today!	http://wix.com
<i>Big Cartel</i> is home to nearly a million clothing designers	http://bigarcatel.com

Appendix E: FAKE Website Example Leveraging Expiring Domains

Another example of a FAKE website under the expiring domain based category is showed below:

<http://www.minecraftdownloads.de/>

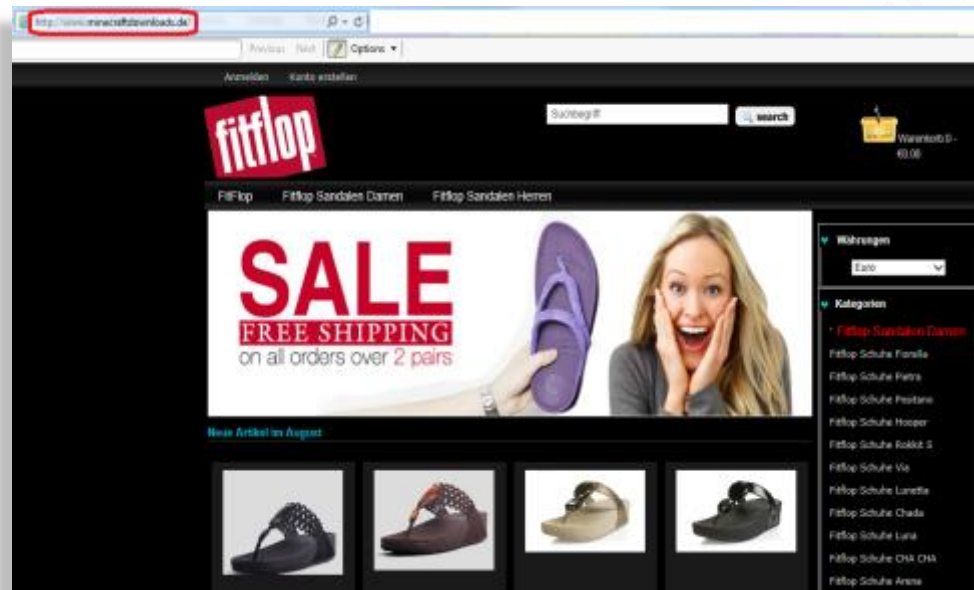
which in 1 Feb 2016 (Internet Archive, 2016) did show minecraft related content as seen in the picture:

Figure 46. Original www.minecraftdownload.de website



At the time of writing this paper the domain is showing a FAKE shop of a famous sandals brand as shown below:

Figure 47. FAKE website leveraging an expiring domain



In this case the actor is leveraging an expired domain as the evidence in the whois registry does show an update on May 9, 2016 based in DomainTools (<https://research.domaintools.com/>)

Figure 49. DomainTools information

— Whois & Quick Stats	
Email	xhq35177@hotmail.com is associated with ~77 domains
Registrant Org	tauleny jens is associated with ~13 other domains
Dates	Updated on 2016-05-09
IP Address	93.174.90.52 - 6 other sites hosted on this server
IP Location	🇬🇧 - English River - Victoria - Novogara Ltd
ASN	🇬🇧 AS29073 QUASINETWORKS , NL (registered May 26, 2003)
Whois History	20 records have been archived since 2013-04-21
Whois Server	whois.denic.de
— Website	
Website Title	🛒 Günstige Fitflop Sandalen Sale, Mode Fitflop Schuhe Outlet
Server Type	Apache/2.2.29 (Unix) mod_ssl/2.2.29 OpenSSL/1.0.1e-fips mod_bwlimited/1.4

The free email account used to register this domain can be used to track additional FAKE websites, as they are usually used by the fraudsters as a batch to either register domains or to be used in the contact forms of the FAKE websites.

