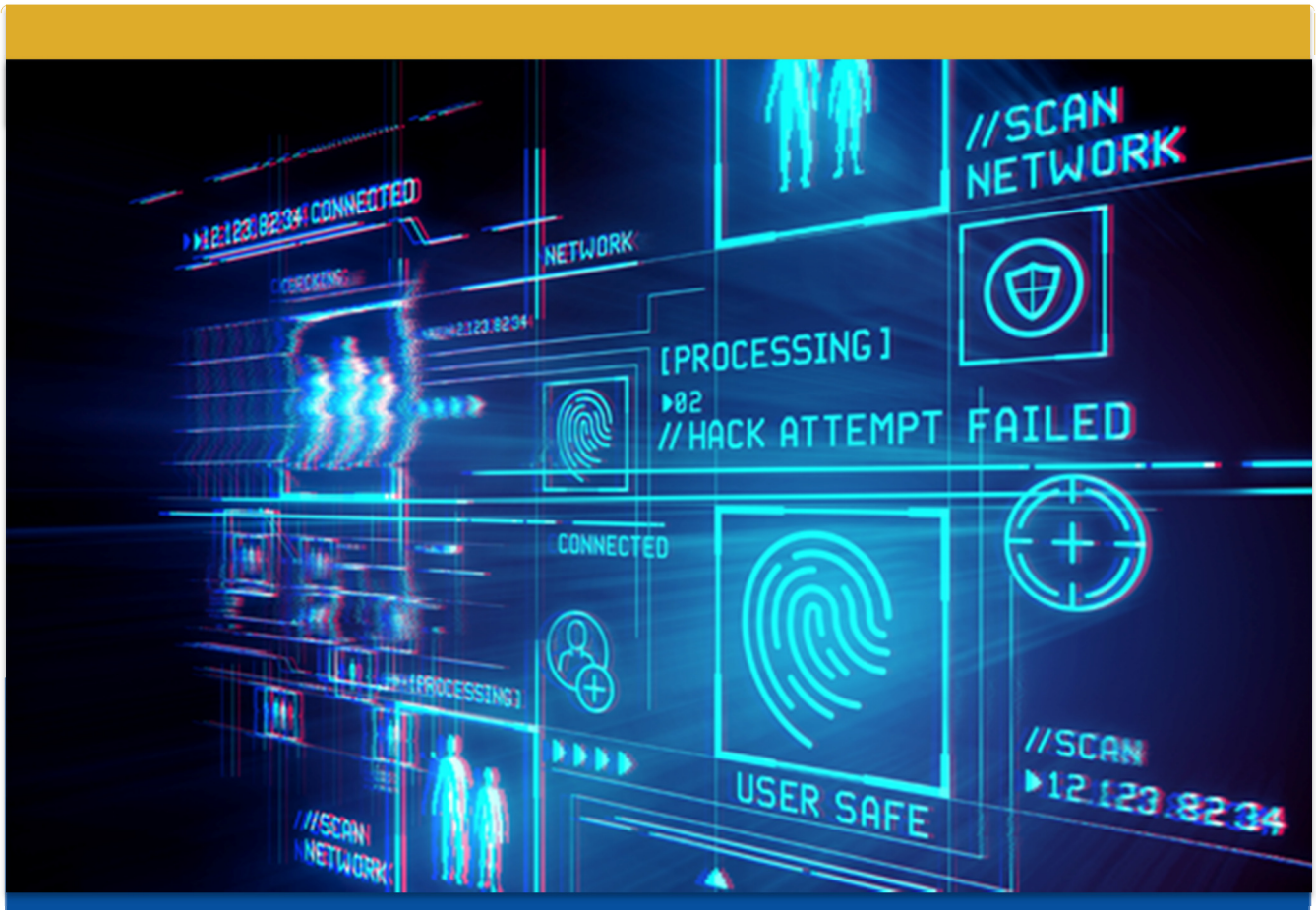
EUIPO
EUROPEAN UNION
INTELLECTUAL PROPERTY OFFICE

www.euipo.europa.eu

IDENTIFICATION AND ANALYSIS OF MALWARE ON SELECTED SUSPECTED COPYRIGHT- INFRINGING WEBSITES



September 2018

ISBN 978-92-9156-254-1 doi: 10.2814/004056 TB-01-18-336-EN-N

Abstract

Suspected copyright-infringing content represents a significant infringement of intellectual property rights. There are some websites that share such content publicly, sometimes even free of charge, without any registration. Along with this content, the websites commonly distribute various kinds of malware and potentially unwanted programs (PUPs), luring users into downloading and launching these files. The study provides an overview of the most up-to-date examples of malware and PUPs found on suspected copyright-infringing websites. These programs use deceptive techniques and social engineering — such as empty game installations and ostensibly 'useful' software — to trick end-users into disclosing their sensitive information. During the study, a variety of PUPs were discovered such as either 'useful' software, fake game installers and clients for video-streaming platforms. This software does not necessarily pose direct dangers to the user's software or hardware. However, through social engineering tricks, a user might be convinced to disclose sensitive personal information or payment card details. In addition, information about the computer itself might be leaked to other parties without explicit user consent.

Research Team

The research team consisted of Francesca Bosco, UNICRI Programme Officer, and Andrii Shalaginov, PhD research fellow in information security at the Department of Information Security and Communication Technology (Digital Forensics Group), Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology.

Disclaimer

In this context, it should be emphasised that the sole aim of the research was to determine the technical characteristics of malware and PUPs that were encountered during the study and could be encountered by internet users looking for suspected copyright-infringing content. The documented malware and PUP samples cannot be considered exhaustive, nor was the aim of the study (or its results) to provide an assessment of the overall likelihood or risk of malware and PUP infection an internet user would encounter when looking for suspected copyright-infringing material.

Contents

Foreword

Suspected online copyright-infringing activities can be financed in a variety of ways, including subscription fees, donations, payment for auxiliary services and income from online display advertising.

However, not all means of financing are as benign as the examples given. For years, dissemination of malware infection and other kinds of potentially unwanted programmes (PUPs) has been of key importance in relation to financing suspected copyright-infringing activities on the internet.

Ordinary internet users are starting to become aware of the risks of infection when accessing suspected copyright-infringing websites or mobile applications.

The EUIPO's 2015 IP Youth Scoreboard showed that 52 % of youngsters consider that safety on a website is important when accessing online content. Altogether, 78 % of youngsters stated that they would think twice if they were aware of a risk that the computer or device could be infected by viruses or malware. Altogether, 84 % stated that they would think twice if they were aware of a risk that credit card details could be stolen.

In the research for this study, the Office set out on a very technically challenging task, namely to detect and document examples of malware and PUPs that an internet user could encounter when trying to access popular pirated films, music, video game and television titles.

In this context, it should be emphasised that the sole aim of the research was to determine the technical characteristics of malware and PUPs that were encountered during the study and that could be encountered by internet users looking for suspected copyright-infringing content. The documented malware and PUP samples cannot be considered exhaustive, nor was the aim of the study (or its results) to provide an assessment of the overall likelihood or risk of malware and PUP infection an internet user would encounter when looking for suspected copyright-infringing material.

The research was carried out in several phases, in close cooperation with the European Cybercrime Centre (EC3) at Europol.

The results show a variety of different malware and PUP threats that an internet user can encounter when looking for suspected copyright-infringing content. Most of the documented malware and PUPs can be described as Trojans or other unwanted software that is able to gain unwarranted access to the personal data of internet users. These examples will be relevant and of interest not only to the IP rights holder community, but also to enforcement authorities and, last but not least, to consumers who are concerned about their personal data being accessed without their authorisation.

List of Acronyms and Abbreviations

ABI	Application binary interface
API	Application programming interface
APK	Android application package
ARM	Advanced RISC machine processor architecture
AWS	Amazon web services
BIOS	Basic input-output system
CRL	Certificate revocation list
C++	Programming language
DDoS	Distributed denial of service attack
DGA	Domain-generation algorithm
DLL	Dynamic-linked library in Microsoft OS
DoS	Denial of service attack
EMAS	Europol Malware Analysis Solution
ENISA	European Union Agency for Network and Information Security
EXE	Executable files in Microsoft OS
EU	European Union
EUIPO	European Union Intellectual Property Office
EULA	End-user licence agreement
FEBETA	A class of malicious activities in EMAS report
GB	Gigabyte
GT	Google Transparency
GUID	Globally Unique IDentifier
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTP GET	A method that uses string values sent in the URL
HTTP POST	A method that uses the body of a request to send values
IP	Internet protocol
ISO	International Standards Organization (Refers to ISO-9660, a package that contains an archive of optical disc)
ISP	Internet Service Provider
LEA	Law enforcement agency
MD5	Message-Digest algorithm 5 (A cryptographic hash sum method)
NSIS	Nullsoft Scriptable Install System
OCSP	Online Certificate Status Protocol
OS	Operating system
PKG	Mac OS installation package
PHA	Potentially harmful applications
PUP	Potentially unwanted program
ROM	Read-only memory (related to games, a read-only memory copy of the game)
SDK	Software development kit
SOAP	Simple Object Access Protocol
UN	United Nations
UNICRI	United Nations Interregional Crime and Justice Research Institute
URL	Uniform resource locator
VPN	Virtual private network
VT	VirusTotal (malware analysis website)
XML	Extensible Mark-up Language
YARA	Yet Another Ridiculous Acronym (a tool for malware detection)
4G	Fourth generation cellular network

Glossary

Adware	Potentially unwanted program that is used to show advertisements actively.
Audio streaming	Access to audio content that plays immediately without needing to download a file to a device.
Backdoor	Functionality of malicious or otherwise unwanted software that allows an attacker to access a user's computer without proper access.
Banking Trojans	A malware family that seeks to steal bank credentials with the purpose of stealing money at a later date.
BitTorrent tracker	Peer-to-peer protocol that allows decentralised file-sharing across multiple users without needing to store the file on a central server.
Certification authority (CA)	A trusted organisation that issues digital certificates.
Copyright	An exclusive legal right of ownership of intellectual property, such as digital content. The content cannot be distributed, reproduced or sold without the author's consent.
Copyright-infringing website	A website that is used to distribute suspected copyright-infringing content without consent from the rights holder.
Click fraud	A type of fraud that seeks to maliciously profit from increasing the number of clicks on legitimate advertisements through manual or automated approaches.
Codec	A specific software program or hardware that encodes and decodes media content streams for storing, compressing and/or reproducing analogue media content.
Cyberlockers	Shared hosting used to store users' files that can later be accessed using unique links.
Digital content	In the context of this report, any type of digital content that can be distributed on copyright-infringing websites (films, TV programmes, music, software, games).
Domain suffix	Top-level domain or the last part of a domain name.
Drive-by downloads	Unintended and probable unauthorised downloading of files to a computer or mobile device from the internet.
Dynamic malware analysis	Analysis of malware that includes executing software and studying behavioural characteristics.
Fake installers	Type of software distributed on copyright-infringing websites that simulates the installation of legitimate software, with the difference that the software is not installed. The purpose is to acquire the user's personal data.
Goodware	Software that was designed and used for a good purpose.
Google Transparency Report	A comprehensive report by Google that reveals various requests by governments and companies over a specific period of time.

Hosting websites	Websites that provide direct access to digital content in addition to a general description of this content on a corresponding web page.
Installer	Specifically designed software package used to perform actions necessary for proper operation of new software being installed (includes copying of files, registry entries, and link creation).
Keylogger	Function or process used to capture all activities from a keyboard for any purpose, such as the collection of passwords or stealing of other credentials.
Linking website	A website that displays links to different suspected copyright-infringing content on other websites. Typically, it does not host any illegal data.
Malicious activities	Any actions by software that may cause harm to a user or a computer system it is being used on.
Malware	Software designed to infect a computer system or cause harm to a user's data through alteration, theft, or deletion.
Malware analysis artefacts	A set of specific characteristics indicating malicious activities, such as files on a disc, network packets and registry keys.
Malvertising	Embedding malicious or otherwise unwanted programs within advertisements on web pages, providing a way to spread such programs using legitimate marketing platforms.
Media player	Software designed to present digital media content such as films, audio files, etc.
Potentially unwanted program	Software that is not necessarily harmful, but may be considered as annoying and unwanted by some users.
Ransomware	A type of malware that encrypts content from the computer and requires a user to pay a ransom in order to regain access to the content, often using bitcoins.
Remote access tools	Software that allows various types of external control over a user's computer; may be benign or malicious (as used in botnets).
Sandbox environment	Specifically designed safe environment for malware analysis that prevents the hosting machine from being exploited or attacked.
Search engine Optimisation	A set of methods devoted to increasing the popularity of a website and its ranking by search engines.
Search result poisoning	Methods to promote malicious websites in search engines, so that specific search keywords result in links to such websites.
Static malware analysis	Malware analysis that examines the static properties of malicious binary files, such as file size and content, without executing them.
Streaming websites	Websites that provide access to media content such as audio and video through streaming and without any need to download.
Tor browser	Browser built upon the Tor anonymisation network to hide the true identity of users.
Torrent	A file that is used in the BitTorrent protocol to distribute files.

Torrent websites	Websites that provide access to torrent files through corresponding web pages that usually include a brief description (text and images) of the content being distributed.
Trojan	A type of malicious or otherwise unwanted software that has a hidden functionality, which is activated under certain conditions.
Unofficial Android markets	Unofficial third-party application markets (as opposed to the official Google Play store), from which applications can be installed for the Android mobile platform.
'Useful' software	Software that can generally be considered by users as 'useful'. It may include functionalities such as cleaning up old files, speeding up one's computer and removing unnecessary system files.
Video streaming/Video on demand (VOD)	Accessing video materials such as films and TV shows without direct downloading of the video files to computers or mobile phones.

1. Executive Summary

The study provides an overview of the most up-to-date examples of malware and potentially unwanted programs (PUPs) found on suspected copyright-infringing websites. These programs use deceptive techniques and social engineering — such as empty game installations and ostensibly 'useful' software — to trick end-users into releasing their sensitive information.

The goal of this study is to discover and document malicious or otherwise unwanted software disseminated on selected websites suspected of infringing copyright and to categorise the samples found in line with various malware taxonomies. In this context, it should be emphasised that the study had the sole aim of determining the technical characteristics of malware and PUPs that were encountered during the research and could be encountered by internet users looking for suspected copyright-infringing content. The documented malware and PUP samples cannot be considered exhaustive, nor was the aim of the research (or its result) to provide an assessment of the overall likelihood or risk of malware and PUP infection an internet user would encounter when looking for suspected copyright-infringing material. For the purpose of this study, TV shows, films, music and video games are considered copyright-protected content.

Outcomes of the Study

Suspected copyright-infringing content represents a significant intellectual property rights violation. There are some websites that share such content publicly, sometimes even free of charge, without any registration. Along with such content, the websites commonly distribute various kinds of malware and PUPs, luring users into downloading and launching such files. During the website identification based on the Alexa Top 500 ranking, in addition to a simulation of average user searches using well-known search engines, such as Google, Yahoo, and Bing, it was found that the set of websites changed between the two rounds of study. This change is probably the result of efforts by search engines to remove links to suspected copyright-infringing websites, while new suspected websites continue to appear. In relation to website identification, one interesting finding related to the fact that the overwhelming majority of the websites are hosted in the United States or have domain names linked to hosting there. On the contrary, only a few are located on servers within the EU. Furthermore, .com and .net are the most frequent top-level domain names used on suspected copyright-infringing websites. This may be caused by the fact that, unlike country-specific domains, these may not require identification of the user with a passport or other identification documents. On average, 20 % of new websites were added, and 20 % of old websites were removed between the two rounds of identification. Moreover, nearly 8 % of the websites identified in both rounds were characterised as malicious by the VirusTotal platform. With the help of various content management systems, it has now become almost effortless to create a website and deliver content to users, even malicious applications.

Before the malware collection, this study engaged in a desk review of malware threats in 2017 and a categorisation of the state of the art. This body of knowledge was further used during the malware analysis to follow community-accepted principles in malware types and family identification. In total, 106 files were collected during both rounds of data collection. These include files downloaded directly from suspected copyright-infringing websites, as well as files that were created during execution of the downloaded files. During the study, a variety of PUPs were discovered, such as either 'useful' software, fake game installers and clients for video-streaming platforms. Such software does not necessarily pose direct dangers to the user's software or hardware. However, through social engineering tricks, a

user might be convinced to disclose sensitive personal information or payment card details. In addition, information about the computer itself might be leaked to other parties without explicit user consent.

The collected malware was analysed initially using open-source tools to understand the internal logic, detect possible malicious activities and evaluate their relevance to the present malware study. In addition to the preliminary analysis using open-source tools, the collected malware samples were analysed by the Europol Malware Analysis Solution (EMAS) platform. This resulted in the detection of a large number of different artefacts and malicious activities. The EMAS reports include a comprehensive analysis of files using four versions of MS Windows, where network traffic, function calls, and disc activities are thoroughly logged for further analysis. In addition, the platform highlights any suspicious activities detected during file execution routines. After analysing all of the reports, 35 types of malicious activities were noted by EMAS that are aggregated in 17 classes of malicious events. These range from general anomalies (such as launching system processes or looking up processes in memories) to unmistakably malicious actions (such as keylogger, rootkit, and network traffic tampering).

Generally, the binary samples of malware and PUPs that were collected revealed a few different general business models: ‘useful’ programs claiming to clean up old files on a user’s computer upon a paid subscription; game installation simulators that require the user’s personal data; and free programs offering access to platforms that distribute pirated content, such as through BitTorrent tracker. The two rounds of website identification and malware collection produced promising results in terms of comprehending the methods of malware dissemination and social engineering in luring out sensitive personal and identifiable information. Furthermore, the increased popularity of mobile devices in recent years is evident in light of the detection of many PUPs for the Android OS, available through the suspected copyright-infringing content-distribution platforms. As a result of correlating the analyses, the conclusion was drawn that the threat landscape for malware distributed via copyright-infringing websites is more sophisticated than it might appear at first glance. Among the software discovered, some can additionally be classified as Trojan, adware, backdoor, and agent. This is compounded by the fact that many specific malware families, such as WisdomEyes, DealPly, and FileRepMalware were also found. Moreover, such a comprehensive categorisation is equally valid for the Android platform, not just Microsoft Windows. There is a wide range of threats to users’ assets, including but not limited to stealing sensitive credentials, personal data, hardware configuration information, and modifying network traffic. Therefore, even though the identified software may be PUPs, they can nevertheless have an impact on users, especially in cases involving an average user who might not be fully aware of basic online security practices and measures.

An example of the study's findings is shown below.

Website 03

The website tricks users into using a fake game installation; the entire process of obtaining a user's sensitive information has changed between the first and second rounds of malware collection.
The user of this service downloads an archive that contains content masked as game-related files and not an explicitly binary executable file that can be detected by any anti-virus as malicious. The encrypted archive grants access only to filenames but not the substantive content of the files.

Website 09

The website offers access to any kind of video content available through torrent trackers with the help of a software tool. This tool requires fewer user interactions in comparison with other BitTorrent trackers.
Only a few clicks are required to download content from unknown sources, while the user is neither protected nor has control of what is being downloaded.

Website 08

(Android) The website provides access to a range of free mobile applications without registration. One application provides unlimited access to streaming of TV shows and films. There is no explicit request to provide a user's sensitive information or payment details for buying access to copyright-protected videos. However, a user needs to disable security settings that will allow installation of applications other than ones from an official application market.

Methodology

In order to perform the research, a sound methodology had to be adopted to deal with the selection of titles and websites, as well as the technically challenging task of detecting and documenting the examples of malware and PUPs found. A brief overview of the methodology is described below:

1. In Phase I of the UNICRI research, in collaboration with the European Observatory on Infringements of Intellectual Property Rights (Observatory), an expert support group was established to provide advice on the research methodology, selection of websites used for analysis and to assess the research undertaken within each phase of project implementation. The expert support group was comprised of representatives from Observatory stakeholders, rights holder organisations, academia, law enforcement, and EU agencies.

2. In parallel, the research team was selected. Within the framework of this report, it was not technically possible[1] to research all EU Member States; therefore, 10 sample countries were randomly selected from the 28 EU Member States in Phase II.

3. In Phase III, popular films, television programmes, songs, and video games were identified. Popularity included worldwide popularity as well as popularity in only one or more of the 10 sample countries as at the start of the data collection period, 23 June 2017. In the subsequent phases of the study, these sample titles were systematically used in online web searches to find

[1] The number of selected countries will have a direct impact (increase) on the number of the selected suspected copyright-infringing websites and corresponding binary files to be analysed. Therefore, it was decided to concentrate only on a sample of countries to be able to successfully perform the practical part of the study within a given time frame.

copyright-infringing websites and mobile applications. Each title met two or more of the following criteria:

- popular at the time of data collection within EU Member States,
- popular at the time of data collection on a global scale,
- popular historically on a global scale, and
- categorised as a film, television programme, song, or video game.

Five film titles, five television titles, five music titles, and five video game titles were selected, resulting in a total of 20 sample titles. Careful consideration was given to the sources used to identify the popularity of a particular title, which involved a systematic selection process to ensure source data would be available for all or most of the Member States.

4. Phase IV identified websites suspected of providing illegal access to copyright-protected material that were popular worldwide and/or among the 10 sample countries as at 26 June 2017 (first round of malware collection). In a later phase of the study, these websites were analysed for the presence of malware and potentially unwanted programs.

The methodology for identifying suspected copyright-infringing websites was developed with the input of the expert support group identified in Phase I, as well as upon a review by UNICRI of the existing literature. It was specifically devised to generate a sample of websites that:

- are popular within different EU Member States, ensuring a wide geographical coverage;
- represent different types of suspected copyright-infringing websites, including streaming websites, linking websites, hosting websites, cyberlockers, and torrent websites;
- represent a broad range of suspected copyright-infringing content, including films, television titles, music, and video games; and
- represent websites that the average internet user would encounter when attempting to access suspected copyright-infringing material.

Five steps were used to select suspected copyright-infringing websites. The first three steps were designed to identify the most popular suspected copyright-infringing websites across EU Member States. This method mimicked those scenarios in which an average user might search for suspected copyright-infringing websites without specifying, for example, the title of a film or a song. The final two steps were designed to identify suspected copyright-infringing websites that an average user might encounter when searching for ways to download a specific popular title without specifying a website. This step was particularly significant, given the presence of suspected malicious websites that engage in search result poisoning, by which they exploit trending topics through search engine optimisation. Together, the two approaches covered the different ways an average internet user would attempt to find suspected copyright-infringing material online.

Emphasis was placed on the concurrent analysis of malware and PUPs specific to mobile applications on devices, such as smartphones and tablets, as one of the key emerging cybercrime threats. Analysis was limited to Android devices due to indications in the existing literature of a greater presence of malware on Android application stores (i.e. Google Play) than on the Apple iTunes store. The methodology was devised to generate a sample of mobile applications that:

- are popular at the time of data collection on a global scale;
- represent different types of applications (to include streaming applications, torrent applications, and hosting applications);

- contain or provide access to a broad range of suspected copyright-infringing content (to include films, television titles, music, and mobile games); and
- represent what an average user of a mobile device will encounter when attempting to download or use an application facilitating access to suspected copyright-protected content.

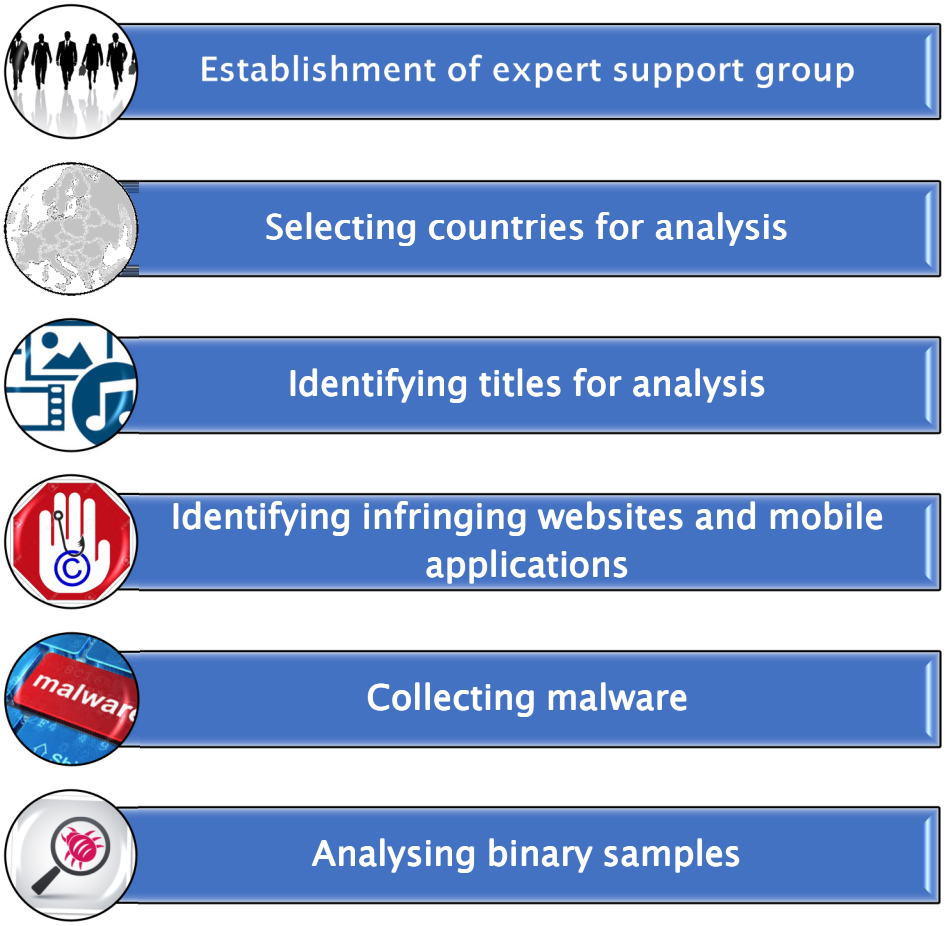
5. Phase V consisted of collection of malware and PUPs in addition to mobile applications on the identified websites, to be examined at a later stage for proper categorisation. The data acquisition phase included two rounds of malware collection and analysis performed during the summer of 2017. The first round of malware collection resulted in 1 054 unique domain names and the second round gave 1 057 unique domain names across 10 selected EU Member States. Malware was collected in both a manual and automated manner in order to simulate an average user's experience.

Manual collection. This method involved manually reviewing the domains identified in the previous phase. Using manual collection, the expert was able to simulate the experience of an average internet user by clicking advertisements and interacting with websites that required prompts.

Automated collection. This method employed an automated web crawler designed by an expert to follow all available links on a designated suspected copyright-infringing website. First, on any given website, the crawler would first collect information from the links on the home page. Second, the crawler would follow each of those links to secondary websites. Third, the crawler would follow each of those links to tertiary websites. At each step, the crawler retrieved binary files that could be of interest for subsequent manual analysis, including potential or suspected malware and potentially unwanted programs. This process continued for up to 1 000 links per website.

6. Once the binaries were collected, they were analysed in a safe computing environment to understand their internal functionality and for proper categorisation. Preliminary analysis was carried out using open-source tools to be able to correlate findings with cyberthreat reports. Collected software samples were then delivered to EMAS for analysis; the EMAS analysis was then compared with the preliminary results.

Overview of the methodology



Detected Malware and PUP samples

As at 28 July 2017, 5 240 websites (1 054 unique) had been automatically checked during the first round of collection, with 617 relevant files (music, video, torrent files and software) retrieved of an overall size of 47 GB. This unsorted batch of files required further analysis to decide which collected files were relevant for the study. The samples of copyright-infringing websites were similar across all 10 sample countries for each of the types of media (television programmes, films, music, and video games). As a result, Belgium was randomly chosen from the sample countries, and all websites identified as copyright-infringing websites for Belgium were manually verified for the presence of malicious or otherwise unwanted software. On 10 August 2017, after the second round of collection, a total of 3 665 files were automatically retrieved from the websites for all countries, with a total size of 167 GB. The overall number of unique URLs extracted for all countries was 1 057 out of the 5 606 websites, which made it unfeasible to check all of them manually.

After a preliminary analysis of the collected files, 106 unique binary files for MS Windows, Android and the Mac OS were extracted as a result of both rounds of malware collection. More specifically, 41 files were selected during the first round and 65 were selected during the second round — in particular: 2 for Mac, 15 for Android and 89 for MS Windows. Out of these files, 21 can be considered as well-known malicious programs as marked by multiple anti-virus vendors as being aggregated by the VirusTotal platform. These include files downloaded directly from selected websites suspected of infringing

copyright, as well as files that were created during execution of the downloaded files. Subsequently, collected software samples were analysed in a sandbox environment and delivered to EMAS for more advanced analysis of possible malicious activities. In overall, 821 distinct malicious events were discovered across four EMAS reports (Windows 7 SP1, Windows7 SP1 64-bit, Windows 10 64-bit, Windows XP SP3) for all binary files. Some of the reports did not have any suspicious activities and some of them had up to 10 previously known malicious activities. During the final stage of the study, the results of the preliminary analysis and from EMAS reports were correlated. The quantitative summary of the results is given in the table below.

	Round 1	Round 2
Date	28 July 2017	10 August 2017
Discovered websites across 10 EU countries	5 240	5 606
Unique websites	1 054	1 057
Relevant files	617	3 665[2]
Size of relevant files, GB	47	167
Delivered to EMAS		
Android	3	12
Mac OS	2	–
MS Windows	36	53
Total size, bytes	175 600 117	522 991 095

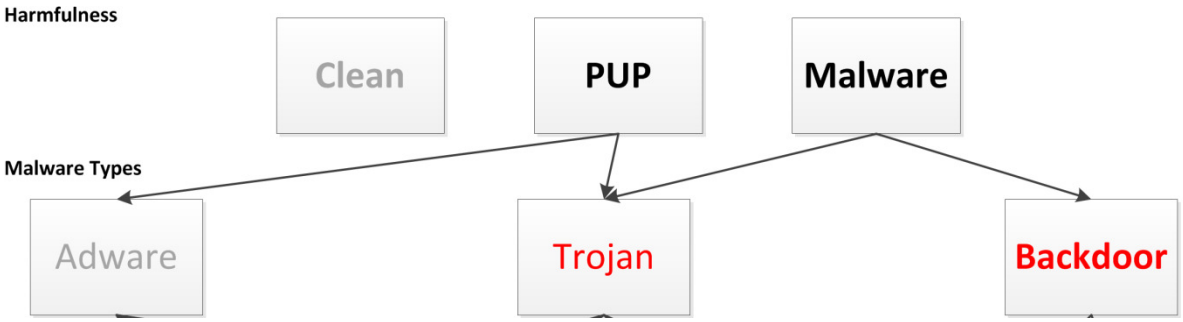
Europol Malware Analysis Solution (EMAS)

The Europol Malware Analysis Solution (EMAS) is a dynamic, automated malware analysis solution provided by Europol to EU Member States. EMAS offers the possibility of creating analysis reports, but its most revolutionary feature is to produce intelligence for police investigators. Automated cross-checks can show links between attacks performed in different countries with the same malware, or with the same criminal organisation behind the same malware family, connecting to the same domains and related to different investigations within or outside the EU. In 2015, EMAS became fully automated to allow direct access to law enforcement parties with which Europol has operational agreements. In 2015: 525 108 files were analysed in EMAS, out of which 356 863 were identified as malicious.

As shown in the figure below, collected binary files can generally be categorised according to harmfulness, as benign (files that do not bring any harm), PUPs and harmful malware. Moreover, PUPs were not only discovered for Microsoft Windows; they were also found for the Android and the Mac OS, which suggests that malware developers try to affect as many users as possible by using different platforms. The PUPs and malware can be further differentiated based on the main malware types, that is, Trojan, adware and backdoor. Most of the software that was found fell into the PUP category. The functioning of PUPs can be associated with one of the following business models: fake game installation requiring personal and bank account details, download of 'useful' programs that force users to buy a subscription to a paid version, or installation of free programs to access copyright-infringing platforms. These applications may compromise users' personal details and computer configuration.

[2] To explain the difference in numbers between Round 1 and Round 2, during Round 2 of automated collection there were websites that published multiple sets of files on each of their web pages.

Through social engineering tricks, various kind of private data, such as payment card details, personally identifiable information and social media account credentials may also be disclosed. Likewise, the research identified 15 Android applications from third-party application markets and, after the preliminary analysis, it was concluded that such applications may be involved in the distribution of copyright-infringing content and in disclosing personal data.



Threats to end-users

During two rounds of website identification and malware analysis, no ransomware binaries were found. Generally, most of the collected malware can be characterised as Trojans, meaning that they might be represented on the websites as benign commonly used or popular software, while in reality they can steal or disclose private information. An inexperienced user might have a high degree of trust in the software and might not be able to notice any abnormalities. In addition, static analysis and dynamic behavioural observations of such software might not reveal the complete functionality without having a source code. Following the preliminary malware analysis, EMAS analysis showed more specific malicious activities. The impact of having this software installed on an end-user's computer might be considerable, causing not only financial losses, but also theft of personal data and other risks of unwanted access and control. These activities may be expected to result in personal information gathering and transmission to third parties in encrypted or open text format. Such data might consist of, for example, bank account credentials from the browser, details of the computer hardware/software configuration, or basically anything typed on the keyboard.

2. Introduction

The distribution of protected digital content on copyright-infringing websites is one form of intellectual property infringement[3] occurring in the European Union[4]. These websites allow users to stream content directly to their computer, download content or torrent content using peer-to-peer file sharing.

In addition to the protected digital content, copyright-infringing websites also play host to malicious or unwanted software designed to carry out unauthorised actions on a computer system. Malicious software (malware) and potentially unwanted programs (PUPs) can be hidden in and disseminated by way of advertisements (malvertising), the digital content itself, media players, or codecs on websites. As previous studies suggest, the prevalence of malvertising, malware and other unwanted software on suspected copyright-infringing sites is far from rare. For example, a 2016 report commissioned by the EUIPO acting through the Observatory found that over half (51 %) of the advertisements present on such websites contained malware[5]. A 2015 report by the Digital Citizens Alliance found that one out of every three websites that distributed copyright-infringing films and television programmes contained malware[6]. Malware is also thought to be present on mobile applications, particularly given the increasing incidence of mobile malware[7].

To better understand this issue, the EUIPO acting through the Observatory has engaged with UNICRI to perform the current research; its purpose is to study the prevalence and types of malware and PUP threats present on popular suspected copyright-infringing websites and mobile applications. Binary samples of such software were collected from advertising, executable files and the digital content itself available on popular suspected copyright-infringing websites in the European Union. The binary samples were then analysed to provide information on the types and seriousness of malware and PUPs.

The results of this study might have potential implications with regard to the behaviour of internet users. There is evidence to suggest that once people are made aware of the dangers of malware in specific scenarios, they are more likely to act cautiously. For example, the 2016 IP Youth Scoreboard found that over three quarters (78 %) of respondents aged 15-24 stated they would 'think twice before using illegal sources' if they knew there were virus or malware-related threats[8].

The study can also have implications for policymakers, law enforcement, civil society, and private businesses in terms of formulating effective strategies to counter the challenge of suspected commercial-scale online copyright infringements connected to dissemination of malware and PUPs.

[3] Europol and EUIPO, *2017 Situation Report on Counterfeiting and Piracy in the European Union*, 2017; retrieved from https://www.europol.europa.eu/sites/default/files/documents/counterfeiting_and_piracy_in_the_european_union.pdf.
[4] EUIPO, *Research on Online Business Models Infringing Intellectual Property Rights*, 2017; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/resources/Research_on_Online_Business_Models_IBM/Research_on_Online_Business_Models_IBM_en.pdf.
[5] WhiteBullet Solutions Ltd, *Digital Advertising on Suspected Infringing Websites*, European Observatory on Infringements of Intellectual Property Rights, Alicante, 2016; retrieved from https://euipo.europa.eu/ohimportal/documents/11370/80606/Digital+Advertising+on+Suspected+Infringing+Websites.
[6] 'How content theft sites and malware are exploited by cybercriminals to hack into internet users' computers and personal data', *Digital Bait*, Digital Citizens Alliance and RiskIQ, December 2015; retrieved from http://www.digitalcitizensalliance.org/clientuploads/directory/Reports/digitalbait.pdf.
[7] Europol, *Internet Organised Crime Threat Assessment (IOCTA) 2016*, Europol, The Hague, 2016; retrieved from https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2016
[8] EUIPO, *Intellectual Property and Youth: Scoreboard 2016*; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/IP_youth_scoreboard_study/IP_youth_scoreboard_study_en.pdf.

3. Methodology of the Study

3.1 Phase I. Establishment of the expert support group

In the first phase of the UNICRI study, in collaboration with the Observatory, an expert support group was established to advise on the research methodology, the selection of websites used for analysis, and to assess the research undertaken within each phase of project implementation. The expert support group was comprised of representatives from Observatory stakeholders, rights holder organisations, academia, law enforcement, and EU agencies.

3.2 Phase II. Selecting countries for analysis

Within the framework of this report, it was not technically possible[9] to research all EU Member States, therefore in Phase II 10 sample countries were randomly selected from the 28 EU Member States. As a result of the randomisation process[10], the sample countries used throughout the study are as follows.

- Eastern Europe: Bulgaria, Croatia, Czech Republic, Hungary, and Lithuania, and
- Western Europe: Belgium, Finland, France, Portugal, and Sweden.

3.3 Phase III. Identifying titles for analysis

Popular films, television programmes, songs, and video games were identified in Phase III. Careful consideration was given to the sources used to identify the popularity of a particular title, which involved a systematic selection process to ensure source data would be available for all or most of the Member States[11]. For a summary of the sample countries, titles, and sources used in the project, refer to *Table 1*[12].

Sample countries

Eastern Europe:	Bulgaria Croatia Czech Republic Hungary Lithuania
Western Europe:	Belgium Finland France

[9] The number of selected countries would have a direct impact (increase) on a number of the selected suspected copyright-infringing websites and corresponding binary files to be analysed. Therefore, given the framework of the current study, it was decided to concentrate only on a sample of countries to be able to successfully perform the practical part of the study within a given time line.
[10] RAND function in Microsoft Excel, a pseudo-random number generator developed by B.A. Wichman and I.D. Hill.
[11] UNICRI investigated the viability of using Amazon, Netflix, or Google Play to gather information on the most popular titles. Numerous Member States do not have dedicated Amazon websites, however, which would make it difficult to ascertain what titles are the most popular. Netflix does not release official lists of its most popular titles. Lastly, Google Play does not appear to share lists of popular titles by country on its website.
[12] The details of the methodology for the selection of titles are in Annex 1.

		Portugal Sweden
	Sample titles	
Films:		1. *Avatar* 2. *Kong: Skull Island* 3. *Beauty and the Beast* 4. *Baywatch* 5. *Pirates of the Caribbean: Dead Men Tell No Tales*
Television:		1. *Game of Thrones* 2. *The Walking Dead* 3. *Pretty Little Liars* 4. *Alias* 5. *The Missing*
Music		1. 'See You Again' by Wiz Khalifa, featuring Charlie Puth 2. 'Love Yourself' by Justin Bieber 3. 'Despacito' by Luis Fonsi and Daddy Yankee, featuring Justin Bieber 4. 'Wild Thoughts' by DJ Khaled 5. 'Θ Macarena' by Damso
Video Games		1. Middle-Earth: Shadow of Mordor (Game of the Year Edition) 2. Playerunknown's Battlegrounds 3. The Sims 4 4. Minecraft 5. The Witcher III: Wild Hunt

Table 1: Summary of the titles selected for use in the project

3.4 Phase IV.A. Identifying suspected copyright-infringing websites for analysis

The methodology for identifying suspected copyright-infringing websites was developed with the input of the expert support group identified in Phase I as well as upon a review by UNICRI of the existing literature. It was specifically devised to generate a sample of websites that:

- are popular within different EU Member States, ensuring a wide geographical coverage;
- represent different types of suspected copyright-infringing websites, including streaming websites, linking websites, hosting websites, cyberlockers, and torrent websites;
- represent a broad range of suspected copyright-infringing content, including films, television titles, music, and video games; and
- represent websites that the average internet user would encounter when attempting to access suspected copyright-infringing material.

3.5 Phase IV.B. Identifying mobile applications for analysis

During the research project, there was increasing interest among the expert support group in conducting concurrent analysis on malware and PUPs specific to mobile applications on devices, such as smartphones and tablets. Europol's *Internet Organised Crime Threat Assessment* (2016) identified mobile malware as one of the key cybercrime threats facing Europe[13]. The incidence of mobile malware

[13] Europol, *Internet Organised Crime Threat Assessment (IOCTA) 2016*, Europol, The Hague, 2016; retrieved from https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2016.

and PUPs is increasing and becoming more complex[14]. Different forms of mobile malware and PUPs mirror their computer-based counterparts and include remote access tools, drive-by downloads, click fraud, banking Trojans, and ransomware. In addition, phishing applications have made their way onto Google Play, the main application store for smartphones and mobile devices using the Android OS[15]. In addition, the Android OS allows third-party applications to be installed, opening opportunities for attackers to exploit this functionality. The applications often purport to be associated with reputable companies (e.g. financial companies and payment service providers) and, when downloaded and accessed by a user, prompt the user with a dialogue box to enter his or her login name and password. Such applications are usually made to appear similar to official applications, but they contain a malicious payload. The application then steals the user's username and password and any other relevant information that has been entered.

3.6 Phase V.A. Collecting malware and PUPs on identified websites and mobile applications

The goal of Phase V was to collect samples of malware and PUPs that an average internet user might encounter when attempting to access suspected copyright-infringing content via suspected websites. A sandbox environment with a Tor browser installed was used to collect the malware. Explicit instructions for creating a safe sandbox environment were given by UNICRI to the expert conducting the collection of binary samples on behalf of UNICRI. The expert conducted the searches in a manner consistent with low security-awareness internet browsing. This included not using an adblocking service, as well as clicking suspicious links and buttons. The Tor browser was configured to simulate the searches as being conducted locally from the respective sample countries. To do this, the configuration file was edited to explicitly define the desired country of Tor exit network node. Additionally, the expert composed an HTTP GET request that included, inter alia, the corresponding country-specific search engine domain name, search keyword and English titles of the digital content searched for. This allowed for the collection of any relevant geo-targeted data for analysis, recognising that the ranking position and popularity of each website differs from one EU Member State to another. Collection of binary samples was carried out over two stages: manual and automated collection using an internet crawler designed by the expert.

3.7 Phase V.B. Analysing the binary samples

Once the binaries were collected, they were analysed in a safe environment for proper categorisation. Preliminary analysis was carried out using open-source tools to be able to correlate findings with cyberthreat reports. It included static and dynamic analysis highlighting the corresponding artefacts that indicate malicious or potentially unwanted activity. Static analysis is related to processing static features of files such as size, specific content, etc., while dynamic analysis may reveal specific functionality logic upon execution of the binary. Desktop and mobile applications were analysed in corresponding virtual environments to reduce the risk of malware infection of the research equipment. To carry out the in-depth analysis, UNICRI sent the binary samples to the Italian postal police, who in turn passed this data on to the European Cybercrime Centre (EC3) at Europol via the Secure Information Exchange Network Application (SIENA) tool. The samples were analysed by EC3 in the Europol Malware Analysis Solution (EMAS) sandbox environment, and the results were returned via SIENA to the Italian police, and subsequently to UNICRI. The origin, threat level, and functionality of each piece of malware or otherwise unwanted software were documented by UNICRI. When analysing the origin of the malware and PUPs, attempts were made to identify the family and campaign and classify whether its purpose

[14] Mulvehill, T., 'The risk from mobile malware is real — and growing', *Security Intelligence*, 25 April 2016; retrieved from https://securityintelligence.com/the-risk-from-mobile-malware-is-real-and-growing/.
[15] Shilko, J., 'Fraudster phishing users with malicious mobile apps', *The PhishLabs Blog*, 25 April 2016; retrieved from https://info.phishlabs.com/blog/fraudster-phishing-users-with-malicious-mobile-apps.

was to disrupt users' systems upon downloading suspected copyright-infringing content or to gather user data for the purpose of obtaining personal information.

Europol Malware Analysis Solution (EMAS)

The Europol Malware Analysis Solution (EMAS) is a dynamic, automated malware analysis solution provided by Europol to EU Member States. EMAS offers the possibility of creating analysis reports, but its most revolutionary feature is to produce intelligence for police investigators. Automated cross-checks can show links between attacks performed in different countries with the same malware, or with the same criminal organisation behind the same malware family, connecting to the same domains and related to different investigations within or outside the EU. In 2015, EMAS became fully automated to allow direct access to law enforcement parties with which Europol has operational agreements. In 2015, 525 108 files were analysed by EMAS, out of which 356 863 were identified as malicious.

Overview of the Methodology

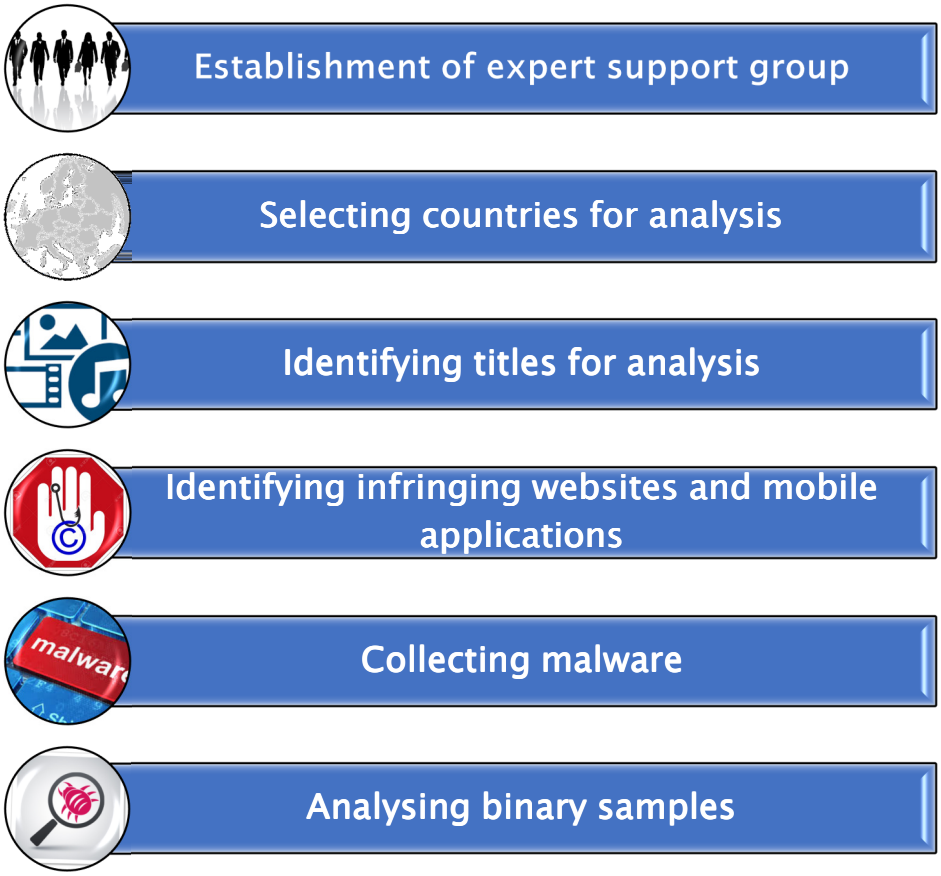


Figure 1: Overview of the methodology

4. Literature Review

4.1 Suspected copyright-infringing websites and mobile applications

4.1.1 Trends in use of suspected copyright-infringing websites and mobile applications

Sharing, downloading and streaming suspected copyright-infringing content from the internet is a widespread practice across the European Union. The 2017 report, *European Citizens and Intellectual Property: Perception, Awareness, and Behaviour* by the EUIPO[16] confirmed the result found in 2013[17], which was that almost 1 in 10 (9 %) Europeans had illegally downloaded or streamed copyright-protected material in the past 12 months. Only 8 % of the Europeans surveyed declare they would not necessarily go for the legal option even if it were available and an affordable option. Altogether, 13 % of respondents aged 15 to 24 would still use the illegal option. However, this proportion has dropped by 4 points compared with 2013, while decreasing by 2 points overall. This decline is consistent with the increase in the use of lawful services by the youngest Europeans. To explain this attitude, 'when asked who benefits most from the protection of IP, only around 5 % of Europeans mention 'consumers like themselves' and much more frequently mention large companies and successful artists as the primary beneficiaries of this set of rules and their enforcement'[18].