Appendix F: Fake Website Example of Free Hosting Based Site Category

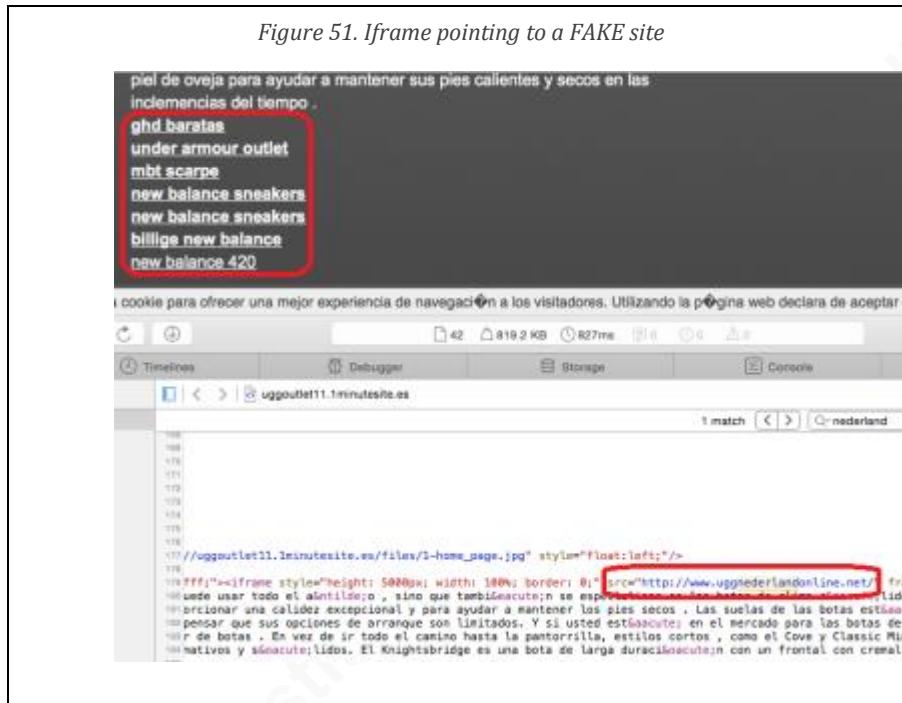
Another example of a FAKE website under the free hosting based category is showed below:

Figure 50. FAKE site related with the UGG brand



The above FAKE site contained an iframe, highlighted in the figure 51, which pointed to the FAKE site to proceed with the purchase as additional links to other pivotal FAKE sites:

Figure 51. Iframe pointing to a FAKE site



Appendix G: Facebook Advertisement Leading to a FAKE Website

Example of a Facebook advertisement leading to a FAKE website. Currently the resources showed here are no longer available.

The typical Facebook advertisement below:



Led to the Facebook event of the figure 53:



Whose website being advertised: www.rblovez.pw was a clear FAKE website based on the methodology exposed here:

Figure 54. FAKE website being promoted in Facebook



hence categorized as such in VirusTotal¹⁴.

Figure 48. FAKE website flagged as suspicious in VirusTotal.



URL:	http://www.rblovez.pw/
Detection ratio:	0 / 68
Analysis date:	2016-03-11 14:08:41 UTC (11 months ago)

Analysis	Additional information	Comments	0	Votes
----------	------------------------	----------	---	-------