The Europol and EUIPO Situation Report from 2017 states that the most prevalent current threat stems from the online dissemination of protected content: 'well-known acts of piracy on the open internet include the sharing of protected content through BitTorrent networks, illegally facilitating downloading or streaming from central sources and (under certain conditions) illegally making links to IPR-protected content freely available without rights holder consent' [19]. As a breach of the copyright holder's intellectual property rights, illegal sharing, downloading and streaming of copyrighted material may cause financial losses to the author, publisher, and disseminators of the copyright-protected work. Many of these websites and networks for sharing suspected copyright-infringing content may also pose a threat to their users in the form of malware or malvertising[20].

A 2016 EUIPO study on online business models infringing intellectual property rights[21] describes one IPR-infringing activity as consisting of digital content sharing and malware dissemination through several types of online digital platforms such as adversary websites, third-party marketplaces, social media, gaming, emails and mobile platforms. The report also highlights how relevant business models

[16] EUIPO, *European citizens and intellectual property: perception, awareness, and behaviour*, EUIPO, Alicante, 2017; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/IPContributionStudy/2017/european_public_opinion_study_web.pdf.

[17] EUIPO, *European citizens and intellectual property: perception, awareness, and behaviour*, Office for Harmonization in the Internal Market, Alicante, 2013; retrieved from https://euipo.europa.eu/ohimportal/documents/11370/80606/IP+perception+study.

[18] EUIPO, *European citizens and intellectual property: perception, awareness, and behaviour*, EUIPO, Alicante, 2017; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/IPContributionStudy/2017/european_public_opinion_study_web.pdf.

[19] Europol and EUIPO, *2017 Situation Report on Counterfeiting and Piracy in the European Union,* 2017; retrieved from https://www.europol.europa.eu/sites/default/files/documents/counterfeiting_and_piracy_in_the_european_union.pdf.

[20] As referred to in the 2017 Situation Report, 'Altogether, 51 % of advertising on 280 suspected piracy websites available in Europe in the summer of 2015 were connected to various types of malware (malvertising)'. Europol and EUIPO, *2017 Situation Report on Counterfeiting and Piracy in the European Union*, 2017; retrieved from https://www.europol.europa.eu/sites/default/files/documents/counterfeiting_and_piracy_in_the_european_union.pdf.

[21] EUIPO, *Research on Online Business Models Infringing Intellectual Property Rights*, EUIPO, Alicante, 2016; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/resources/Research_on_Online_Business_Models_IBM/Research_on_Online_Business_Models_IBM_en.pdf.

can be identified. The 2017 Europol and EUIPO Situation Report[22] on counterfeiting in the EU, mentioned above, mentions that malware can be also embedded in website content itself, mostly in the form of executable files containing software and digital game downloads.

4.1.2 Types of suspected copyright-infringing websites and mobile applications

Suspected copyright-infringing websites and mobile applications offer access to suspected copyright-infringing content, such as films, television titles, music, and video games, using a variety of means. Operators of such websites generally implement a chain of links that leads from the original website to a specific, more hidden content-sharing service. This obscures access to content and allows the website or application to disseminate online advertisements to its users for financial gain. The main goal of these services is financial profit or gaining access to users' sensitive data through malicious or otherwise unwanted software. The existing literature recognises two major taxonomies that are used to categorise suspected copyright-infringing websites and mobile applications.

The first designates five categories based on content that is shared or hosted by the website[23] or mobile application[24]:

1. video (films, television programmes, etc.),
2. music (songs, music albums, etc.),
3. games ('cracked' versions of CD/DVD-based games or downloadable games),
4. mobile applications: specific software that contains content or provides an access to specific resources on the internet, and
5. software: programs that are used on personal computers or other hardware.

The second way of categorising copyright-infringing websites and mobile applications is based on how the content is distributed to or accessed by users. In this regard, there are five major methods used to distribute or allow access to suspected copyright-infringing content[25]:

1. BitTorrent trackers or torrent trackers: peer-to-peer file sharing that uses decentralised architecture to distribute content of any size. The websites usually offer customisable searches and extensive descriptions of the content being shared. The content itself is stored on users' computers and then shared through downloading specific torrent files or using magnet links[26].

2. Linking websites: web services that aggregate information about content and provide corresponding links to other websites that distribute the content. Linking websites do not usually store any copyright-infringing content themselves; instead, they provide descriptions and illustrations of content, along with links to other sites that host the content.

[22] Europol and EUIPO, *2017 Situation Report on Counterfeiting and Piracy in the European Union*, 2017; retrieved from https://www.europol.europa.eu/sites/default/files/documents/counterfeiting_and_piracy_in_the_european_union.pdf.
[23] The Statistics Portal, 'Online and mobile content which internet users paid for in the past month as at 4th quarter 2014', 2014; retrieved from https://www.statista.com/statistics/388215/paid-online-mobile-content/.
[24] Mullan, E., 'What is Digital Content?', *EContent*, 29 December 2011; retrieved from http://www.econtentmag.com/Articles/Resources/Defining-EContent/What-is-Digital-Content-79501.htm.
[25] EUROPOL and EUIPO, *2017 Situation Report on Counterfeiting and Piracy in the European Union*, EUIPO, Alicante, 2017; retrieved from https://www.europol.europa.eu/sites/default/files/documents/counterfeiting_and_piracy_in_the_european_union.pdf.
[26] A magnet link is a type of hyperlink that enables downloading of files and data from P2P sharing networks, particularly torrent networks. It works in a serverless environment and contains all the information a torrent client requires to download a specific file. The magnet link was designed to replace and upgrade torrent file specifications. Retrieved from https://www.techopedia.com/definition/28464/magnet-link.

3. Direct website hosting: the copyright-protected content is stored on a central server(s) associated with the website and can be directly accessed or downloaded by users. After the content has been downloaded, users can save the copyright-protected content to their device (e.g. a personal computer or mobile phone).

4. Streaming services: centralised server architectures that offer access to video or music though web browsers or specific software. The content is transmitted or 'streamed' to the user's device but cannot be saved to or stored on the user's device.

5. Cyberlockers or shared file hosting: allow users to upload files of almost any size to a server. Users can access the uploaded files at a later date and can share links to the uploaded files with other individuals.

4.1.3 Profit models of copyright-infringing websites and mobile applications

As mentioned previously, a goal of many suspected copyright-infringing websites and mobile applications is to generate income. This may seem counterintuitive given that many such websites and applications offer access to suspected copyright-infringing content free of charge. However, there are a variety of other means by which they are able to solicit financial income[27].

The first and most common method is through digital display advertising. Suspected copyright-infringing websites and mobile applications often dedicate substantial space to advertisements. To fill the advertising space, it is common to partner with intermediary advertising networks, which use automated processes to fill available advertising space with advertisements that may be appealing to users of the website or application. The website or mobile application generates revenue every time a user clicks on one of the advertisements[28]. This often leads to aggressive advertising practices. A report prepared by WhiteBullet Solutions Ltd for the EUIPO detailed the common types of advertisements on suspected copyright-infringing websites. They included pop-up advertisements, pop-under advertisements, mid-page units, skyscrapers (vertical advertisements along the sides of a web page), and banners (horizontal advertisements at the top or bottom of a web page)[29]. When aggressive advertising coincides with advertising content that is actually malware, users may become victims. The advertisement appears, for example, after attempting to download a suspected copyright-protected video.

The second income generating model is through distributing PUPs, programs that employ duplicitous methods to trick users into downloading them. These programs advertise themselves as useful software, such as anti-virus programs, operating system cleaners, video players, or download helpers. Instead, many of them are adware or more malicious forms of malware. According to a Digital Citizens Alliance report from 2015, over half (55 %) of malware infections could be traced back to user-initiated downloads[30]. Most of these incidents were comprised of cases in which the users were lured by the 'usefulness' of the software being offered for download.

[27] INCOPRO, *The Revenue Sources for Websites Making Available Copyright Content without Consent in the EU*, INCOPRO, London, 2015; retrieved from http://www.incoproip.com/resources-news-events/case-studies-reports/.
[28] WhiteBullet Solutions Ltd, *Digital Advertising on Suspected Infringing Websites*. European Observatory on Infringements of Intellectual Property Rights, Alicante, 2016; retrieved from https://euipo.europa.eu/ohimportal/documents/11370/80606/Digital+Advertising+on+Suspected+Infringing+Websites.
[29] WhiteBullet Solutions Ltd, *Digital Advertising on Suspected Infringing Websites*, European Observatory on Infringements of Intellectual Property Rights, Alicante, 2016; retrieved from https://euipo.europa.eu/ohimportal/documents/11370/80606/Digital+Advertising+on+Suspected+Infringing+Websites.
[30] 'How content theft sites and malware are exploited by cybercriminals to hack into internet users' computers and personal data', *Digital Bait*, Digital Citizens Alliance and RiskIQ, December 2015; retrieved from http://www.digitalcitizensalliance.org/clientuploads/directory/Reports/digitalbait.pdf.

The third income generating model is by providing users with the option of 'premium access'. These so-called privileged subscriptions offer users faster downloading speeds and allow for downloading numerous files concurrently, among other extras. Users must often create an account using personal data or buy a time-limited package.

4.1.4 Threats to users

At present, suspected copyright-infringing websites and streaming services are not normally considered to be dominant sources of malware or otherwise unwanted software distribution. However, considering the increasing popularity of streaming services, increased bandwidth of broadband networks, and the deployment of 4G networks, it cannot be ruled out that they may pose a growing risk moving forward. Given the potential threat to users and the increasingly digitally connected world, it is important to deepen our understanding of the threat of malicious and otherwise unwanted software found on suspected copyright-infringing websites and mobile applications.

4.2 Types of malware

Malware is an umbrella term for 'malicious or otherwise unwanted software' that is designed to carry out unauthorised actions on a computer system. While malware can take many forms, it is noteworthy that a considerable amount of malware requires users' willingness to install files that they receive from the internet or other sources. The exception is drive-by downloads, which do not require any interaction on the part of users to be downloaded[31].The major, recognisable forms of malware include viruses, worms, Trojans and ransomware, although, as will be discussed, there are different taxonomies of malware used by the specialists. As malware increases in sophistication and impact, it becomes ever more important to study and understand its distribution methods in order to ensure readiness and reinforce protective mechanisms. Significantly, the European Union Agency for Network and Information Security (ENISA) named malware as the top cyberthreat of 2016[32].

The specific attack vector and exploits used in malware depend highly on the market share of the operating systems installed on personal computers, servers, tablets, mobile devices, and other electronic devices. Given that the goal of malware is often financial gain, it makes financial sense to create or spread malware that is tailored to today's computing environments. Attackers also consider known vulnerabilities that can be exploited for different operating systems. As an example, almost half (49 %) of Microsoft Windows users still use Windows, an operating system released in 2009, despite the availability of newer, more secure operating systems. The recent WannaCry ransomware specifically targeted vulnerabilities present in Windows 7; nearly 98 % of the computers infected by the recent WannaCry ransomware outbreak were running the operating system[33].

4.2.1 Malware threats in 2017

There are no official worldwide estimates on the prevalence of malware[34]. A report by Malwarebytes found that there were one billion instances of malware detection during 2016[35]. Symantec reports that

[31] 'How content theft sites and malware are exploited by cybercriminals to hack into internet users' computers and personal data', *Digital Bait*, Digital Citizens Alliance and RiskIQ, December 2015; retrieved from http://www.digitalcitizensalliance.org/clientuploads/directory/Reports/digitalbait.pdf.

[32] European Union Agency for Network and Information Security, *ENISA Threat Landscape Report 2016: 15 top cyber-threats and trends*, Heraklion, ENISA, 2017; retrieved from https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2016.

[33] Brandom, R., 'Almost All WannaCry Victims Were Running Windows 7', *The Verge*, 19 May 2017; retrieved from https://www.theverge.com/2017/5/19/15665488/wannacry-windows-7-version-xp-patched-victim-statistics.

[34] For the purpose of the study, examples of data and estimates from relevant private sector actors will be given and sources quoted to check the methodology and scope of the report mentioned. The examples are by no means meant to be exhaustive nor comprehensive.

malware is becoming more complex and exploiting more zero-day vulnerabilities, defined as vulnerabilities that take advantage of a newly-discovered and unpatched security flaw[36]. The IT security company found 357 million malware samples in 2016 alone.

One of the major threats recognised by most studies is that of ransomware. The abovementioned Symantec report found a noticeable increase in ransomware detections in 2016, from approximately 26 000 to 48 000 per month. In addition, the average ransom demanded from users increased from USD 294 in 2015 to USD 1 077 in 2016[37]. McAfee reported similar increases in the prevalence of ransomware: the number of ransomware incidents grew from nearly four million during the third quarter of 2015 to nearly nine million in the third quarter of 2016[38]. Malwarebytes identified two reasons why ransomware is such a frequently used type of malware: it results in faster and more reliable financial gain and it is frequently offered in the form of ransomware-as-a-service, to allow people with little technical skill to attack with relative ease[39].

A second major threat is the increasing amount of malware targeting mobile devices. In its annual report, Europol assessed that mobile malware was increasing in number and in sophistication[40]. Mobile malware now includes many of the same types as its computer-based counterparts, including ransomware, Trojans, and drive-by-downloads[41]. Over half of the mobile malware samples analysed by Symantec in 2016 targeted the Android OS[42].

A third major threat is the increasing prevalence of malware targeting the Mac OS[43]. McAfee Labs reported a significant increase in the presence of Mac OS malware over the second half of 2016, from nearly 50 000 new malware samples in the third quarter to 325 000 malware samples in the fourth quarter of 2016[44].

4.2.2 Taxonomies for categorising malware

There is no unified standard for categorising and naming malware; as a result, a single example of malware may have more than one identifier. ‘Categorising’ means assigning one or more category or subcategories (usually known as types or families) to a malware program, while ‘naming’ describes the manner by which this malware is named (descriptive machine name) by anti-virus vendors.

[35] Malwarebytes, *State of Malware Report 2017*, Malwarebytes Labs, Santa Clara, CA, 2017; retrieved from https://www.malwarebytes.com/pdf/white-papers/stateofmalware.pdf.
[36] Symantec, ‘Internet Security Threat Report: 2017’, *ISTR*, vol. 22, April 2017, Symantec, Mountain View, CA, 2017; retrieved from https://www.symantec.com/security-center/threat-report.
[37] Symantec, ‘Internet Security Threat Report: 2017’, *ISTR*, vol. 22, April 2017, Symantec, Mountain View, CA, 2017; retrieved from https://www.symantec.com/security-center/threat-report.
[38] Sophos, *Looking Ahead: SophosLabs 2017 malware forecast*, Sophos, Abingdon, 2017; retrieved from https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-2017-malware-forecast-report.pdf?la=en.
[39] Malwarebytes, *State of Malware Report 2017*, Malwarebytes Labs, Santa Clara, CA, 2017; retrieved from https://www.malwarebytes.com/pdf/white-papers/stateofmalware.pdf.
[40] Europol, *Internet Organised Crime Threat Assessment (IOCTA) 2016*, Europol, The Hague, 2016; retrieved from https://www.europol.europa.eu/sites/default/files/documents/europol_iocta_web_2016.pdf.
[41] Europol, *Internet Organised Crime Threat Assessment (IOCTA) 2016*, Europol, The Hague, 2016; retrieved from https://www.europol.europa.eu/sites/default/files/documents/europol_iocta_web_2016.pdf.
[42] Symantec, ‘Internet Security Threat Report: 2017’, *ISTR*, vol. 22, April 2017, Symantec, Mountain View, CA, 2017; retrieved from https://www.symantec.com/security-center/threat-report.
[43] Sophos, *Looking Ahead: SophosLabs 2017 malware forecast*, Sophos, Abingdon, 2017; retrieved from https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-2017-malware-forecast-report.pdf?la=en.
[44] McAfee Labs, *2017 Threats Predictions*, McAfee, Santa Clara, CA, 2016; retrieved from https://www.mcafee.com/us/resources/reports/rp-threats-predictions-2017.pdf.

The following section details the most common approaches to categorising malware or suspected malware.

(1) The broadest method for categorising suspected malware considers whether or not the suspected malware is harmful. Suspected malware is categorised as clean (meaning it poses no threat to the user and performs benign functions) harmful, or a PUP[45]. PUPs are those that are not necessarily malicious but may be unwanted, depending on the user or environment.

(2) A second common taxonomy classifies malware according to its functional classification, dissemination methods, and behavioural aspects[46]. This is the approach most commonly used by anti-virus vendors, researchers, and malware labs[47]. For example, Microsoft lists 34 types of malware[48].

(3) A third common taxonomy classifies malware into 'families'. Members of a malware family may share a functionality, exploitation, or code. The names of some malware families may be recognisable to those familiar with widespread cyberattacks in recent years, for example, WannaCry and Petya[49]. A 2016 study reported 10 362 malware families in the first 10 archives of VirusShare[50], an electronic repository of malware samples[51]. Conversely, Malware Wiki, one of the most comprehensive collections of malware types and families, includes approximately only 3 000 malware families[52].

(4) A fourth common taxonomy is used by anti-virus vendors to indicate the danger posed by a particular piece of software. For example, Microsoft Security Essentials differentiates between the following security levels: severe, high, medium and low.

(5) A fifth common taxonomy is based on forensic traces that malware leaves in the system[53]. Anti-virus software uses these traces to detect the presence of malware. There are three main types of system artefacts: disc-, network- and memory-related.

(6) There are also subtypes of the virus category of malware. Hardikar (2008) suggested the following categorisation of viruses based on their functionality aspects: memory-based, payload-based, obfuscation-technique-based and target-based[54].

5.2.1 Malware and potentially unwanted program infection and dissemination on suspected copyright-infringing websites

Some suspected copyright-infringing websites offer PUPs when a user follows the 'final' link to download or watch a video. Instead of content, however, the user is supplied with a small executable file that is often advertised as a tool to simplify the process of watching the video or some other

[45] F-Secure, 'Classification: Categories'; retrieved from https://www.f-secure.com/en/web/labs_global/classification
[46] Kane, C., *Malware Taxonomy and Terminology*, University of Cincinnati, Cincinnati, 2017; retrieved from http://class.malware.re/lecture-slides/lecture-w03-1.pdf.
[47] Erdélyi, G., *Malware Taxonomy*, F-Secure, Helsinki, 2010; retrieved from http://www.cse.tkk.fi/fi/opinnot/T-110.6220/2010_Spring_Malware_Analysis_and_Antivirus_Tchnologies/luennot-files/Erdelyi-Introduction_to.pdf.
[48] Microsoft, 'Naming Malware'; retrieved from https://www.microsoft.com/en-us/wdsi/help/malware-naming.
[49] McAfee Labs, 'Ransomware | McAfee', 2017; retrieved from https://www.mcafee.com/us/security-awareness/articles/ransomware.aspx.
[50] VirusShare, 'VirusShare.com | About'; retrieved from https://virusshare.com/about.4n6.
[51] Shalaginov, A., Grini, L.S, Franke, K., 'Understanding Neuro-Fuzzy on a Class of Multinomial Malware Detection Problems', *Proceedings of the International Joint Conference on Neural Networks 2016*, October 2016, IEEE.
[52] Malware Wiki, 'Category of Malware', 2017; retrieved from http://malware.wikia.com/wiki/Category:Category_of_Malware.
[53] Lee, A., Varadharajan, V., Tupakula, U., (2013). 'On Malware Characterization and Attack Classification', *Proceedings of the First Australasian Web Conference*, vol. 144, Australian Computer Society, Darlinghurst, 2013.
[54] Hardikar, A., *Malware 101 — viruses*, SANS Institute, 2008; retrieved from http://amanhardikar.com/papers/malware101viruses.pdf.

misleading process. RiskIQ[55] studied how malware is delivered to the end-user. Among the malware studied, 45 % of files were downloaded as a background process via so-called drive-by-downloads, while 55 % of malware ended up on the user's computer by luring him or her into acquiring some kind of 'useful' software such as Flash Player, anti-virus software, etc. The user interface design is often very similar to popular benign software, and might even simulate the activity of a real program. In addition, the report indicated that 54 % of malware being downloaded is Trojan, 29 % adware, 5 % toolbars for browsers and 3 % botnets (and 9 % others).

Check Point researchers[56] discovered a new way that malware has been distributed recently: attackers have crafted malicious video subtitle files that are available through subtitle repositories. Furthermore, users access those subtitle files using trusted platforms and several popular media players. Reportedly, 200 million users may have been affected by this type of malicious activity.

A common means of PUP dissemination is through malvertisements and pop-up messages that draw the user's attention by offering to install or to download some kind of popular and well-known software, such as Java, as described in a report by OPSWAT[57]. Once a user installs such an application, it might suggest or force the installation of many more of the same kind, thereby increasing the difficulty of removing the malicious applications.

Another model of malware distribution in relation to suspected copyright-infringing content is through additional software needed to play video, such as codecs. Ducklin (2016)[58] describes the way in which users are forced into receiving additional software when they try to watch videos acquired from The Pirate Bay. The link to the PUP was embedded in Windows Media Video.

In relation to software piracy, the National University of Singapore (NUS)[59], in a study commissioned by Microsoft, examined how malware affects pirated software on computers and non-genuine software available for downloading in Asia. In fact, 92 % of the computers studied had malware embedded, with 61 % of CDs and DVDs containing infected software. The NUS study revealed that the main target group of devices[60] susceptible to viruses and PUPs are those using the Microsoft Windows OS, which uses a portable executable 32-bit file format for.EXE and.DLL files. In this context, these make up the target binaries that may function in a variety of ways, from adware to Trojan droppers.

With regard to the business model, in the EUIPO's 2016 report[61], a business model of malware dissemination from a website making unauthorised use of trade marks was presented[62]. According to this model, in most cases, the digital identifier or domain name misuse, combined with phishing, malware dissemination, and fraud are distributed via a website controlled by the infringer. In addition to

[55] 'How content theft sites and malware are exploited by cybercriminals to hack into internet users' computers and personal data', *Digital Bait*, Digital Citizens Alliance and RiskIQ, December 2015;
retrieved from http://www.digitalcitizensalliance.org/clientuploads/directory/Reports/digitalbait.pdf.
[56]Check Point Research Team, 'Hacked in Translation — from Subtitles to Complete Takeover', 23 May 2017; retrieved from https://blog.checkpoint.com/2017/05/23/hacked-in-translation/.
[57] Matthews-Winn, S., 'Why Avoid Film Piracy? It's Illegal and Dangerous', 2 December 2014, *OPSWAT*; retrieved from https://www.opswat.com/blog/why-avoid-film-piracy-its-illegal-and-dangerous.
[58] Ducklin, P., 'Will a visit to The Pirate Bay end in malware?', 6 May 2016, *Naked Security by Sophos*, 2016; retrieved from https://nakedsecurity.sophos.com/2016/05/06/will-a-visit-to-the-pirate-bay-end-in-malware/.
[59] National University of Singapore, *Cybersecurity Risks from Non-Genuine Software*, Microsoft Corp., Singapore, 2017; retrieved from https://ncmedia.azureedge.net/ncmedia/2017/10/Whitepaper-Cybersecurity-Risks-from-Non-Genuine-Software.pdf.
[60]In the methodology of the NUS report, it is explained that 90 samples of personal computers and laptops from 8 countries from south east Asia were procured (Bangladesh, Indonesia, Malaysia, the Philippines, South Korea, Sri Lanka, Thailand, and Vietnam), together with 165 CDs and DVDs containing software.
[61] EUIPO, *Research on online business models infringing intellectual property rights,* EUIPO, Alicante, 2016; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/resources/Research_on_Online_Business_Models_IBM/Research_on_Online_Business_Models_IBM_en.pdf.
[62] Canvas No 19 of the EUIPO 2016 report.

this, in another identified model[63], an Android application offering access to pornographic content was identified as ransomware. When the app is launched, it stealthily captures the user's picture, locks the phone and demands a ransom to be paid for decryption. Highly deceptive models are applied to convince users to download or launch downloaded malware.

[63] Canvas No 18 of the EUIPO 2016 report.

5. Qualitative and Quantitative Analysis of Research Findings

5.1 Binary collection — Round I

5.1.1 Phase IV.A. Suspected copyright-infringing websites identified during Round I

Following the selection of the countries and the titles, the following phase, Phase IV, sought to select a sample of websites (A) and mobile applications (B) for the collection and analysis of malware samples. The selection of suspected copyright-infringing websites occurred over five steps using the respective country's Alexa Top 500 lists[64]. The Alexa Top 500 list compiles a monthly list of the most popular websites globally, regionally and by country. The collection of malware samples occurred in two rounds (weeks 26-29 and 30-32 of 2017)[65].

Table 2 reflects the number of website domains selected after performing each step in Phase IV.A. Step 2 includes cross-checking the country-specific lists of websites from the Alexa Top 500 against the regional EU list. During Step 3, website lists are checked against the Google Transparency Report to remove websites without reported copyright infringements. Step 4 shows the number of new domains that were not in the Alexa Top 500 list per country, yet were identified during searches for copyright-infringing content using search engines. The domains of Netflix, Facebook, and Twitter, inter alia, were excluded from the search results. Furthermore, all selected domains are also in the Google Transparency database of requests to delete the content. Step 5 shows the number of domains selected during week 26 to be examined later for possible malware species.

	Step 2	Step 3	Step 4	Step 5
Belgium	500	387	213	600
Bulgaria	500	316	117	433
Croatia	500	308	123	431
Czech Republic	500	318	204	522
Finland	500	323	213	536
France	500	397	253	650
Hungary	500	325	194	519

[64] The details of the methodology are contained in Annex 1 and a more extended version of the Qualitative and Quantitative Analysis of Research Findings in Annex 2. During analysis of the selected suspected copyright-infringing websites, the IP address for each corresponding domain name was retrieved. However, further information from the WHOIS record of the DNS servers or hosting providers was not explored or checked.

[65] Regarding the timeframe, the following sequence should be considered: week 26 — website identification, week 27 — manual & automated malware collection; week 28 — malware analysis, week 30 — website identification, week 31 — manual & automated malware collection, week 32 — malware analysis. Corresponding dates: 3 -21 July, first round of research; 22-30 July break; 31 July-18 August, second round of research; 19-31 August, analysis of the reports obtained from EMAS.

	Step 2	Step 3	Step 4	Step 5
Lithuania	500	318	209	527
Portugal	500	385	212	597
Sweden	500	336	219	555

Table 2: Identification of the selected websites for each step in Phase IV.A (Round I)

The figures below present the distribution of the countries of websites hosting locations and domain-name suffixes. To find out the country of web hosting location, each domain name was resolved to retrieve the corresponding IPv4 address of the web hosting. Then, each address was checked against the GeoIP database provided by Maxmind[66]. It can be seen that the overwhelming majority of the websites are hosted in the United States. There is also a significant prevalence of .com domains with almost no websites having country-specific domain suffixes, such as .bg or .be. This might be explained by the fact that .com is a commercial domain that had been open for general public registration. In this way, virtually everybody can register and use it. However, country-specific top-level domains might require residence, a registered trade mark or an organisation in the country.

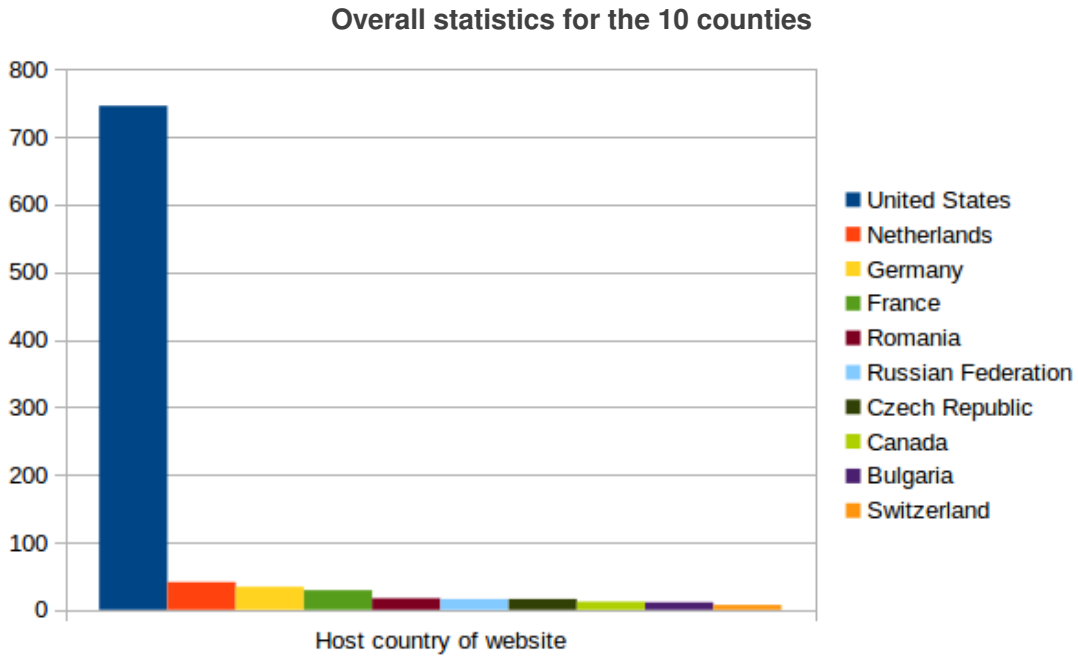


Figure 2: Distribution of host countries of websites added during Round I of malware collection for all countries

[66] Maxmind, 'GeoIP products'; retrieved from https://dev.maxmind.com/geoip/.

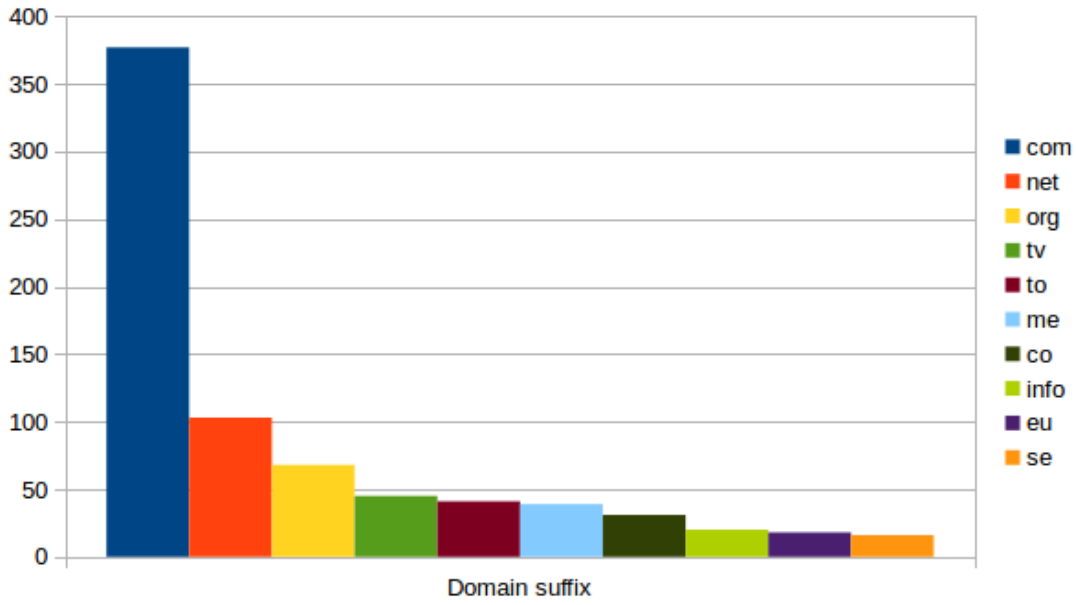


Figure 3: Distribution of domain suffixes of websites identified in Round I of data collection for all countries

5.1.2 Phase IV.B. Mobile application identification

After performing analysis of some of the mobile applications retrieved, it was found that they use similar dissemination methods — that is, third-party application markets. Android OS has a default application market — Google Play — that is a safe and trusted channel for mobile application distribution. All applications are verified before being published there. So-called third-party application markets deceive users by advertising access to popular applications that can also be found on the official Google Play store. Examples of such third-party Android application markets are presented below.

Website01

This Vietnamese web service offers access to different content that can be downloaded by users, including games and other applications. It offers applications for iOS and Windows Phone, in addition to Android. The interface of the website looks similar to application markets, ranks different software and offers free access. In order to do so, a user is required to retrieve a Website01 application that serves as a gateway to other applications listed on the website. The application requires a wide range of permissions that may appear suspicious. For example, it can manage user accounts, access internet, modify and delete files on an SD card, read phone identities and access phone logs. Such an extensive list of permissions can be unusual and not related to an application's advertised functionality.

Website01 is not an official marketplace and is not represented on Google Play. Therefore, it is logical that the application requires such permissions to be able to handle installation. However, such information is at risk of being exposed to a third party because the user freely gives all these permissions. The Android installation package from this application market has been identified by eight anti-virus solutions as suspicious or as a Trojan. In this sense, there is a high level of confidence that the file is unwanted and may cause harm to the user's personal and sensitive information. Moreover, the installation file of the application is available for downloading without any registration. To install it, however, a user must enable 'third-party source' in the Android settings.

Website02

This website offers access to popular applications that are also available through the official Google Play store. There are unique subdomains, which are designed specifically for the application on offer. However, there is a single application that is being downloaded across all of the subdomains. Upon downloading this app, the user is asked to enable the option 'Unknown sources' in the security settings of the Android OS. Corresponding instructions are clearly stated on the website that offer free access to the Android installation file.

This is not normally required for applications accessed from the official market. The number of requested permissions is large and very suspicious. For example, the application that provides access to other applications does not usually need to read the phone's state or access contacts, the camera or call logs. Apparently, it was designed to collect such information that is further used by developers for their own purposes. The application has very few activities and services. Obviously, judging by the requested permission settings, one can conclude that sensitive user information will be at a high risk.

5.1.3 Phase V.A. Binary collection

In the binary collection phase, the intention was to retrieve any potentially malicious software and PUPs that average users may encounter when looking for ways to access content on suspected copyright-infringing websites. Considering the large number of domain names per country, it is difficult to anticipate the number of binary samples that will be collected and particular types of malicious or otherwise unwanted software. The suspected copyright-infringing websites investigated in this study usually create specifically crafted web pages to make consumers think that they are on a legitimate page. To ensure reasonable collection of relevant data, the binary collection was separated into two stages:

Manual collection. This approach includes manual surfing of the internet while looking for content with selected titles in search results. This allows the researcher to mimic the average user experience and select links that are either malvertising or require direct human interaction to access. Moreover, this enables the researcher to take screenshots of the information displayed and analyse the relevance of the website content.

Automated collection. Contrary to the manual approach, automated malware collection is fast and does not require user interaction to browse and follow links on a web page. The first step of the collection employs an intelligent crawler that follows any available links on the suspected copyright-infringing websites. This is so-called breadth-first search, where the links are checked first on a target web page, then links on the secondary web pages that this web page refers to, etc. The second step of the collection retrieves files that might be of interest for later manual analysis. The number of links checked per website was limited to 1 000 to ensure timely execution. In addition, a crawler does not process JavaScript because this can only occur on a fully operational user's browser. To mitigate this challenge, manual collection was performed.

As at 28 July 2017, 5 240 websites had been automatically checked, with 617 relevant files retrieved of an overall size of 47 GB. This unsorted batch of files requires further analysis to decide which collected files are relevant for the study.

5.1.4 Phase V.B. Binary analysis

During this phase, the files that were collected on suspected copyright-infringing websites, according to the corresponding country, were comprehensively checked to see if they contained any malicious payload or caused any harm to the user's sensitive information or computer.

Website03

This website offers access to cracked games, ISO files, and other relevant data that users may look for on the internet. Games are sorted according to genres and users can also use a website search to find a particular game. Each web page provides multiple links to the game's files that the end-user needs to download. In most cases, the files that are being downloaded are of a small size (only a few megabytes), named as the original game and compressed using the zip archive format. Preliminary analysis indicates that they are PUPs/adware. The user experience starts by opening a web page designed for a particular game, which includes a relevant description and information.

By hitting the corresponding download button, the user receives a small zip file that contains an executable application. Even though the system warns of an unknown publisher and potentially harmful software, this does not stop users from executing the file. The graphical user interface of the installation software looks like a legitimate piece of software with a title and picture of the corresponding game. The only suspicious element at this stage is an icon of the software that might look rather general and not related specifically to this game.

Furthermore, the software shows an end-user licence agreement (EULA) that may appear legitimate to an unsuspecting user who does not read the terms that are mentioned. In addition, the user is required to confirm that he or she accepts the terms of the licence.

The file installation process is well tuned and shows not only the files being downloaded, but also the status of the whole process. It is worth mentioning that there was no high-bandwidth network traffic activity from the application registered during this installation process. It also takes some time until the software moves to the next step, which tricks the user into thinking that the installation process is actually taking place.

Upon completion of the installation, the software requests a licence key, which is a logical step. The licence key is usually purchased together with the game and is required in order to be able to use this software product. It can be seen that the interface offers both the option of downloading or confirming the key and an explanation of how to obtain the key. Clicking the licence key 'Download now' button opens an overlay window with links to a survey. It is claimed that the file will be available once the user chooses to complete one of the surveys. This appears to be a normal procedure, similar to standard file-sharing services.

Website04

This website offers streaming of different TV series from a variety of streaming web services. The interface is in German. Links to various streaming services are provided. The page that contains, for example, the link to 'Game of Thrones' has an interface that includes a description of the episode, pictures and streaming links. While accessing FlashX, the player window displays an overlay button reading 'Download Now' that further leads to an external web page.

Website05

This website has a very similar interface to a Microsoft web page, including the location of the elements, the text and many of the colours. This may trick users to think that they are on a legitimate web page and that the software is benign. Furthermore, advertised software on the website offers a functionality to fix the common problem of unknown drivers. This problem appears when a user uses hardware for which the drivers are not found in MS Windows. Therefore, file driver-updater-setup.exe is downloaded after pressing the 'Download' button. The installation process appears realistic and displays a logo that fraudulently reads 'Microsoft Partner'.

After the installation is complete, the Website05 Driver Updater starts downloading information from the internet and scans drivers that are installed in the system. Once the diagnostics are complete, the user is alerted to the fact that there are some outdated drivers, for which 'useful' software is offered that purports to improve the system's performance and resolve any outstanding problems.

After claims about updating corresponding drivers and installing the recommended software to improve system stability, the process completes the installation of several new programs with apparently legitimate names and a credible design that looks like professional software that can be trusted. Therefore, by installing Driver Updater, the user is tricked into accepting three additional pieces of software of questionable usefulness.

Website06

This website provides access to content that is described as user-posted. Therefore, the website's disclaimer states that it has no responsibility for such materials. Each game's web page has a common set of features: pictures, description, video preview, and a download button.

It also appears that these small files are crafted to look like a game installer by having the corresponding name, a built-in picture of the game, and relevant information. Once the .EXE file has been extracted, the user is prompted to choose the language for the game that is being installed. The next four steps are consistent with a typical installation procedure: welcome message, folder selection, link creation, and installation. The program appears to be professionally designed and looks like realistic game installation software because it uses the picture of a game, its name and its relevant description. For unexperienced users, all software dialogue windows may look trustworthy. It basically simulates the downloading process. However, no high-bandwidth activity was registered during the malware analysis.

What is interesting is that, during the installation, the software creates an ISO file of the game, which bears the corresponding name and has quite a large file size. Moreover, once the file is created, it has a size of only a few megabytes. During the Downloading process shown above, the size of this file increases until it reaches the size of the legitimate game. According to the article regarding the size of the DiRT 4 installation, the game occupies 32.26 GB on the disc. The file that has been created is 33.6 GB. This is very similar and indeed seems to be a true game, except for the fact that nothing has been downloaded. Finally, the simulation of the downloading process takes around 40 minutes, therefore it looks trustworthy to an average user. The actual time can vary based on environmental checks that the software may perform if considered appropriate by malware developers.

The web page enables users to download a licence key file with a size of 0.1 MB using the link provided. However, before this, a user has to go through a survey web page to unlock the file.

Finally, on the survey page, the user is required to provide a mobile phone number to receive an SMS and complete registration. It is claimed that this provides mobile virus protection. There is a phone number validity check that does not let the user go to the next step unless the phone number is correct.

Website07 — Mac OS

This online service offers a wide variety of TV series that can be watched and downloaded using the website. Each of the web pages represents a specific episode of the TV series being accessed. However, there are several links that simply do not work. In addition, the download button redirects users to a different advertisement web page that offers complimentary 'useful' software. Once the

'Download in HD' button has been pressed, one of the advertisements offers a Firefox plug-in to be added to the browser.

Another advertisement leads to the sounding web page that includes also the self-explainable name of the OS for which it is designed. It claims that the software can clean the Mac OS and has been downloaded by millions of users. It has an attractive interface and claims that it is the number one Mac utility in the world. Once the installation .pkg file has been downloaded and launched, the software proceeds through several steps of the installation process, including the general information regarding the software, EULA and privacy policy. A range of useful features can be viewed, such as Internet Security, Memory Cleaner, and Data Encryptor. Notably, these features look more like general names than specific applications or known trade marks. Further, the installation takes some time and performs additional suspicious activities.

Although the software offers 'useful' functionality, it has been known to trick users to buy the extended version that neither improves the performance of the system nor provides a full set of advertised functionalities. According to one article, there had been legal actions against a company because the program named problems on the computer that did not exist and generated false notifications that tricked users into buying the upgraded 'better' version of a certain software package.

Website08 — Android

This website offers a range of applications for Android free of charge. These can be downloaded easily using the links provided. This particular application offers access to a specific application that has streaming capabilities of recent TV series, films, etc. This amounts to copyright infringement since the application gives free access to copyright-protected content without the authorisation of the rights holder. The download link for the application is highlighted on the web page and unexperienced users may consider it to be a legitimate installation. However, the option 'unknown applications' in Android has to be enabled because the website is not an official Google Play store. Android provides additional security assurance to users by disabling this option.

For the analysis, Google SDK with Nexus and the API 18 emulator were used, running Android 4.3 using ARM. Of particular note is the fact that the application cannot be installed on the x86 platform and indicates an error in which a suitable ABI[67] has not been found. Once installed, the software presents a variety of categories of content that can be downloaded or streamed. The content is categorised according to TV series, episodes, etc. The user can also watch trailers and read different supplementary materials. For example, there is the option to download an episode of *Game of Thrones* directly to the phone in HD quality. The downloading process takes some time and the file is saved in the download folder on the Android device. Furthermore, the metadata of the files shows that it is an H264 compressed video file that is easily played on a computer and it has content being advertised in the application. It can be concluded that the software offers completely free access to copyright-infringing content without requesting any user-specific information.

[67] ABI — application binary interface; the Android application checks the type of hardware being used. For example, it can be configured to target only specific models of smartphones or tablets from a particular manufacturer.

5.2 Binary collection — Round II

The second round of the website identification and binary collection was conducted shortly after the first round. Nonetheless, there are differences in the websites identified for use in data collection between the two rounds. This is attributable to updates to the Google Transparency Report, changes in the search engine results, websites shutting down, and the introduction of new malware campaigns, among other factors.

The following sections provide detailed information on:

1. websites identified during the second round of data collection,
2. a comparison of the results of website identification between the first and second round of data collection,
3. the malware collection process during the second round of data collection, and
4. analysis of malware identified during the second round of data collection.

5.2.2 Phase IV.A. Suspected copyright-infringing websites identified during Round II

The second round of website identification was conducted in the same manner as the first round, and as described in Phase IV.A of the methodology[68].

Table 3 reflects the number of website domains selected after performing each step of Phase IV.A during the second round of website identification. Step 2 includes cross-checking country-specific lists of websites from the Alexa Top 500 against the regional EU list. During Step 3, website lists are checked against the Google Transparency Report to remove websites without reported copyright infringement. Step 4 shows the number of new domains that were not in the Alexa Top 500 list per country, but were identified during searches for copyright-infringing content using search engines. Domain names that belong to well-known companies, such as Netflix, Facebook, Twitter, and others were excluded from the search results. Moreover, selected domain names also came from the Google Transparency database of requests for deletion from the Google search engine results in relation to copyright-protected content. Step 5 shows the number of domains selected during weeks 30-31 to be examined later for possible malware. Overall, this process resulted in a total of 5 606 websites, including 1 057 unique websites for all 10 countries.

	Step 2	Step 3	Step 4	Step 5
Belgium	500	389	+219	608
Bulgaria	500	318	+142	460
Croatia	500	310	+137	447
Czech Republic	500	320	+239	559
Finland	500	325	+229	554
France	500	400	+265	665
Hungary	500	327	+246	573
Lithuania	500	320	+241	561
Portugal	500	387	+231	618
Sweden	500	338	+223	561

Table 3: Identification of selected websites for each step of Phase IV.A (Round II)

[68] Details of the methodology are contained in Annex 1.

Two sets of figures are presented below showing the number of websites identified for each sample country during the second round of data collection. The first graph shows, for all sample countries, the distribution of websites based on the hosting country of the website. The second graph shows the distribution of websites for all sample countries by domain suffix.

Overall statistics for 10 counties

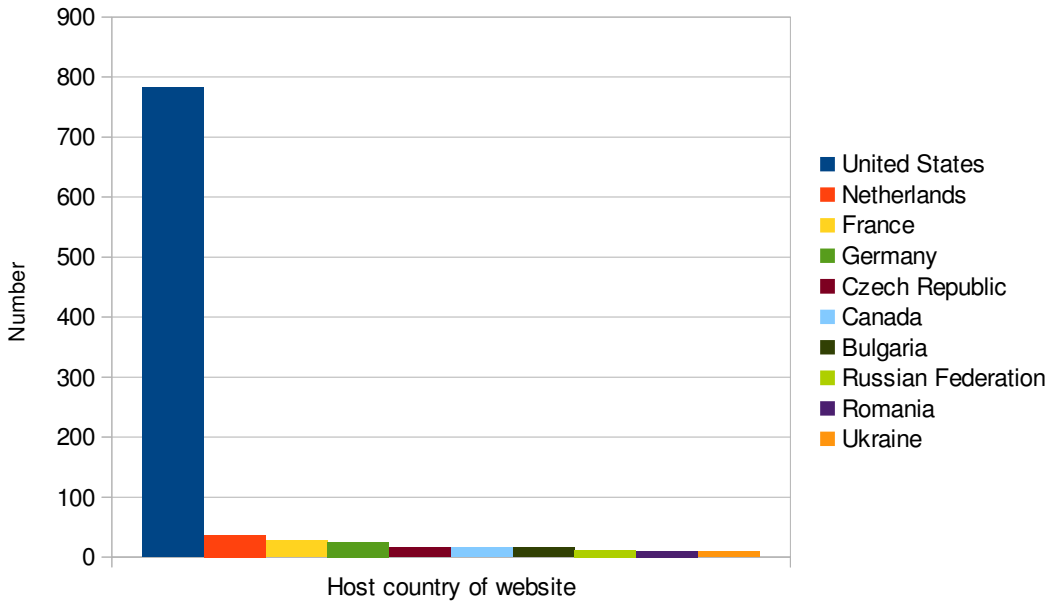


Figure 4: Distribution of host countries of websites identified in Round II of data collection for all 10 countries

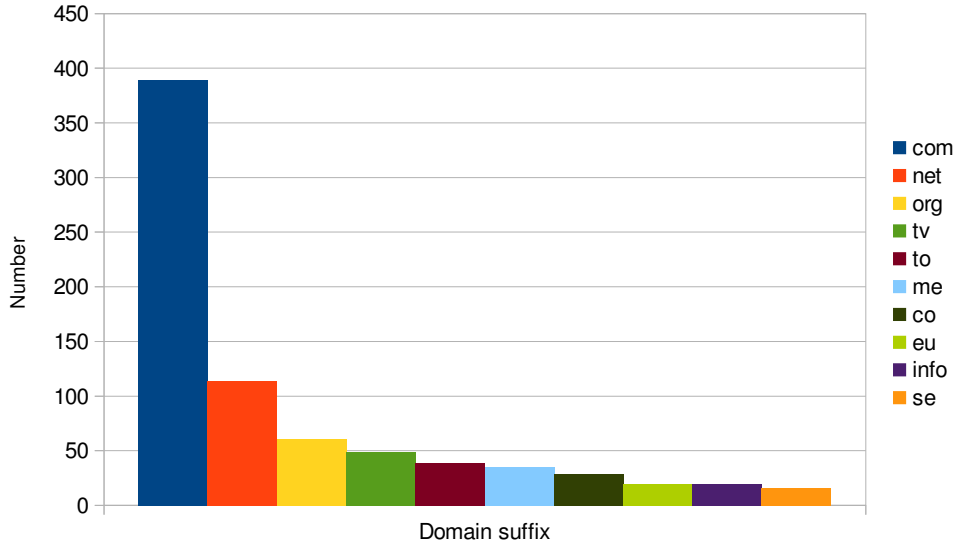


Figure 5: Distribution of domain suffixes of websites identified in Round II of data collection for all 10 countries

5.2.3 Phase IV.A. Comparison between two rounds of suspected website identification

A comparison between the two rounds of data collection is provided below. It is important to see how the malware and otherwise unwanted software distributed via the suspected copyright-infringing websites changed even over a short period of time. As the average user's awareness during this period changed as a result of information security incidents occurring around the world, malware evasion techniques also improved. In other words, attackers came to utilise new methods to infect users' computers and to obtain sensitive or private information from compromised systems.