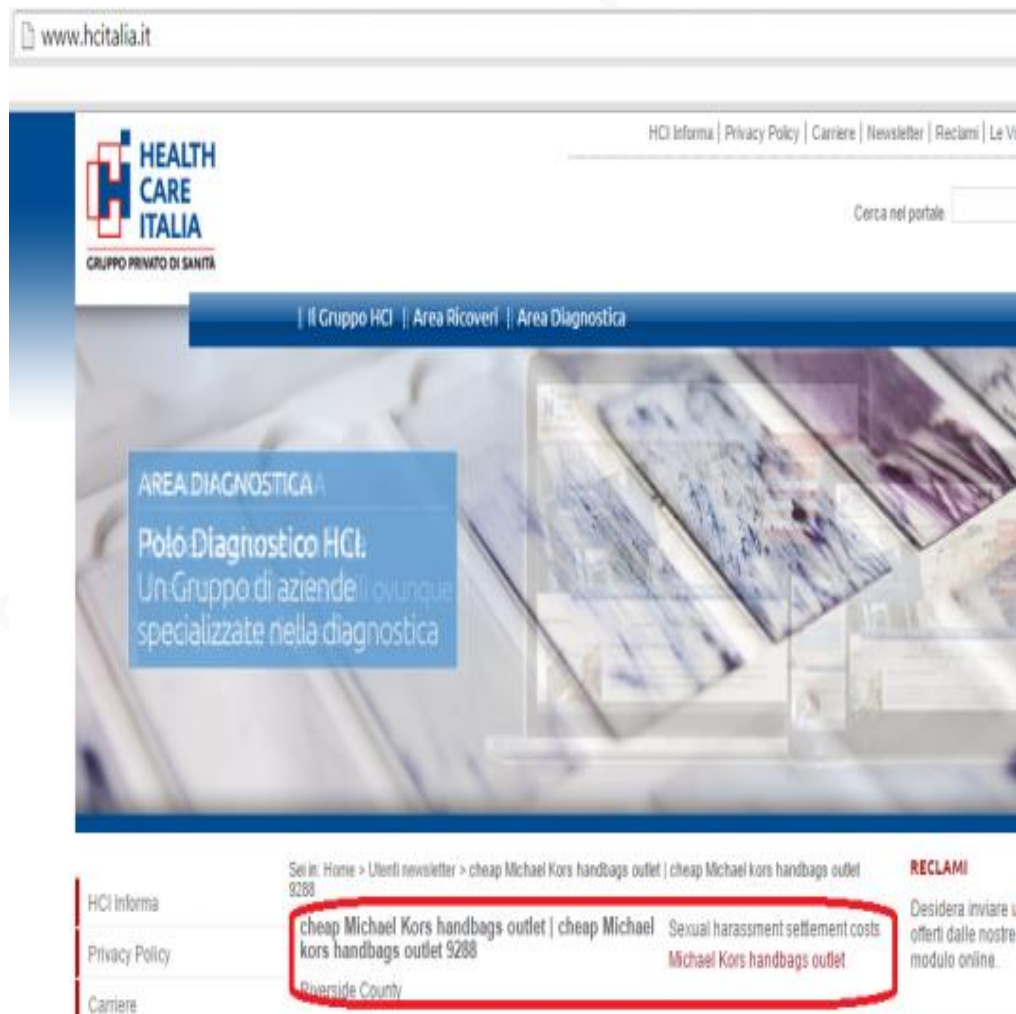
URL Scanner	Result
desenmascara.me	Suspicious site
ADMINUSLabs	Clean site
AegisLab WebGuard	Clean site
AlienVault	Clean site

¹⁴ <https://www.virustotal.com/en-gb/url/e651e1c5c9e31be8152b9ef28111f9cf0a4db1473b0f5d1830ba6ef2270449eb/analysis/1457705321/>

Appendix H: Legitimate Website Compromised Leading to a FAKE Website

Legitimate Italian private healthcare group website compromised with links leading to a FAKE website.

Figure 49. Compromised website linking to a FAKE MK site



Appendix I: PassiveDNS Information

Historic data obtained with a passiveDNS search of an IP extracted from a FAKE website. This information allows us to collect additional FAKE websites.