Comparison of Google Transparency Report. The Google Transparency Report is frequently updated to reflect new websites that have been accused of distributing or hosting copyright-protected content and old websites that have been removed.

Table 4 compares information on the domains (website names submitted for removal by Google from search results), requests for removal from search engine results (specific URLs, reporting organisation, Lumen database ID), and the number of URLs with no action taken (the URLs that were not removed) between the Google Transparency Reports used in the first round of website identification (version 25 June 2017) and the second round of website identification (version 24 July 2017). There was a significant increase in the number of domains and requests between the first and second round of website identification. Considering the low cost of deploying websites and distributing copyright-infringing content, new websites can generally be expected to appear in the Google Transparency Report database. One of the most efficient ways to reduce harm to consumers from copyright-infringing websites is to remove their links from search engine results.

	25 June 2017	**24 July 2017**
Domains	170 118 970	174 049 634
URLs with no action taken	105 376 095	106 347 188

Table 4: Comparison of Google Transparency Report between the first and second round of identification

Comparison of number of websites identified during Round I and Round II. The five steps in Phase IV.A resulted in the identification of a large number of websites based on the Alexa Top 500 and systematic keyword searches on online search engines. There are considerable differences between the results of the two rounds of website identification. Some of the copyright-infringing websites were removed from the Google Transparency Report or no longer functioned, while new websites appeared, and others promoted their services to achieve a higher ranking in the search engine results. This illustrates how malware dissemination is a process that evolves over time. For each sample country, there was a net increase in the number of websites identified between the first and second round of website identification.

	Round I	**Round II**	**Added***	**Removed***
Belgium	600	608	116	108
Bulgaria	433	460	80	53
Croatia	431	447	74	58
Czech Republic	522	559	118	81
Finland	536	554	122	104
France	650	665	139	124

	Round I	Round II	Added*	Removed*
Hungary	519	573	127	73
Lithuania	527	561	127	93
Portugal	597	618	120	99
Sweden	555	561	116	110

Table 5: Comparative statistical information of differences between Round I and Round II of malware collection

Round II in comparison with Round I

Top 10 websites hosting countries and domain suffixes added during the second round of website identification in comparison with the first round of websites identification

Overall statistics for 10 countries

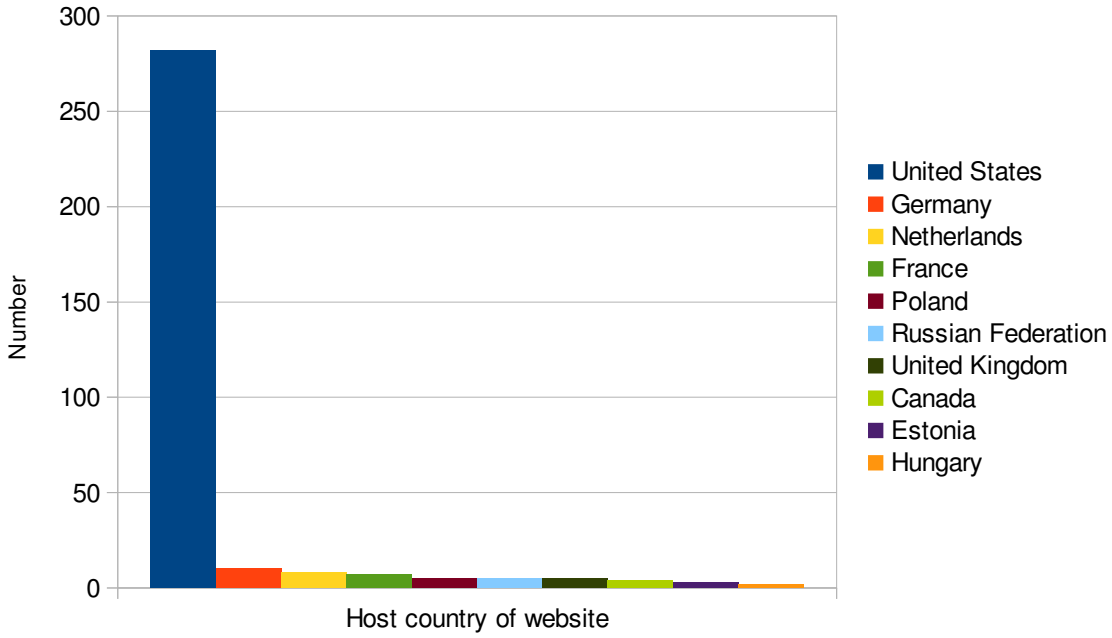


Figure 6: Distribution of host countries of websites added during Round II of malware collection for all countries

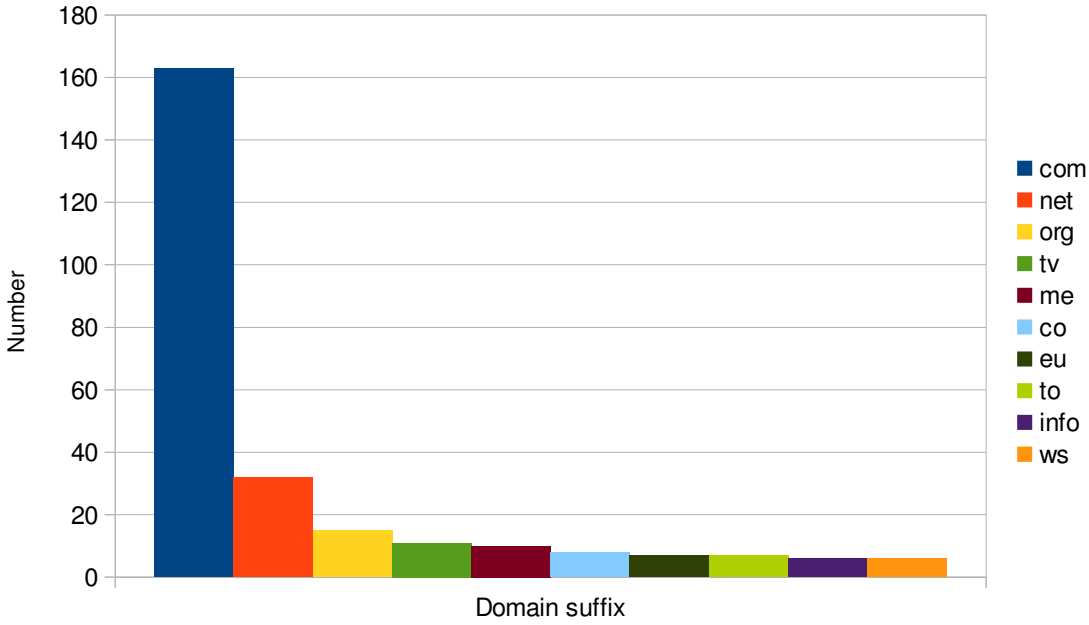


Figure 7: Distribution of domain suffixes of websites identified in Round II of data collection for all countries

5.2.4 Phase V.A. Binary collection

The binary collection that followed the second round of website identification was performed using two approaches: manual and automated. The manual approach simulated the experience of the average user, characterised by human-understandable reasoning in the identification of covert links and suspicious files on websites. Manual analysis is slow and only a fraction of all websites can be processed by such means. Automated analysis allows for all the URLs on a web page to be inspected much more quickly; however, the disadvantages of the automated approach are that it is not able to process JavaScript and it cannot bypass security mechanisms, such as Captcha or other similar tools.

The overall number of unique URLs extracted for all countries was 1 057 out of the 5 606 websites, which made it unfeasible to check all of them manually. The samples of copyright-infringing websites were similar across all 10 sample countries for each of the types of media (films, music, television programmes, and video games). As a result, Belgium was randomly selected from the sample countries, and all websites identified as copyright-infringing websites for Belgium were manually checked for the presence of malicious or otherwise unwanted software. A total of 3 665 files were automatically collected from the websites for all countries, with a total size of 167 GB.

5.2.5 Phase V.B. Binary analysis

A number of samples of malware or otherwise unwanted software identified during the second round of data collection are discussed below.

Website09 is a website that distributes a BitTorrent-based client, which offers access to various types of video content available through torrent trackers. The tool has a very simple graphical user interface and only a few settings accessible to the user. Upon selecting a film, the film is downloaded and played immediately. Website09 requires fewer user interactions compared with other BitTorrent trackers. Only a few clicks are required to download the content from unknown sources. This can include video files as well as other malicious payloads. In this way, the user is neither protected nor has control of what is being downloaded. Moreover, the software can be considered as a perfect distribution platform for

copyright-infringing content. The installation of the program is quick and no understanding of advanced settings is required.

Once installed and launched, the user has two options: he or she can select films or television programmes. The available titles are sorted by genre. IMDB sorting for films is also available. Additionally, the program features a reminder for users to employ VPNs to avoid any detection issues. Each film page on Website09 contains a general description of the film, its rating, and video quality options that need to be selected before watching the video. Once the user clicks the 'Watch Now' button, the file is downloaded to the user's computer via a BitTorrent protocol, with the corresponding information displayed on the screen, including the speed and status of the download, as well as the number of peers, seeds, and leechers[69]. After the download has completed, the requested film is available for viewing. The overall process of accessing copyright-infringing content is fast and can be carried out by almost anybody.

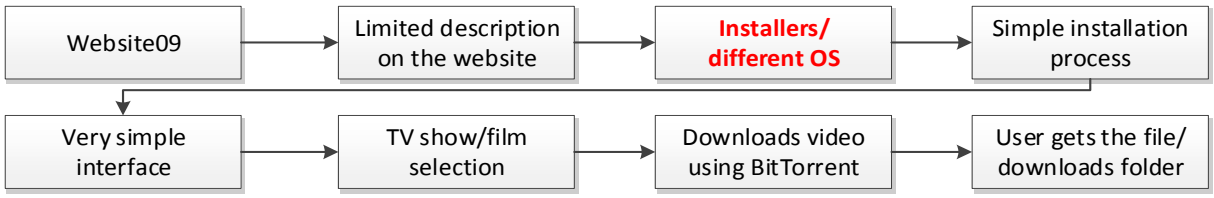


Figure 8: Distribution model of Website09 Software

Website03

Website03 was also discovered during the first round of malware collection and analysis. It is noticeable that the website owner(s) have considerably changed how malware is distributed. Previously, it was a tiny executable file for MS Windows that, after simulated installation, directed users to a website to receive a licence key for a game. Even though one of the downloading links on the web page reads 'Google Drive', the link itself points to another website, which is owned by the attacker. Moreover, torrent files can be found there with an 'ultra seed' option (more peers with better speed) and general torrent files with normal or compressed files.

The design of the downloading page is professional looking and is formatted in a manner that is similar to popular file-sharing services. The option is offered to download using special software or through a web browser. The suggested file does not bear a specific game's name and was downloaded using browser options.

Distribution Model. The whole process of getting a user's sensitive information has changed since the last round of malware collection. The user of this service downloads an archive, which contains content masked as game-related files and not an explicit binary executable that can be detected as malicious by anti-virus programs. The encrypted archive only grants access to filenames, not to the substantive content of the files. Furthermore, due to the use of encryption there is no way the exact nature of the content of the archive can be ascertained. In this case, an analyst can look at the raw content and hypothesise about whether the content is a text, an image, or some other content. However, it may be the case that the files simply contain dummy content, although this cannot be confirmed.

[69] BitTorrent is organised as a decentralised file-exchange protocol that includes users (peers) who either share files or download them. Seeds denote users who have fully available files on their computer ready for downloading through BitTorrent, while leechers denote users (peers) who do not yet have the full set of file segments.

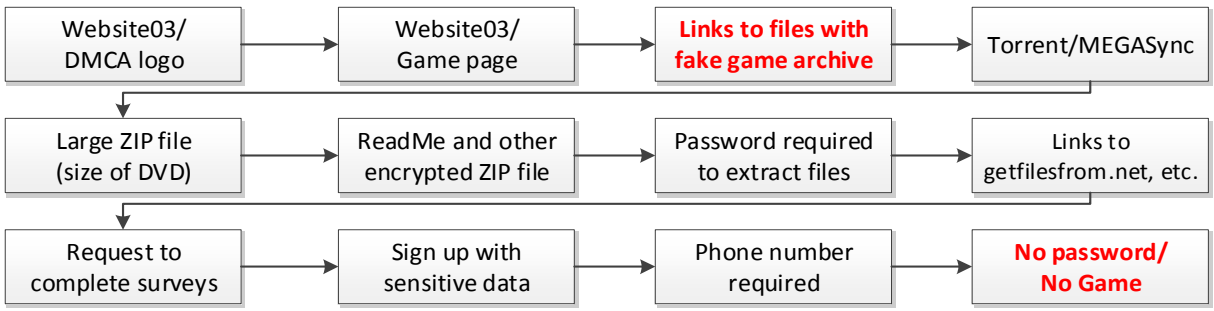


Figure 9: Distribution model of Website03 software

6. Malware and Potentially Unwanted Programs Discovered on Suspected Copyright-Infringing Websites: Categorisation and EMAS Analysis

6.1 Results of the two rounds of website identification and malware collection

During the two rounds of identification and analysis of copyright-infringing website identification and analysis, several corresponding domain names for each of the 10 sample countries were collected. Furthermore, to determine whether these websites provided only benign services or had some malicious indications, VirusTotal (VT)[70] reports were used to identify whether any of those websites had been seen before or were suspected of either performing malicious activities or distributing malicious or otherwise unwanted software targeting the end-user. VT offers scan results from over 60 well-known anti-virus vendors. The results from these investigations are outlined in *Table* 6: Statistics of websites characterised as malicious based on the VirusTotal reports for each of the 10 selected countries

It can be seen that nearly 8 % of the identified websites were characterised as malicious by anti-virus vendors.

Country	Round 1 — malicious %	Round 2 — malicious %
Belgium	42/600 (7 %)	42/608 (7 %)
Bulgaria	33/433 (8 %)	38/460 (8 %)
Croatia	34/431 (8 %)	38/447 (9 %)
Czech Republic	36/521 (7 %)	36/558 (6 %)
Finland	37/535 (7 %)	36/553 (7 %)
France	43/650 (7 %)	43/664 (6 %)
Hungary	34/518 (7 %)	39/572 (7 %)
Lithuania	37/526 (7 %)	39/560 (7 %)
Portugal	42/597 (7 %)	46/617 (7 %)
Sweden	38/555 (7 %)	38/561 (7 %)

Table 6: Statistics of websites characterised as malicious based on the VirusTotal reports for each of the 10 selected countries

In addition, during the two rounds of malware collection from the identified copyright-infringing websites, several malicious and suspected-of-being-malicious files were collected. The details of both rounds of malware collection are presented in *Table 7*. These were files directly downloaded from the websites. In addition, several files were acquired upon installation of the directly downloaded files. Those included any kind of side packages, software libraries, and other files that can pose threats to end-users wanting

[70] VirusTotal.com, https://www.virustotal.com; VirusTotal reports represent the most complete collection of more than 60 antivirus databases, so that these are often used to cross-check malware or the URL of a website against 1.5 billion entries in the VirusTotal collection. In the malware research community, it is common practice to validate results against the state-of-the-art malware classification found in VirusTotal.

to use them. In fact, in some cases, a single downloaded file created and installed several additional pieces of software that were not originally advertised. Such activities are called drive-by downloads and are very commonly used by malware developers. In addition, the files for the second round contain only new unique files that were not discovered during the first round of malware collection.

	Round 1	Round 2
Android	3	12
Mac OS	2	-
MS Windows	36	53
Total size bytes	175 600 117	522 991 095

Table 7: Statistics of the files acquired during the two rounds of malware collection

Summary of malware and potentially unwanted program dissemination techniques discovered on suspected copyright-infringing websites

Website03

Website03 is a website that distributes PUPs, advertising itself as providing free licence keys for copyright-protected games. After the simulated game installation, the user is prompted to provide sensitive personal information, including his or her name, address, and email address. That information is then automatically transferred to online surveys and discount campaigns without the user's consent. No licence for the desired game is provided. Some anti-virus vendors have already flagged the files associated with Website03 as suspicious or adware; however, there are only 4 detections out of 63 on the VirusTotal malware scanning service. The sequence of events that occurs if an individual installs Website03 is illustrated in Figure 10 below.

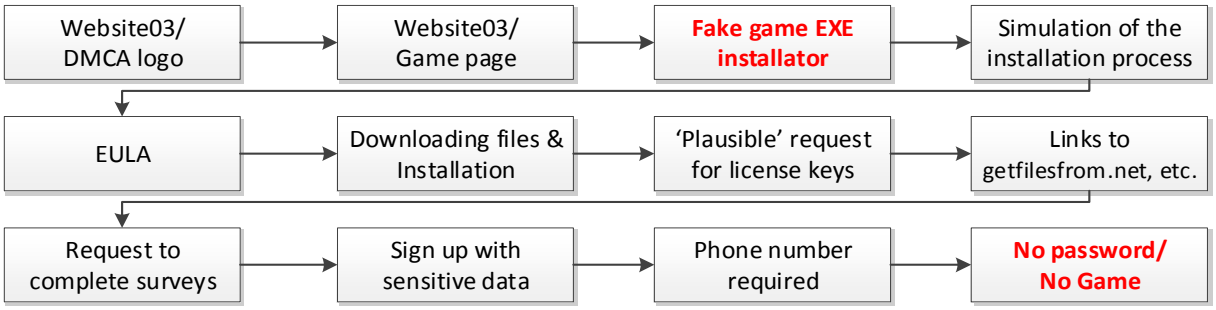


Figure 10: Action flow of the potentially harmful program (PUP) on Website03

Website04

Website04 is a website that distributes PUPs. It advertises itself on copyright-infringing websites as a tool for downloading that provides access to copyright-protected television programmes. In addition, it offers to install a number of other 'useful' tools, including anti-adware software and software to improve the user's computer-processing speed. During the installation process, it sends the user's system and browser details to a remote server over plain text — that is, without encryption — using SOAP and XML. The sequence of events that occurs if an individual installs bs.to is illustrated in the figure below.

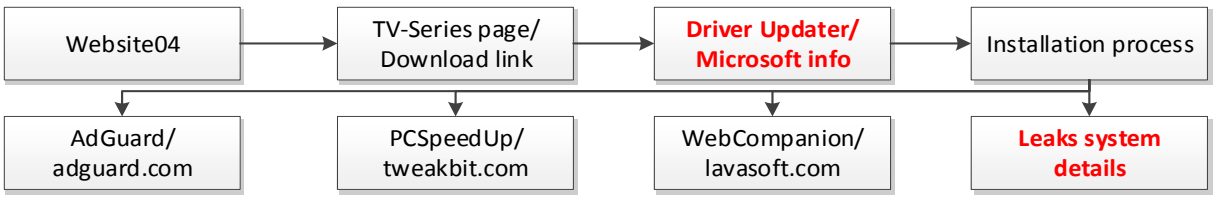


Figure 11: Action flow of malware on Website04

Website06

Website06 offers free software that can be classified as a PUP. It advertises itself as providing installation files and licence keys for a copyright-protected game. The software does not create any files on the user's hard drive except for a simulation of downloading an ISO image of the desired game. Upon installation, the user is prompted to provide a phone number, validate the phone number, and complete a survey. The sequence of events that occurs if an individual installs software from Website06 is illustrated below.

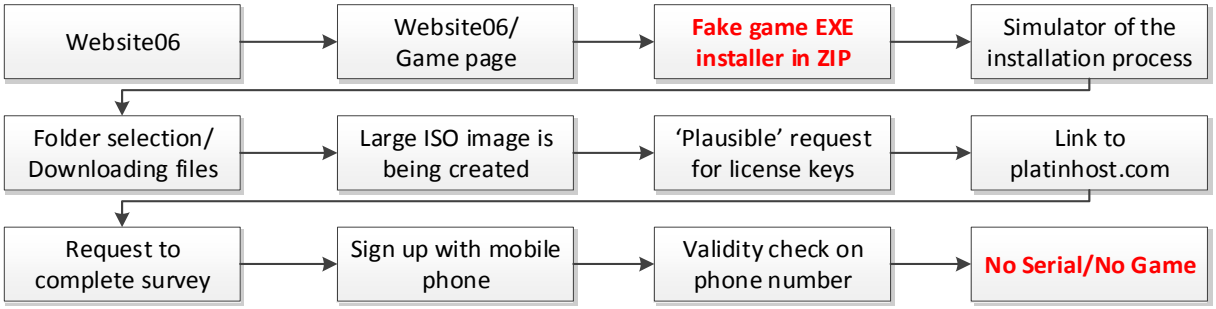


Figure 12: Action flow of malware on Website06

Website07

Website07 distributes PUPs that specifically target the Mac OS. The software is distributed through an advertisement on a copyright-infringing streaming website. The advertisement is attached or linked to the 'Download' button of the film being accessed. The copyright-protected video was not available, so for the user this button and software are the only options for accessing the copyright-protected material. Once the software has been installed, it performs a system check and identifies 'problems' that can purportedly be solved by buying additional components of the software or upgrading it to a paid pro version. The software offers a paid subscription for users. The Website07 malware is classified as 'unwanted software' by 11 out of 57 available anti-virus vendors. There is no additional information available on Virus Total, and there is only a very limited set of metadata. The sequence of events that occurs if an individual installs software from movieonline.io is illustrated in the figure below.

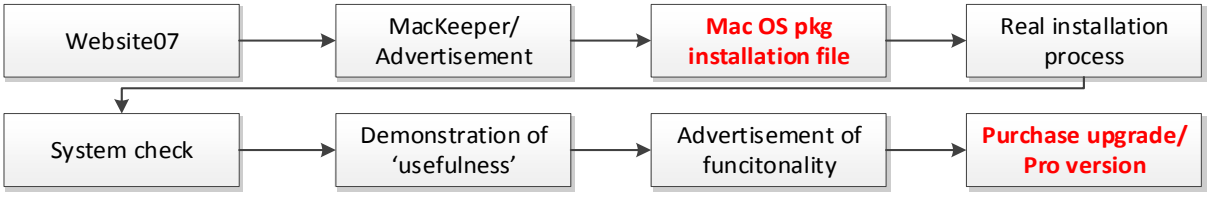


Figure 13: Action flow of Mac OS-specific malware on Website07

Website08

Website08 website offers a range of Android OS applications for downloading and streaming copyright-protected material free of charge. The application requests access to the user's files. There is no required registration, nor does it request personal data or payment information from the user. After installing the application, the user is provided access to copyright-protected television series and films. The sequence of events that occurs if an individual installs the application from Website08 is illustrated below.

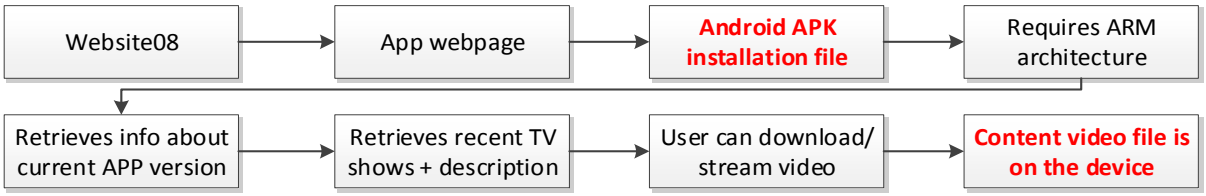


Figure 14: Action flow of Android OS-specific malware on Website08

6.2 EMAS

During the final stages of the project, an analysis was conducted of collected suspicious malicious or otherwise unwanted pieces of software by the EMAS[71]. The main function of this system is to detect all types of malicious activities in the memory, on the disc and in the network. In addition, there are a number of network services that can be simulated in an isolated environment to detect any anomalies in the network traffic.

Delivered packages with malicious or otherwise unwanted software

To facilitate the processing of the collected files by the EMAS platform, they were archived and encrypted in 8 separate packages. This was done to meet the platform requirement of having a single package of no more than 100 MB (104 857 600 bytes) in total[72].

Along with the submitted files, there was a need to define the corresponding internal submission codes to be used to describe the sources of information and their reliability, as described below.

Evaluation Code is X2

- evaluation of the information: 'information known personally to the source but not known personally to the official passing it on.'
- evaluation of the source: 'the reliability of the source cannot be assessed.'

Handling Code is H2: 'this information must not be disseminated without the permission of the provider.'

At present, only four versions of operating systems are supported by the EMAS solution (which limits the availability of the behavioural analysis for other platforms), including:

- Windows XP SP3
- Windows 7 32-bit SP1
- Windows 7 64-bit SP1
- Windows 10 64-bit.

[71] This platform required police officers to serve as intermediaries. For the purpose of the study, the Italian Polizia Postale kindly acted as an intermediary.
[72] On 9 October 2017, at 14:11:00, the packages were delivered to an Italian Law Enforcement Agency (LEA) for further processing.

EMAS supports the analysis of dozens of file types, such as documents, media content, scripting languages, etc.

6.3 Malware categorisation and analysis of collection findings and EMAS reports

In order to create a taxonomy of the malware collected that was distributed via the copyright-infringing websites, there was a need to co-analyse and integrate the findings made from two rounds of malware collection, as well as the EMAS analysis reports. Upon additional analysis of more than 60 anti-virus reports, which were made available for each file by VirusTotal, we have come to the following conclusions.

Harmfulness. During malware collection, the following types of distributed software were discovered.

- Benign — software that does not cause any harm to users and is designed for specific good purposes, such as content-distribution platforms or office programs.
- Potentially unwanted program (PUP) — software that provides advertisements, etc.
- Malware — harmful software that tampers and steals personal data and accesses files on the computer without proper authorisation.
- Malware/PUP — a piece of software that can be included equally in both categories.

The distribution of specific types of harmfulness is shown in *Figure 15*. It can be seen that most of the software found fell into the PUP category, while the benign category is the smallest group of software discovered.

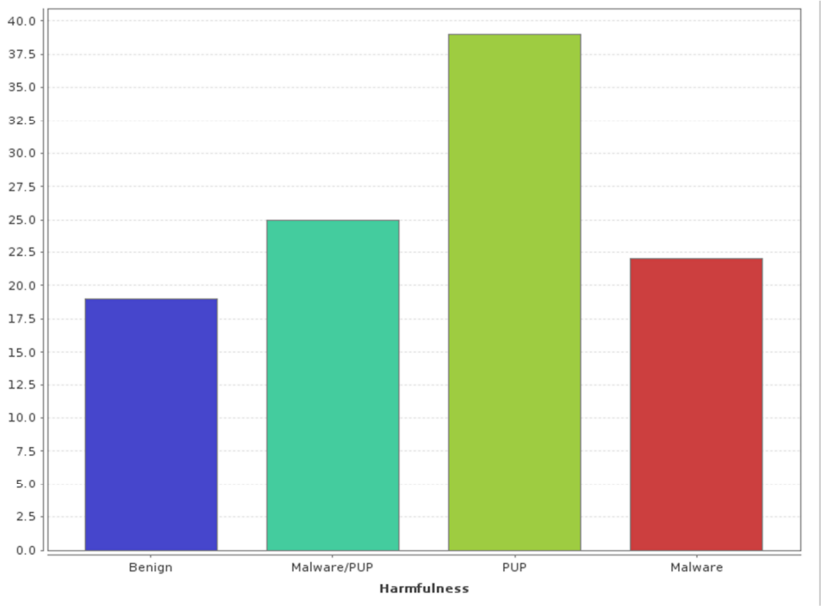


Figure 15: Distribution of harmfulness of the collected binaries

Functionality. All collected pieces of software are categorised as follows.

- Fake installers — software that lures users into disclosing personal information or providing payment card details by simulating game installation processes.

- Streaming — software that provides free access to pirated video or audio content.
- 'Useful' software — programs that may or may not improve something, yet promote a functionality that may be perceived as useful by some users.

The distribution of the specific functionality of software found is shown in *Figure 16*. Most of the programs are known as 'useful' software, which advertises various benefits to end-users, such as installing missing drivers and cleaning old files from PCs. Fake game installers and streaming services follow with a smaller share, yet one that is still considerable in comparison with the rest of the analysed programs.

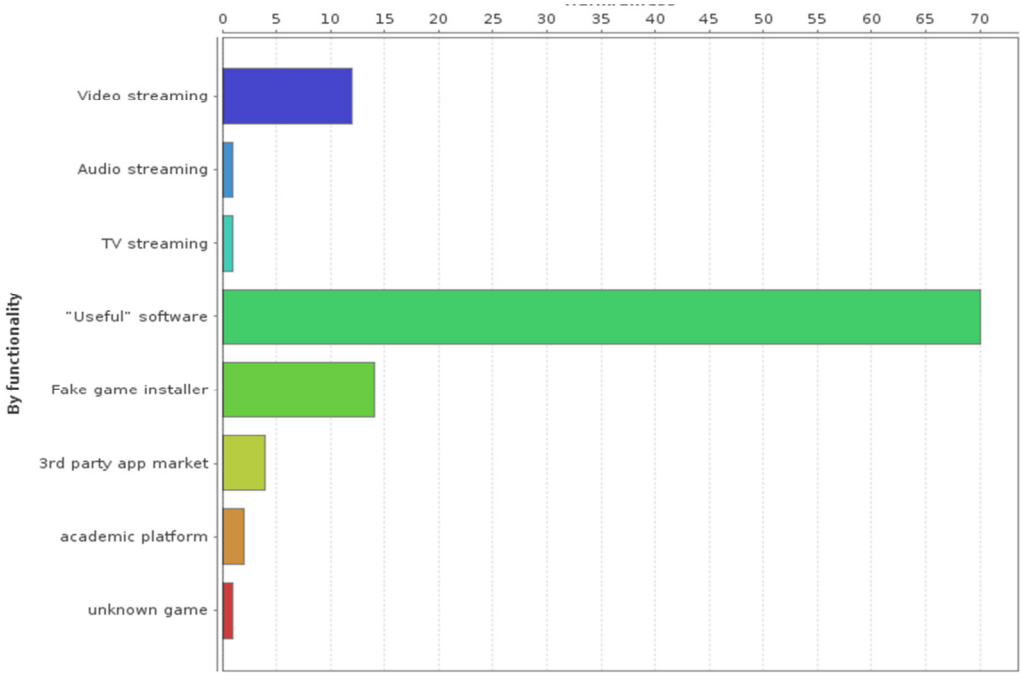


Figure 16: Categorisation of the collected files with respect to their purpose and demonstrated functionality

Malware Types. Four general categories can be distinguished: Trojan, adware, backdoor, and agent. Additionally, '-', in the figure below, means that there was no information available on community-accepted malware type even though multiple anti-virus vendors marked files as malicious. In this case, the labelling includes following general keywords such as 'not trusted', 'unsafe', 'unwanted', etc., which does not provide any additional semantic information about specific functionality or characteristics of malware. Therefore, in this study, such files were considered as generally malicious without a specific type.

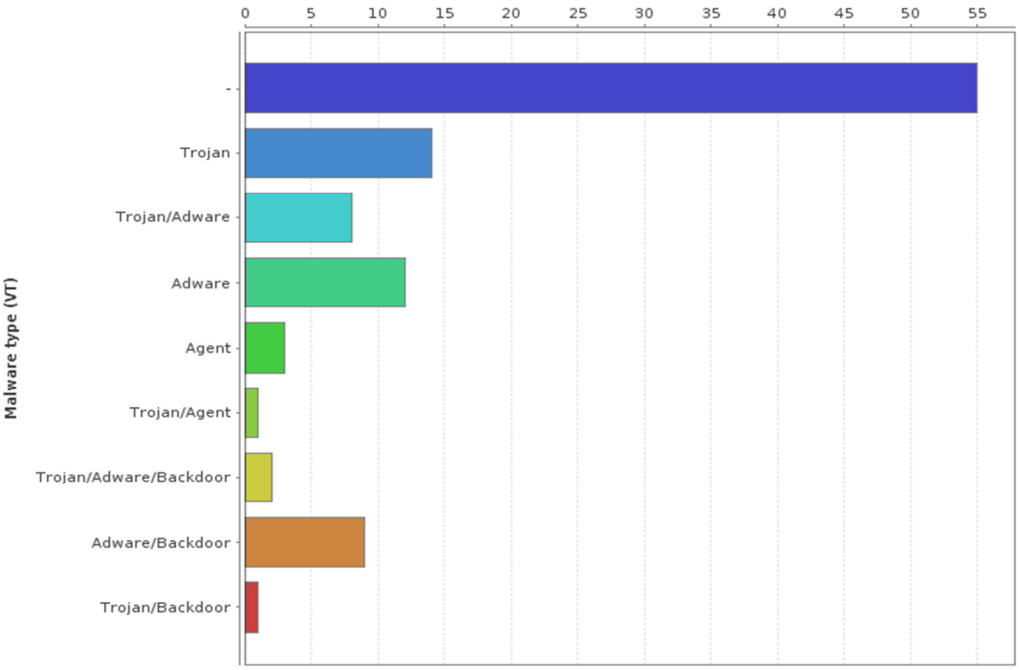


Figure 17: Distribution of the general malware types

Malware Families. There were several distinct specific malware families discovered during the analysis; '-' means that no type was assigned either because the information was not available or because it was not characterised by malware analysts before. In regard to malware families, this means that the malware labels assigned by anti-virus vendors included general types like 'Trojan', 'backdoor' or 'agent', without naming specific malware families to which the samples belonged. Therefore, it was not possible to determine a malware family based on previous classifications, only a general type as mentioned above.

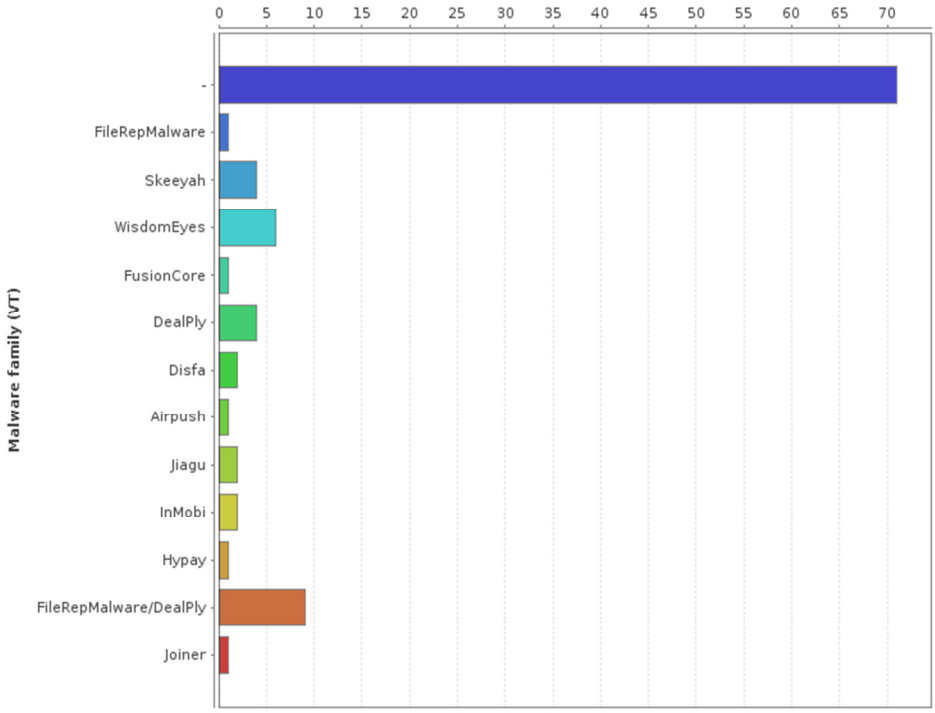


Figure 18: Distribution of more specific malware families from VirusTotal reports

Microsoft Windows. After correlating the manual analysis of the files and the analysis of the suspicious activities performed by EMAS, categorisation of the Windows malware disseminated via the copyright-infringing websites was identified. A summary based on cross-correlation of the existing categories and findings from EMAS reports is presented below in
Figure 19. The boxes with red text show categories of malware that are considered harmful and can potentially have an impact on users' personal information and software installed on the computer.

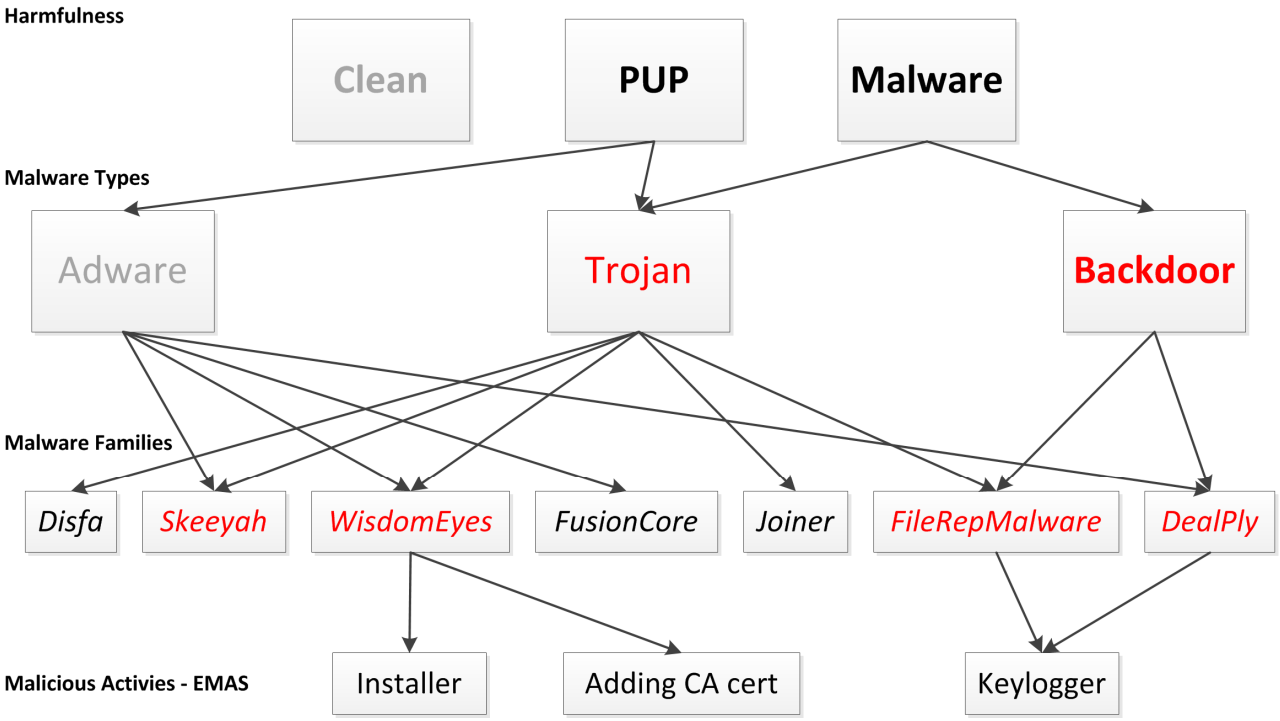


Figure 19: Taxonomy of the malicious or otherwise unwanted software distributed through copyright-infringing websites that target the Microsoft Windows desktop platform

Mac OS. A single malware sample was found that is to a certain degree a PUP, which was also marked as a Trojan by some of the anti-virus vendors. No specific family was assigned to it. In addition, EMAS did not have the capabilities to process Mac OS files.

Android. A number of Android applications were discovered. The variety of malware families is considerably smaller in comparison with those found for PCs because installation of third-party software requires specific permission on the Android platform. Moreover, this Android OS differs significantly from desktop computers, which also affects the attack vectors that are used. The boxes with red text show categories of malware that are considered harmful and can potentially have an impact on users' personal information and software installed on their computers.

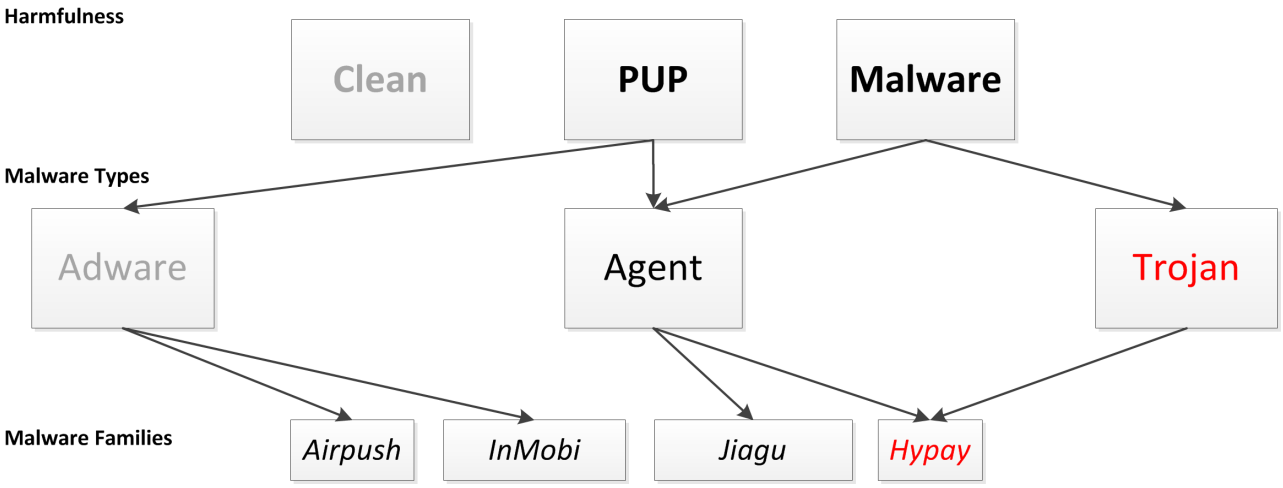


Figure 20: Taxonomy of the malicious or otherwise unwanted software distributed through copyright-infringing websites that target the Android mobile platform

Malicious Activities detected by EMAS. During the dynamic behavioural analyses, several of the most common malicious activities could be observed in the collected malware: rootkit, keylogger, adding CA certificate, NSIS installer, and network tampering and domain-generation activities. These also depend on the version of MS Windows used to perform the behavioural analysis. Most of the malware was successfully launched on both 32-bit and 64-bit MS Windows, implying that attackers still target 32-bit OS as a primary goal and very few specifically designed 64-bit applications. During the behavioural analysis, a malicious file was launched for 300 seconds and all the events during this time frame were recorded for future analysis. As can be seen from the EMAS reports, such events are divided into two classes: normal and anomalous (malicious). Furthermore, the class of malicious or anomalous events can include several subclasses with more specific descriptions of the anomalies.

6.4 Threats to end-users

During two rounds of website identification and malware analysis, no profoundly harmful malware samples, such as ransomware, botnets or others, were found. Generally, most of the collected malware were characterised as Trojans, meaning that they might be represented on the websites as benign commonly used or popular software, whereas in reality they can steal or disclose private information. What is more complicated is that some of the software that was found may be of multiple malicious types, such as Trojan, adware, and/or backdoor. This means that there is no clear separation in this case between these types, and malware developers have created sophisticated tools to trick the inexperienced user. Such users might have a high degree of trust in the software and might not be able to notice any abnormalities. However, without having the source code of the collected malware, it may be very difficult to fully understand the exact functionality and possible harm it would cause to the user in terms of privacy and security. In this regard, the main sources of information — static analysis and dynamic behavioural observations — might not reveal the complete picture to malware analysts.

Following the preliminary malware analysis, EMAS analysis showed more specific malicious activities and YARA rules that were detected for each software sample. At the same time, several cases included multiple malicious activities, such as keylogger, network activity tampering and rootkits. In other words, the impact of having this software installed on the end-user's computer might be considerable, causing not just financial losses, but also theft of personal data. These activities result in personal information gathering and transmission to third parties in encrypted or open text format. Such data can be bank account credentials from the browser, details of the computer hardware/software configuration or

basically anything typed on the keyboard. It was also found that malware can perform background checks: for example, it can check the version of the MS Windows OS and generate different activities and probably exploit specific vulnerabilities or weaknesses in the OS protection. Among the malware samples that were collected, Windows XP was found to be more vulnerable than Windows 10 and Windows XP also exhibited more malicious events being detected during malware execution. In addition, the 64-bit version of Windows exhibited fewer malicious events than those detected for the 32-bit version.

7. References

Brandom, R., 'Almost All WannaCry Victims Were Running Windows 7', *The Verge*, 19 May 2017 (https://www.theverge.com/2017/5/19/15665488/wannacry-windows-7-version-xp-patched-victim-statistics)

Bucher, A., 'Class Action Lawsuit: ZeoBIT Dupes Users into Buying MacKeeper Upgrade'. *Top Class Actions*, 7 May 2014 (https://topclassactions.com/lawsuit-settlements/lawsuit-news/26392-class-action-lawsuit-zeobit-dupes-users-buying-mackeeper-upgrade/)

Budd, C., 'Even more problems with apps and malware', *Trend Micro*, 23 June 2016 (http://blog.trendmicro.com/even-more-problems-with-apps-and-malware/)

Check Point Research Team, 'Hacked in Translation — from Subtitles to Complete Takeover' , *Check Point Blog*, 23 May 2017 (https://blog.checkpoint.com/2017/05/23/hacked-in-translation/)

Ducklin, P., 'Will a visit to The Pirate Bay end in malware?', 6 May 2016, *Naked Security by Sophos*, (https://nakedsecurity.sophos.com/2016/05/06/will-a-visit-to-the-pirate-bay-end-in-malware/)

Erdélyi, G., *Malware Taxonomy*. F-Secure, Helsinki, 2010 (http://www.cse.tkk.fi/fi/opinnot/T-110.6220/2010_Spring_Malware_Analysis_and_Antivirus_Tchnologies/luennot-files/Erdelyi-Introduction_to.pdf).

EUIPO, *European Citizens and Intellectual Property: Perception, Awareness, and Behaviour*, Office for Harmonization in the Internal Market, Alicante 2013 (https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/IPContributionStudy/25-11-2013/european_public_opinion_study_web.pdf)

EUIPO, *Intellectual Property and Youth: Scoreboard 2016* (https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/IP_youth_scoreboard_study/IP_youth_scoreboard_study_en.pdf)

EUIPO, *Research on Online Business Models Infringing Intellectual Property Rights*, EUIPO, Alicante 2016 (https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/resources/Research_on_Online_Business_Models_IBM/Research_on_Online_Business_Models_IBM_en.pdf)

European Union Agency for Network and Information Security, *ENISA Threat Landscape Report 2016: 15 top cyber-threats and trends,* Heraklion, ENISA, 2017 (https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2016).

Europol, *Internet Organised Crime Threat Assessment (IOCTA) 2016*, Europol, The Hague, 2016 (https://www.europol.europa.eu/sites/default/files/documents/europol_iocta_web_2016.pdf).

Europol and EUIPO, *2017 Situation Report on Counterfeiting and Piracy in the European Union: A joint project between Europol and the European Union Intellectual Property Office,* EUIPO, Alicante, 2017 (https://www.europol.europa.eu/sites/default/files/documents/counterfeiting_and_piracy_in_the_european_union.pdf).

F-Secure, 'Classification: Categories', 2017

(https://www.f-secure.com/en/web/labs_global/classification)

Gamerheadquarters, 'Dirt 4 install size', 2017 (http://articles.gamerheadquarters.com/article803dirt4installsize.html)

Google, *Google Transparency Report*, 2017 (https://www.google.com/transparencyreport/removals/copyright/?hl=en)

Hardikar, A., *Malware 101 — Viruses*, SANS Institute, 2008 (http://amanhardikar.com/papers/malware101viruses.pdf)

'How content theft sites and malware are exploited by cybercriminals to hack into internet users' computers and personal data', *Digital Bait*, Digital Citizens Alliance and RiskIQ, December 2015 (http://www.digitalcitizensalliance.org/clientuploads/directory/Reports/digitalbait.pdf)

IMDB, 'Internet Movie Database: Movies, TV, and Celebrities' (http://www.imdb.com/)

INCOPRO, *The Revenue Sources for Websites Making Available Copyright Content without Consent in the EU,* INCOPRO, London, 2015 (http://www.incoproip.com/resources-news-events/case-studies-reports/)

Kane, C., *Malware Taxonomy and Terminology*, University of Cincinnati, Cincinnati, 2017 (http://class.malware.re/lecture-slides/lecture-w03-1.pdf)

Lee, A., Varadharajan, V., Tupakula, U., 'On Malware Characterization and Attack Classification', *Proceedings of the First Australasian Web Conference*, vol. 144. Australian Computer Society, Darlinghurst, 2013

Lu, L., Perdisci, R., Lee, W., 'SURF: Detecting and Measuring Search Poisoning', *Proceedings of the 18th ACM conference on Computer and Communications Security,* ACM, New York, 2011

Malware Wiki, 'Category of Malware', 2017 (http://malware.wikia.com/wiki/Category:Category_of_Malware)

Malwarebytes, *State of Malware Report 2017*, Malwarebytes Labs, Santa Clara, CA, 2017 (https://www.malwarebytes.com/pdf/white-papers/stateofmalware.pdf)

Matthews-Winn, S., 'Why Avoid Film Piracy? It's Illegal and Dangerous', 2 December 2014, *OPSWAT,* (https://www.opswat.com/blog/why-avoid-film-piracy-its-illegal-and-dangerous)

Maxmind, 'Geoip products', 2017 (https://dev.maxmind.com/geoip/)

McAfee Labs, *2017 Threats Predictions*, McAfee, Santa Clara, CA, 2016 (https://www.mcafee.com/us/resources/reports/rp-threats-predictions-2017.pdf)

Microsoft, 'Naming Malware', 2017 (https://www.microsoft.com/en-us/wdsi/help/malware-naming)

Mullan, E., 'What is Digital Content?', *EContent*, 29 December 2011 (http://www.econtentmag.com/Articles/Resources/Defining-EContent/What-is-Digital-Content-79501.htm)

Mulvehill, T., 'The risk from mobile malware is real — and growing'. *Security Intelligence*, 25 April 2016 (https://securityintelligence.com/the-risk-from-mobile-malware-is-real-and-growing/)

National University of Singapore, *Cybersecurity Risks from Non-Genuine Software,* Microsoft Corp., Singapore, 2017 (https://ncmedia.azureedge.net/ncmedia/2017/10/Whitepaper-Cybersecurity-Risks-from-Non-Genuine-Software.pdf)

NetMarketshare, 'Mobile/Tablet Operating System Market Share', 2017 (https://www.netmarketshare.com/operating-system-market-share.aspx)

Nullsoft scriptable install system. (http://nsis.sourceforge.net/)

Rafique, Z. et al., 'It's free for a reason: exploring the ecosystem of free live streaming services', *Proceedings of the 23rd Network and Distributed System Security Symposium*, NDSS, San Diego, CA, 2016

Shalaginov, A., Grini, L.S., Franke, K., 'Understanding neuro-fuzzy on a class of multinomial malware detection problems', *Proceedings of the International Joint Conference on Neural Networks 2016*, October 2016, IEEE.

Shilko, J., 'Fraudster phishing users with malicious mobile apps', *PhishLabsBlog*, 25 April 2016 (https://info.phishlabs.com/blog/fraudster-phishing-users-with-malicious-mobile-apps)

Sophos, *Looking Ahead: SophosLabs 2017 Malware Forecast*, Sophos, Abingdon, 2017 (https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophoslabs-2017-malware-forecast-report.pdf?la=en)

StatCounter, *GlobalStats: Search Engine Market Share in Europe: May 2016 to May 2017*. (http://gs.statcounter.com/search-engine-market-share/all/europe)

Symantec, 'Internet security threat report 2017', *ISTR*, vol. 22, April 2017, Symantec, Mountain View, CA, 2017 (https://www.symantec.com/security-center/threat-report).

Techopedia (https://www.techopedia.com/definition/28464/magnet-link)

The Statistics Portal, 'Online and mobile content which internet users paid for in the past month as at 4th quarter 2014', 2014 (https://www.statista.com/statistics/388215/paid-online-mobile-content/)

United Nations, 'United Nations Regional Groups of Member States', 2014 (http://www.un.org/depts/DGACM/RegionalGroups.shtml)

VirusShare, 'VirusShare.com | About', 2017 (https://virusshare.com/about.4n6)

WhiteBullet Solutions Ltd, *Digital Advertising on Suspected Infringing Websites*, European Observatory on Infringements of Intellectual Property Rights, Alicante, 2016 (https://euipo.europa.eu/ohimportal/documents/11370/80606/Digital+Advertising+on+Suspected+Infringing+Websites)

Wikipedia, 'Best-Selling Digital Singles', Wikipedia, 2017 (https://en.wikipedia.org/wiki/List_of_best-selling_singles#Best-selling_digital_singles)

Wikipedia, 'List of Best-Selling PC Games', Wikipedia, 2017 (https://en.wikipedia.org/wiki/List_of_best-selling_PC_games)

Zhang, V., (2016, June 21). "'GODLESS' mobile malware uses multiple exploits to root devices", *TrendLabs Security Intelligence Blog*, 21 June 2016 (http://blog.trendmicro.com/trendlabs-security-intelligence/godless-mobile-malware-uses-multiple-exploits-root-devices/)

8. List of Tables

9. List of Figures

10. Annex 1: Methodology of the Study (extended version)

10.1 Phase I. Establishment of expert support group and selection of the expert

In the first phase of the UNICRI study, in collaboration with the Observatory, an expert support group was established to give advice on the research methodology, the selection of websites used for analysis, and to assess the research undertaken within each phase of project implementation. The expert support group was comprised of representatives from Observatory stakeholders, rights holder organisations, academia, law enforcement, and EU agencies[73].

Selection of the Expert. Andrii Shalaginov is a PhD research fellow in information security at the Department of Information Security and Communication Technology (Digital Forensics Group), Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU). The NTNU is the largest public university in Norway that has multiple research fields and study programme, also including cybersecurity. The NTNU Digital Forensics Group (in collaboration with police and industrial partners) conducts research in digital forensics and artificial intelligence applications for analytics problems related to big data. Over the last six years, Mr Shalaginov has been working on malware research. His primary expertise is in static and dynamic malware analysis, development of machine learning-aided intelligent computer virus detection models and similarity-based categorisation of malware types and families.

10.2 Phase II. Selecting countries for analysis

Within the framework of this report, it was not technically possible[74] to research all EU Member States, therefore in Phase II, 10 sample countries were randomly selected from the 28 EU Member States. Before the random selection process, UNICRI classified the EU Member States according to the regional groups used by the United Nations (UN): Eastern Europe and Western Europe[75]. The UN Eastern Europe regional group includes the following 11 EU Member States: Bulgaria, Croatia, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, and Slovenia. For the purposes of this study, Cyprus was also included in the Eastern Europe regional group[76]. The UN Western Europe regional group includes 16 EU Member States: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Portugal, Spain, Sweden, and the United Kingdom. UNICRI used random assignment to select 5 Member States from each regional group[77].
As a result of the randomisation process[78], the sample countries used throughout the study are as follows.

[73] Special thanks go to Mo Ali, IP Coordinator at UKIE, Glenn Deen, Sr Director, Networking & Distribution Technology, at NBCUniversal, and Lars Underbjerg, Security Manager at Nordic Content Protection.
[74] The number of selected countries will have a direct impact (increase) on the number of the selected suspected copyright-infringing websites and corresponding binary files to be analysed. Therefore, it was decided to concentrate only on a sample of countries to be able to successfully perform the practical part of the study within a given time frame.
[75] The United Nations refers to the Western European group as 'Western European and Others Group', in its 'United Nations Regional Groups of Member States'; retrieved from http://www.un.org/depts/DGACM/RegionalGroups.shtml
[76] Cyprus is included in the Asia-Pacific Group of the United Nations.
[77] First, each Member State was assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first five results were selected as sample countries.
[78] RAND function in Microsoft Excel, a pseudo-random number generator developed by B.A. Wichman and I.D. Hill.

- **Eastern Europe**: Bulgaria, Croatia, the Czech Republic, Hungary, and Lithuania, and
- **Western Europe**: Belgium, Finland, France, Portugal, and Sweden.

10.3 Phase III. Identifying titles for analysis

In Phase III, popular films, television programmes, songs, and video games were identified. Popularity included worldwide popularity as well as popularity in only one or more of the 10 sample countries as at the start of the data collection period, 23 June 2017. In the subsequent phases of the study, these sample titles were systematically used in online web searches to find copyright-infringing websites and mobile applications. Each title met two or more of the following criteria:

- popular at the time of data collection within EU Member States,
- popular at the time of data collection on a global scale,
- popular historically on a global scale, and
- categorised as a film, television programme, song, or video game.

Five film titles, five television titles, five music titles, and five video game titles were selected, resulting in a total of 20 sample titles. Careful consideration was given to the sources used to identify the popularity of a particular title, which involved a systematic selection process to ensure source data would be available for all or most of the Member States[79].

Selection of film titles

The five film titles include a mixture of films that were popular worldwide and in specific selected countries from Eastern and Western Europe respectively. The titles included three films that were reflecting worldwide trends (both over time and at the time of the data collection), one film title that was popular in Eastern Europe at the time of data collection, and one film that was popular in Western Europe at that time. The specific film titles and sources are below.

(1) **Highest-grossing film of all time (worldwide)**. UNICRI gathered this information from Box Office Mojo[80]. Box Office Mojo is a website operated by Amazon that compiles information on the all-time highest-grossing films worldwide. The site indicates the highest-grossing film worldwide (as measured by worldwide box office sales) as at 23 June 2017 was 20th Century Fox's *Avatar*, which grossed more than USD 2.8 billion in box office sales.

(2) **Most popular film recently released on DVD or for downloading, as measured by sales (worldwide)**. UNICRI gathered this information from Amazon[81], the third largest retailer worldwide, at the time of data collection[82]. As at 23 June 2017, the most popular film recently released on DVD or for downloading was *Kong: Skull Island*.

[79] UNICRI investigated the viability of using Amazon, Netflix, or Google Play to gather information on the most popular titles. Numerous Member States do not have dedicated Amazon websites, however, which makes it difficult to ascertain what titles are most popular. Netflix does not release official lists of its most popular titles. Lastly, Google Play does not appear to share lists of popular titles by country on its website.

[80] Box Office Mojo, 'All Time Box Office: Worldwide Grosses 2017'; retrieved from http://www.boxofficemojo.com/alltime/world/ Box Office Mojo is a website operated by Amazon that compiles information on box office trends worldwide.

[81] Amazon, 'Amazon Best Sellers in Movies & TV 2017'; retrieved from https://www.amazon.com/gp/bestsellers/movies-tv/ref=sv_mov_1.

[82] Gensler, L., 'The World's Largest Retailers 2017: Amazon & Alibaba Are Closing in on Wal-Mart', *Forbes*, 24 May 2017; retrieved from https://www.forbes.com/sites/laurengensler/2017/05/24/the-worlds-largest-retailers-2017-walmart-cvs-amazon/#21db9dfc20b5. Amazon maintains a regularly updated list of the bestselling film titles available for purchase on DVD, Blu-Ray, or as a digital download.

(3) **Highest-grossing film of 2017 at the time of data collection (worldwide)**. UNICRI gathered this information from Box Office Mojo[83]. As at 23 June 2017, the highest-grossing film of the year was *Beauty and the Beast*.

(4) **Highest-grossing film still in theatres at the time of data collection (Eastern Europe)**. A random number generator was used to select one of the sample countries in Eastern Europe[84]. Lithuania was selected as a result of this process. The highest-grossing film still in cinemas in Lithuania at the time of data collection was included as a sample title. As at 23 June 2017, *Baywatch* was still the highest-grossing film in cinemas in Lithuania, according to Box Office Mojo[85].

(5) **Highest-grossing film still in theatres at the time of data collection (Western Europe)**. A random number generator was used to select one of the sample countries from Western Europe[86] and Finland was selected as a result. The highest-grossing film still in cinemas in Finland at the time of data collection was included as a sample title. As at 23 June 2017, *Pirates of the Caribbean: Dead Men Tell No Tales* was still the highest-grossing film in cinemas in Finland, according to Box Office Mojo[87].

Selection of television titles

A total of five television titles were selected for inclusion in the research. The titles include a mixture of television programmes that were popular worldwide and in Eastern and Western Europe respectively. The titles include three television programs that reflect worldwide trends (both historically and as at data collection), one television programme that was popular in Eastern Europe at the time of data collection, and one television programme that was popular in Western Europe at that time. The specific television programmes and sources are below.

(1) **2016 most popular television programme (worldwide)**. UNICRI gathered this information from analysis by Parrot Analytics[88], a media analytics firm[89]. Parrot Analytics indicated the most popular television programme of 2016 globally was HBO's *Game of Thrones* series.

(2) **2016 second most popular television programme (worldwide)**. The second most popular television programme of 2016 globally was AMC's *The Walking Dead* series, according to Parrot Analytics.

(3) **2016 third most popular television programme (worldwide)**. The third most popular television programme of 2016 globally was ABC's *Pretty Little Liars* series, according to Parrot Analytics.

[83] Box Office Mojo, 'Yearly Box Office Worldwide 2017'; retrieved from http://www.boxofficemojo.com/yearly/?view2=worldwide&view=releasedate&p=.htm.
[84] First, each of the five Eastern European sample countries was assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first result was selected.
[85] Box Office Mojo, 'Box Office Mojo International, Current Results by Territory 2017'; retrieved from http://www.boxofficemojo.com/intl/.
[86] First, each of the five Western European sample countries was assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first result was selected.
[87] Box Office Mojo, 'Box Office Mojo International, Current Results by Territory 2017', retrieved from http://www.boxofficemojo.com/intl/.
[88] Parrot Analytics, 'Demand Data | Parrot Analytics 2017'; retrieved from https://www.parrotanalytics.com/demand-data/
[89] Lubin, G., 'Data Reveals the 20 Most Popular TV Shows of 2016', *Business Insider*, 30 December 2016; retrieved from http://www.businessinsider.com/most-popular-tv-shows-2016-12?IR=T. In 2016, Parrot Analytics released a report naming the most popular television programmes for the year, as measured by the 'demand rating' for each show. The 'demand rating' is a weighted function of the popularity of on-air television programmes, the popularity of streaming shows, activity on social media surrounding television programmes, activity on fan sites, viewer-generated ratings, the presence of active Wiki sites, as well as how frequently television programmes are downloaded and streamed on copyright-infringing websites.

(4) **Most popular television programme at the time of data collection (Eastern Europe)**. Hungary was randomly selected from the Eastern European sample countries[90]. Initially, the Hungarian Apple iTunes store was going to be used as a source for the most popular television programme; however, as at 23 June 2017, the Hungarian iTunes store did not allow users to download television programmes or series. IMDB[91] was used as an alternative source[92]. IMDB, also known as the Internet Movie Database, is an online database of television programmes and films. The site lists *Alias* (2001-2006) as the most popular television series in Hungary as at 23 June 2017.

(5) **Most popular television programmes at the time of data collection (Western Europe)**. Belgium was randomly selected from the Western European sample countries[93]. Initially, the Belgian Apple iTunes store was going to be used as a source for the most popular television programme; however, as at 23 June 2017, the Belgian iTunes store did not allow users to download television programmes. IMDB[94] was used as an alternative source.[95] IMDB lists *The Missing* (2014-present) as the most popular television series in Belgium.

Selection of song titles

A total of five song titles were selected for inclusion in the study. The five song titles include a mixture of songs that were popular worldwide (both historically and at the time of data collection), one song that was popular in Eastern Europe, and one song that was popular in Western Europe. The specific song titles and sources are below.

(1) **Bestselling digital single of all time (worldwide)**. UNICRI gathered this information from Wikipedia[96], which maintains a regularly updated list of the bestselling digital singles of all time worldwide[97]. As at June 2017, Wikipedia indicated the bestselling digital single worldwide of all time was 'See You Again' by Wiz Khalifa, featuring Charlie Puth, which was released in 2015.

(2) **Most popular single of 2016, as measured by radio airplay, sales data, and streaming activity data (worldwide)**. UNICRI gathered information on the most popular digital single of 2016 from the *Billboard Hot 100 Songs of 2016* list. The list measures popularity 'across all genres, as ranked by radio airplay, audience impressions as measured by Nielsen Music, sales data as compiled by Nielsen Music, and streaming activity provided by online music sources'[98]. *Billboard* indicates the most popular song of 2016 worldwide was 'Love Yourself' by Justin Bieber.

(3) **Most popular single, as measured by radio airplay, sales data, and streaming activity data (worldwide)**. UNICRI gathered information on the most popular digital single from the *Billboard*

[90] First, each of the five Eastern European sample countries was assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first result was selected.
[91] IMDB, 'Internet Movie Database: Movies, TV, and Celebrities'; retrieved from http://www.imdb.com/
[92] Alexa, 'Imdb.com Traffic Statistics'; retrieved from http://www.alexa.com/siteinfo/imdb.com; IMDB was the 57th most popular website worldwide on Alexa as at July 2017.
[93] First, each of the five Western European sample countries were assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first result was selected.
[94] IMDB, 'Internet Movie Database: Movies, TV, and Celebrities'; retrieved from http://www.imdb.com/
[95] Alexa, 'Imdb.com Traffic Statistics'; retrieved from http://www.alexa.com/siteinfo/imdb.com
[96] Wikipedia, 'Best-Selling Singles'; retrieved from https://en.wikipedia.org/wiki/List_of_best-selling_singles#Best-selling_digital_singles
[97] *Billboard* maintains a running list of the best-selling singles of all time; however, the list includes both digital and non-digital singles. As a result, the best-selling singles were songs from decades prior, which were not considered appropriate for this particular study.
[98] Billboard, 'Year End Charts: Hot 100 Songs 2016'; retrieved from http://www.billboard.com/charts/year-end/2016/hot-100-songs.

Hot 100 list, which compiles 'this week's most popular songs across all genres, as ranked by radio airplay, audience impressions as measured by Nielsen Music, sales data as compiled by Nielsen Music, and streaming activity provided by online music sources'[99]. The most popular single worldwide on 23 June 2017 was 'Despacito' by Luis Fonsi and Daddy Yankee, featuring Justin Bieber.

(4) **Most popular single, as measured by streaming activity data (Eastern Europe)**. A random number generator was used to select one of the sample countries in Eastern Europe. Bulgaria was selected as a result of this process[100]. The music single identified as the most streamed song in Bulgaria at the time of data collection was included as a sample title. As at 23 June 2017, 'Wild Thoughts' by DJ Khaled was the most streamed song in Bulgaria. UNICRI gathered this information from the Spotify Charts, a website that maintains a regularly updated list of the most streamed songs on Spotify[101]. Spotify was the 143rd most visited website globally, according to Alexa rankings as at 23 June 2017[102].

(5) **Most streamed single (Western Europe)**. A random number generator was used to select one of the sample countries in Western Europe. France was selected as a result of this process[103]. The music single identified as the most streamed song in France at the time of data collection was included as a sample title. As at 23 June 2017, 'Θ Macarena' by Damso was the most streamed song in France. UNICRI gathered this information from Spotify Charts[104].

Selection of video game titles

Altogether, five video game titles were selected for the research. All five titles reflect worldwide trends (both historically and at the time of data collection)[105].

(1) **Most popular video game, as measured by sales on Steam (worldwide)**. UNICRI gathered this information from Steam, a major online retailer of video games for the Windows, Mac, and Linux operating systems[106]. Steam was the most popular website globally among gaming-related websites[107] and was the 201st most popular website globally, according to Alexa rankings[108]. Steam maintains a regularly updated list of its bestselling video games. As at 23 June 2017, 'Middle-earth: Shadow of Mordor Game of the Year Edition' was the top selling game on Steam.

(2) **Most popular video game, as measured by sales on Humble Bundle (worldwide)**. UNICRI gathered this information from Humble Bundle, a major online retailer of video games for the Windows and Mac operating systems[109]. Humble Bundle was listed as the second most popular

[99] Billboard, 'The Hot 100 | How It Works'; retrieved from http://www.billboard.com/charts/hot-100.
[100] First, each of the five Eastern European sample countries was assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first result was selected.
[101] Spotify, 'Spotify Charts | Regional'; retrieved from https://spotifycharts.com/regional.
[102] Alexa, 'Spotify.com Traffic Statistics', retrieved from http://www.alexa.com/siteinfo/spotify.com.
[103] First, each of the five Western European sample countries was assigned a random number using a random number generator in Excel. Second, the Member States were sorted numerically according to the random numbers. The first result was selected.
[104] Spotify, 'Spotify Charts | Regional'; retrieved from https://spotifycharts.com/regional.
[105] All of the titles represent worldwide trends, as an open-source review did not reveal any reliable sources of information regarding the most popular computer games for each Member State.
[106] Valve, 'Welcome to Steam' Steam Store; retrieved from http://store.steampowered.com
[107] Alexa, 'Top Sites by Category | Games'; retrieved from http://www.alexa.com/topsites/category/Top/Shopping/Toys_and_Games/Games
[108] Alexa, 'Steampowered.com Traffic Analysis'; retrieved from http://www.alexa.com/siteinfo/steampowered.com.
[109] Humble Bundle, "The Humble Store'; retrieved from https://www.humblebundle.com/store.

website globally among gaming-related websites[110] and was the 560th most popular website globally, according to Alexa rankings[111]. Humble Bundle maintains a regularly updated list of its bestselling video games. As at 23 June 2017, 'Playerunknown's Battlegrounds' was the top-selling game on Humble Bundle.

(3) **Most popular video game, as measured by sales on GameStop (worldwide)**. UNICRI gathered this information from GameStop, a major online and physical retailer of video games for Windows and Mac operating systems[112]. GameStop was listed as the third most popular website globally among gaming-related websites[113] and was the 1 496th most popular website globally, according to Alexa rankings[114]. GameStop maintains a regularly updated list of its bestselling video games. As at 23 June 2017, 'The Sims 4' was the top selling game on GameStop's website.

(4) **Bestselling video game of all time (worldwide)**. UNICRI gathered this information from Wikipedia[115], which maintains a regularly updated list of the bestselling video games of all time worldwide[116]. Wikipedia indicated the bestselling video game worldwide as at 23 June 2017 was 'Minecraft,' which was released in 2011 and had sold 26 million copies.

(5) **Bestselling video game of 2016, as measured by sales on Steam (worldwide)**. UNICRI gathered this information from Steam[117] and Alexa[118] rankings[119]. Steam indicates the bestselling video game of 2016 worldwide was 'The Witcher III: Wild Hunt,' which was released in 2015[120].

For a summary of the sample countries, titles, and sources used in the project, refer to *Table 8.*

Sample countries

Eastern Europe	Bulgaria Croatia Czech Republic Hungary Lithuania
Western Europe	Belgium Finland France Portugal Sweden

Sample titles

Films	*6. Avatar*

[110] Alexa, 'Top Sites by Category | Games'; retrieved from http://www.alexa.com/topsites/category/Top/Shopping/Toys_and_Games/Games
[111] Alexa, 'Humblebundle.com Traffic Statistics'; retrieved from http://www.alexa.com/siteinfo/humblebundle.com
[112] GameStop, 'GameStop'; retrieved from https://www.gamestop.com/.
[113] Alexa, 'Top Sites by Category | Games'; retrieved from http://www.alexa.com/topsites/category/Top/Shopping/Toys_and_Games/Games.
[114] Alexa, 'GameStop.com Traffic Statistics'; retrieved from http://www.alexa.com/siteinfo/gamestop.com
[115] Wikipedia, 'List of Best-Selling PC Games'; retrieved from https://en.wikipedia.org/wiki/List_of_best-selling_PC_games
[116] An open-source search for information on all-time global sales of PC games did not reveal any additional, comprehensive sources except for Wikipedia.
[117] Valve, 'Welcome to Steam', Steam Store; retrieved from http://store.steampowered.com.
[118] Alexa, 'Top Sites by Category | Games'; retrieved from http://www.alexa.com/topsites/category/Top/Shopping/Toys_and_Games/Games.
[119] Alexa, 'Steampowered.com Traffic Analysis'; retrieved from http://www.alexa.com/siteinfo/steampowered.com.
[120] Valve, 'Top 100: Best Sellers of 2016', Steam Store; retrieved from http://store.steampowered.com/sale/2016_top_sellers/.

	7. *Kong: Skull Island* 8. *Beauty and the Beast* 9. *Baywatch* 10. *Pirates of the Caribbean: Dead Men Tell No Tales*
Television	6. *Game of Thrones* 7. *The Walking Dead* 8. *Pretty Little Liars* 9. *Alias* 10. *The Missing*
Music	6. 'See You Again' by Wiz Khalifa, featuring Charlie Puth 7. 'Love Yourself' by Justin Bieber 8. 'Despacito' by Luis Fonsi and Daddy Yankee, featuring Justin Bieber 9. 'Wild Thoughts' by DJ Khaled 10. 'Θ Macarena' by Damso
Video games	6. Middle-earth: Shadow of Mordor (Game of the Year Edition) 7. Playerunknown's Battlegrounds 8. The Sims 4 9. Minecraft 10. The Witcher III: Wild Hunt

Table 8: Summary of the titles selected for use in the project

10.4 Phase IV.A. Identifying suspected copyright-infringing websites for analysis

Phase IV identified websites suspected of providing unlawful access to copyright-protected material that were popular worldwide and/or among the 10 sample countries as at 26 June 2017. In a later phase of the study, these websites were analysed for the presence of malware and potentially unwanted programs.

The methodology for identifying suspected copyright-infringing websites was developed with the input of the expert support group identified in Phase I, as well as after a review by UNICRI of the existing literature. It was specifically devised to generate a sample of websites that:

- are popular within different EU Member States, ensuring a wide geographical coverage;
- represent different types of suspected copyright-infringing websites, including streaming websites, linking websites, hosting websites, cyberlockers, and torrent websites[121];
- represent a broad range of suspected copyright-infringing content, including films, television titles, music, and video games; and
- represent websites that the average internet user would encounter when attempting to access suspected copyright-infringing material.

Five steps were followed to select the suspected copyright-infringing websites. The first three steps were designed to identify the most popular suspected copyright-infringing websites across EU Member States. This method mimicked those scenarios in which an average user might search for suspected copyright-infringing websites without specifying, for example, the title of a film or a song. The final two

[121] EUIPO (2016). , *Research on Online Business Models Infringing Intellectual Property Rights.* , EUIPO, Alicante, 2016; retrieved from https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/resources/Research_on_Online_Business_Models_IBM/Research_on_Online_Business_Models_IBM_en.pdf.

steps were designed to identify suspected copyright-infringing websites that an average user might encounter when searching for ways to download a specific popular title without specifying a website. This step was particularly significant, given the presence of suspected malicious websites that engage in search result poisoning, where they exploit trending topics[122] through search engine optimisation[123]. Together, the two approaches covered the different ways an average internet user would attempt to find suspected copyright-infringing material online[124].

First, to mimic a scenario in which an average internet user searches for a website where suspected copyright-infringing material can be accessed, UNICRI generated a list of the 500 most popular suspected websites for each of the 10 sample countries (sample list). The websites were identified using the Alexa Top 500 list for the respective country. The Alexa Top 500 list compiles a monthly list of the most popular websites globally, regionally, and by country. Its rankings are based on page views and the average number of daily visitors a website receives[125]. The country or countries associated with each website were denoted in the sample list to ensure each sample country was represented in the final sample of websites.

During the second step, the sample list was cross-referenced against the Alexa Top 500 list for the entire EU. Websites suspected of providing access to copyright-infringing content that were not included on the country-specific Alexa Top 500 lists were added to the sample list.

The third stage involved UNICRI manually reviewing the sample list of websites suspected of providing access to suspected copyright-infringing material. Websites that did not appear to provide access to suspected copyright-infringing material were removed from the sample list. When necessary, Google Transparency Report was used as a guiding tool to determine whether or not a domain name was appropriate for inclusion on the sample list[126]. Google Transparency Report offers a searchable database of all domains for which Google has received requests to remove suspected copyright-infringing content.[127]

Fourth, to approximate a scenario in which an average user searches for a specific copyright-protected title in a search engine, targeted keyword searches were carried out using popular search engines for each of the 20 sample titles identified in Phase III. The three search engines with the highest market share in Europe were used for the keyword searches: Google (91.92 % market share), Yahoo (1.64 %), and Bing (3.63 %)[128]. Search results tend to be geographically targeted, so each keyword search was performed using a proxy server[129] located in the target sample country[130]. The keyword searches were performed using both the main page for each search engine (e.g. www.google.com), as well as the

[122]Lu, L., Perdisci, R., Lee, W, 'SURF: Detecting and Measuring Search Poisoning', *Proceedings of the 18th ACM conference on Computer and Communications Security*, ACM, New York, 2011, pp. 467-476.

[123] The order of search results is far from random; rather, it is the result of a search engine's highly complex algorithm. Cybercriminals may attempt to take advantage of these algorithms (e.g. by including the name of popular films, music, etc. in web page titles) to maximise the ranking of malicious websites. This is known as search engine poisoning.

[124] This methodology is an adaptation of: a) Rafique, Z. et al., 'It's free for a reason: exploring the ecosystem of free live streaming services', *Proceedings of the 23rd Network and Distributed System Security Symposium*, NDSS, San Diego, CA, 2016, pp. 1-15; and b) WhiteBullet Solutions Ltd, *Digital Advertising on Suspected Infringing Websites*. European Observatory on Infringements of Intellectual Property Rights, Alicante, 2016; retrieved from https://euipo.europa.eu/ohimportal/documents/11370/80606/Digital+Advertising+on+Suspected+Infringing+Websites

[125] Amazon, 'Alexa Top 500 List: by Country', 2017; retrieved from http://www.alexa.com/topsites/countries.

[126] The data from Google Transparency Reports was collected on 26 June 2017.

[127] Google, *Google Transparency Report*; retrieved from https://www.google.com/transparencyreport/removals/copyright/?hl=en.

[128] StatCounter, *GlobalStats: Search Engine Market Share in Europe*: *May 2016 to May 2017*; retrieved from http://gs.statcounter.com/search-engine-market-share/all/europe; Alexa Top 500 List, Europe(2017), Amazon; retrieved from http://www.alexa.com/topsites/category/Regional/Europe.

[129] A Tor browser with an explicit definition of country of exit node was used to conduct the searches. However, Tor network connectivity and reliability is low, meaning that a small fraction of results may be missing.

[130] WhiteBullet Solutions Ltd, *Digital Advertising on Suspected Infringing Websites*, European Observatory on Infringements of Intellectual Property Rights, Alicante, 2016.

country-specific search engine domain, if available (e.g. www.google.de)[131]. IP addresses and ISP details were logged for future reference and documentation.

Each keyword search included one of the 20 sample titles, as well as one keyword or phrase. For the films, television, and music categories, the keywords included: 'pirate,' 'download,' 'stream,' 'free copy,' and 'watch online.' For the video game category, the keywords included: 'pirate,' 'download', 'cracked', 'free copy' and 'ROM.' The video game requires a specific 'crack' to be run on a computer and also comes in a ROM form copied from a specific device, therefore making the list of selected keywords different, Google Trends (GT) was used to validate the popularity of keywords across regions and countries in the EU[132]. GT offers a functionality to test how popular a specific phrase is across a defined country, as well as in a defined period of time. All titles and keywords were in English because suspected copyright-infringing websites generally keep the original title, even though the description may be in a country's own language.

This process was carried out using an expert-assisted crawler[133] and, manually, by an expert. The top 10 results of each keyword search were checked against the existing list of sample websites[134]. Any website not already present was added to the sample list. The country or countries associated with each website were also noted. In the event that search results contained only websites known to be legitimate or websites already contained in the list, the next 10 most relevant websites in the search were included.

As a fifth step, the lists of website domains from the third and fourth steps were combined together, creating a final dataset of suspected copyright-infringing websites that were a) popular in the sample countries, and b) represented common search results in the three major search engines. These domains were later searched for malware samples.

Table 9 below reflects the number of suspected copyright-infringing domains for each sample country after each of the five steps.

	Step 2	Step 3	Step 4	Step 5
Belgium	500	387	213	600
Bulgaria	500	316	117	433
Croatia	500	308	123	431
Czech Republic	500	318	204	522
Finland	500	323	213	536
France	500	397	253	650
Hungary	500	325	194	519
Lithuania	500	318	209	527
Portugal	500	385	212	597
Sweden	500	336	219	555

Table 9: Statistics of the selected websites per each step in Phase IV.A (Round I)

[131] Country-specific search engine domains for Bing and Yahoo could not be located for Croatia and Bulgaria. As a result, only the '.com' versions of Bing and Yahoo were used for those two sample countries.
[132] Google Trends (2017); retrieved from https://trends.google.com/trends/.
[133] A computer program that is used to process and analyse web pages, however, under specific guidance of a human expert and a predefined algorithm.
[134] Results from Google, YouTube, Bing, Yahoo, Amazon, Blogger, HBO, Netflix, Spotify, Twitter, Facebook and Microsoft were filtered out from the search results.

10.5 Phase IV.B. Identifying mobile applications for analysis

During the research project, there was increasing interest among the expert support group in conducting concurrent analysis on malware and PUPs specific to mobile applications on devices, such as smartphones and tablets. Europol's Internet Organised Crime Threat Assessment (2016) identified mobile malware as one of the key cybercrime threats facing the EU[135]. The incidence of mobile malware and PUPs is increasing and becoming more complex[136]. Different forms of mobile malware and PUPs mirror their computer-based counterparts and include remote access tools, drive-by downloads, click fraud, banking Trojans, and ransomware. In addition, phishing applications have made their way onto Google Play, the main application store for smartphones and mobile devices using the Android operating system[137]. In addition, the Android OS allows third-party applications to be installed, opening opportunities for attackers to exploit this functionality. The applications often purport to be associated with reputable companies (e.g. financial companies and payment services providers) and, when downloaded and accessed by a user, prompt the user with a dialogue box to enter his or her login name and password. These applications are usually made to appear similar to official applications, but contain a malicious payload. The application then steals the user's username and password and any other relevant information that has been entered. After receiving input from the expert support group identified in Phase I and after a review by UNICRI of the existing literature on mobile malware, the following methodology was used to identify mobile malware applications on Android devices for inclusion in the analysis. Analysis was limited to Android devices due to indications in the existing literature of a greater presence of malware on Android application stores (i.e. Google Play) than on the Apple iTunes store[138]. The methodology was devised to generate a sample of mobile applications that:

- are popular at the time of data collection on a global scale,
- represent different types of applications (to include streaming applications, torrent applications, and hosting applications),
- contain or provide access to a broad range of suspected copyright-infringing content (to include films, television titles, music, and mobile games), and
- represent what an average user of a mobile device will encounter when attempting to download or use an application facilitating access to suspected copyright-protected content.

Four steps were followed to select the suspected copyright-infringing mobile applications. The first two steps were designed to identify the most popular suspected copyright-infringing Android mobile applications globally. These steps involved mimicking scenarios in which an average user might search for suspected copyright-infringing applications without indicating a specific title of a copyright-protected work. The final two steps were designed to identify suspected copyright-infringing applications that an average user might encounter when attempting to access a specific popular title without indicating a

[135] Europol, *Internet Organised Crime Threat Assessment (IOCTA) 2016*, Europol, The Hague, 2016; retrieved from https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2016

[136] Mulvehill, T., 'The risk from mobile malware is real — and growing', *Security Intelligence*, 25 April 2016; retrieved from https://securityintelligence.com/the-risk-from-mobile-malware-is-real-and-growing/.

[137] Shilko, J., 'Fraudster phishing users with malicious mobile apps', *The PhishLabs Blog*, 25 April 2016; retrieved from https://info.phishlabs.com/blog/fraudster-phishing-users-with-malicious-mobile-apps.

[138] Zhang, V., "'GODLESS' mobile malware uses multiple exploits to root devices", *TrendLabs Security Intelligence Blog*, 21 June 2016: retrieved from http://blog.trendmicro.com/trendlabs-security-intelligence/godless-mobile-malware-uses-multiple-exploits-root-devices/; Shilko, J., 'Fraudster phishing users with malicious mobile apps', *The PhishLabs Blog*, 25 April 2016; retrieved from https://info.phishlabs.com/blog/fraudster-phishing-users-with-malicious-mobile-apps; Budd, C., 'Even more problems with apps and malware', *Trend Micro*, 23 June 2016; retrieved from http://blog.trendmicro.com/even-more-problems-with-apps-and-malware/; Brandom, R. (2016, November 30). 'App-installing malware found in over 1 million Android phones', *The Verge*, 30 November 2016; retrieved from https://www.theverge.com/2016/11/30/13792846/googlian-android-malware-install-app-security.

specific application. Together, these two methods approximated the different ways an average user would attempt to find suspected copyright-infringing content on mobile applications[139].

First, UNICRI generated a list (sample applications list) of the top 300 mobile applications available for downloading on the Google Play store under the 'Entertainment' category and the top 300 mobile applications available for downloading on the Google Play store under the 'Games' category[140]. The list was then filtered manually for applications that could be suspected of providing access to suspected copyright-infringing content. Usually, an application has two types of names on Google Play market: (i) a human-readable title (name) such as 'Popular Music' and (ii) a machine name such as 'com.popular.music'. Normally, an average user looks for the first one, while the second one is used as a unique identifier. Even though there might be a few applications with the same title, their machine name will be different. Therefore, it is also important to use the machine name of the applications in searches to correlate dissemination of versions of popular applications altered to include malware and PUPs.

UNICRI also conducted targeted keyword searches in the applications section of the Google Play store. Each keyword search included two components: a) one of the following keywords: 'pirate', 'download', 'stream,' and 'free copy,' and b) one title from the list of sample titles. This resulted in four keyword searches per sample title. In addition to this, a list of selected machine names mentioned above is used to perform searches in popular search engines such as Google, Bing and Yahoo. These application names often appear on unofficial markets, promoting free access to popular official applications. Therefore, such applications can be regarded as suspicious.

The top 10 applications that were returned as the result of each keyword search were checked against the sample application list. Any application not already on the list was added.

As a final step, UNICRI manually checked the sample list of applications again with a view to ensuring that they were eligible for inclusion in this study and that they had all been added appropriately.

10.6 Phase V.A. Collecting malware and potentially unwanted programs on identified websites and mobile applications

The goal of Phase V was to collect samples of malware and PUPs that an average internet user might encounter when attempting to access suspected copyright-infringing content via suspected websites. A sandbox environment[141] with a Tor browser[142] installed was used to collect the malware. Explicit instructions for creating a safe sandbox environment were given by UNICRI to the expert conducting the collection of binary samples on behalf of UNICRI. The expert conducted the searches in a manner consistent with low security-awareness internet browsing. This included not using an adblocking[143] service, as well as clicking suspicious links and buttons. The Tor browser was configured to simulate the searches as being conducted locally from the respective sample countries. To do this, the configuration file was edited to explicitly define the desired country of the Tor exit network node[144]. Additionally, the expert composed an HTTP GET request[145] that included, inter alia, the corresponding

[139] See footnote 124.
[140] The list of the 'Most Popular' list of entertainment applications on the Google Play store only ranks applications up to the 300th most popular.
[141] A safe and protected virtual environment used to execute and study malware samples.
[142] Browser built upon the Tor anonymisation network to hide the true location of the user.
[143] Adblocking is a general term that describes a functionality of a browser designed to hide or block digital advertisements found on web pages. It can either be embedded in the internet browser or provided by an external plug-in that needs to be installed additionally.
[144] Using Tor did not have an impact on the results of the study because HTTPS protocol was used to access the search engines.
[145] HTTP GET is a Hypertext Transfer Protocol request method that is designed to retrieve a predefined set of information from a specified source.

country-specific search engine domain name, search keyword and English titles of the searched digital content. This allowed for the collection of any relevant geo-targeted data for analysis, recognising that the ranking position and popularity of each website differs from one EU Member State to another. Binary samples were collected over two stages.

Manual collection. This method involved manually reviewing the domains identified in Phase IV. Using manual collection, the expert was able to simulate the experience of an average internet user (in general less aware about online security threats) by clicking advertisements and interacting with websites that required prompts. Additionally, screenshots were taken over the course of the manual binary sample collection for additional analysis.

Automated collection. This method employed an automated web crawler designed by the expert to follow all available links on a designated suspected copyright-infringing website. First, on any given website, the crawler would first collect information from the links on the home page. Second, the crawler would follow each of those links to secondary websites. Third, the crawler would follow each of those links to tertiary websites. At each step, the crawler would retrieve binary files that could be of interest for subsequent manual analysis, including potential or suspected malware and potentially unwanted programs.

This process continued for up to 1 000 links per website, in order to eliminate any possibility of a crawler not working correctly due to an overwhelming number of links. The web crawler was unable to process JavaScript, which is occasionally used to hide links or other content on web pages.

As at 28 July 2017, 5 240 websites were automatically checked. Altogether, 617 binary files (both executables and media content that might hide malware and PUPs) of a total size of 47 GB were retrieved. This unsorted batch of binary files required further analysis to determine whether the collected binary files were relevant for the research. For the collection of mobile applications, a similar procedure was performed using an extracted list of application names and suspected copyright-infringing domain names.

As at 28 July 2017, it performed 2 425 951 web page visits and retrieved 490 unsorted files (including both binaries, video and audio files) with a total size of 19.1 GB.

UNICRI carried out average user searches for each of the 20 sample titles on the websites suspected of infringing copyright[146].

10.7 Phase V.B. Analysing the binary samples.

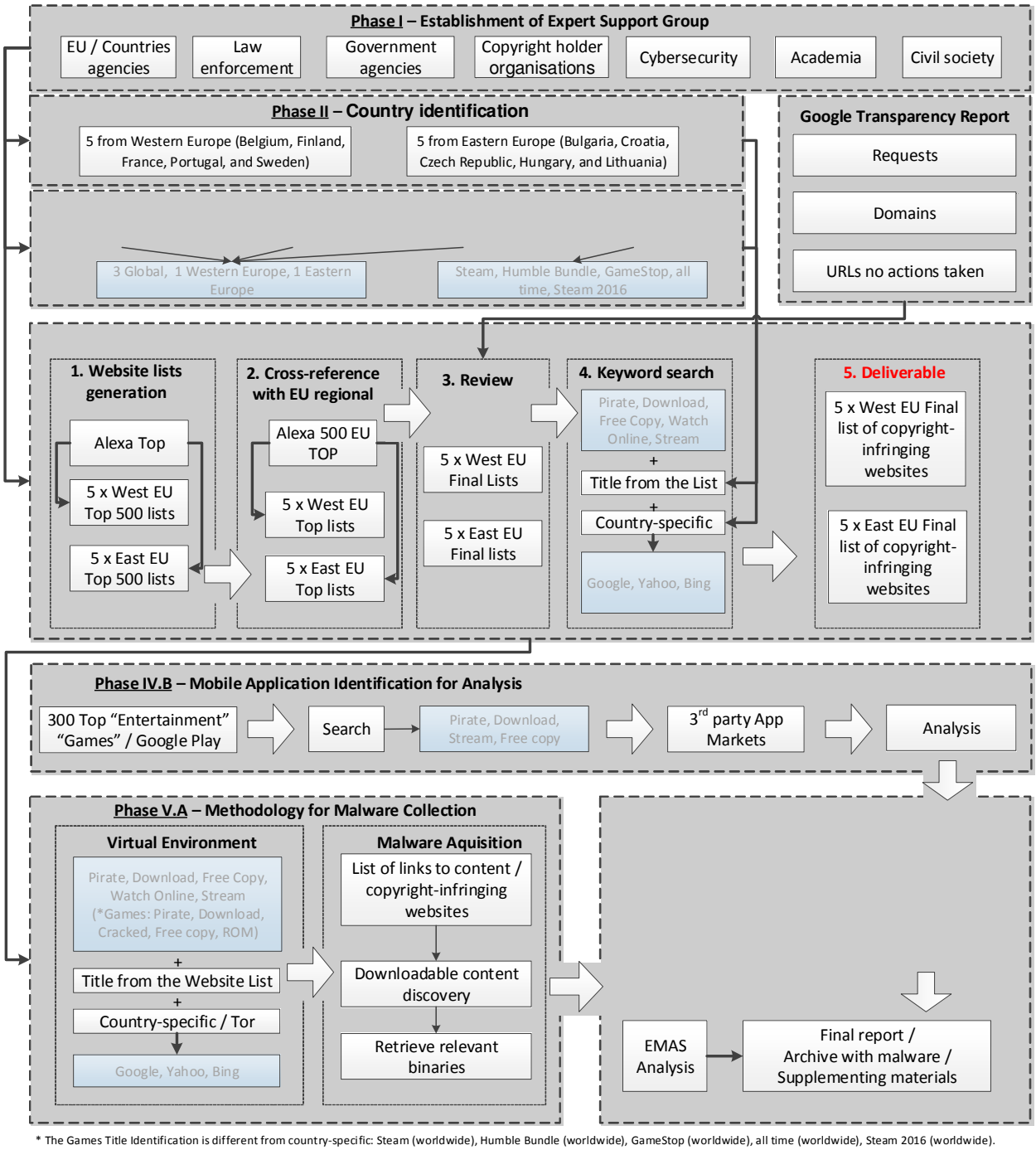
Once the binaries were collected, they were analysed in a safe environment for proper categorisation. Preliminary analysis was carried out using open-source tools to be able to correlate findings with cyberthreat reports. It included static and dynamic analysis, highlighting the corresponding artefacts that indicate malicious or potentially unwanted activity. Static analysis is related to processing static features of files such as size, specific content, etc., while dynamic analysis may reveal specific functionality logic upon execution of the binary. Desktop and mobile applications were analysed in corresponding virtual environments to reduce the risk of malware infection of the research equipment. To carry out the in-depth analysis, UNICRI sent the binary samples to the Italian postal police, who in turn passed this data on to the European Cybercrime Centre (EC3) at Europol via the SIENA tool. The samples were analysed by EC3 in the EMAS sandbox environment, and the results were returned via

[146] Regarding computer operating systems and web browsers, Windows 8 and Internet Explorer were used as a baseline. UNICRI explored the possibility of enlarging the search to OS and web browser options. The language of the searches (and therefore of the OS and browser) were also considered. Web browser security settings in terms of allowing third-party cookies and allowing redirects (e.g. pop-up windows) were defined, as well as using Java and Flash to simulate a standard PC.

SIENA to the Italian police, and subsequently to UNICRI. The origin, threat level, and functionality of each piece of malware or otherwise unwanted software were documented by UNICRI. When analysing the origin of the malware and PUPs, attempts were made to identify the family and campaign and classify whether its purpose was to disrupt users' systems upon downloading suspected copyright-infringing content, or if the intention was to gather user data in order to obtain personal information.

10.8 Overview of methodology

A detailed visual overview of the six stages of the methodology can be found below in *Figure 20*.



* The Games Title Identification is different from country-specific: Steam (worldwide), Humble Bundle (worldwide), GameStop (worldwide), all time (worldwide), Steam 2016 (worldwide).

Figure 21: Overview of the project methodology with corresponding flow of information between different phases

11. Annex 2: Qualitative and Quantitative Analysis of Research Findings (Extended Version)

11.1 Binary collection — Round I: weeks 26-29 of 2017

11.1.1 Phase IV.A. URL collection from the Alexa Top 500

Countries: Belgium, Bulgaria, Croatia, Czech Republic, Finland, France, Hungary, Lithuania, Portugal, and Sweden.

Google Transparency (GT): the datasets retrieved on 26 June 2017 include collections of removal requests, URLs, and domain names.

Step 1. Manual analysis of the Alexa Top 500 websites with statistics across EU countries and overall ranking was performed to select top ranking websites for each of 10 countries. Croatia was not in the original dataset, so the ranking was based on the number of EU Member States in which a particular website is popular. The websites in the Top list were sorted in two ways: ascending global popularity index and descending number of EU Member States where this website was denoted as popular. This is due to the fact that some of the country indexes have numbers that are 9999999 because of insufficient data.