Figure 57. PassiveDNS commands and results

```
[redacted@redacted dnssdb]# python dnssdb_query.py -i 137.175.57.14
051n.cn. IN A 137.175.57.14
0f2.cl820.cn. IN A 137.175.57.14
www.1m2.cl820.cn. IN A 137.175.57.14
www.rii.cl820.cn. IN A 137.175.57.14
www.cl820.cn. IN A 137.175.57.14
dhl.amjsdc3.in. IN A 137.175.57.14
bjhnsn.com. IN A 137.175.57.14
www.bjhnsn.com. IN A 137.175.57.14
dejrb.com. IN A 137.175.57.14
www.dejrb.com. IN A 137.175.57.14
esrbb.com. IN A 137.175.57.14
www.esrbb.com. IN A 137.175.57.14
nrbel.com. IN A 137.175.57.14
www.nrbel.com. IN A 137.175.57.14
oakfa.com. IN A 137.175.57.14
www.oakfa.com. IN A 137.175.57.14
rbadu.com. IN A 137.175.57.14
www.rbadu.com. IN A 137.175.57.14
rbafk.com. IN A 137.175.57.14
www.rbafk.com. IN A 137.175.57.14
rbahk.com. IN A 137.175.57.14
www.rbahk.com. IN A 137.175.57.14
www.rbdeq.com. IN A 137.175.57.14
rbdov.com. IN A 137.175.57.14
www.rbdov.com. IN A 137.175.57.14
rbeii.com. IN A 137.175.57.14
www.rbeii.com. IN A 137.175.57.14
rberd.com. IN A 137.175.57.14
www.rberd.com. IN A 137.175.57.14
rbfen.com. IN A 137.175.57.14
www.rbfen.com. IN A 137.175.57.14
rbfid.com. IN A 137.175.57.14
www.rbfid.com. IN A 137.175.57.14
rbhma.com. IN A 137.175.57.14
www.rbhma.com. IN A 137.175.57.14
rbjnl.com. IN A 137.175.57.14
www.rbjnl.com. IN A 137.175.57.14
www.rbkhi.com. IN A 137.175.57.14
rbloq.com. IN A 137.175.57.14
www.rbloq.com. IN A 137.175.57.14
rbmwc.com. IN A 137.175.57.14
www.rbmwc.com. IN A 137.175.57.14
rbnee.com. IN A 137.175.57.14
www.rbnee.com. IN A 137.175.57.14
rbofe.com. IN A 137.175.57.14
www.rbofe.com. IN A 137.175.57.14
rbqde.com. IN A 137.175.57.14
www.rbqde.com. IN A 137.175.57.14
rbs-k.com. IN A 137.175.57.14
www.rbs-k.com. IN A 137.175.57.14
rbsai.com. IN A 137.175.57.14
www.rbsai.com. IN A 137.175.57.14
rbuom.com. IN A 137.175.57.14
```

Appendix J: Plaintiff Against FAKE Websites

The below example is just an overview of some examples of FAKE sites identified as FAKE by worldwide-anticounterfeiting programs managed by some brands.

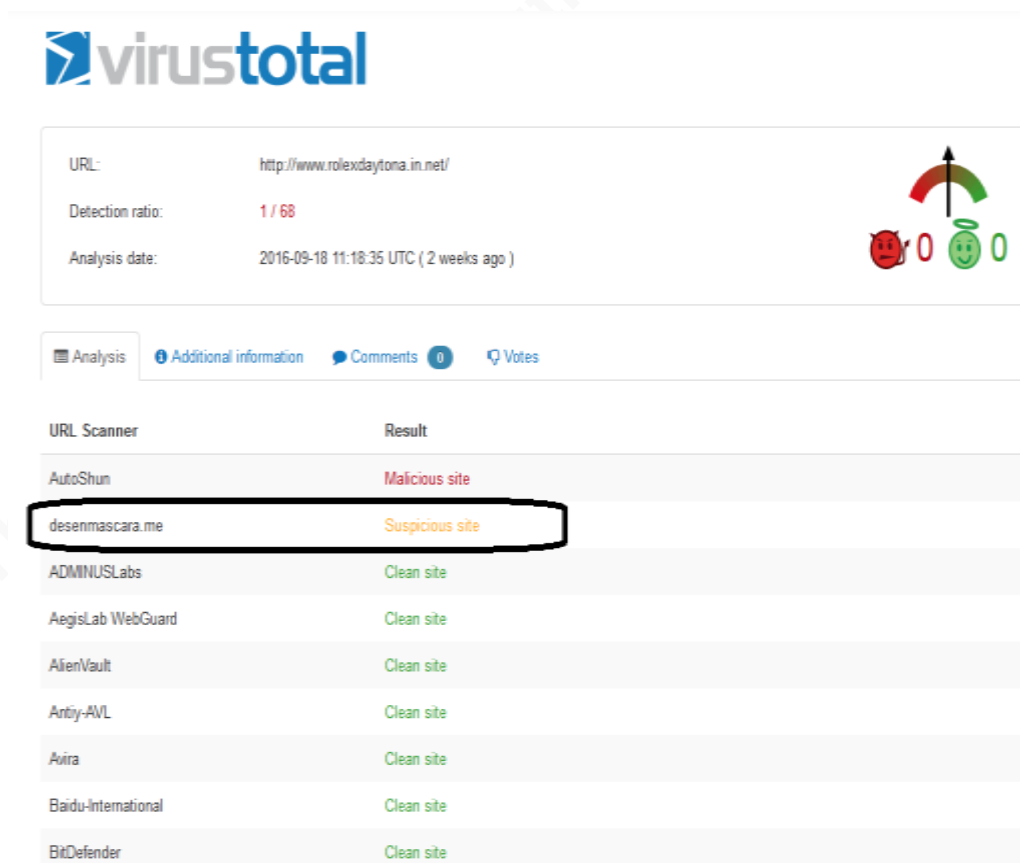
Figure 58. Thousands of FAKE sites targeting to Ray-Ban and UGG brands