Step 2. 500 website domains were extracted from the previous step resulting in 10 datasets with the most used domain names per country and also an additional dataset for the most popular across the EU.

Step 3. Each of the domains in the list for each country had been checked against Google Transparency. Google Transparency contains information about removal requests with corresponding domains for which removal has been requested. The statistic files also contain a number of requests for each domain that can be found in the GT domain database.

Step 4. During this step, additional domain names associated with the content titles were gathered using search results from popular search engines, including Google, Yahoo, and Bing. An expert-assisted crawler was used together with a manual search for selected titles and selected keywords in the abovementioned search engines. **In total, 300 search result pages** *(20 titles x 5 keywords x 3 search engines = 300) per country were found. On the search result page of each search engine, the top 10 search terms and their associated links are listed. In addition, the initial filtering of these results excluded the following resources: Google, YouTube, Bing, Yahoo, Amazon, Blogger, HBO, Netflix, Spotify, and Microsoft, since these are considered to be non-copyright-infringing websites.*

For the collection, a Tor browser was used with an explicit definition of country of exit node to similar user activity from *the target country. In addition, the corresponding country-specific search engine domain was used to perform searches. IP addresses and ISP details were logged for future study. However,* Tor network connectivity and reliability is low and it frequently reconnected, meaning that a small fraction of results could be missing because of this.

***Croatia**:* during the searches, it was discovered that hr.search.yahoo.com is neither registered nor represented in this country, while bing.hr stands for some local company. The search results on.com domains of these search engines show no specific attribution to the country.

***Bulgaria**: similar to Croatia,* the searches revealed that bg.search.yahoo.com is neither registered nor represented in this country, while bing.bg stands for some local company. The search results on.com domains of these search engines show no specific attribution to the country.

Step 5. List of domains from Steps 3 and 4 were compiled together, which resulted in a final set of domains of interest that are both popular in selected EU Member States and also represented in search results in three major search engines.

Table 10 reflects a number of website domains selected after performing each step in Phase IV.A. Step 4 shows the number of new domains that were not in the Alexa 500 Top List per country, yet were identified during searches for copyright-infringing content using search engines. The domains of Netflix, Facebook, and Twitter, inter alia, were excluded from the search results. Moreover, selected domains are also in the Google Transparency database of requests to delete content. Step 5 shows the number of domains selected during week 26 to be searched later for possible malware species.

	Step 2	Step 3	Step 4	Step 5
Belgium	500	387	213	600
Bulgaria	500	316	117	433
Croatia	500	308	123	431
Czech Republic	500	318	204	522
Finland	500	323	213	536
France	500	397	253	650
Hungary	500	325	194	519
Lithuania	500	318	209	527
Portugal	500	385	212	597
Sweden	500	336	219	555

Table 10: Identification of selected websites for each step in Phase IV.A (Round I)

Top 10 countries and domain suffixes per country

The figures below present the distribution of countries of website hosting locations and domain name suffixes. To find out the country of web hosting location, each domain name was resolved to retrieve the corresponding IPv4 address of the web hosting. Then, each address was checked against the GeoIP database provided by Maxmind[147]. It can be seen that the overwhelming majority of the websites are hosted in the US. This might be explained, in part, by the fact that the major corporations that offer web hosting services tend to store websites in data centres located in both continents, whereby Amazon has six AWS regions in the US and only three in Europe[148]. There is also a significant prevalence of .com domains with almost no websites having country-specific domain suffixes, such as .bg or .be. This might be explained by the fact that .com is a commercial domain that had been open for general public registration. In this way, virtually everybody can register and use it. However, country-specific top-level domains might require residence, a registered trade mark or an organisation in the country.

[147] Maxmind, 'Geoip products'; retrieved from https://dev.maxmind.com/geoip/.
[148] Amazon, 'AWS Global Infrastructure'; retrieved from https://aws.amazon.com/about-aws/global-infrastructure/.

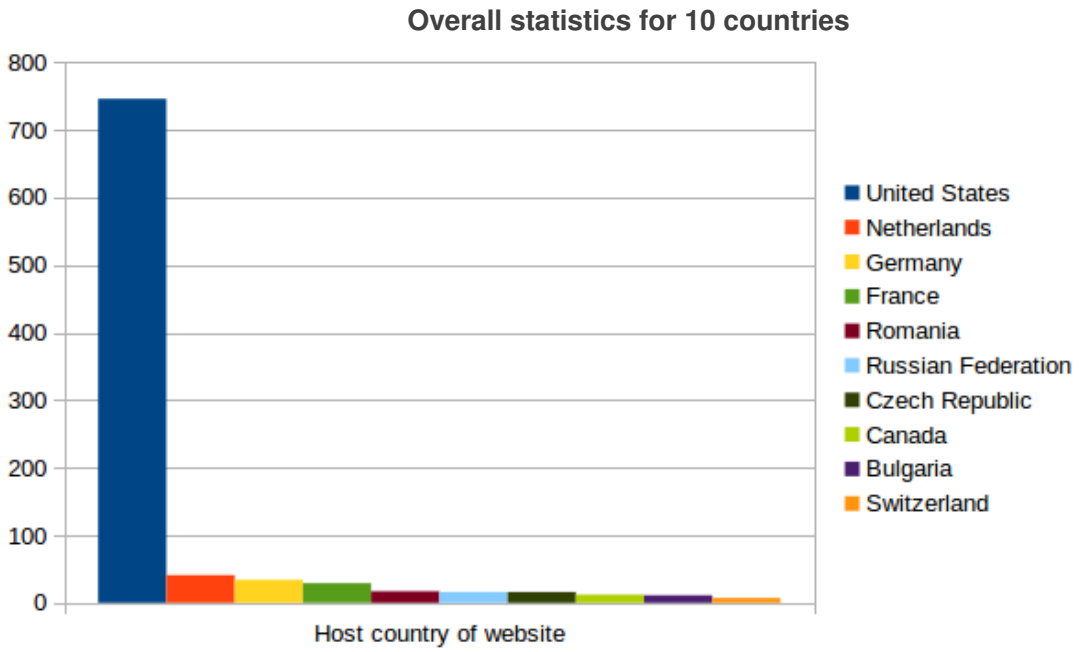


Figure 22: Distribution of host countries of websites added during Round I of malware collection for all countries

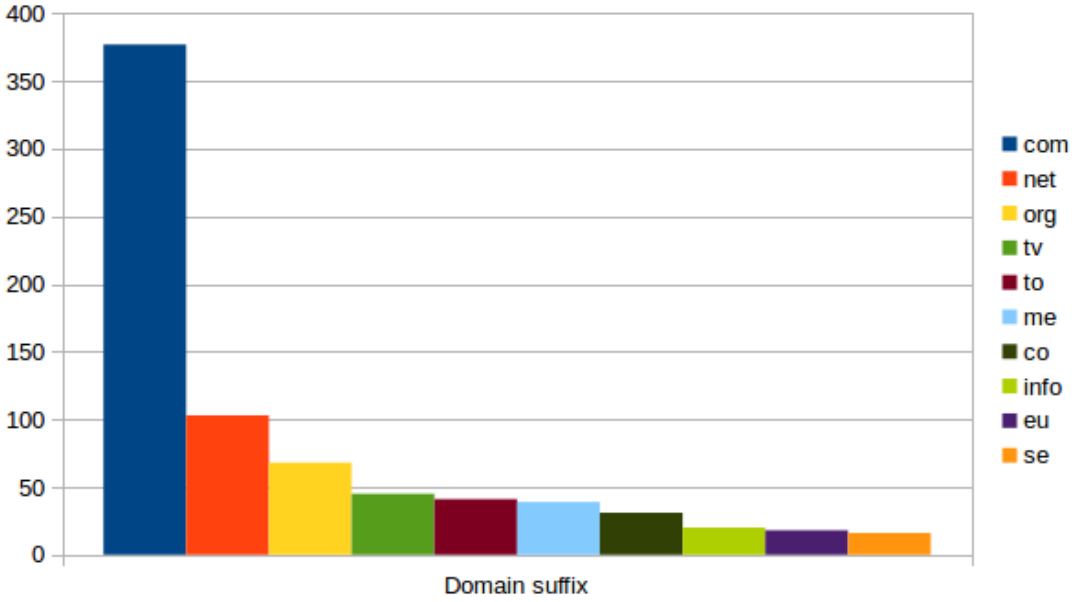


Figure 23: Distribution of domain suffixes of websites identified in Round I of data collection for all countries

Belgium

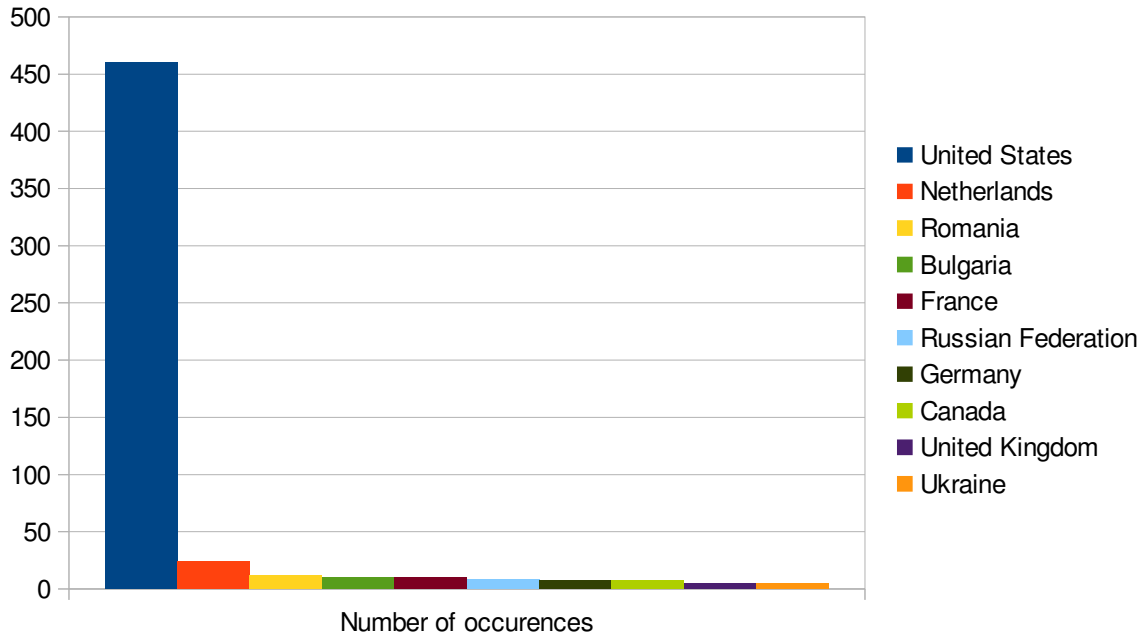


Figure 24: Distribution of host countries of websites identified in Round I of data collection for Belgium

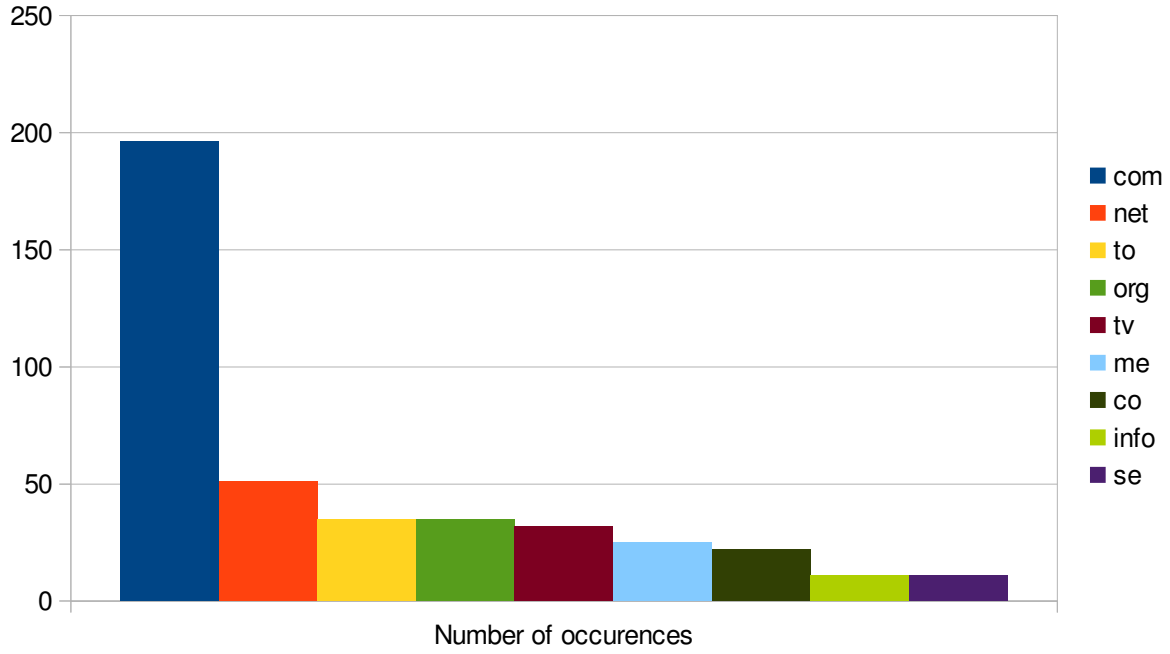


Figure 25: Distribution of domain suffixes of websites identified in Round I of data collection for Belgium

Bulgaria

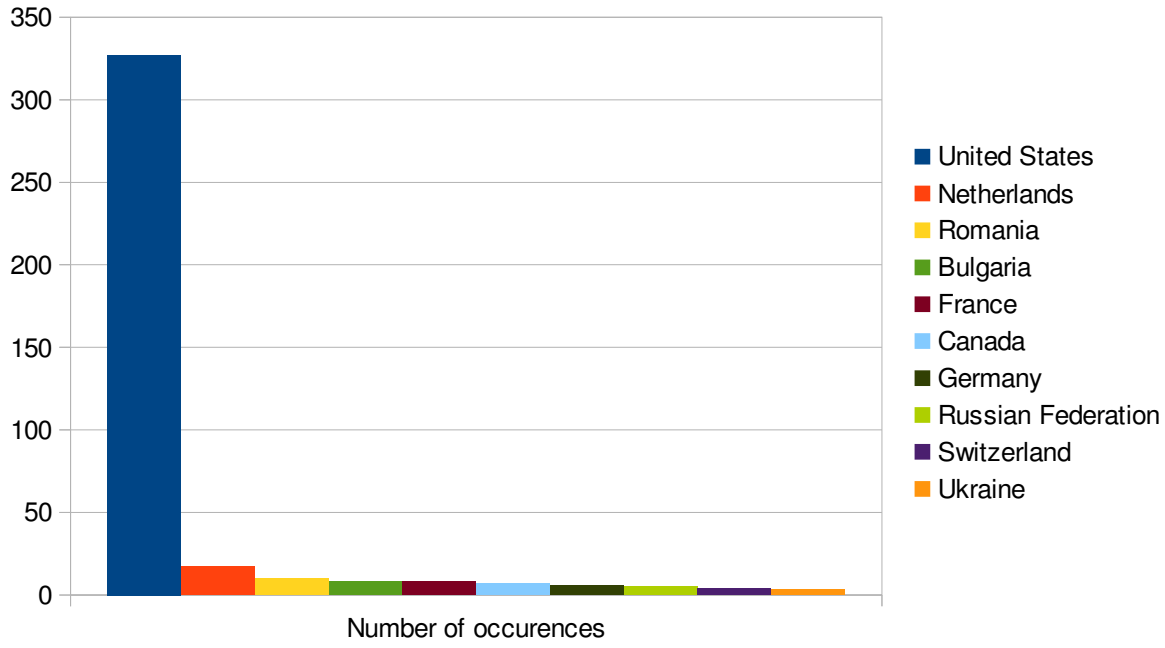


Figure 26: Distribution of host countries of websites identified in Round I of data collection for Bulgaria

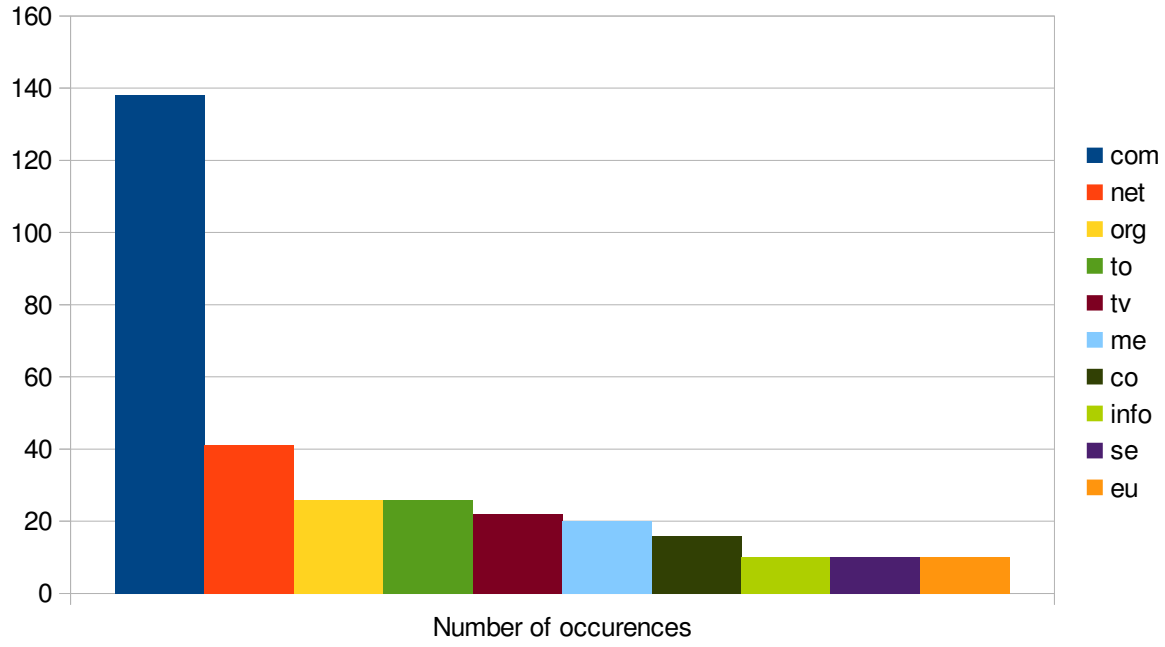


Figure 27: Distribution of domain suffixes of websites identified in Round I of data collection for Bulgaria

Croatia

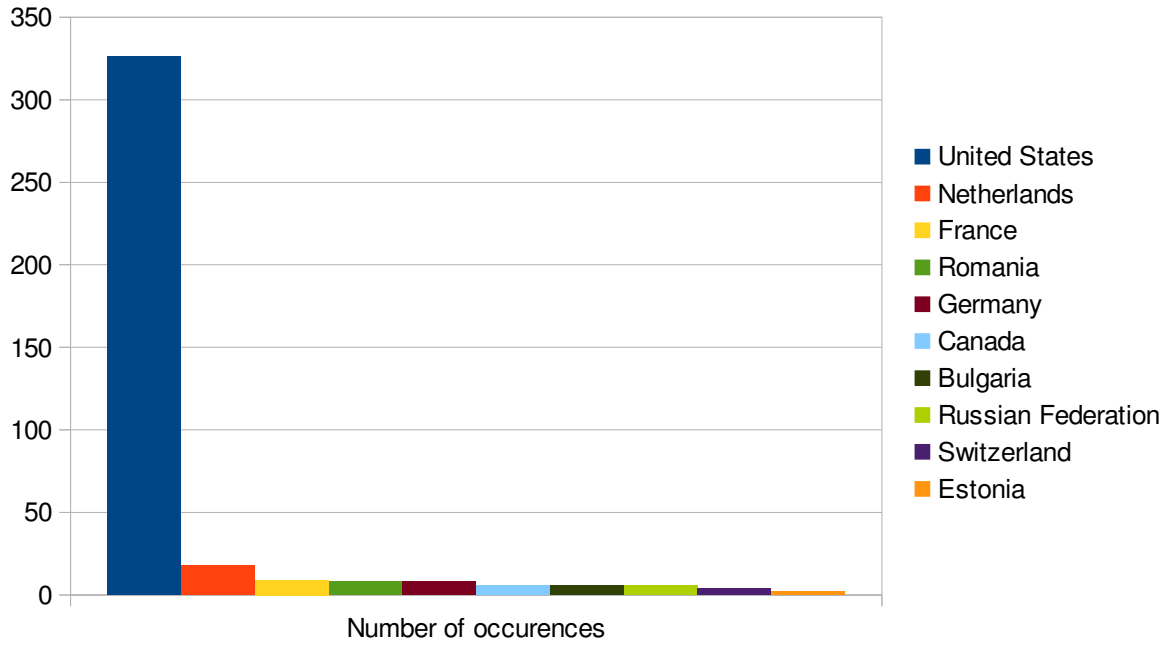


Figure 28: Distribution of host countries of websites identified in Round I of data collection for Croatia

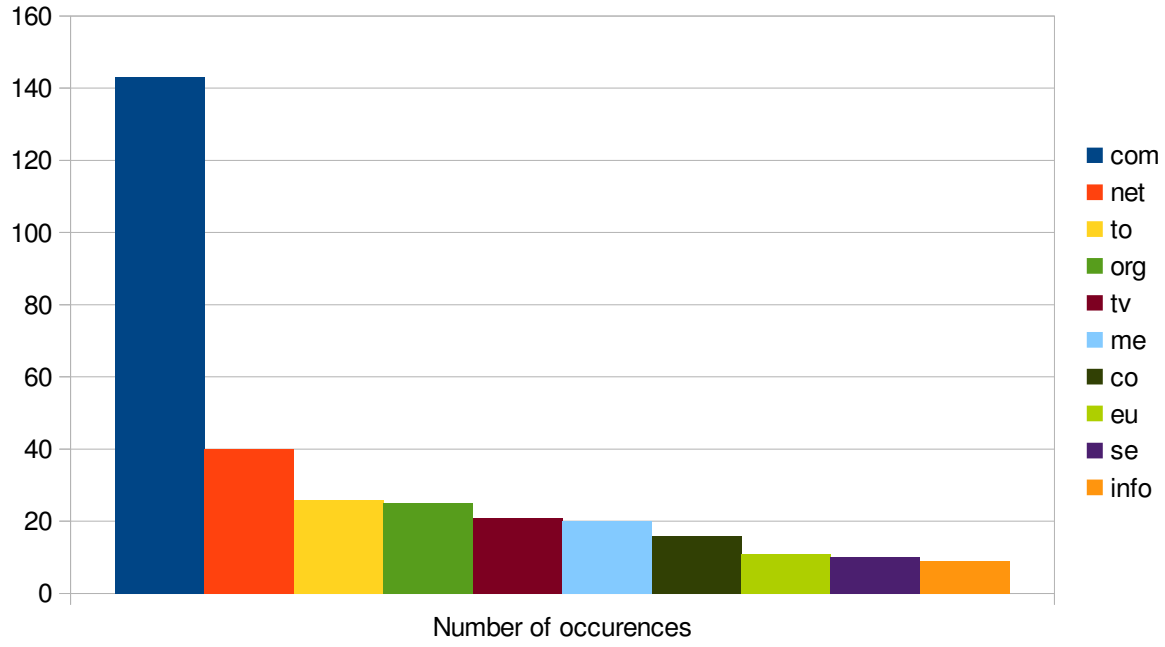


Figure 29: Distribution of domain suffixes of websites identified in Round I of data collection for Croatia

Czech Republic

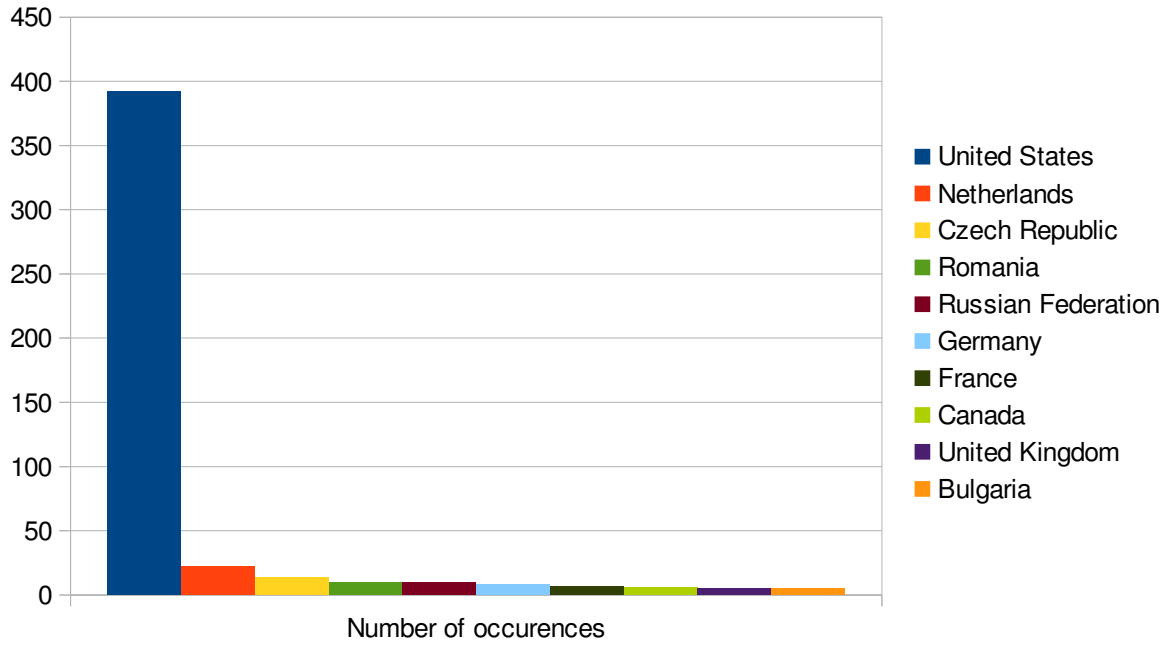


Figure 30: Distribution of host countries of websites identified in Round I of data collection for the Czech Republic

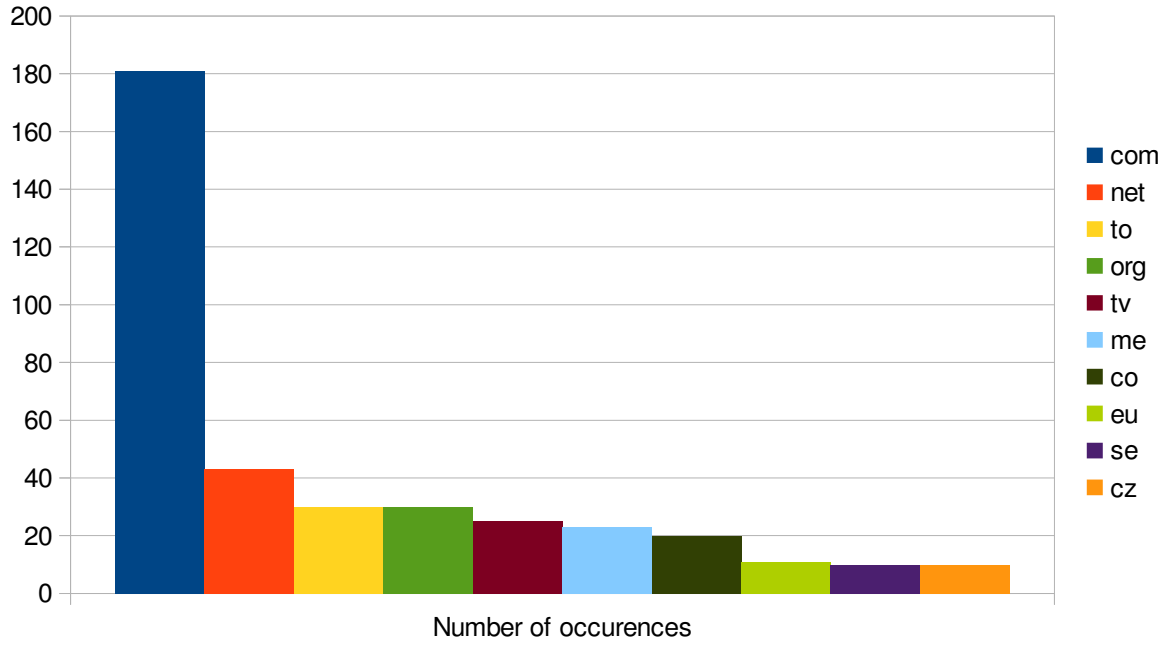


Figure 31: Distribution of domain suffixes of websites identified in Round I of data collection for the Czech Republic

Finland

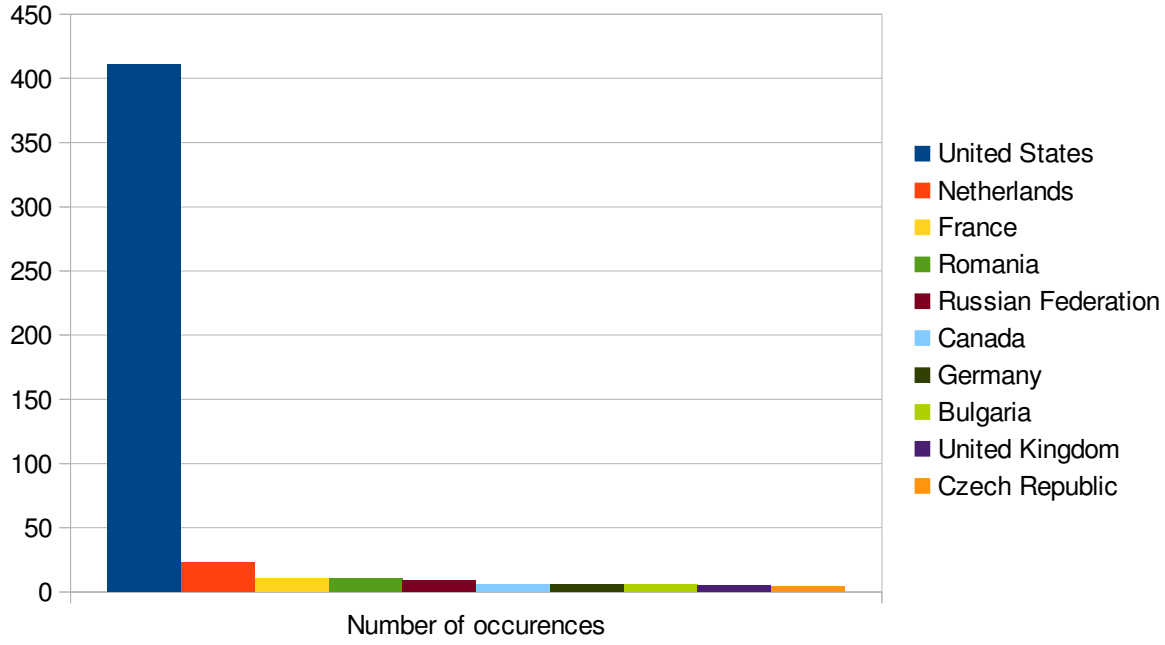


Figure 32: Distribution of host countries of websites identified in Round I of data collection for Finland

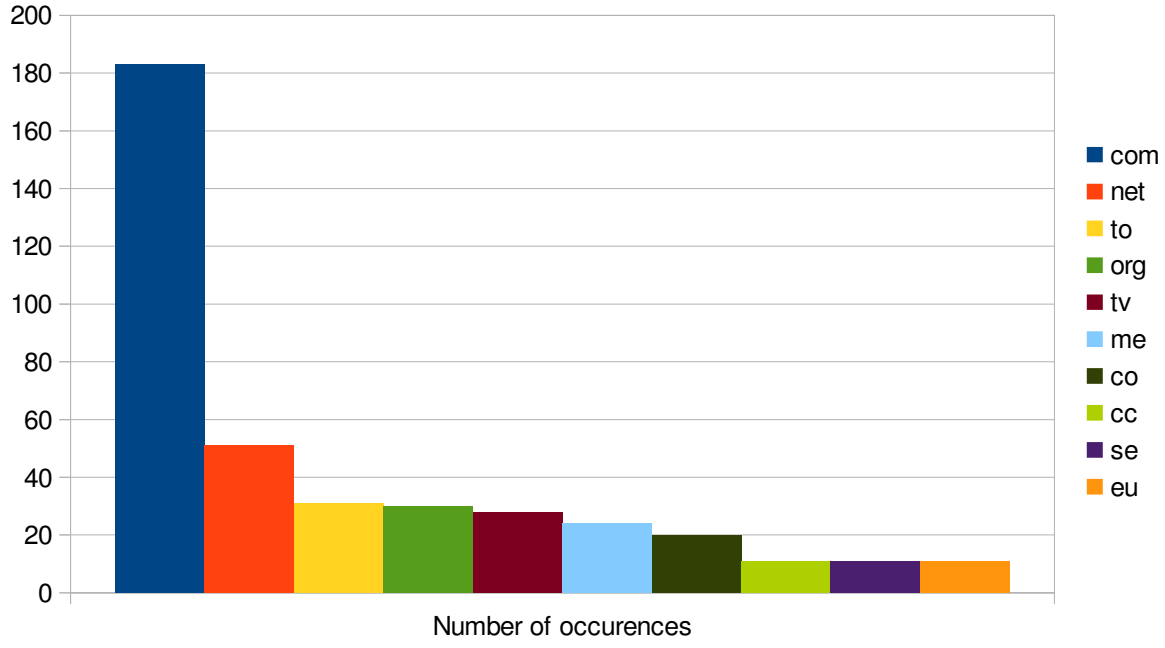


Figure 33: Distribution of domain suffixes of websites identified in Round I of data collection for Finland

France

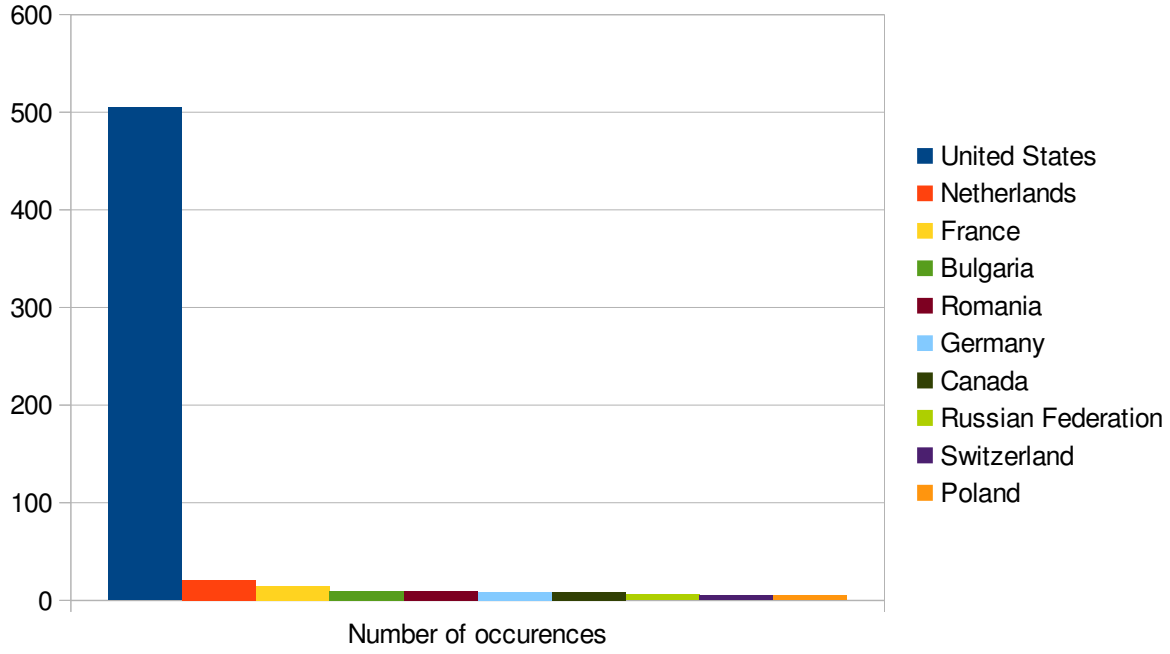


Figure 34: Distribution of host countries of websites identified in Round I of data collection for France

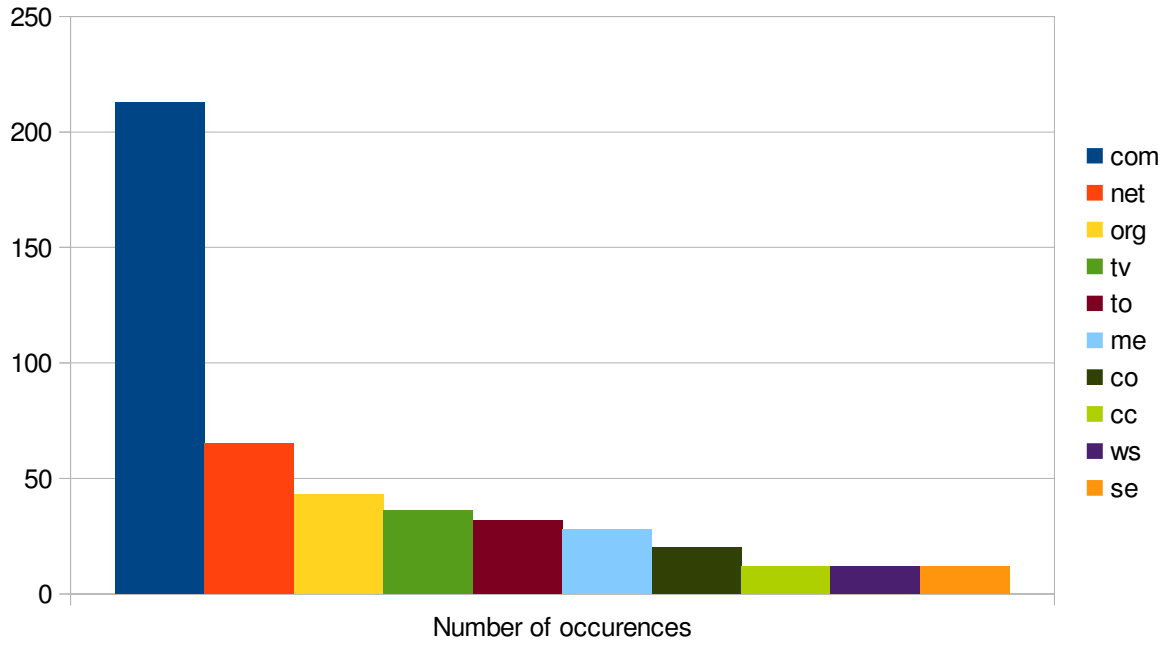


Figure 35: Distribution of domain suffixes of websites identified in Round I of data collection for France

Hungary

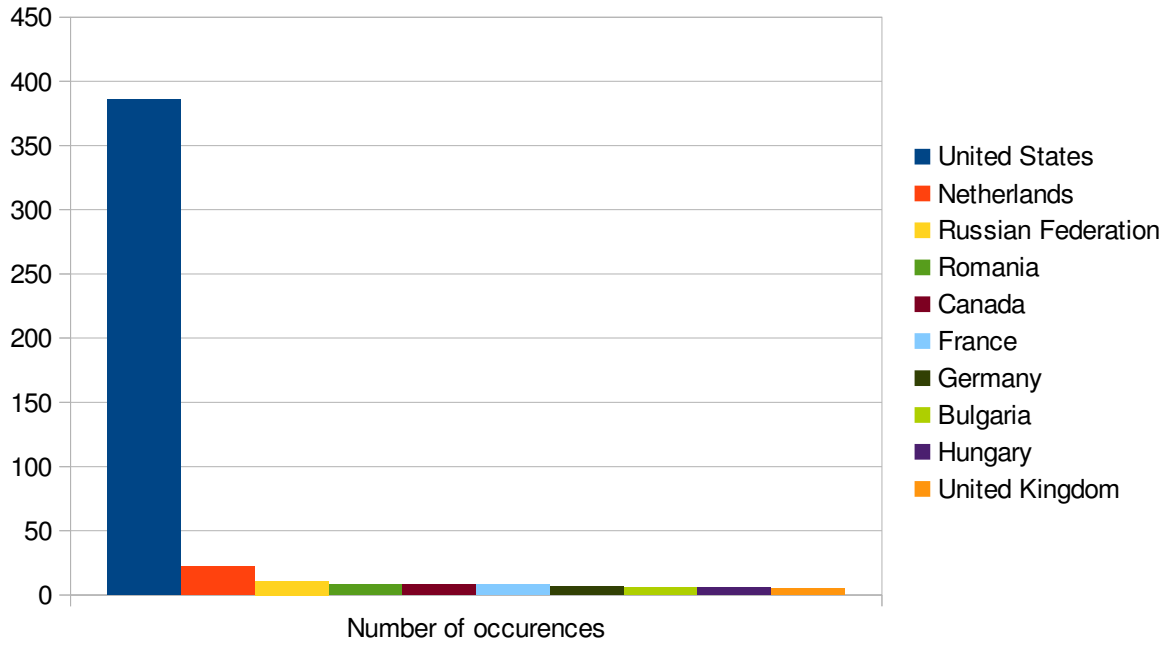


Figure 36: Distribution of host countries of websites identified in Round I of data collection for Hungary

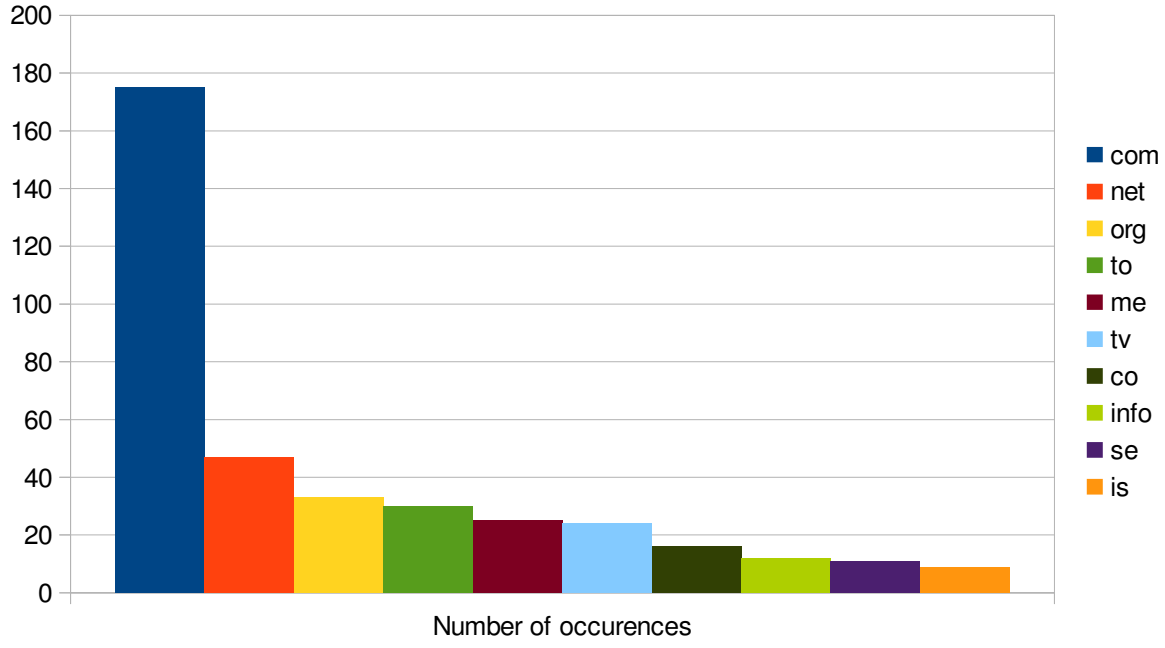


Figure 37: Distribution of domain suffixes of websites identified in Round I of data collection for Hungary

Lithuania

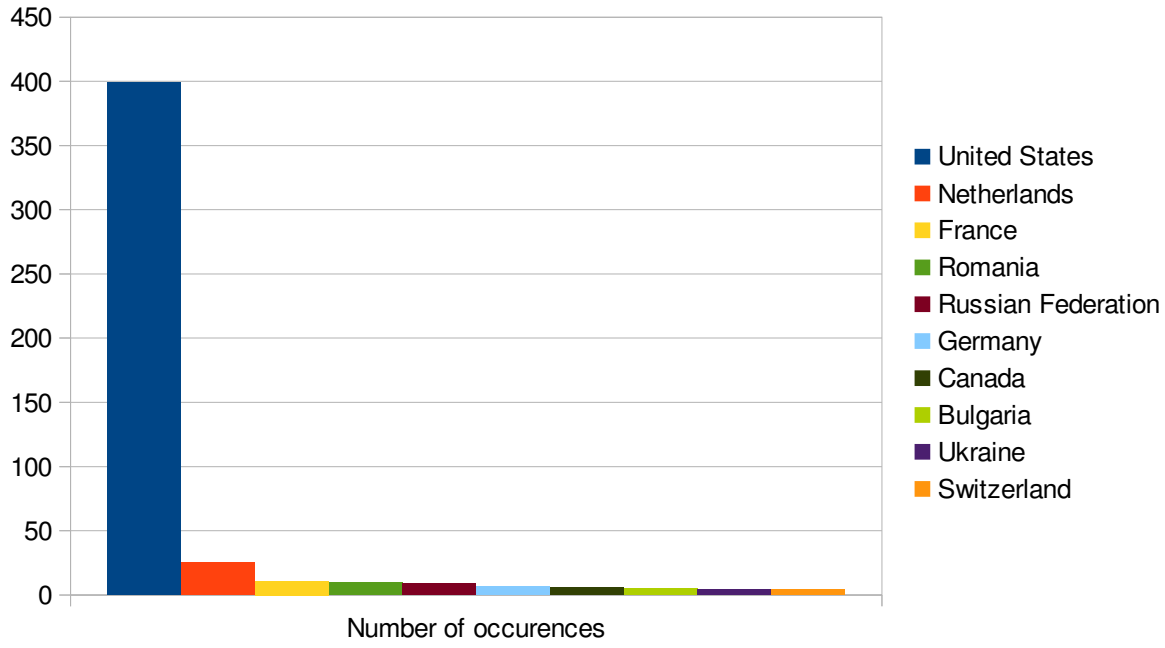


Figure 38: Distribution of host countries of websites identified in Round I of data collection for Lithuania

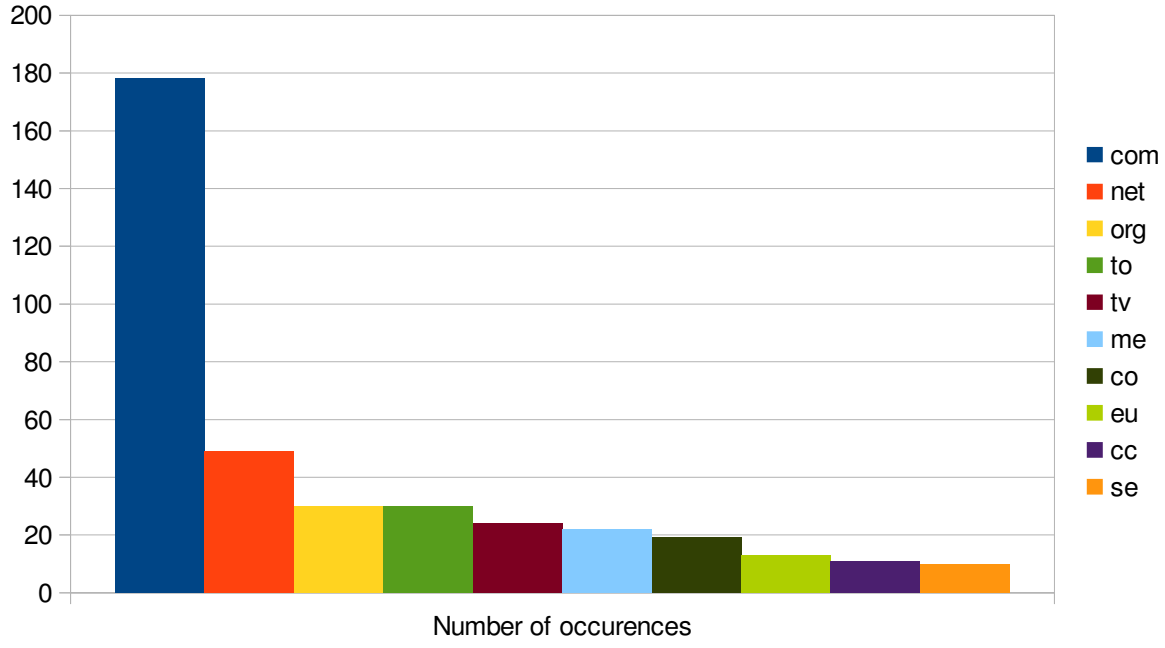


Figure 39: Distribution of domain suffixes of websites identified in Round I of data collection for Lithuania

Portugal

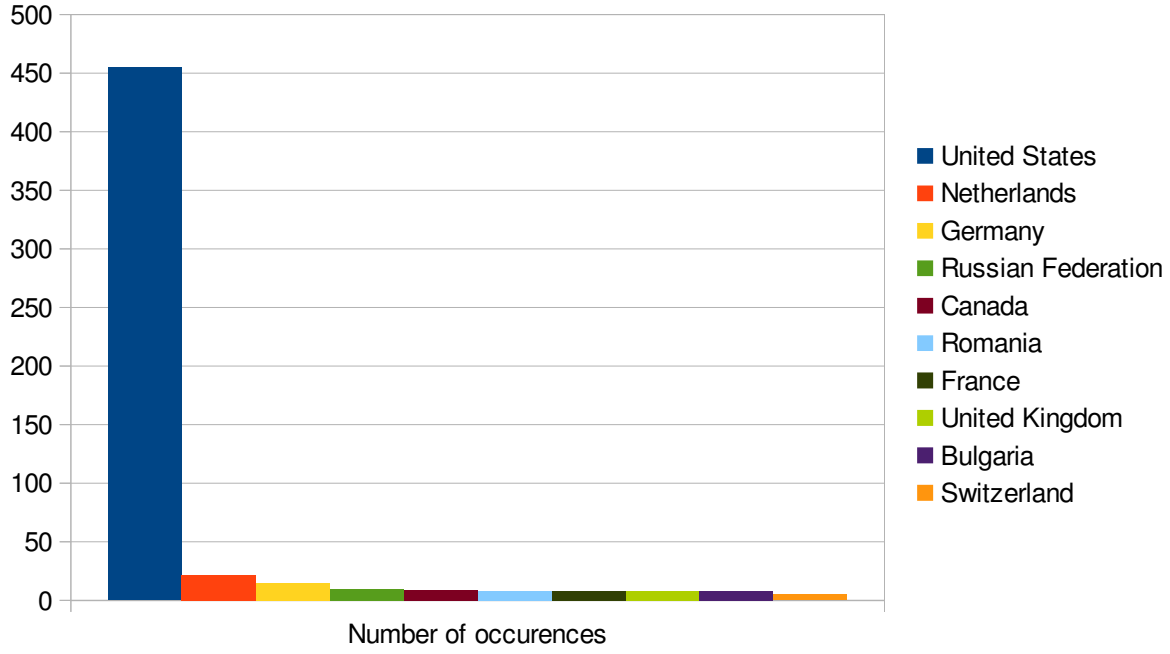


Figure 40: Distribution of host countries of websites identified in Round I of data collection for Portugal

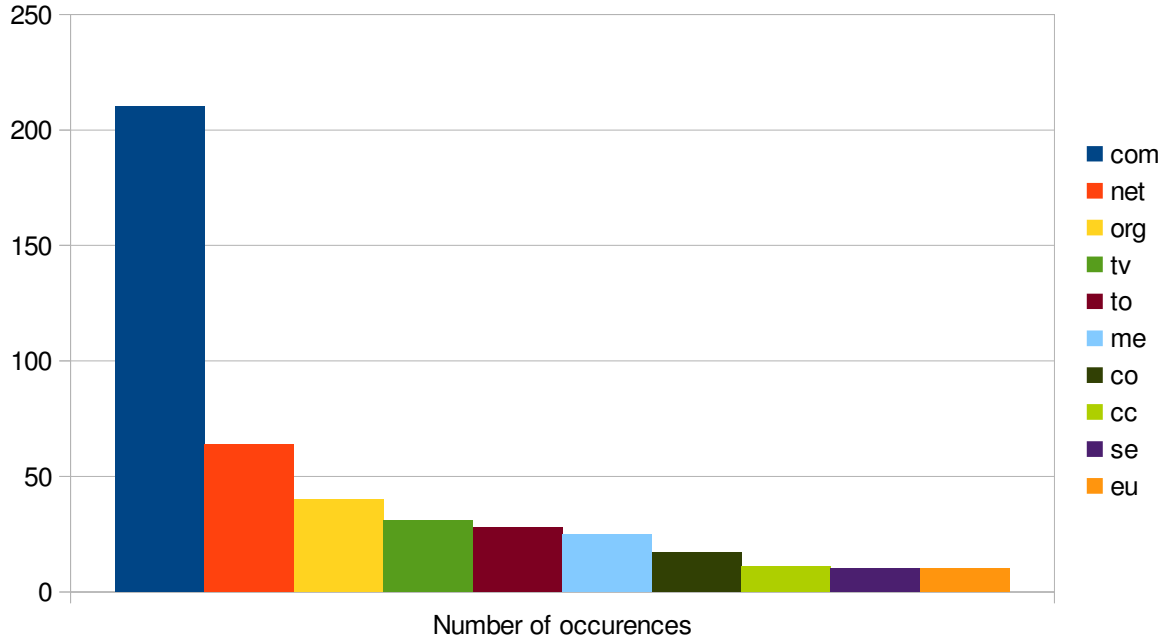


Figure 41: Distribution of domain suffixes of websites identified in Round I of data collection for Portugal

Sweden

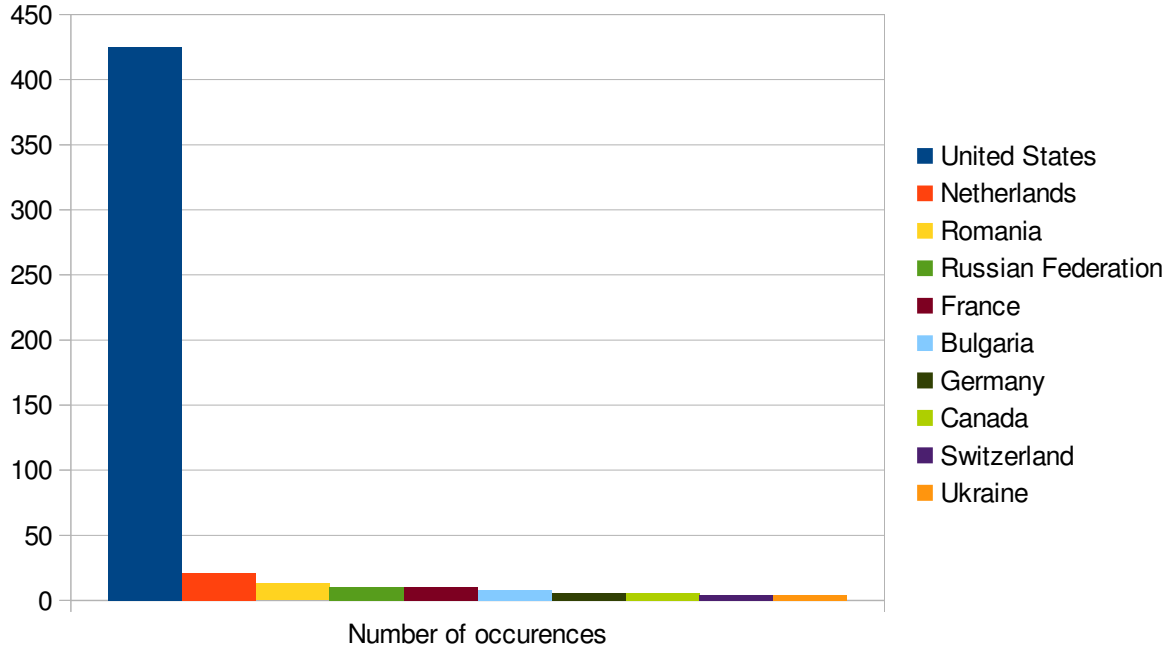


Figure 42: Distribution of host countries of websites identified in Round I of data collection for Sweden

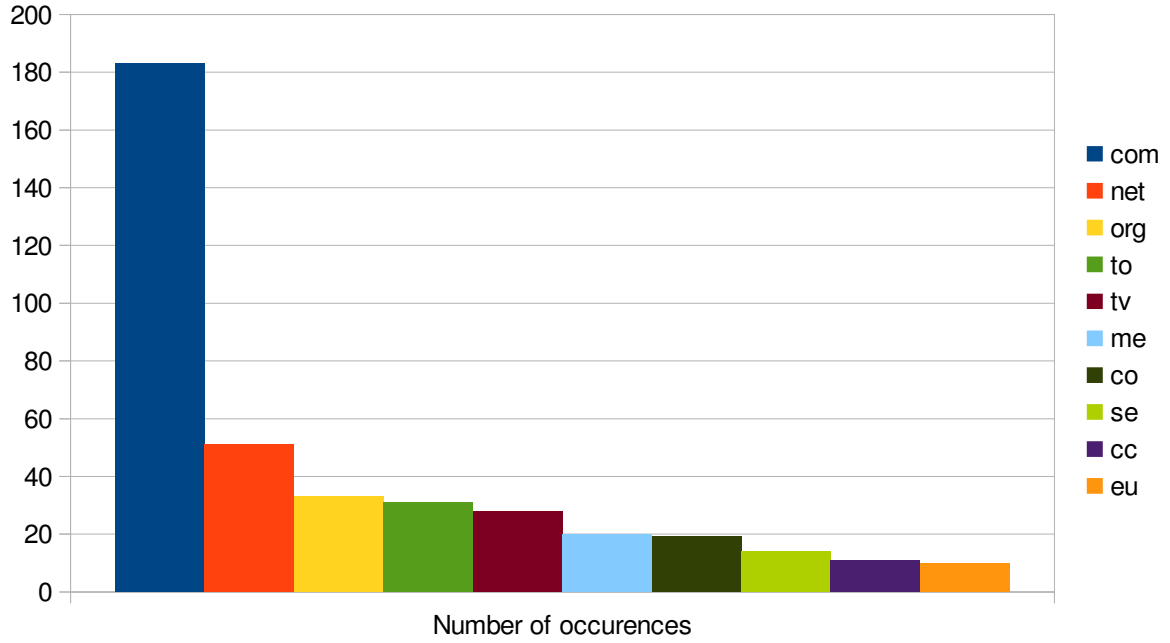


Figure 43: Distribution of domain suffixes of websites identified in Round I of data collection for Sweden

11.1.2 Phase IV.B. Mobile application identification

Step 1. During this step, 300 application names from the 'Entertainment' category were included in addition to 65 application names from the 'Games' category. In addition, 49 application names were retrieved from the 'Stream' category. In total, this resulted in 407 unique application names that are available on the official Google Play store.

Step 2. In this step, searches in Google, Bing, and Yahoo were performed to identify unofficial application markets and other relevant websites that provide access to the applications found during Step 1. In total, 763 unique domains were gathered that appeared in search results upon requesting an application name.

Step 3. Finally, a web crawler was used to retrieve any relevant files from the domains found. As at 28 July 2017, it performed 2 425 951 web page visits and retrieved 490 files with a total size of 19.1 GB.

Mobile binary analysis

After analysing some of the retrieved mobile applications, it was found that they use similar dissemination methods — that is, third-party application markets. They advertise access to popular applications from the official Google Play store, while these third-party applications found are not represented on the official market at all.

Website01

Filesize: 15 580 446 bytes
Acquired: 26 July, 23.29
Analysed: 28 July, 6.28

This Vietnamese web service offers access to different content that can be downloaded by users, including games and other applications. It offers applications for iOS and Windows Phone, as well as Android. The interface of the website looks similar to application markets, ranks different software and offers free access. In order to do so, a user is required to retrieve a Website01 application that serves as a gateway to other applications listed on the website. The application requires a wide range of permissions that may appear suspicious. For example, it can manage user accounts, access internet, modify and delete files on an SD card, read phone identity and access phone logs. Such an extensive list of permissions can be unusual and not related to an application's advertised functionality.

Website01 is not an official market place and is not represented on Google Play. Therefore, it is logical that the application requires such permissions to be able to handle installation. However, such information is at risk of being exposed to a third party because the user freely gives all these permissions. The Android installation package for this application market has been identified by eight anti-virus solutions as suspicious or as a Trojan. In this sense, there is a high level of confidence that the file is unwanted and may cause harm to the user's personal and sensitive information. Moreover, the installation file of the application is available for downloading without any registration. To install it, however, a user must enable 'third-party source' in the Android settings.

Website02

Filesize: 16 796 692 bytes
Acquired: 27 July, 1.32
Analysed: 28 July, 7.01

This website offers access to popular applications that are also available through the official Google Play store. There are unique subdomains, which are designed specifically for the application being offered. However, there is a single application that is being downloaded across all of the subdomains. Upon downloading this app, the user is asked to enable the option 'Unknown sources' in the security settings of the Android OS. Corresponding instructions clearly state on the website that free access to the Android installation file is offered.

This is not normally required for applications accessed from the official market. The number of requested permissions is large and very suspicious. For example, the application that provides access to other applications usually does not need to read the phone's state or access contacts, the camera or call logs. Apparently, it was designed to collect such information that is further used by developers for their own purposes. The application has very few activities and services. Obviously, judging by the requested permission set, one can conclude that sensitive user information will be at a high risk.

11.1.3 Phase V.A. Binary collection

In the binary collection phase, the intention was to retrieve any potentially malicious and PUPs that average users may encounter when looking for ways to access content on suspected copyright-infringing websites. Considering the large number of domain names per country, it is difficult to anticipate the number of binary samples that will be collected and particular types of malicious or otherwise unwanted software. The suspected copyright-infringing websites investigated in this study usually create specifically crafted web pages to make consumers think that they are on a legitimate page. To ensure reasonable collection of relevant data, the binary collection was separated into two stages.

Manual collection. This approach includes manual surfing of the internet while looking for content with selected titles in search results. This allows the researcher to mimic the average user experience and select links that are either malvertising or require direct human interaction to access. Moreover, this enables the researcher to take screenshots of the information displayed and analyse the relevance of the website content.

Automated collection. Contrary to the manual approach, automated malware collection is fast and does not require user interaction to browse and follow links on a web page. The first step of the collection employs an intelligent crawler that follows any available links on the suspected copyright-infringing websites. This is known as breadth-first search, where the links are checked first on a target web page, then links on the secondary web pages that this web page refers to, etc. The second step of the collection retrieves files that might be of interest for later manual analysis. The number of links checked per website was limited to 1 000 to ensure timely execution. Furthermore, a crawler does not process JavaScript because this can be done only on a fully operational user's browser. To mitigate this challenge, manual collection was performed.

As at 28 July 2017, 5 240 websites have been automatically checked, with 617 relevant files retrieved of an overall size of 47 GB. This unsorted batch of files requires further analysis to decide which collected files are relevant for the study.

After the list of copyright-infringing websites was identified, the web crawler performed searches for any relevant content on those websites. In total, this process resulted in the following statistics per country:

Belgium:	1 377 108 visits
Bulgaria:	1 013 305 visits
Finland:	1 307 036 visits
France:	1 249 964 visits

Croatia:	1 020 322 visits
Czech Republic:	1 262 995 visits
Hungary:	1 146 467 visits
Lithuania:	1 239 016 visits
Portugal:	1 311 809 visits
Sweden:	1 245 618 visits.

11.1.4 Phase V.B. Binary analysis

During this phase, the files that were collected on suspected copyright-infringing websites, according to the corresponding country, were comprehensively checked to see if they contained any malicious payload or caused any harm to the user's sensitive information or computer.

Website03

File size: 1 377 792 bytes
Acquired: 5 July 2017, 14.15 CEST
Analysed: 10 July 2017, 5.17 CEST

This website offers access to cracked games, ISO files, and other relevant data that users may look for on the internet. Games are sorted according to genres and users can also use the website search to find a particular game. Each web page provides multiple links to the game's files that the end-user needs to download. In most cases, the files that are being downloaded are of a small size (only a few megabytes), named as the original game and compressed using the zip archive format. Preliminary analysis shows that they are PUPs/adware. The user experience starts by opening a web page designed for a particular game, which includes all relevant descriptions and information.

By hitting a corresponding download button, one receives a small zip file that contains an executable application. Even though the system warns of an unknown publisher and potentially harmful software, it does not stop the users from executing the file. The graphic user interface of the installation software looks like a legitimate piece of software with a title and picture of the corresponding game. The only suspicious element at this stage is an icon of the software that might look rather general and not related specifically to this game.

Furthermore, the software shows an EULA that may appear legitimate to an unsuspecting user who does not read the terms that are mentioned. The user is also required to confirm that he or she agrees with the licence.

The file installation process is well-tuned and shows not only the files being downloaded, but also the status of the whole process. It is worth mentioning that there was no high-bandwidth network traffic activity from the application registered during this installation process. Moreover, it takes some time until the software moves to the next step, which tricks the user into thinking that the installation process is actually taking place.

Upon completion of the installation, the software requests a licence key, which is a logical step. The licence key is usually purchased together with the game and is required in order to be eligible to use this software product. It can be seen that the interface offers both the option to download or confirm the key and an explanation of how to obtain the key. Clicking the licence key 'Download now' button opens an overlay window with links to a survey. It is claimed that the file will be available once the user chooses to complete one of the surveys. This appears to be a normal procedure, similar to standard file-sharing services.

Website04

Filesize: 572 064 bytes
Acquired: 11 July 2017, 13.37 CEST
Analysed: 12 July 2017, 12.10 CEST

This website offers streaming of different TV series from a variety of streaming web services. The interface is in German. Links to various streaming services are provided. The page that contains, for example, the link to *Game of Thrones* has an interface that includes a description of the episode, pictures and streaming links. While accessing FlashX, the player window displays an overlay button reading 'Download Now' that further leads to an external web page.

Website05 has a very similar interface to a Microsoft web page, including the location of the elements, the text and many of the colours. This may trick users to think that they are on a legitimate web page and that the software is benign. Furthermore, advertised software on the website offers a functionality to fix the common problem of unknown drivers. This problem appears when a user uses hardware for which the drivers are not found in MS Windows. Therefore, file driver-updater-setup.exe is downloaded after pressing the 'Download' button. The installation process appears realistic and displays a logo that illicitly reads 'Microsoft Partner'.

After the installation is complete, the Website05 Driver Updater starts downloading information from the internet and scans drivers that are installed in the system. Once the diagnostics are complete, the user is alerted to the fact that there are some outdated drivers, for which 'useful' software is offered that purports to improve the system's performance and resolve any outstanding problems.

After claims about updating corresponding drivers and installing the recommended software to improve system stability, the process completes the installation of several new programs with apparently legitimate names and a credible design that looks like professional software that can be trusted. So, by installing Driver Updater, the user is tricked into accepting three additional pieces of software of questionable usefulness.

Website06

File size: 359 424 bytes
Acquired: 7 July 2017, 10.15 CEST
Analysed: 13 July 2017, 19.21 CEST

This website provides access to content that is described as user-posted. Therefore, the website's disclaimer says that it has no responsibility for such materials. Each game's web page has a common set of features: picture, description, video preview, and download button.

It also appears that these small files are crafted to look like a game installer by having the corresponding name, built-in picture of the game, and relevant information. Once the .EXE file has been extracted, the user is prompted to choose the language for the game that is being installed. The next four steps are consistent with a typical installation procedure: welcome message, folder selection, link creation, and installation. The program appears to be professionally designed and looks like realistic game installation software because it uses the picture of a game, its name and its relevant description. For unexperienced users, all software dialogue windows may look trustworthy. It basically simulates the downloading process. However, no high-bandwidth activity was registered during the malware analysis.

What is interesting is that, during the installation, the software creates an ISO file of the game, which bears the corresponding name and has quite a large file size. Moreover, once the file has been created, it has a size of only a few megabytes. During the downloading process shown above, the size of this file

increases until it reaches the size of the legitimate game. According to the article on the size of the DiRT 4 installation[149], the game occupies 32.26 GB on the disc. The file that has been created is 33.6 GB. This is very similar and indeed seems to be a true game, except for the fact that nothing has been downloaded. Finally, the simulation of the downloading process takes around 40 minutes[150], so it looks trustworthy to an average user. The actual time can vary based on environmental checks that the software may perform if considered appropriate by malware developers.

The web page enables users to download a licence key file with a size of 0.1 MB using the provided link. However, before this, a user has to go through a survey web page to unlock the file.

Finally, on the survey page, the user is required to provide a mobile phone number to receive an SMS and complete registration. It is claimed that this provides mobile virus protection. There is a mobile phone number validity check that does not let the user go to the next step unless the check is correct.

Website07— Mac OS

Filesize: 319 160 bytes
Acquired: 13 July 2017, 5.02 CEST
Analysed: 20 July 2017, 17.02 CEST

This online service offers a wide variety of TV series that can be watched and downloaded using the website. Each of the web pages represents a specific episode of the TV series being accessed. However, there are several links that simply do not work. In addition, the download button redirects users to a different advertisement web page that offers complementary 'useful' software. Once the 'Download in HD' button is pressed, one of the advertisements offers a Firefox plug-in to be added to the browser.

Another advertisement leads to the sounding web page that also includes the self-explainable name of the OS for which it is designed. It claims that the software can clean the Mac OS and has been downloaded by millions of users. It has an attractive interface and claims that it is the number one Mac utility in the world. Once the installation pkg file has been downloaded and launched, the software proceeds through several steps of the installation process, including the general information about the software, EULA and privacy policy. A range of useful features can be viewed, such as Internet Security, Memory Cleaner, and Data Encryptor. Notably, these features look more like general names rather than specific applications or known trade marks. Furthermore, the installation takes some time and performs additional suspicious activities.

Although the software offers a 'useful' functionality, it has been known to trick users to buy the extended version that neither improves the performance of the system nor provides a full set of advertised functionality. According to one article[151], there had been legal actions[152] against a company because the program named problems on the computer that did not exist and generated false notifications that tricked users into buying the upgraded 'better' version of a certain software.

[149] Gamerheadquarters, 'Dirt 4 install size'; retrieved from http://articles.gamerheadquarters.com/article803dirt4installsize.html
[150] '40 minutes' is a preprogrammed time that is included in the simulation.
[151] Bucher, A., 'Class Action Lawsuit: ZeoBIT Dupes Users into Buying MacKeeper Upgrade', *Top Class Actions*, 7 May 2014; retrieved from https://topclassactions.com/lawsuit-settlements/lawsuit-news/26392-class-action-lawsuit-zeobit-dupes-users-buying-mackeeper-upgrade/.
[152] Bucher, A., 'Class Action Lawsuit: ZeoBIT Dupes Users into Buying MacKeeper Upgrade', *Top Class Actions*, 7 May 2014; retrieved from https://topclassactions.com/lawsuit-settlements/lawsuit-news/26392-class-action-lawsuit-zeobit-dupes-users-buying-mackeeper-upgrade/.

Website08 — Android

File size: 14 146 848 bytes
Acquired: 13 July 2017, 5.31 CEST
Analysed: 21 July 2017, 9.08 CEST

This website offers a range of applications for Android free of charge. These can be downloaded easily using the links provided. This particular application offers access to a specific application that has streaming capabilities of recent TV series, films, etc. This amounts to copyright infringement since the application gives free access to copyright-protected content without the authorisation of the rights holder. The download link for the application is highlighted on the web page and unexperienced users may consider it to be a legitimate installation. However, the 'unknown applications' option in Android has to be enabled because the website is not an official Google Play store. Android provides additional security assurance to users by disabling this option.

For the analysis, Google SDK with Nexus and the API 18 emulator were used, running Android 4.3 using ARM. Of particular note is the fact that the application cannot be installed on the x86 platform and indicates an error in which a suitable ABI[153] has not been found. Once installed, the software presents a variety of categories of content that can be downloaded or streamed. The content is categorised according to TV series, episodes, etc. The user can also watch trailers and read different supplementary materials. For example, there is the option to download an episode of *Game of Thrones* directly to the phone in HD quality. The downloading process takes some time and the file is saved in the download folder on the Android device. Further, the metadata of the files shows that it is an H264 compressed video file that is easily played on a computer and it has advertising content included in the application. It can be concluded that the software offers completely free access to copyright-infringing content without requesting any user-specific information.

11.2 Binary Collection — Round II: weeks 30-32 of 2017

The second round of the website identification and binary collection was conducted shortly after the first round. Nonetheless, there are differences in the websites identified for use in data collection between the two rounds. This is attributable to updates to the Google Transparency Report, changes in the search engine results, websites shutting down, and the introduction of new malware campaigns, among other factors. The following sections provide detailed information on:

1. websites identified during the second round of data collection,
2. a comparison of the results of website identification between the first and second round of data collection,
3. the malware collection process during the second round of data collection, and
4. analysis of malware identified during the second round of data collection.

11.2.1 Phase IV.A. Suspected websites identified during Round II

The second round of website identification was conducted in the same manner as the first round, and as described in Phase IV.A of the methodology.

Table 11 reflects the number of website domains selected after performing each step of Phase IV.A during the second round of website identification. Step 2 includes cross-checking of country-specific lists of websites from the Alexa Top 500 with the regional EU list. During Step 3, website lists are checked against Google Transparency Report to remove websites without any reported copyright

[153] ABI – Application Binary Interface. The Android application checks the type of hardware being used. For example, it can be done to target only specific models of smartphones or tablets from a particular manufacturer.

infringement. Step 4 shows the number of new domains that were not in the Alexa Top 500 list per country, but were identified during searches for copyright-infringing content using search engines. Domain names that belong to well-known companies, such as Netflix, Facebook, Twitter, and others, were excluded from the search results. Moreover, selected domain names also came from the Google Transparency database of requests for deletion from the Google search engine results in relation to copyright-protected content. Step 5 shows the number of domains selected during weeks 30-31 to be examined later for possible malware. Overall, this process resulted in 5 606 total websites, including 1 057 unique websites for all 10 countries.

	Step 2	Step 3	Step 4	Step 5
Belgium	500	389	+219	608
Bulgaria	500	318	+142	460
Croatia	500	310	+137	447
Czech Republic	500	320	+239	559
Finland	500	325	+229	554
France	500	400	+265	665
Hungary	500	327	+246	573
Lithuania	500	320	+241	561
Portugal	500	387	+231	618
Sweden	500	338	+223	561

Table 11: Identification of selected websites for each step of Phase IV.A (Round II)

Two sets of figures are presented below showing the number of websites identified for each sample country during the second round of data collection. The first graph shows the distribution of websites for all sample countries based on the hosting country of the website. The second graph shows the distribution of websites for all sample countries by domain suffix.

Overall statistics for 10 countries

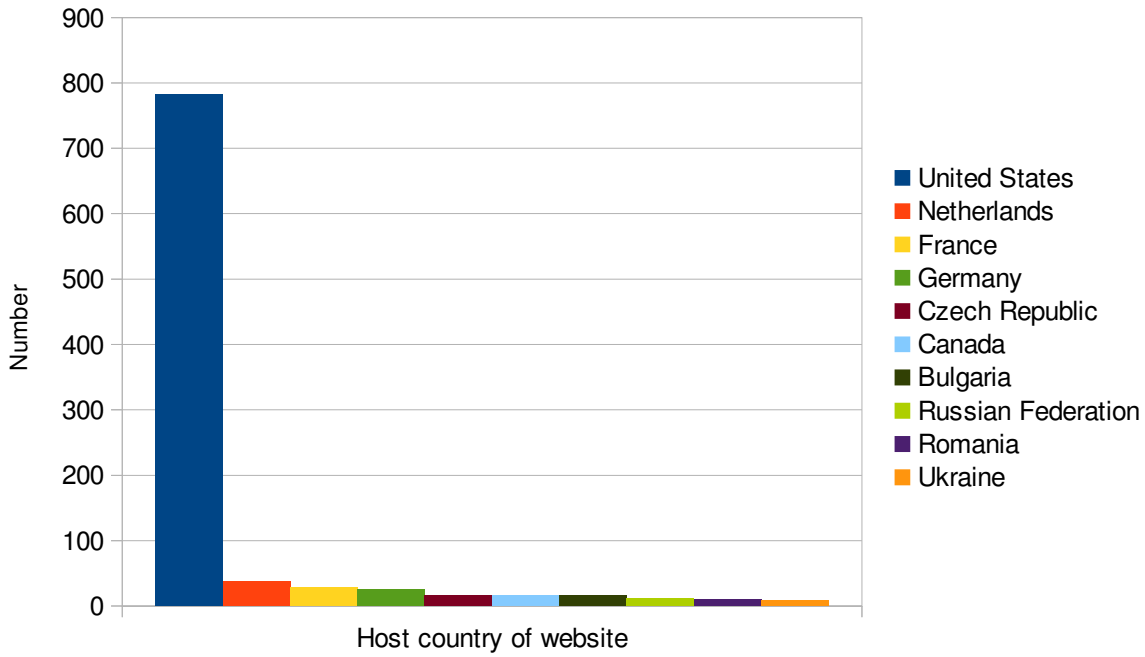


Figure 44: Distribution of host countries of websites identified in Round II of data collection for all 10 countries

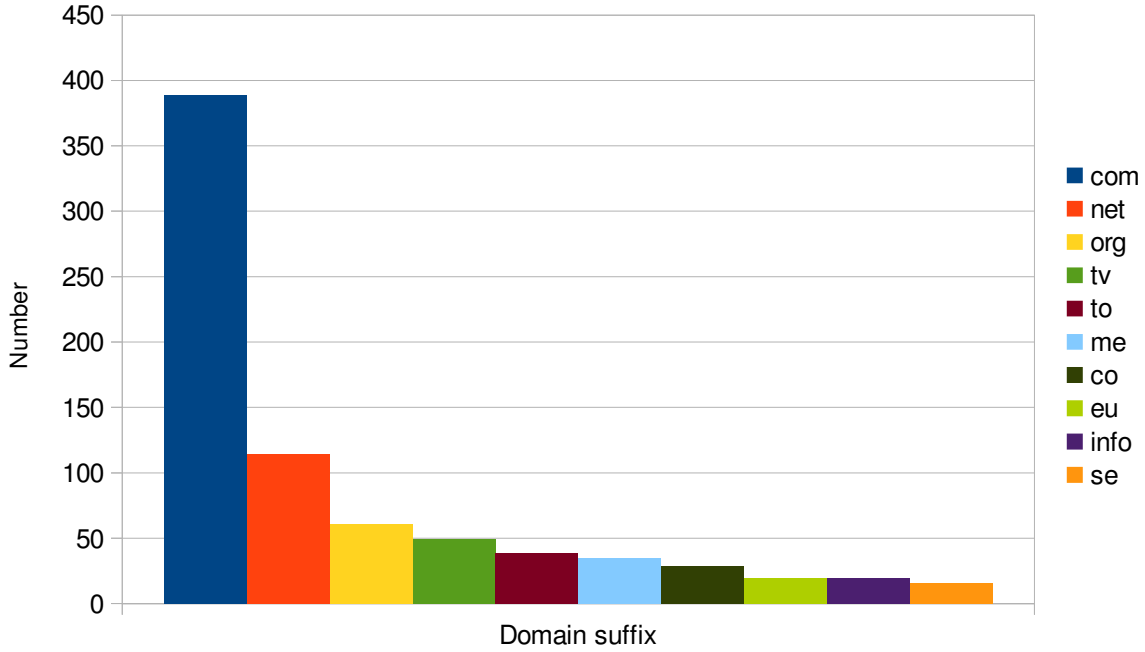


Figure 45: Distribution of domain suffixes of websites identified in Round II of data collection for all 10 countries

Belgium

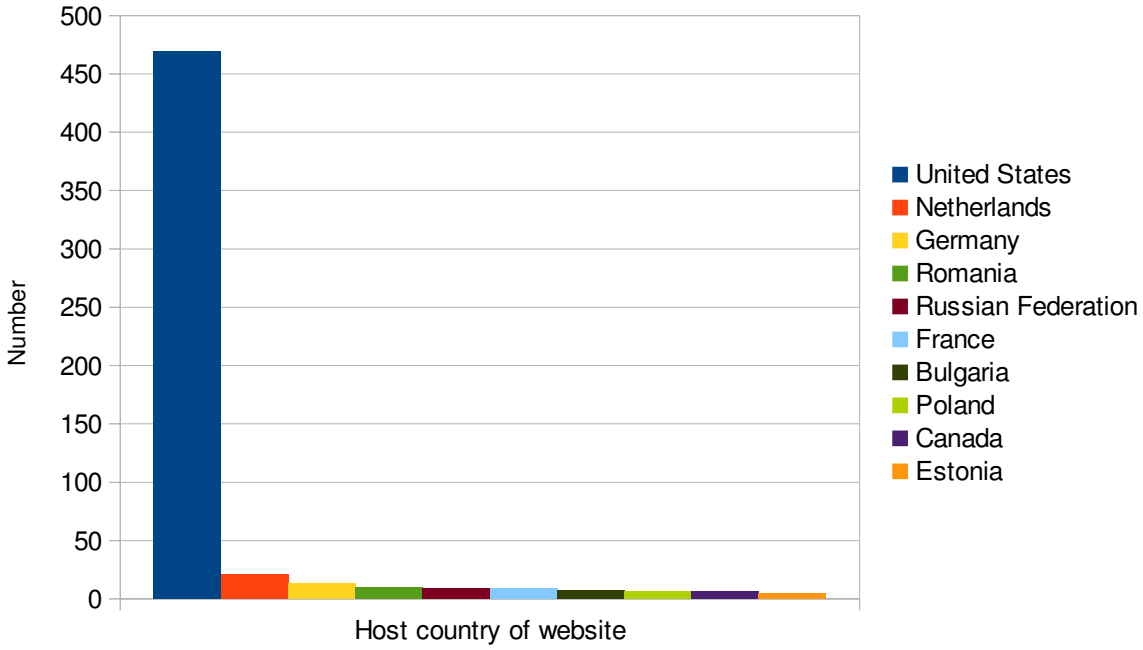


Figure 46: Distribution of host countries of websites identified in Round II of data collection for Belgium

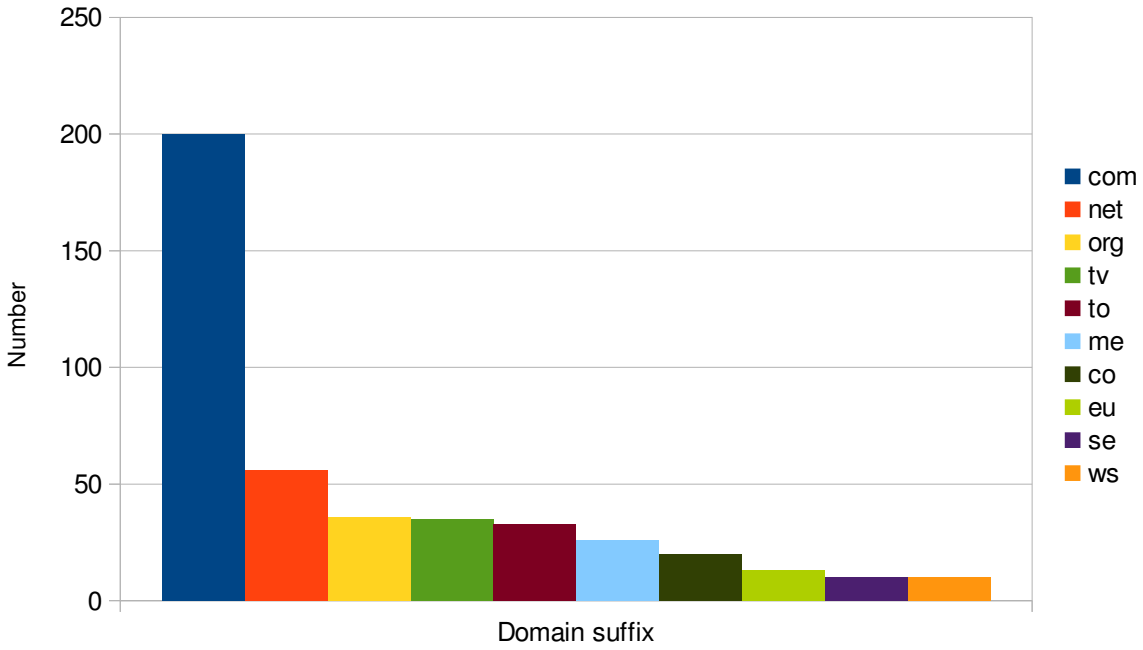


Figure 47: Distribution of domain suffixes of websites identified in Round II of data collection for Belgium

Bulgaria

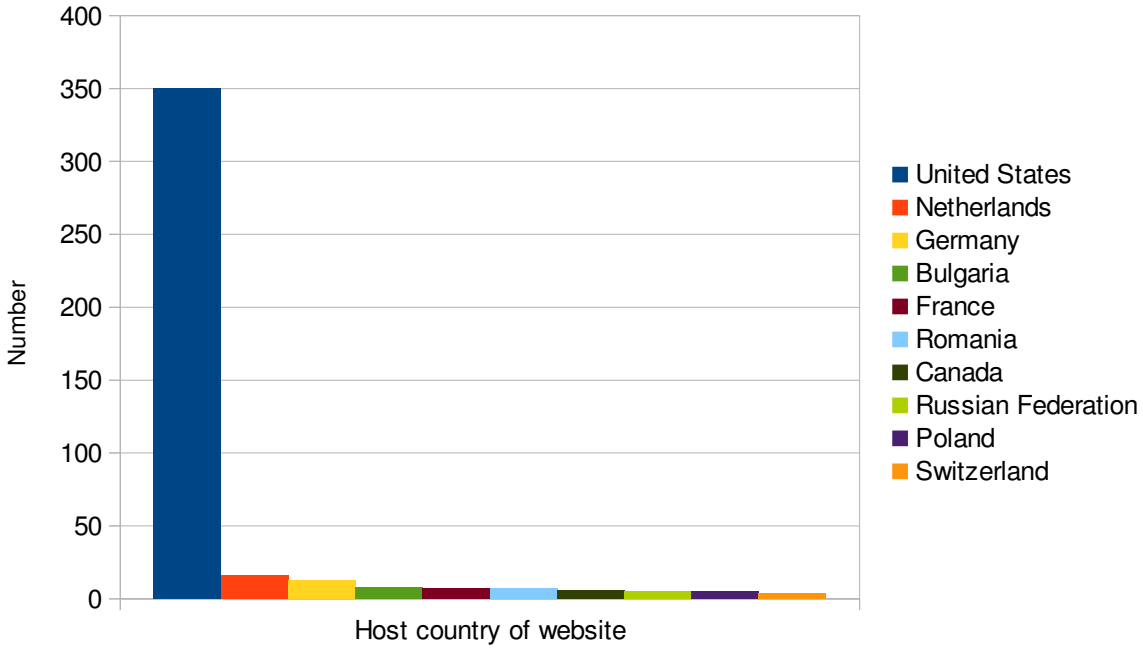


Figure 48: Distribution of host countries of websites identified in Round II of data collection for Bulgaria

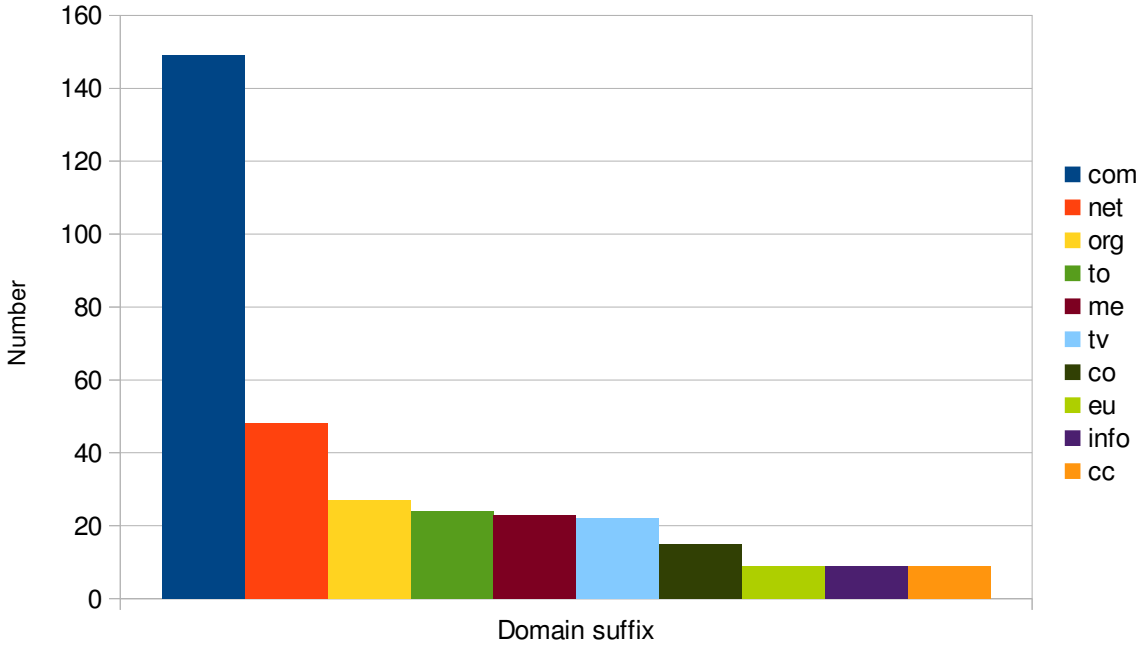


Figure 49: Distribution of domain suffixes of websites identified in Round II of data collection for Bulgaria

Croatia

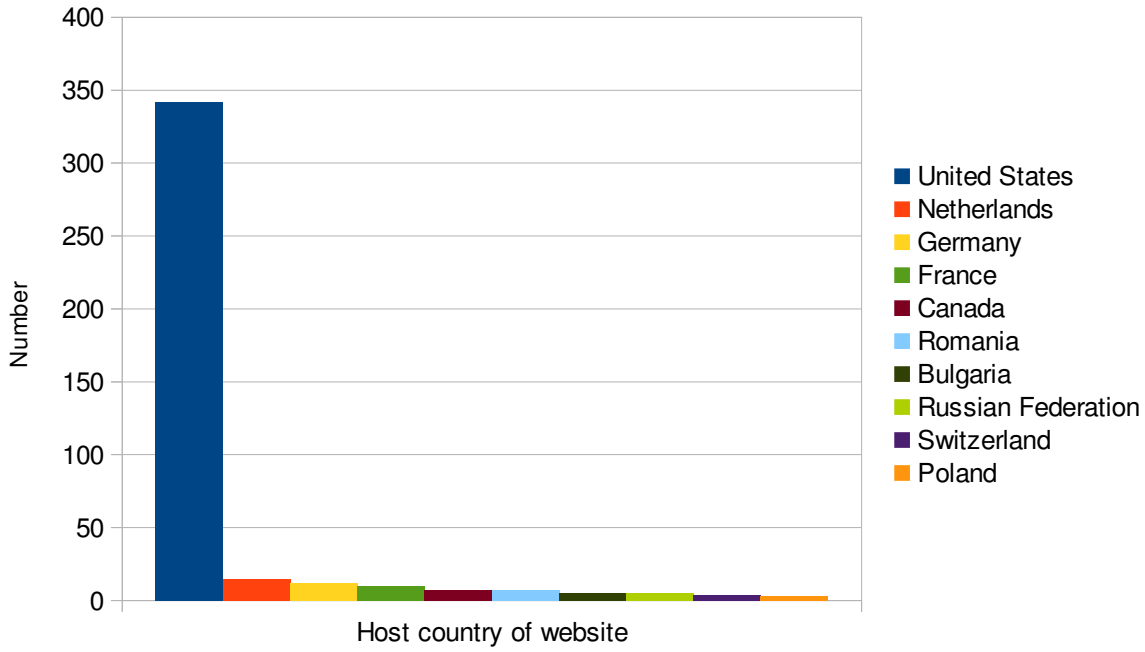


Figure 50: Distribution of host countries of websites identified in Round II of data collection for Croatia

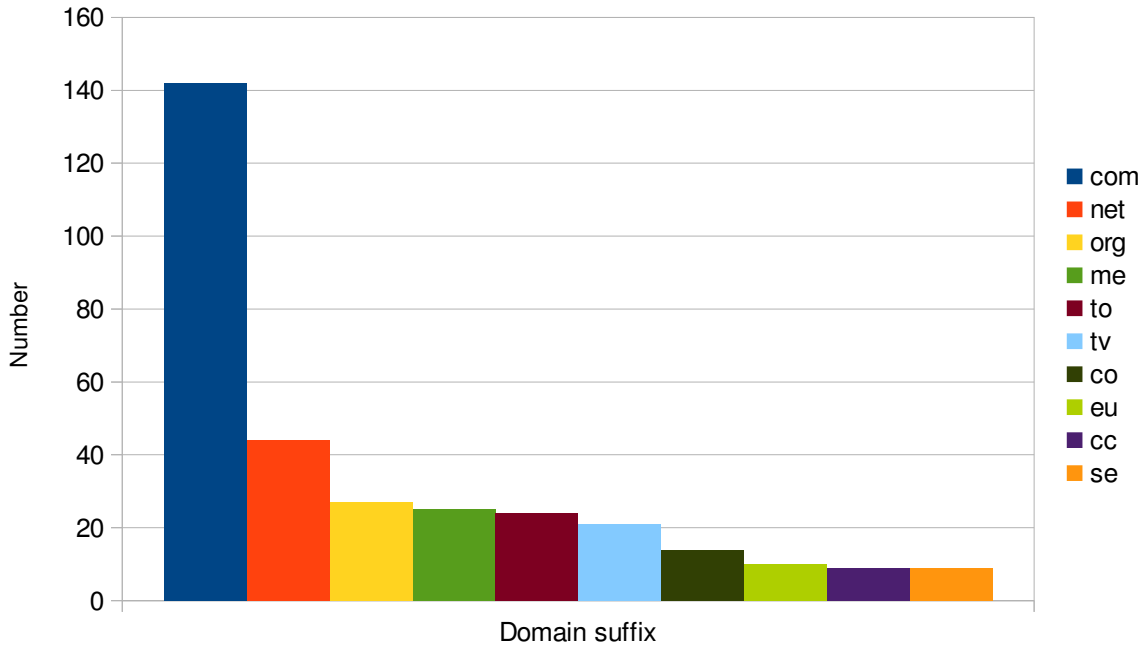


Figure 51: Distribution of domain suffixes of websites identified in Round II of data collection for Croatia

Czech Republic

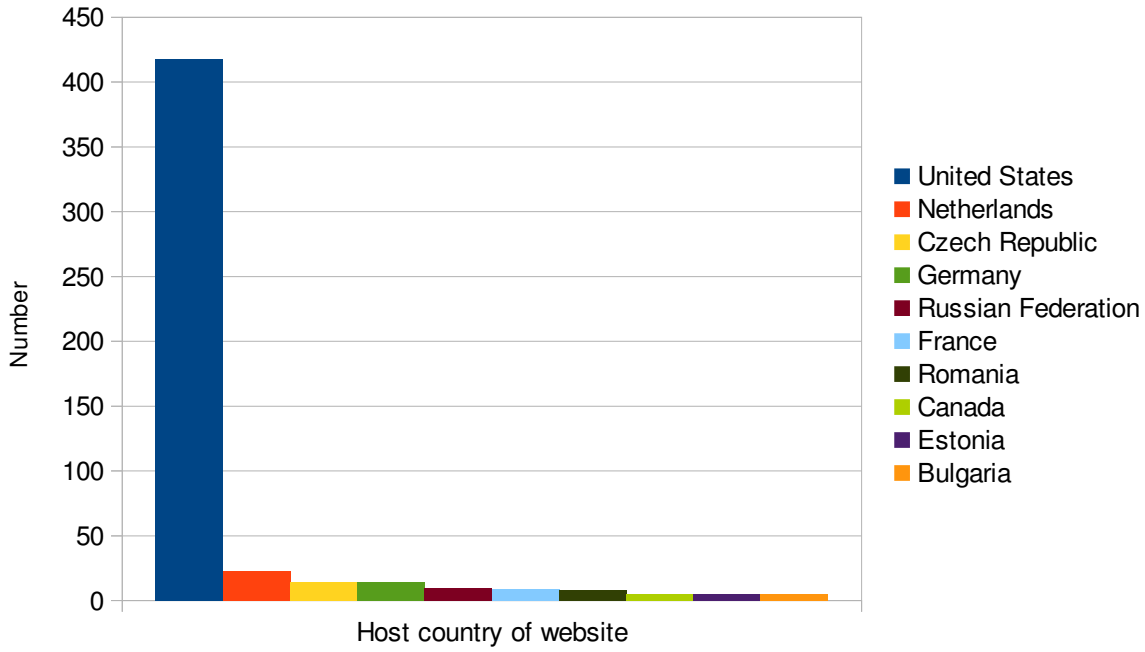


Figure 52: Distribution of host countries of websites identified in Round II of data collection for the Czech Republic