1625	rbnew-usa.com
1626	rbspn.com
1627	rbtop-ca.com
1628	rbtops-ca.com
1629	rbuqs.com
1630	rbv-uk.com
1631	rbx-uk.com
1632	rbz-uk.com
1633	salomoneworld.com
1634	ugg-xshop.com
1635	rb-esale.com
1636	rybnases.site
1637	rb2017.com
1638	rb2fb.com
1639	rb2me.com
1640	rbacj.com
1641	rbacj.com
1642	rbacm.com
1643	rbacp.com
1644	rbacq.com
1645	rbacr.com
1646	rbacu.com
1647	rbacv.com

Appendix K: Desenmascara.me Integrated Into VirusTotal

The desenmascara.me integration with VirusTotal will show the “suspicious site” message whenever a website scanned with VirusTotal turns out is flagged as FAKE by desenmascara.me. In this specific case, as already pointed out in the paper, there might be cases where a website is FAKE and at the same time hosting malware.

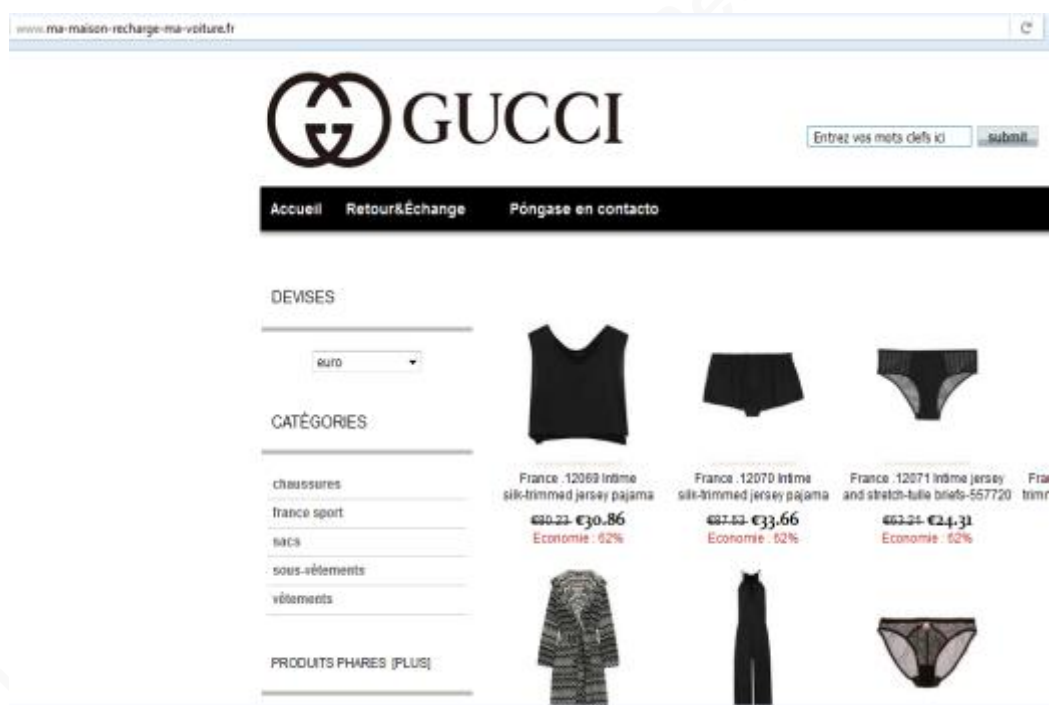
Figure 59. desenmascara.me FAKE site detection integrated with VirusTotal