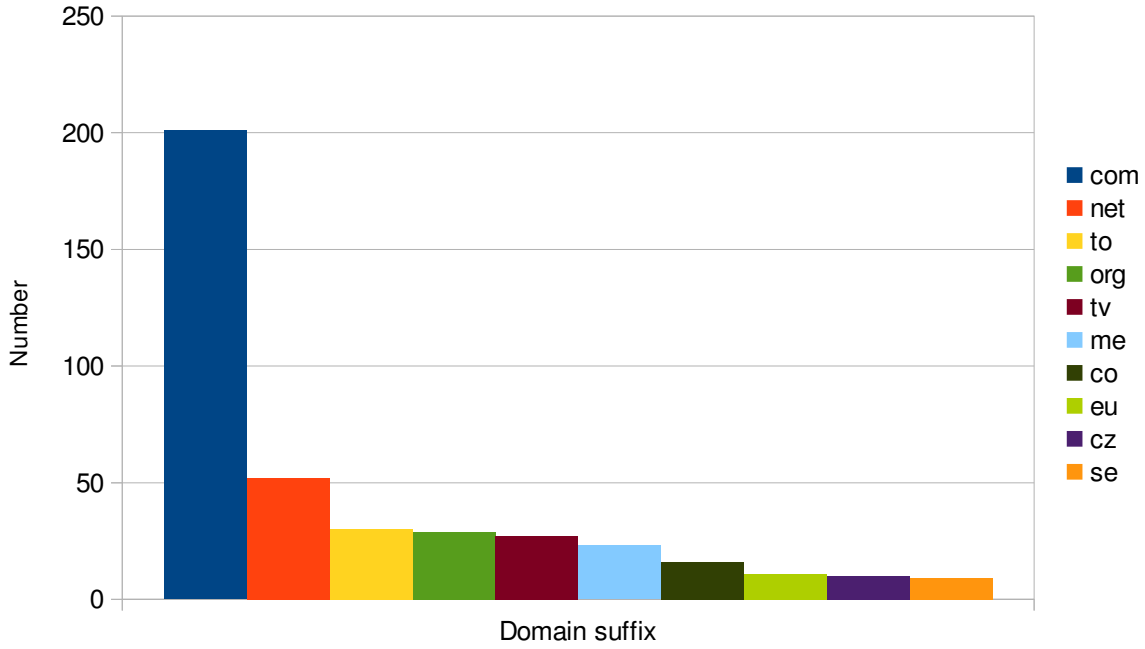


Figure 53: Distribution of domain suffixes of websites identified in Round II of data collection for the Czech Republic

Finland

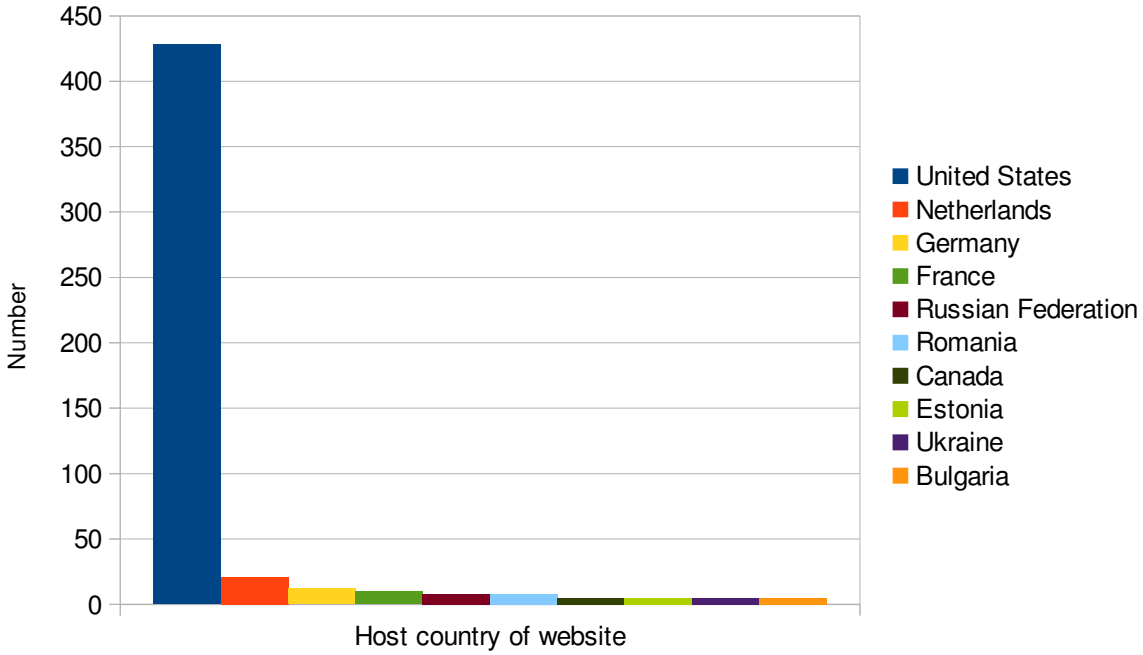


Figure 54: Distribution of host countries of websites identified in Round II of data collection for Finland

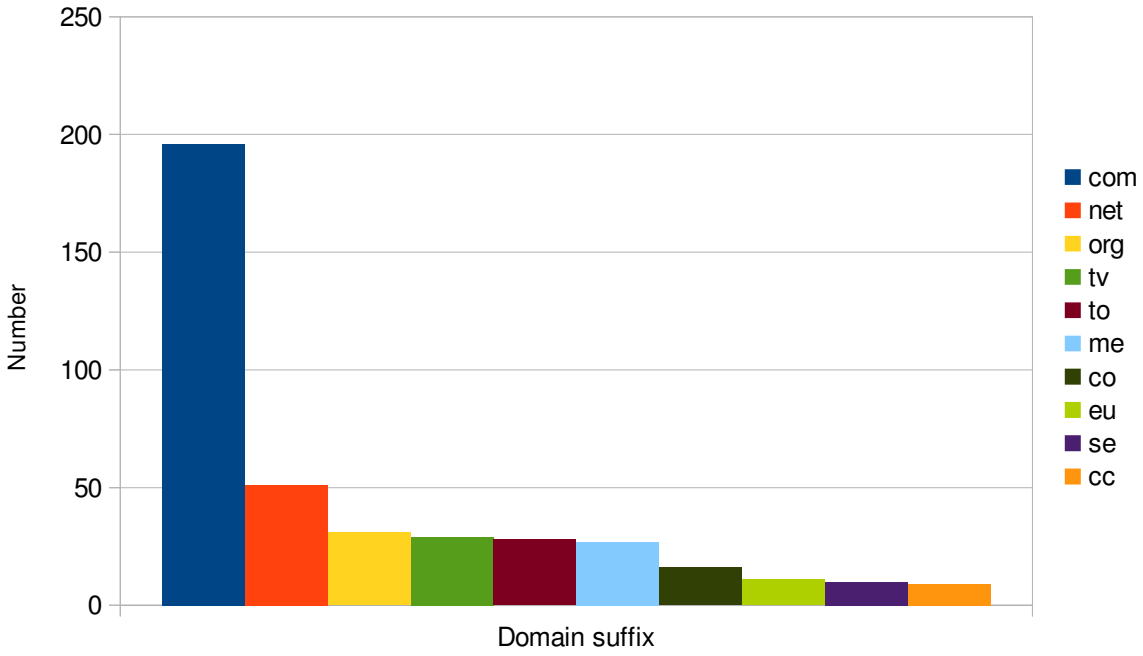


Figure 55: Distribution of domain suffixes of websites identified in Round II of data collection for Finland

France

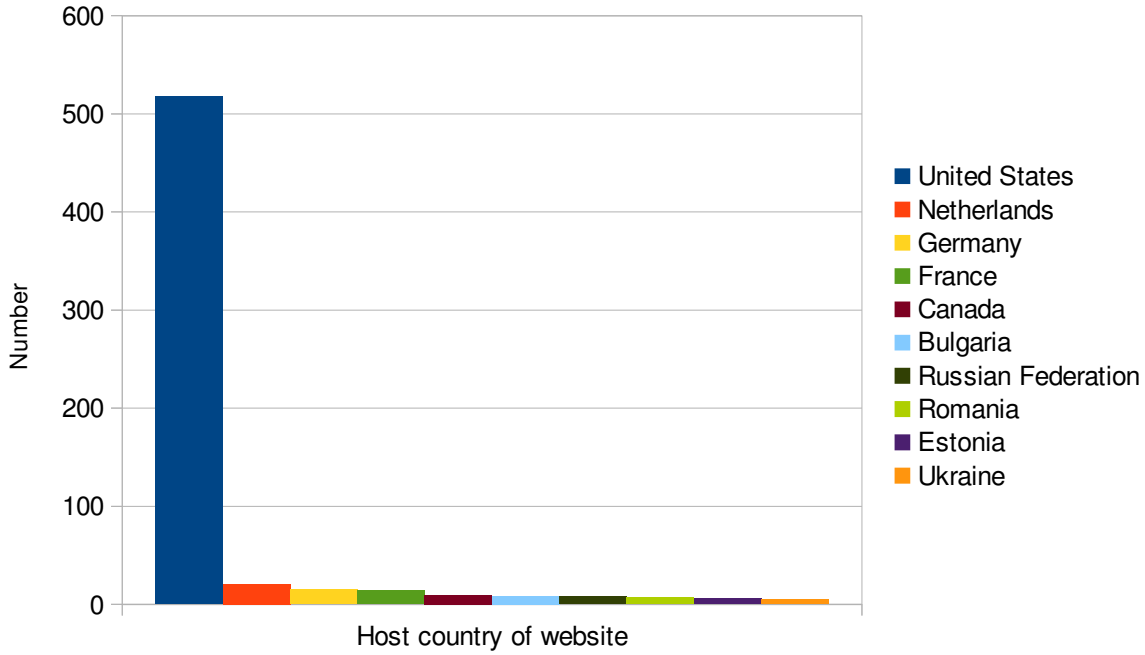


Figure 56: Distribution of host countries of websites identified in Round II of data collection for France

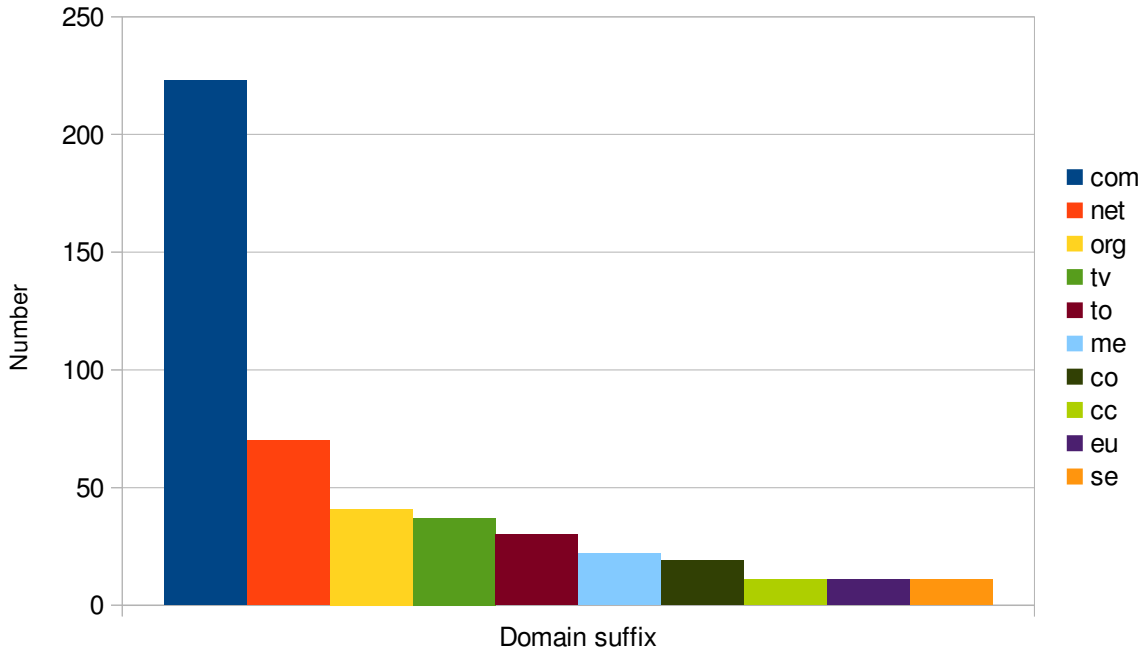


Figure 57: Distribution of domain suffixes of websites identified in Round II of data collection for France

Hungary

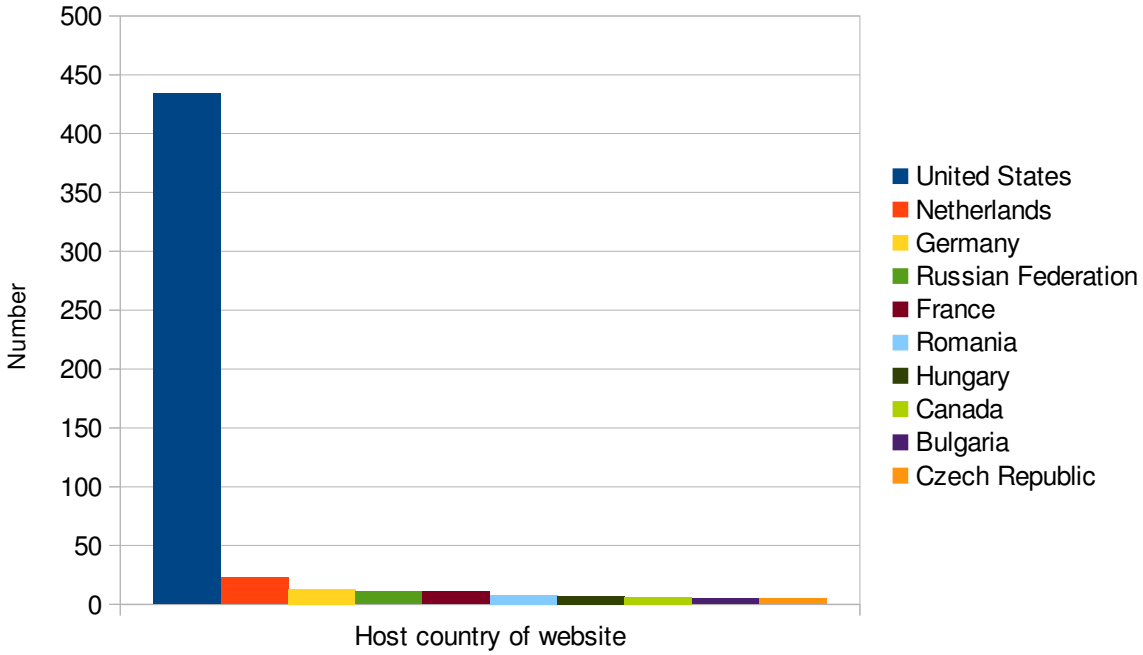


Figure 58: Distribution of host countries of websites identified in Round II of data collection for Hungary

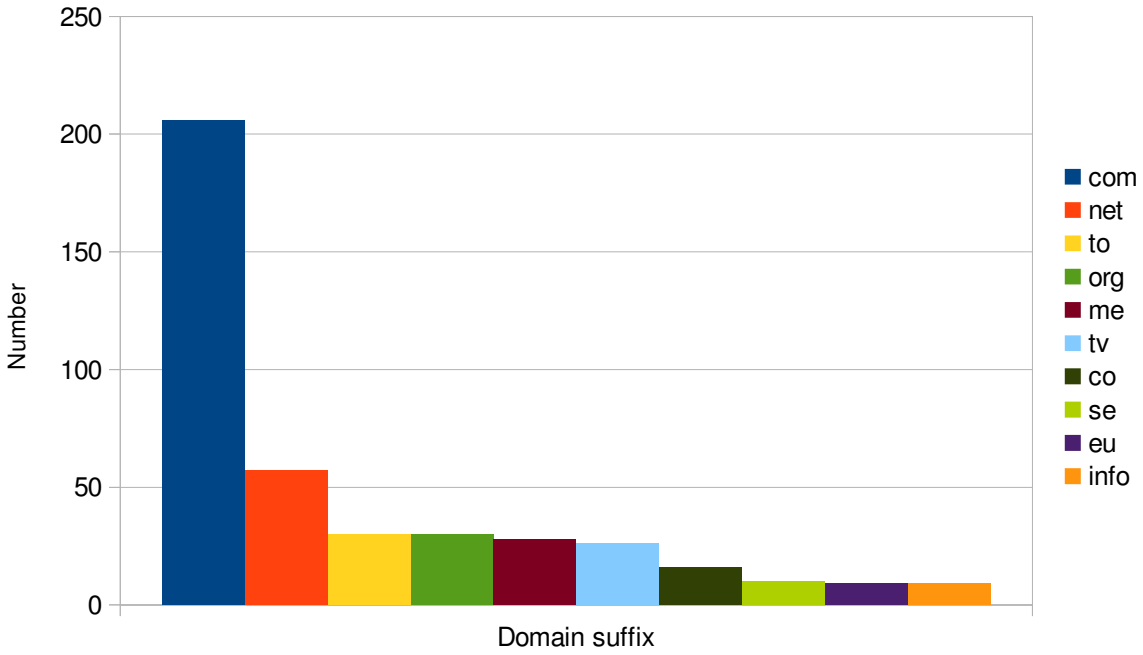


Figure 59: Distribution of domain suffixes of websites identified in Round II of data collection for Hungary

Lithuania

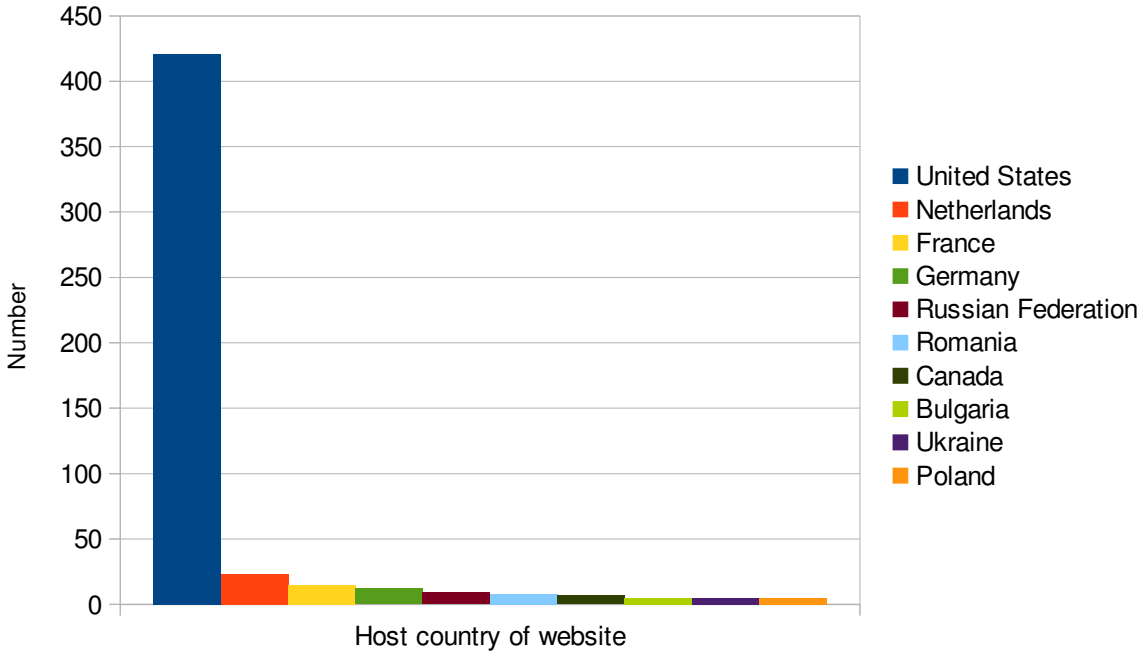


Figure 60: Distribution of host countries of websites identified in Round II of data collection for Lithuania

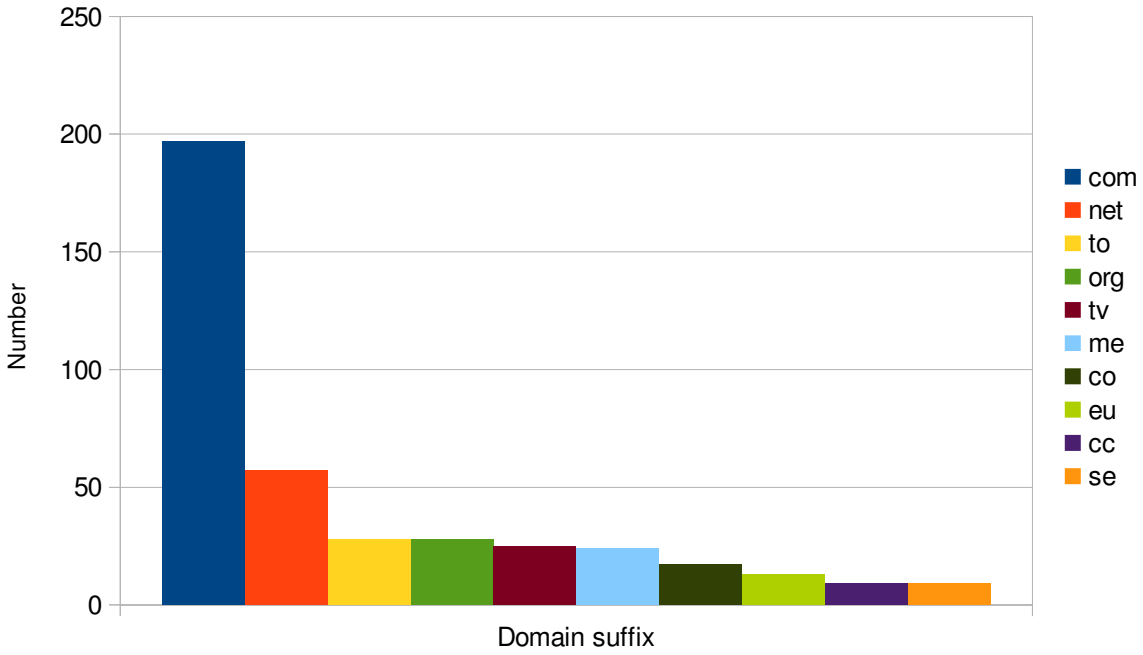


Figure 61: Distribution of domain suffixes of websites identified in Round II of data collection for Lithuania

Portugal

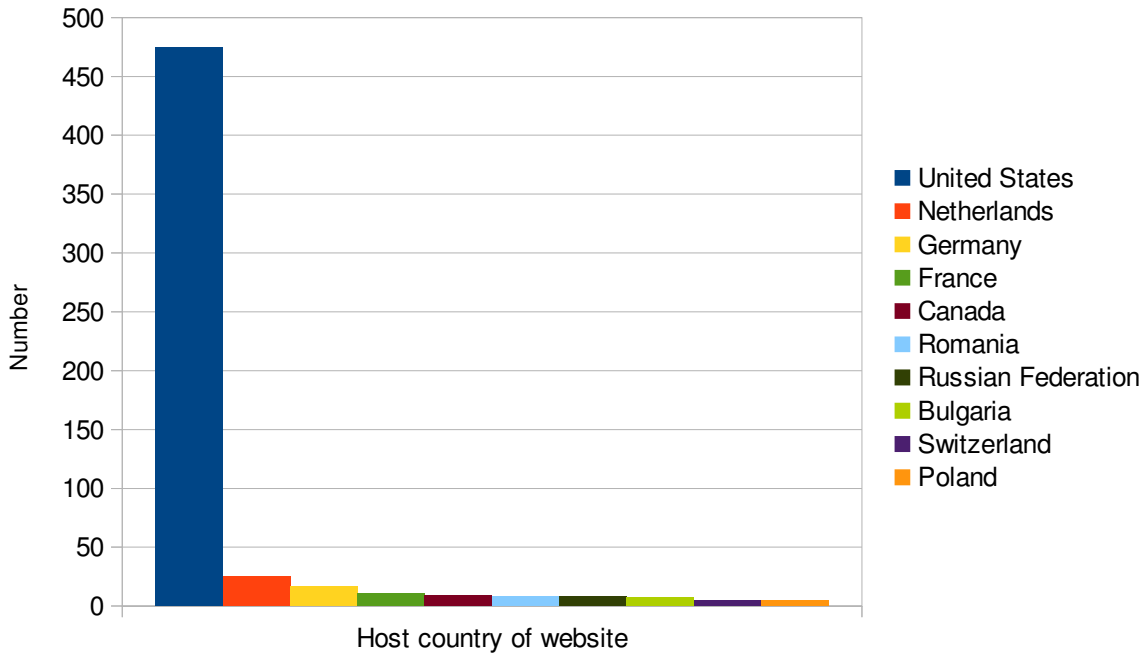


Figure 62: Distribution of host countries of websites identified in Round II of data collection for Portugal

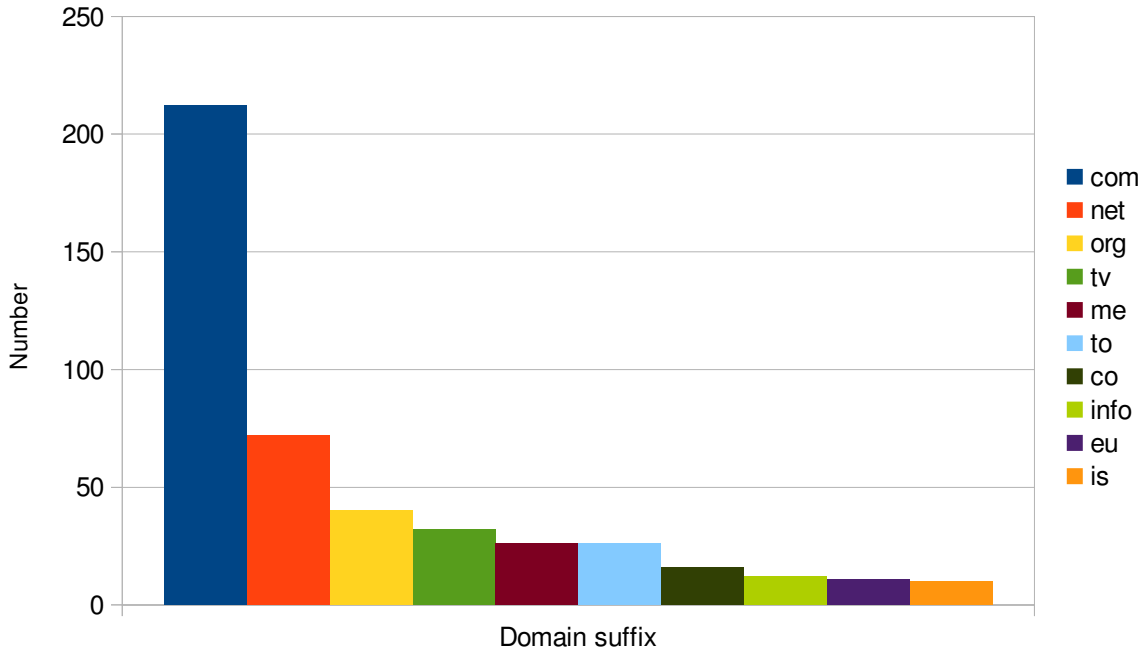


Figure 63: Distribution of domain suffixes of websites identified in Round II of data collection for Portugal

Sweden

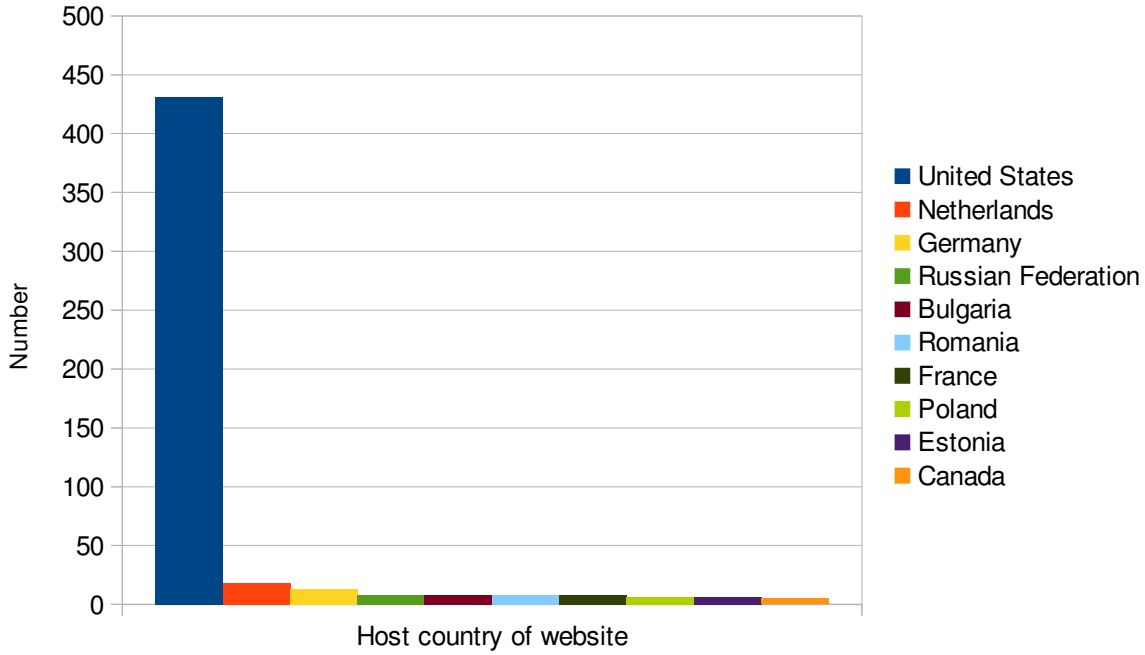


Figure 64: Distribution of host countries of websites identified in Round II of data collection for Sweden

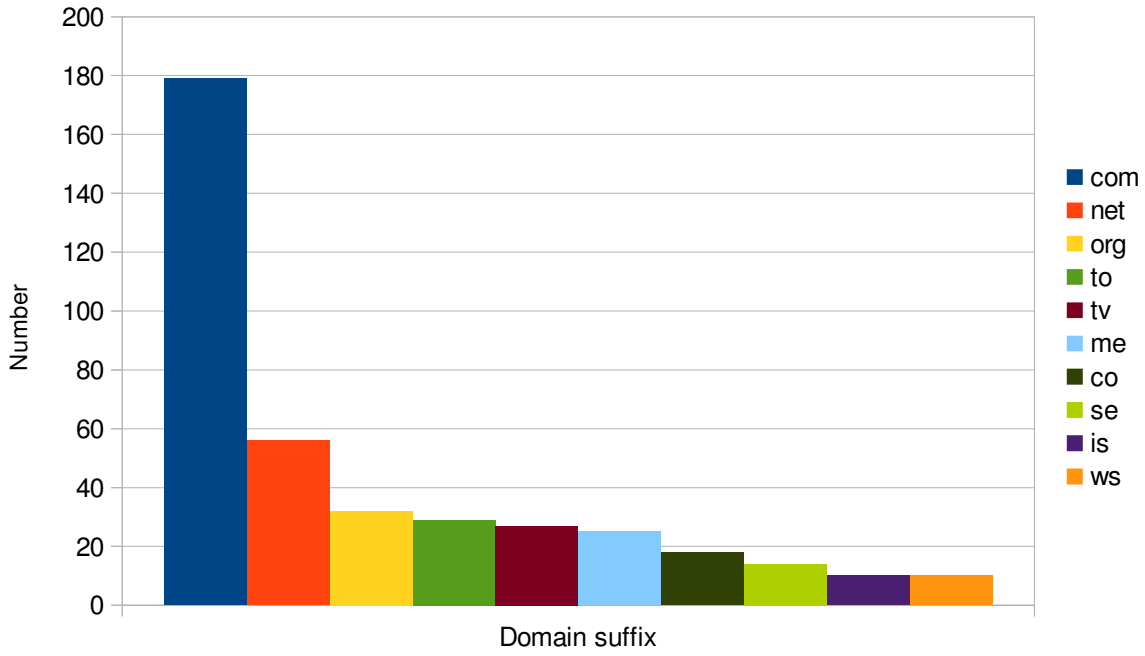


Figure 65: Distribution of domain suffixes of websites identified in Round II of data collection for Sweden

11.2.2 Phase IV.A: Comparison between the two rounds of suspected website identification

A comparison between the two rounds of data collection is provided below. It is important to see how the malware and otherwise unwanted software distributed via the suspected copyright-infringing websites changed even over a short period of time. As the average user's awareness during this period changed as a result of information security incidents occurring around the world, malware evasion techniques also improved. In other words, attackers came to utilise new methods to infect users' computers and to obtain sensitive or private information from compromised systems.

Comparison of Google Transparency Report

The Google Transparency Report is frequently updated to reflect new websites that have been accused of distributing or hosting copyright-protected content and old websites that have been removed. *Table 12* compares information on the domains (website names submitted for removal by Google from search results), requests for removal from search engine results (specific URL, reporting organisation, Lumen database ID), and the number of URLs with no action taken (URLs that were not removed) between the Google Transparency Reports used in the first round of website identification (version as at 25 June 2017) and the second round of website identification (version as at 24 July 2017). There was a significant increase in the number of domains and requests between the first and second round of website identification. Considering the low cost of deploying websites and distributing copyright-infringing content, new websites can generally be expected to appear in the Google Transparency Report database. One of the most efficient ways to reduce harm to consumers from copyright-infringing websites is to remove their links from search engine results.

	25 June 2017	**24 July 2017**
Domains	170 118 970	174 049 634
URLs with no action taken	105 376 095	106 347 188

Table 12: Comparison of Google Transparency Report between the first and second round of identification

Comparison of number of websites identified during Round I and Round II.

The five steps in Phase IV.A resulted in the identification of a large number of websites based on the Alexa Top 500 and systematic keyword searches on online search engines. There are considerable differences between the results of the two rounds of website identification. Some of the copyright-infringing websites were removed from the Google Transparency Report or no longer functioned, while new websites appeared, and others promoted their services to achieve a higher ranking in the search engine results. This illustrates how malware dissemination is a process that evolves over time. For each sample country, there was a net increase in the number of websites identified between the first and second round of website identification.

	Round I	**Round II**	**Added***	**Removed***
Belgium	600	608	116	108
Bulgaria	433	460	80	53
Croatia	431	447	74	58
Czech Republic	522	559	118	81
Finland	536	554	122	104
France	650	665	139	124

	Round I	Round II	Added*	Removed*
Hungary	519	573	127	73
Lithuania	527	561	127	93
Portugal	597	618	120	99
Sweden	555	561	116	110

Table 13: Comparative statistical information on differences between Round I and Round II of malware collection

* Round II in comparison with Round I.

Top 10 website hosting countries and domain suffixes added during the second round of website identification in comparison with the first round of website identification.

Overall statistics for 10 countries

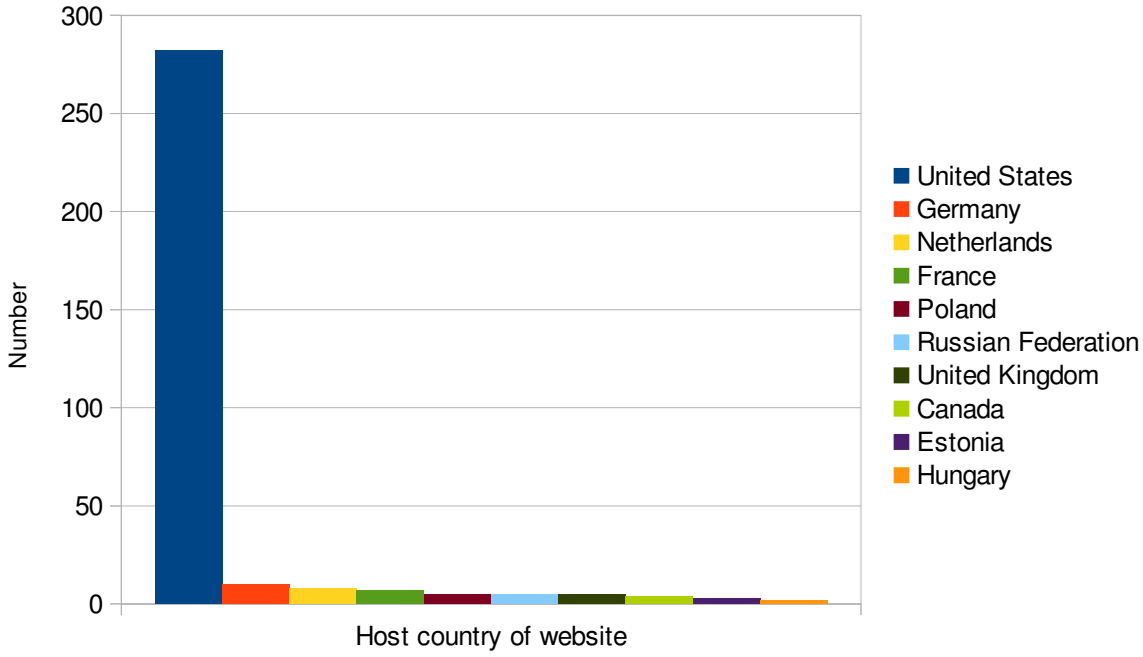


Figure 66: Distribution of host countries of websites added during Round II of malware collection for all countries

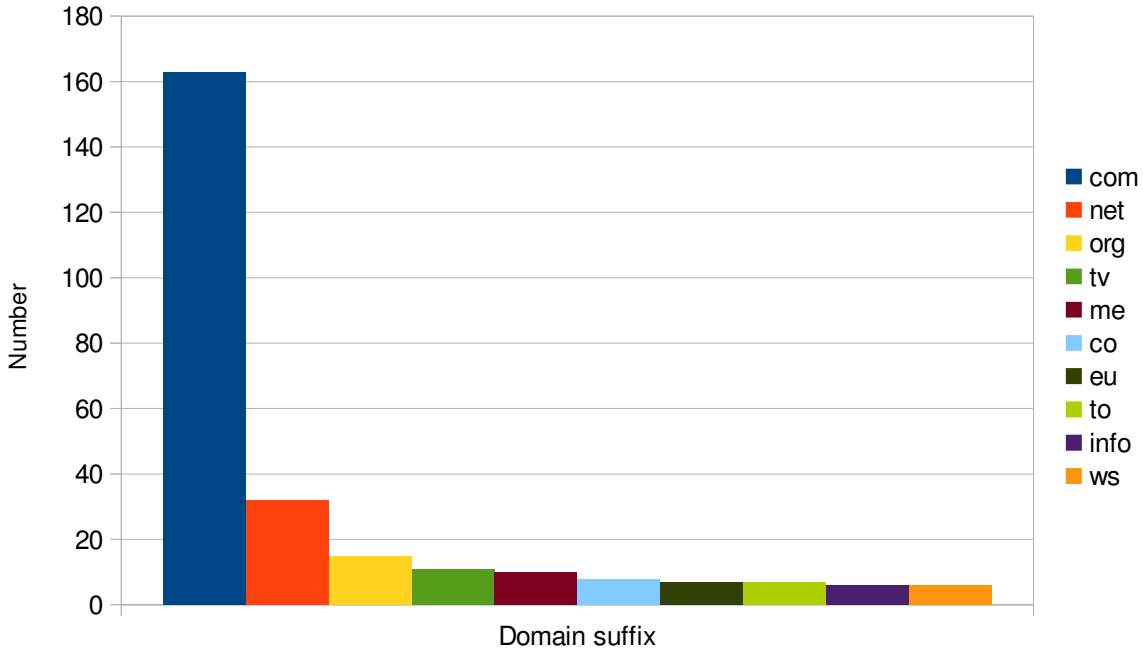


Figure 67: Distribution of domain suffixes of websites identified in Round II of data collection for all countries

Belgium

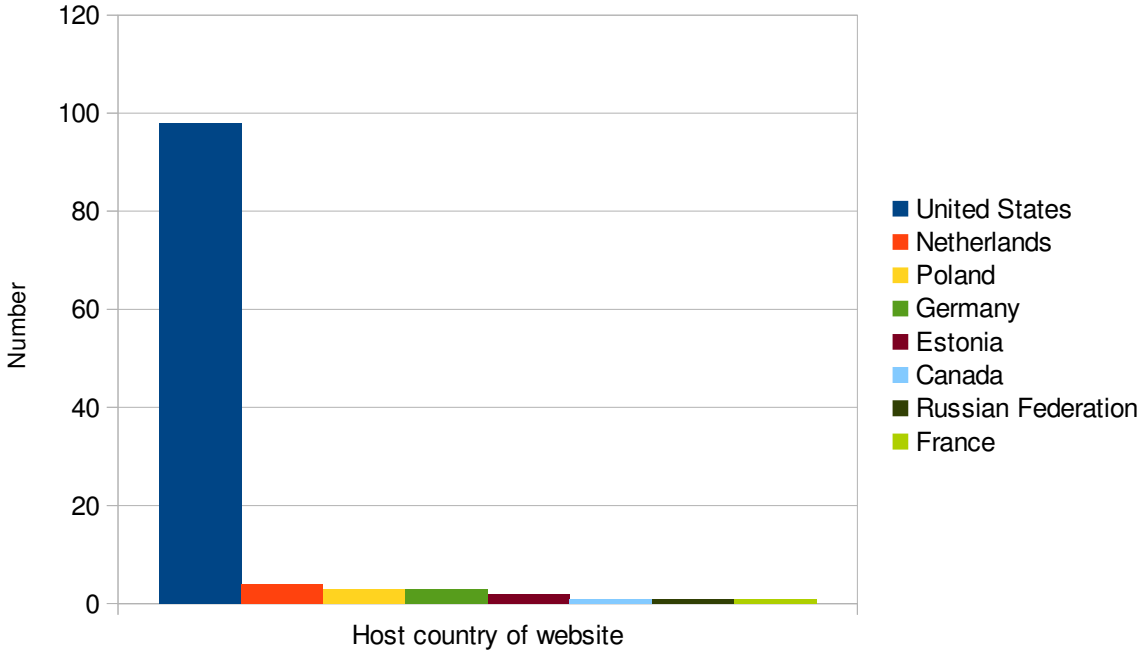


Figure 68: Distribution of host countries of websites added during Round II of malware collection for Belgium

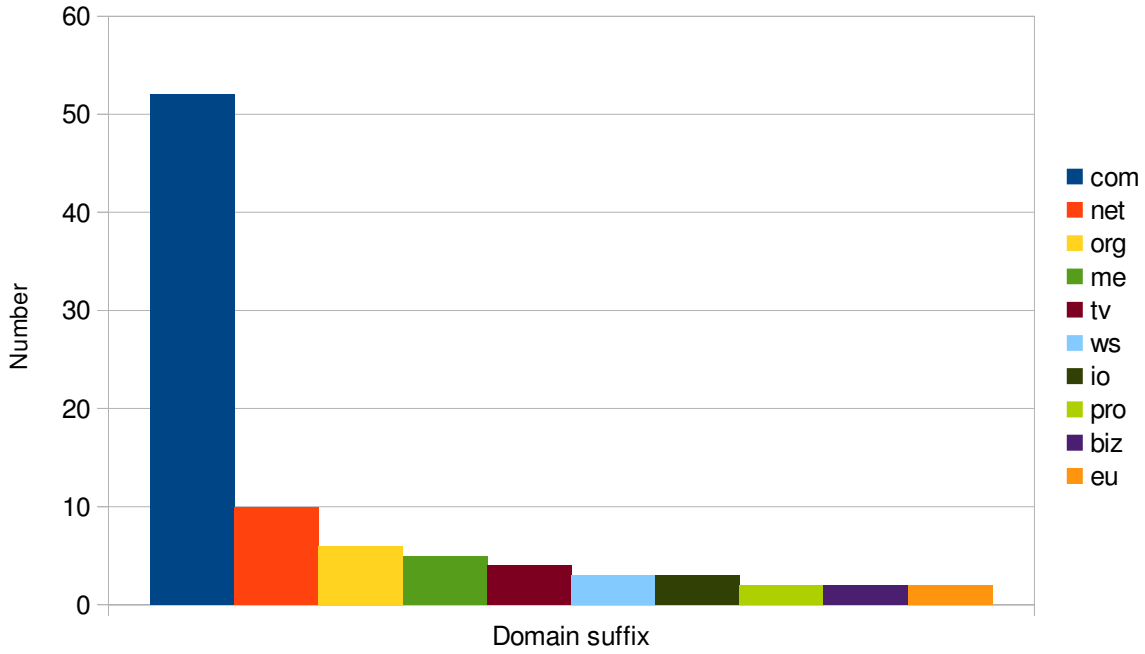


Figure 69: Distribution of domain suffixes of websites added during Round II of malware collection for Belgium

Bulgaria

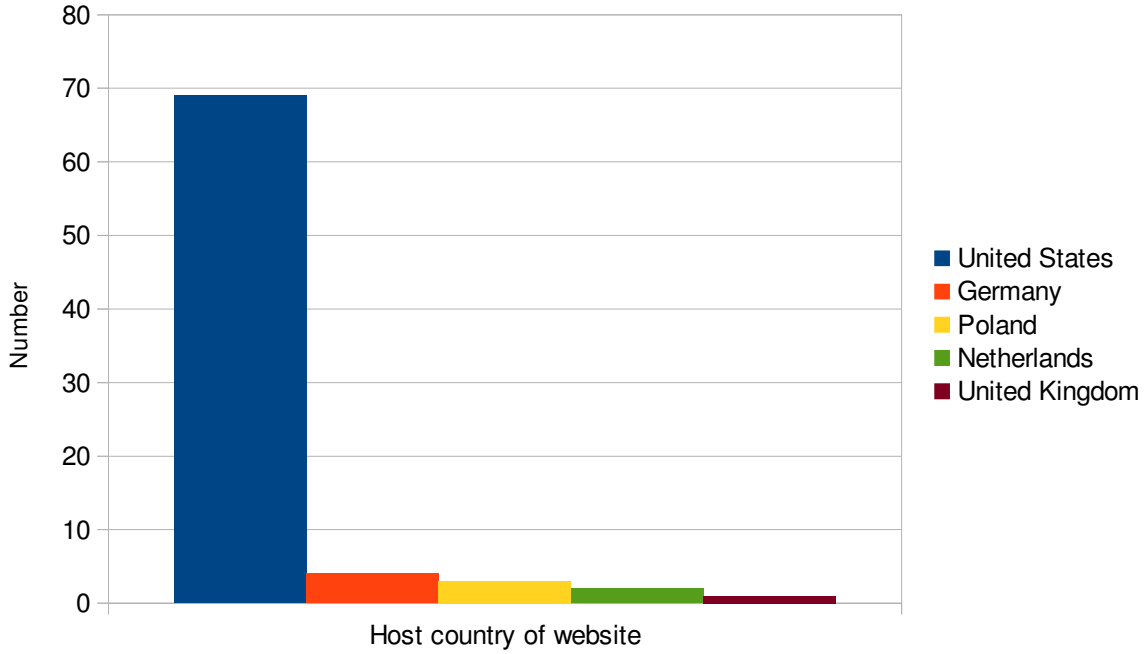


Figure 70: Distribution of host countries of websites added during Round II of malware collection for Bulgaria