Appendix L: Compromised Website Showing a FAKE Shop

The below FAKE Gucci website was setup over the compromised domain: www.ma-maison-recharge-ma-voiture.fr. These kinds of FAKE sites have a short lifetime until the legitimate website owner realized to his surprise that has been compromised.

Figure 60. Compromised domain used to show a FAKE shop



Appendix M: Feed of counterfeit-related websites

With all the information collected here a new threat intelligence feed in relation with counterfeit related websites could be generated. For instance let's see an example:

Figure 61. Feed of FAKE sites

```
URL;BRAND;CURRENT STATUS;LAST TIME CHECKED
http://www.handbagsoutlet.store; vuitton; 200; March 10, 2017, 11:10 p.m.
http://thomassabosaleuk.org; thomassabo; 200; March 10, 2017, 11:10 p.m.
http://www.nikeairmax.us.org; nike; 200; March 10, 2017, 11:10 p.m.
http://moncleroutletsale.us; moncler; 200; March 10, 2017, 11:10 p.m.
http://belstaffjacketsale.top; belstaff; 200; March 10, 2017, 11:10 p.m.
http://www.shopshoes.es; nike; 200; March 9, 2017, 11:07 p.m.
http://www.watches.ac.cn; replica watches; 200; March 9, 2017, 11:07 p.m.
http://www.jimmychoopumps.cn; jimmy choo; 200; March 9, 2017, 11:07 p.m.
http://www.highreplicawatches.cn; rolex; 200; March 9, 2017, 11:07 p.m.
http://nikeshoes.us.com; nike; 200; March 9, 2017, 11:07 p.m.
http://coach--outlet.us.com; Generic; 200; March 9, 2017, 11:07 p.m.
http://www.coachoutletonlinecoachfactoryoutlet.us.com; coach ; 200; March 9, 2017, 11:07 p.m.
http://www.coach--outlet.us.com; coach ; 200; March 9, 2017, 11:07 p.m.
http://zapatillasofficial.com; Generic; 200; March 9, 2017, 11:07 p.m.
http://www.wwwlouisvuittonoutlet.us.com; vuitton; 200; March 9, 2017, 11:07 p.m.
http://www.monclerjacketoutlets.us.com; jackets outlet; 200; March 9, 2017, 11:07 p.m.
http://www.veganpursesandbags.com; michael kors; 200; March 9, 2017, 11:07 p.m.
http://hotshoesbrand.com; Generic; 200; March 9, 2017, 11:07 p.m.
http://www.women-moncler.co; jackets outlet; 200; March 9, 2017, 11:07 p.m.
http://www.uktimberland-sale.com; timberland; socket error; March 9, 2017, 11:07 p.m.
http://www.pataegaonia.com; patagonia; 200; March 9, 2017, 11:07 p.m.
http://www.pandoraeurostop.com; pandora; 200; March 9, 2017, 11:07 p.m.
http://www.nikeseushop.com; nike; 200; March 9, 2017, 11:07 p.m.
http://www.nikeseuromall.com; nike; 200; March 9, 2017, 11:07 p.m.
http://www.nikeseubuy.com; nike; 200; March 9, 2017, 11:06 p.m.
http://www.montblancpens.me; mont blanc; 200; March 9, 2017, 11:06 p.m.
http://www.louisvuittonoutletonlineshop.com; vuitton; 200; March 9, 2017, 11:06 p.m.
```

The feed contains 5 fields separated by semicolon:

- The first field is the FAKE site detected by following this methodology
- The second field is the brand targeted by the online counterfeiters
- The third field is the HTTP code status of the fake site
- The fourth field is the last seen time