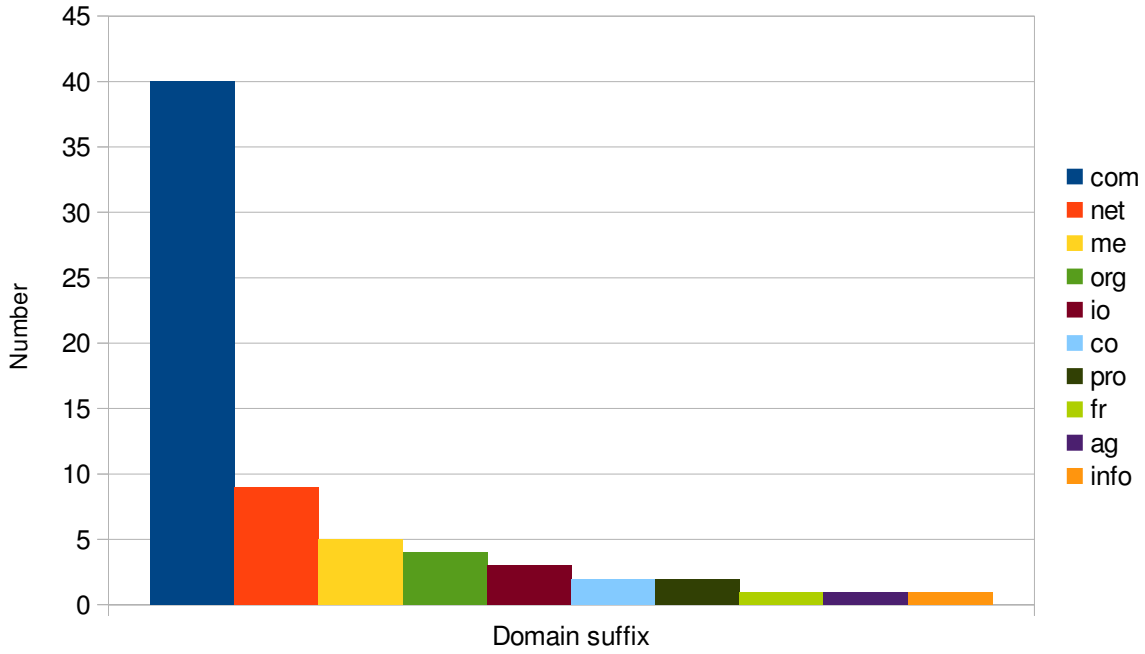


Figure 71: Distribution of domain suffixes of websites added during Round II of malware collection for Bulgaria

Croatia

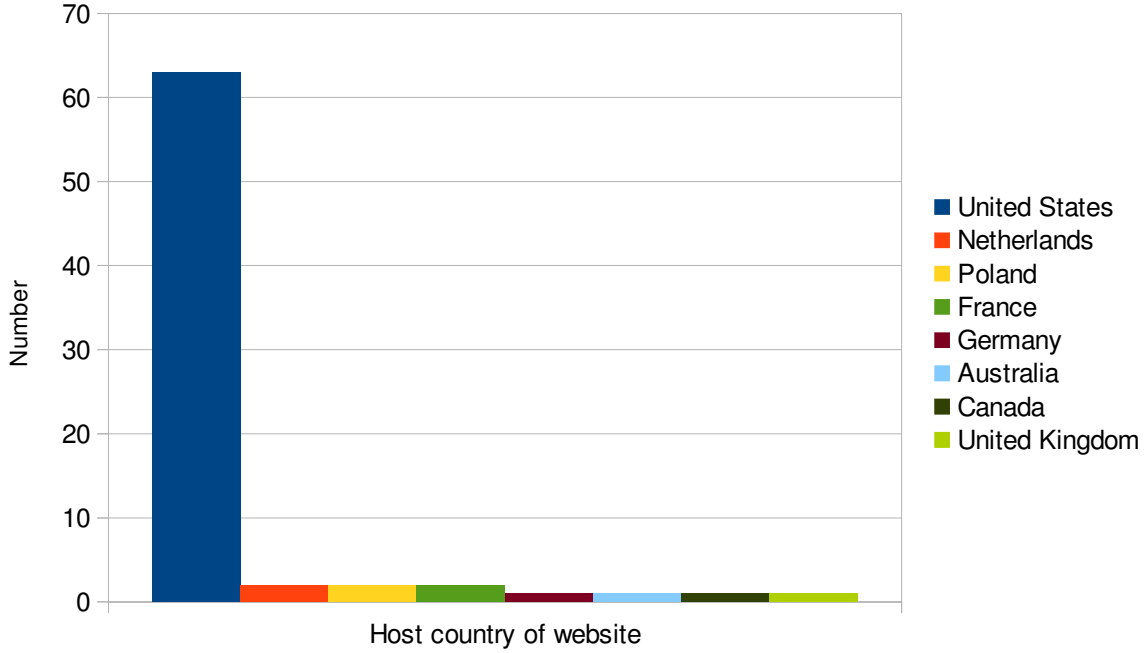


Figure 72: Distribution of host countries of websites added during Round II of malware collection for Croatia

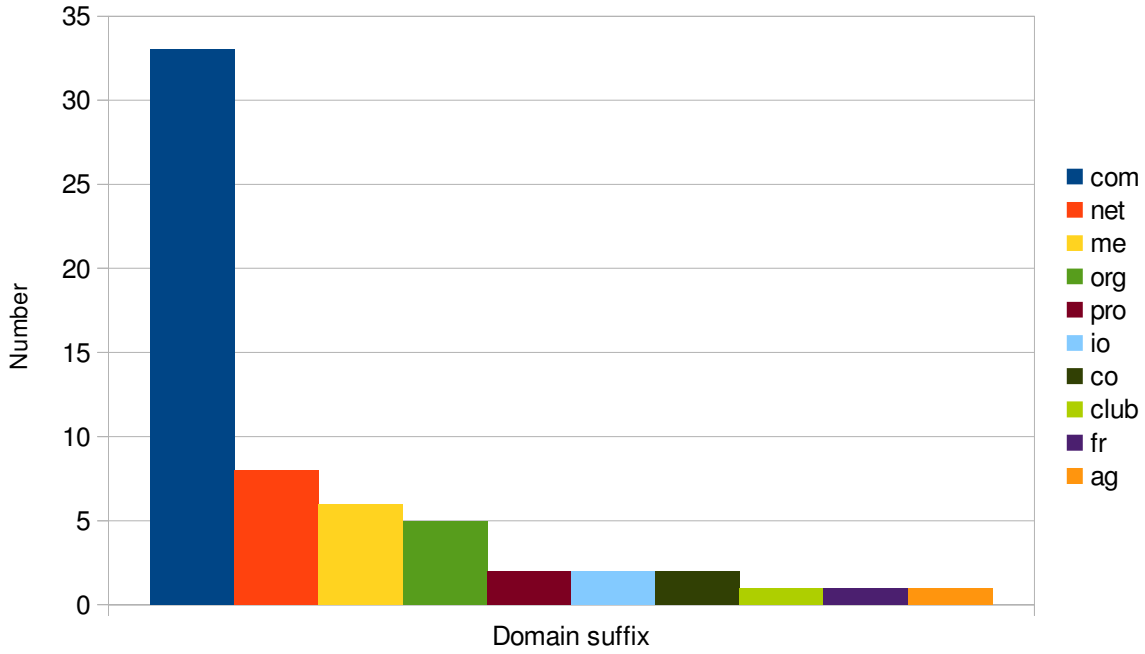


Figure 73: Distribution of domain suffixes of websites added during Round II of malware collection for Croatia

Czech Republic

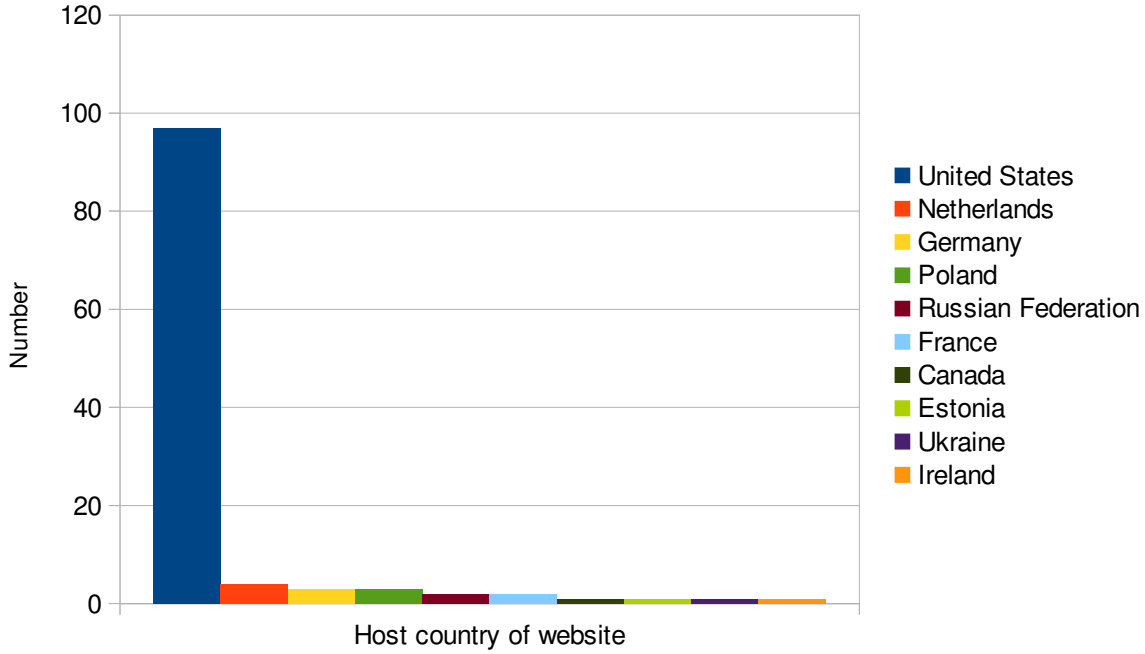


Figure 74: Distribution of host countries of websites added during Round II of malware collection for the Czech Republic

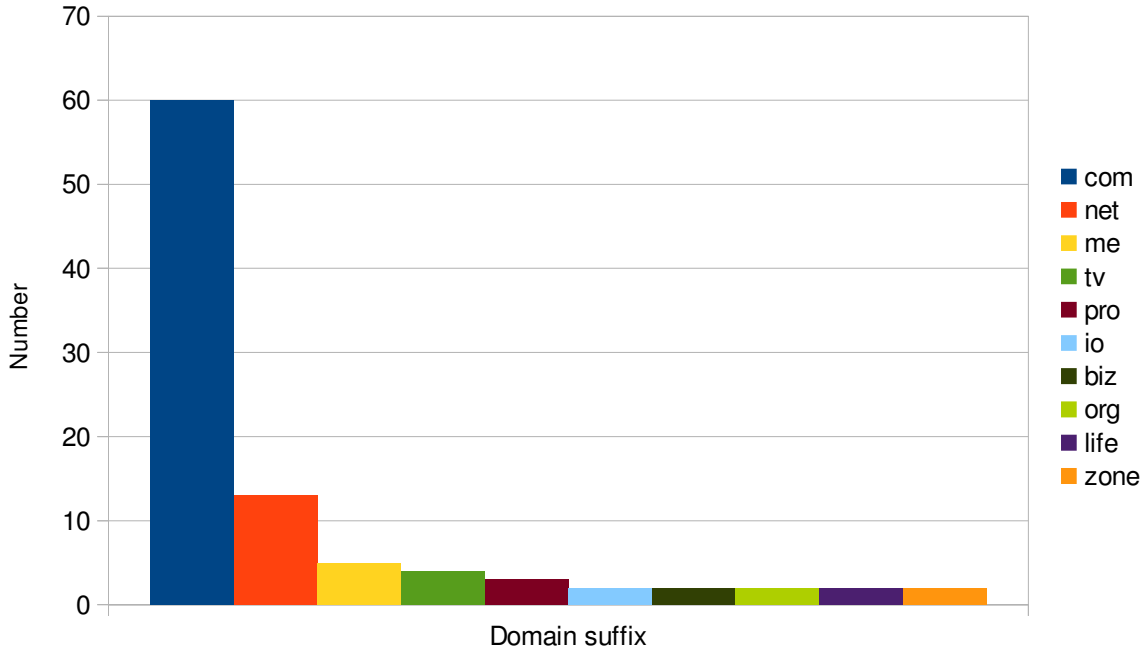


Figure 75: Distribution of domain suffixes of websites added during Round II of malware collection for the Czech Republic

Finland

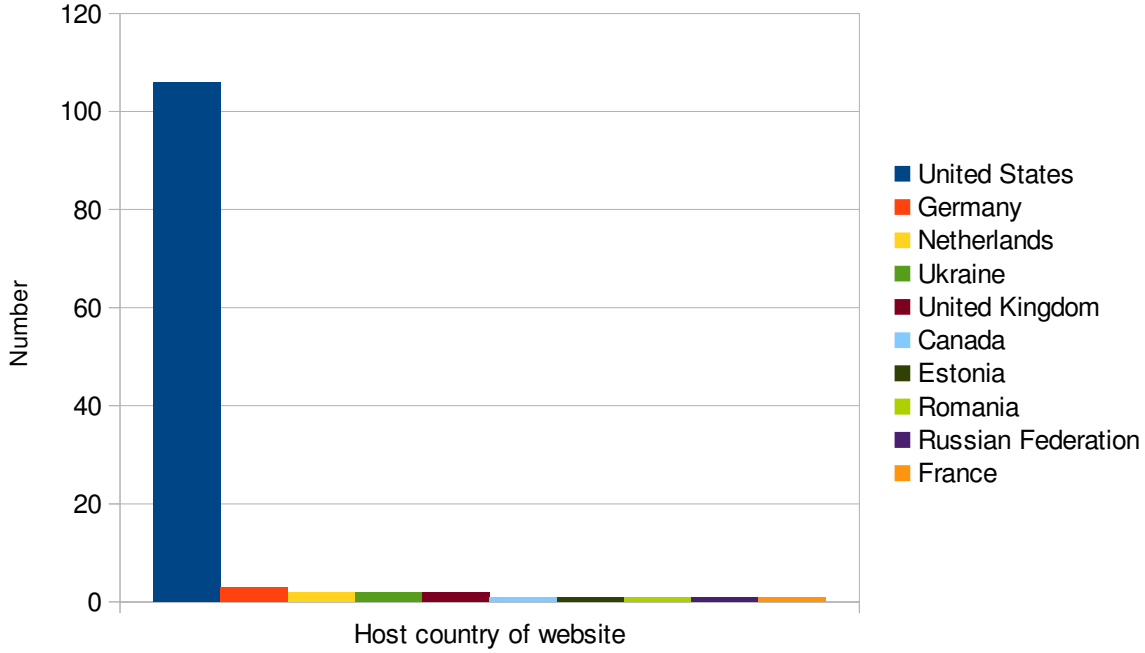


Figure 76: Distribution of host countries of websites added during the Round II of malware collection for Finland

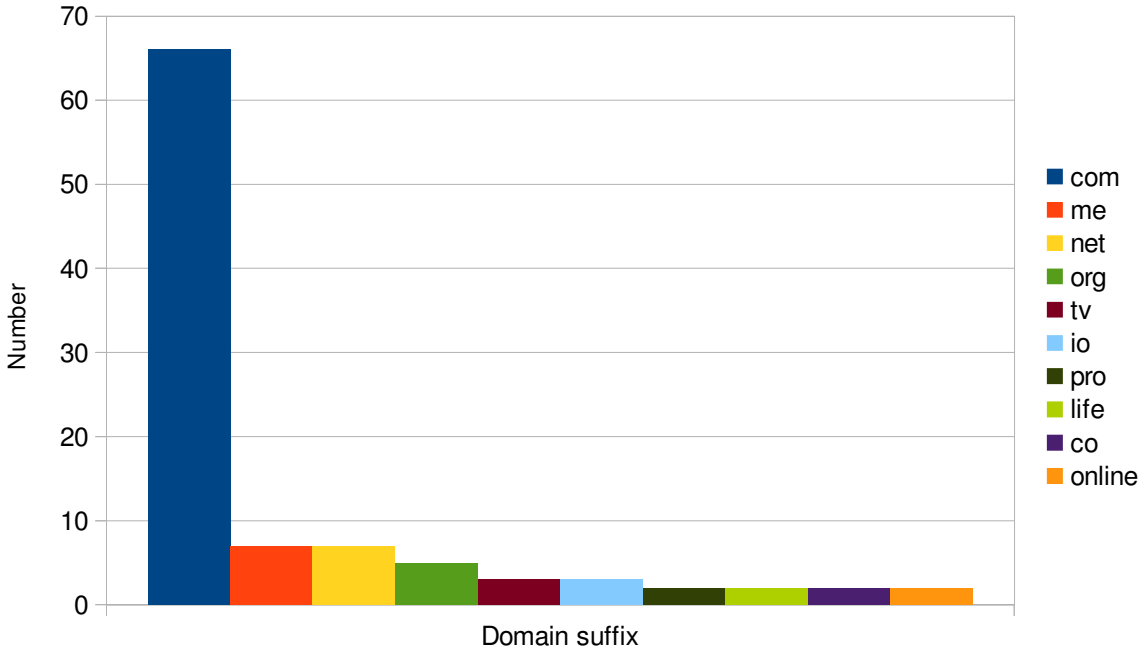


Figure 77: Distribution of domain suffixes of websites added during Round II of malware collection for Finland

France

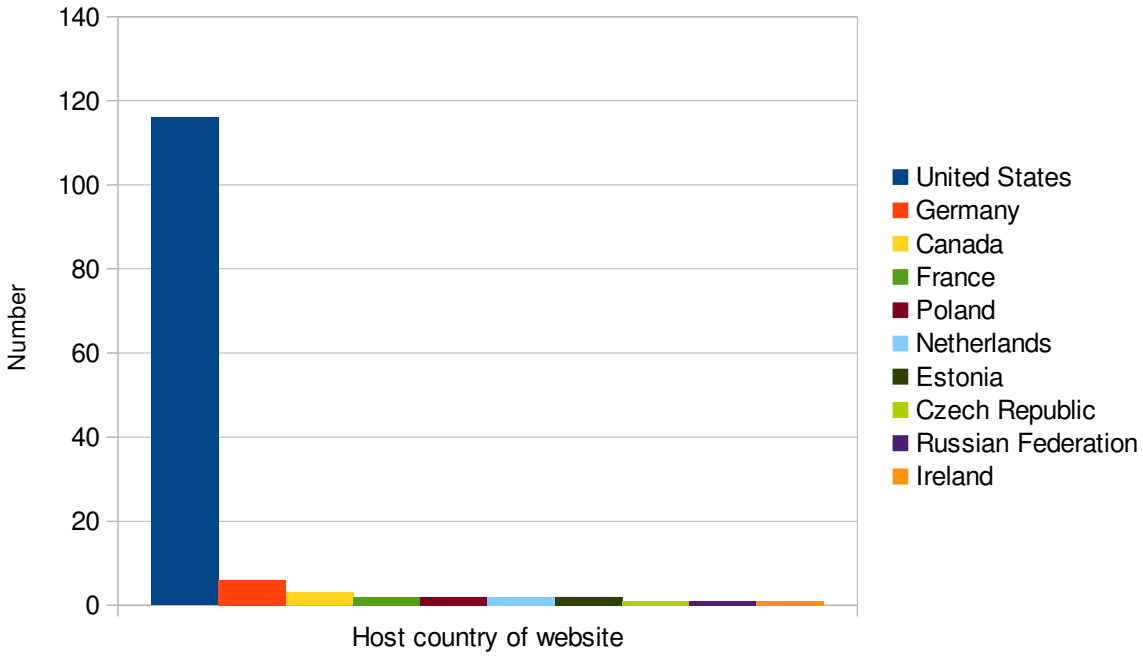


Figure 78: Distribution of host countries of websites added during Round II of malware collection for France

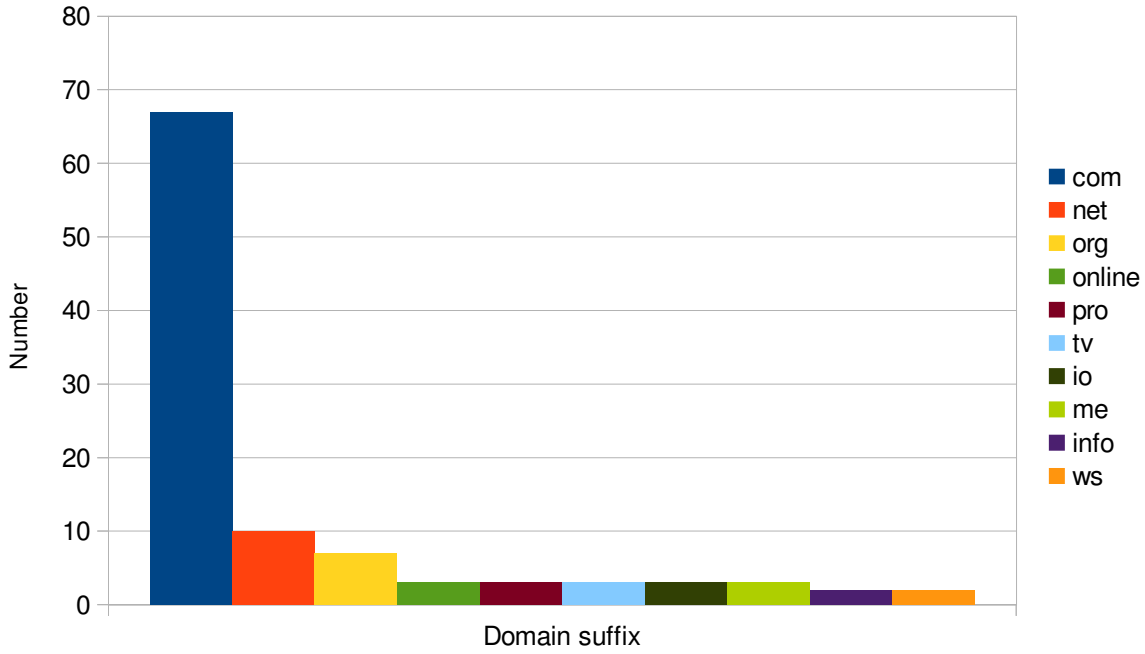


Figure 79: Distribution of domain suffixes of websites added during Round II of malware collection for France

Hungary

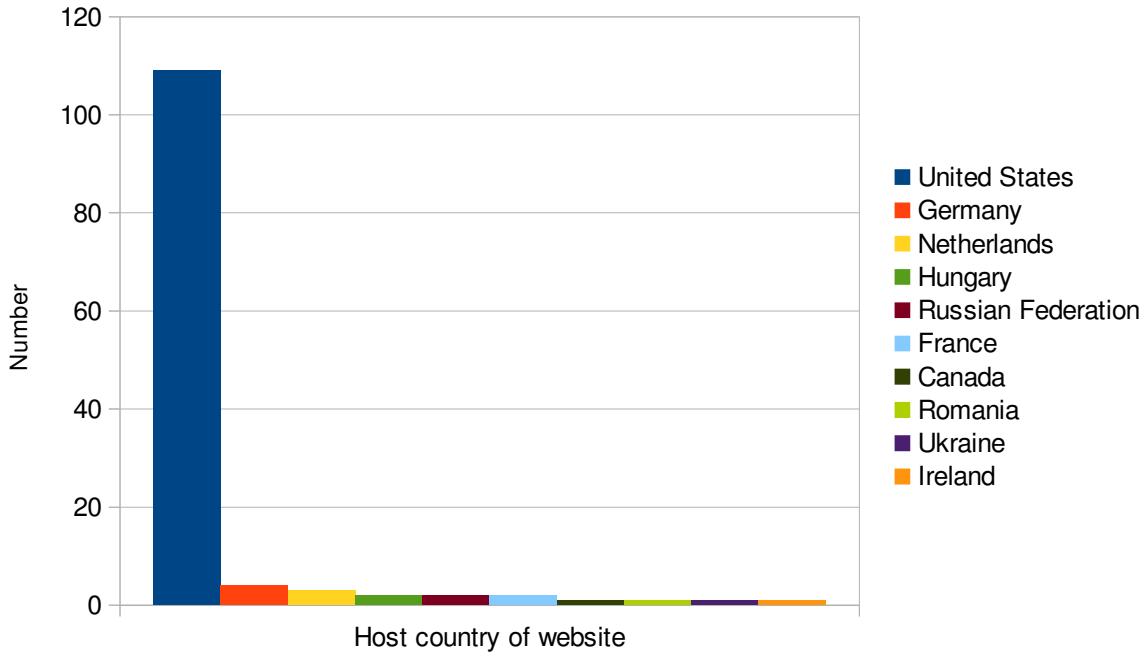


Figure 80: Distribution of host countries of websites added during Round II of malware collection for Hungary

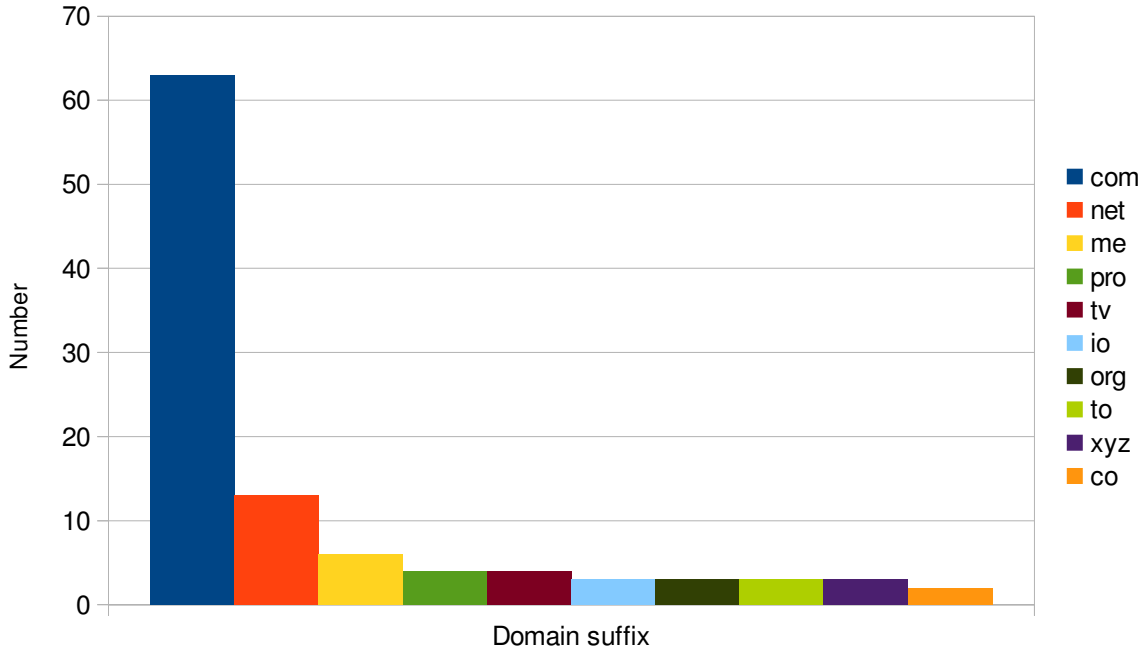


Figure 81: Distribution of domain suffixes of websites added during Round II of malware collection for Hungary

Lithuania

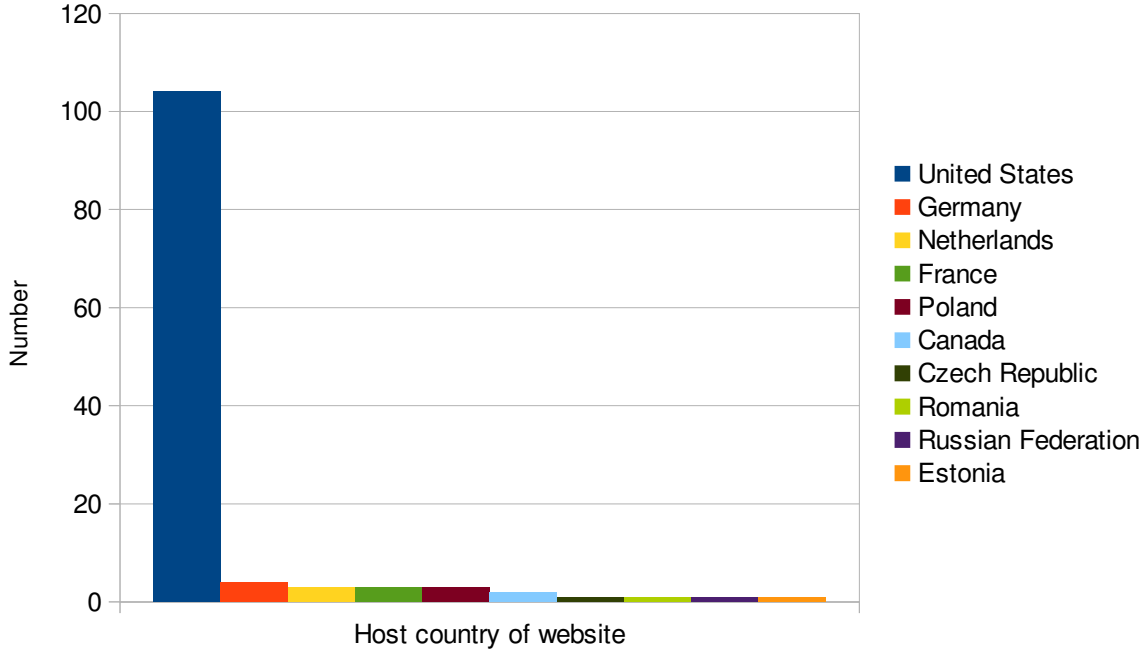


Figure 82: Distribution of host countries of websites added during Round II of malware collection for Lithuania

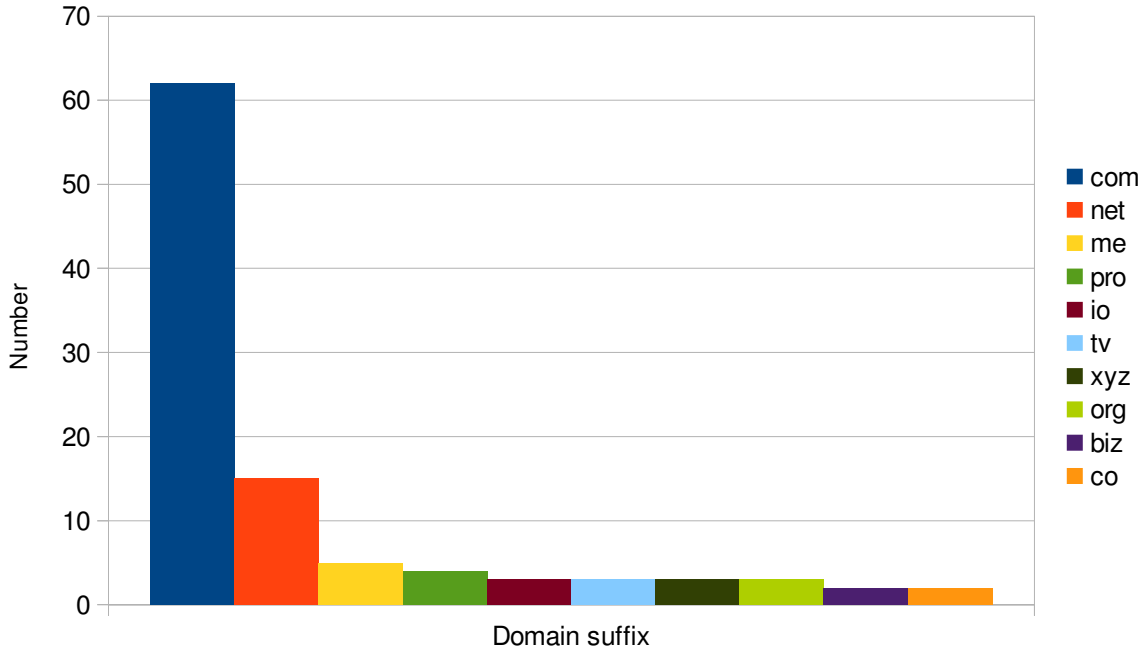


Figure 83: Distribution of domain suffixes of websites added during Round II of malware collection for Lithuania

Portugal

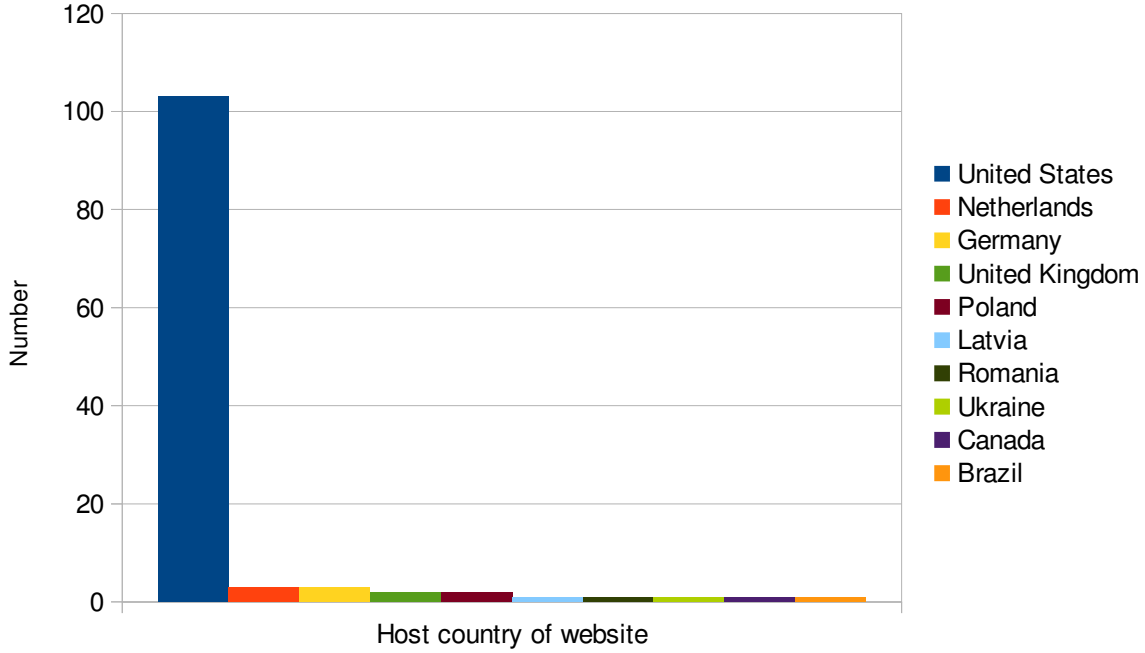


Figure 84: Distribution of host countries of websites added during Round II of malware collection for Portugal

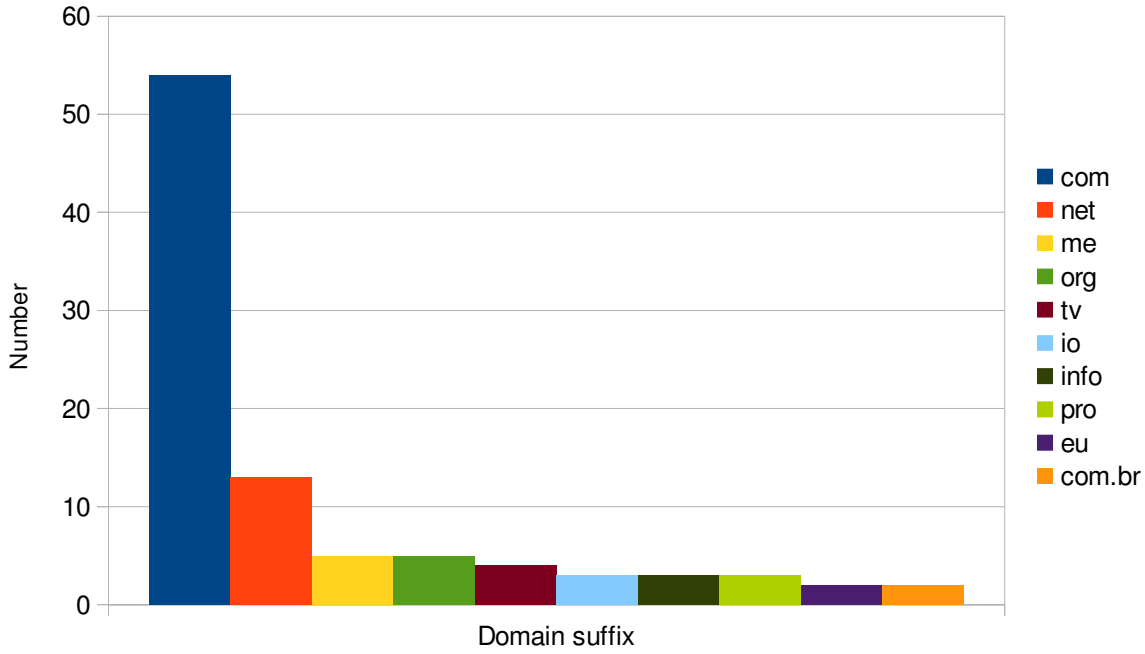


Figure 85: Distribution of domain suffixes of websites added during Round II of malware collection for Portugal

Sweden

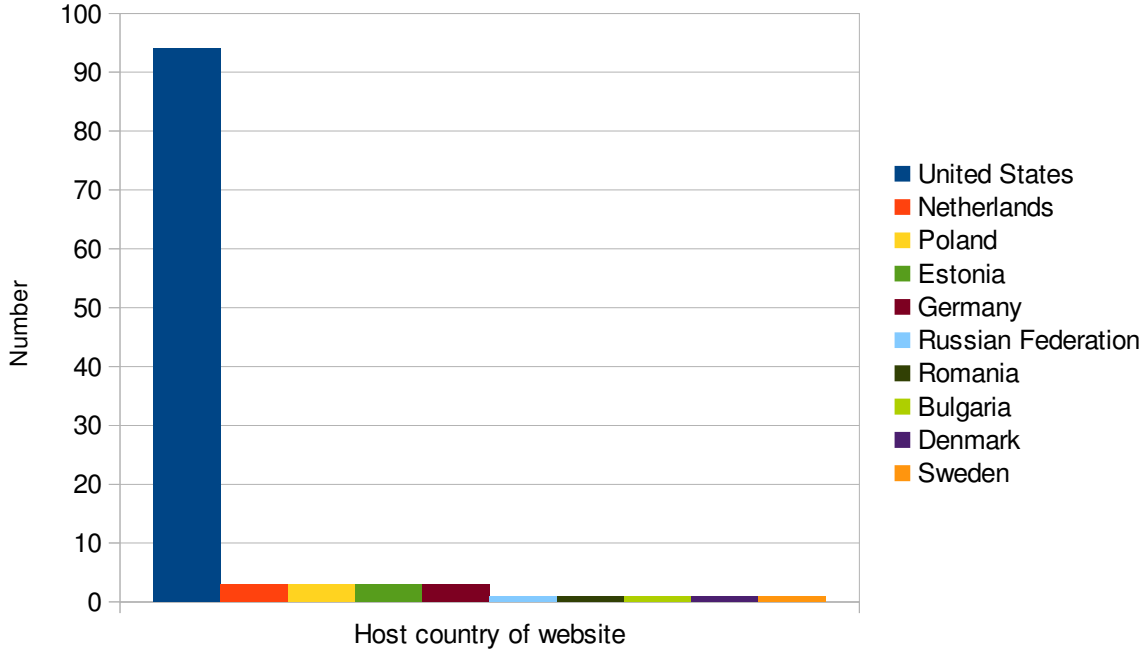


Figure 86: Distribution of host countries of websites added during Round II of malware collection for Sweden

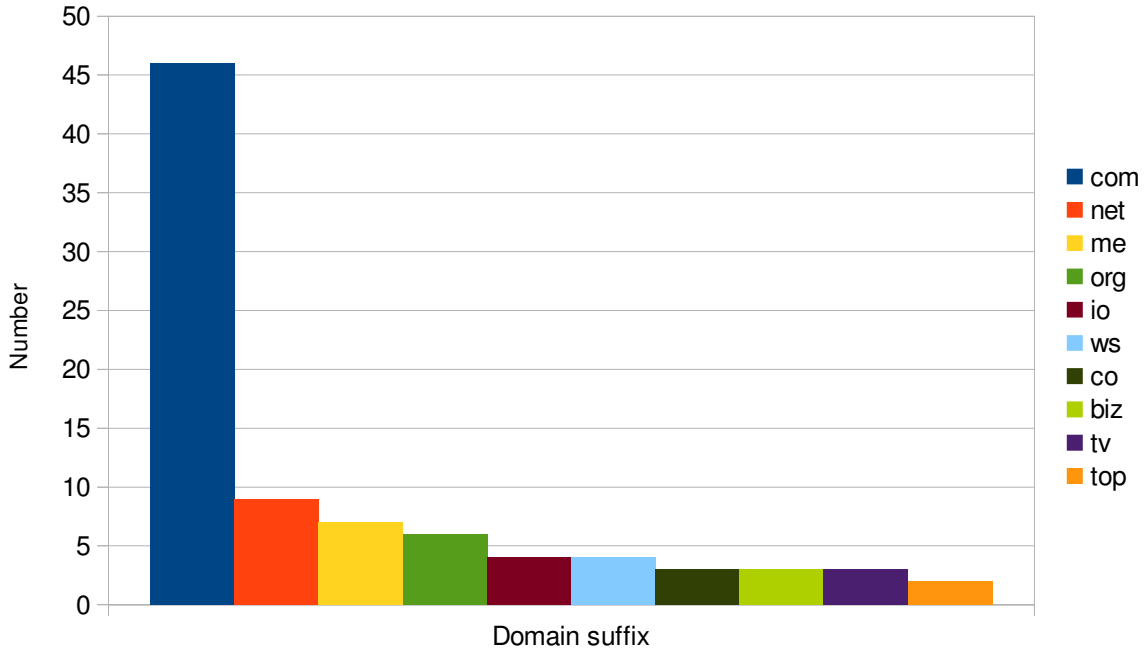


Figure 87: Distribution of domain suffixes of websites added during Round II of malware collection for Sweden

11.2.3 Phase V.A. Binary collection

The binary collection that followed the second round of website identification was performed using two approaches: manual and automated. The manual approach simulated the experience of the average user, characterised by human-understandable reasoning in the identification of covert links and suspicious files on websites. Manual analysis is slow and only a fraction of all websites can be processed by such means. However, automated analysis allows for all the URLs on a web page to be inspected much more quickly. The disadvantages of the automated approach are that it is not able to process JavaScript and it cannot bypass security mechanisms, such as Captcha or other similar tools.

The overall number of unique URLs extracted for all countries was 1 057 out of the 5 606 websites, which made it unfeasible to check all of them manually. The samples of copyright-infringing websites were similar across all 10 sample countries for each of the types of media (films, music, television programmes, and video games). As a result, Belgium was randomly selected from the sample countries, and all websites identified as copyright-infringing websites for Belgium were manually verified for the presence of malicious or otherwise unwanted software. A total of 3 665 files were automatically collected from the websites for all countries, with a total size of 167 GB.

11.2.4 Phase V.B. Binary analysis

A number of samples of malware or otherwise unwanted software identified during the second round of data collection are discussed below.

Website09 is a website that distributes a BitTorrent-based client, which offers access to various kinds of video content available through torrent trackers. The tool has a very simple graphical user interface and only a few settings accessible to the user. Upon selecting a film, the film is downloaded and played immediately. Website09 requires fewer user interactions compared to other BitTorrent trackers. Only a

few clicks are required to download the content from unknown sources. This can include video files as well as other malicious payloads. In this way, the user is neither protected nor has control of what is being downloaded. Moreover, the software can be considered as a perfect distribution platform for copyright-infringing content. The installation of the program is quick and no understanding of advanced settings is required.

Once installed and launched, the user has two options: he or she can select films or television programmes. The available titles are sorted by genre. IMDB sorting for films is also available. Additionally, the program features a reminder for users to employ VPNs to avoid any detection issues.

Each film page in Website09 contains a general description of the film, its rating, and video quality options that need to be selected before watching the video. Once the user clicks the 'Watch Now' button, the file is downloaded to the user's computer via a BitTorrent protocol, with the corresponding information displayed on the screen, including the speed and status of the download, as well as the number of peers, seeds, and leechers[154]. After the download is complete, the requested film is available for viewing. The overall process of accessing copyright-infringing content is fast and can be done by virtually anybody.

Distribution model

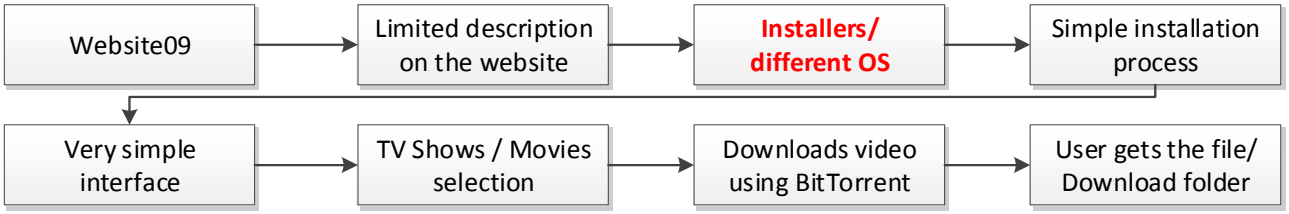


Figure 88: Distribution model of Website09 software

Virus Total: There were no hits for Website09 in the VirusTotal database. This means that the software probably does not cause harm to users' systems. It does not modify any system files or make edits to the registry settings database. Therefore, from the perspective of malware categorisation, the program can be considered not harmful. However, it is a very simple and efficient platform for distributing copyright-infringing content. The user is able to access any film or television programme that is available on BitTorrent trackers.

Technical details. The program left very few artefacts on the disc. The folder created in 'Program Files' included some libraries and supplementary plug-ins needed for proper functioning. The program's executable files, 'notifier.exe' and 'updater.exe', receive no hits in the VirusTotal database.

Website03

Website03 was also discovered during the first round of malware collection and analysis. It is noticeable that the website owner(s) have considerably changed how malware is distributed. Previously, it was a tiny executable file for MS Windows that, after simulated installation, directed users to a website to receive a licence key for a game. Even though one of the downloading links on the web page reads 'Google Drive,' the link itself points to another website. The latter is owned by the attacker. In addition, torrent files with an 'ultra seed' option can be found there (more peers with better speed) and general torrent files with normal or compressed files.

[154] BitTorrent is organised as a decentralised file exchange protocol that includes users (peers) who either share files or download them. Seeds denote users who have fully available files on their computer and ready for downloading through BitTorrent, while leechers denote users (peers) who do not have the full set of file segments yet.

The design of the download page is professional-looking and is formatted in a manner similar to popular file-sharing services. The option is offered to download using special software or through a web browser. The suggested file does not bear a specific game's name and was downloaded using browser options.

Distribution Model

The entire process of getting a user's sensitive information has changed since the last round of malware collection. The user of this service downloads an archive, which contains content masked as game-related files and not an explicit binary executable that can be detected as malicious by anti-virus programs. The encrypted archive grants access only to filenames, not the substantive content of the files. Furthermore, due to the use of encryption there is no way the exact nature of the content of the archive can be ascertained. In this case, an analyst can look at the raw content and make a hypothesis if the content is a text, an image, or some other content. However, it may be the case that the files simply contain dummy content, although this cannot be confirmed.

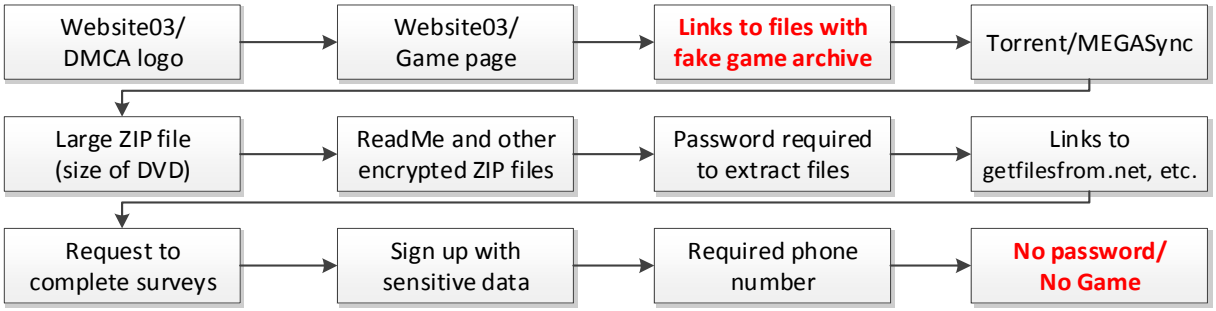


Figure 89: Distribution model of Website03 software

VirusTotal: the size of the download is greater than 4 GB, which far exceeds the default limit of VirusTotal (150 MB). It is worth mentioning that, by using such a distribution model, the attacker is able to avoid detection because part of the archive is encrypted and the other part only contains a few harmless text files with a description of the actions.

12. Malicious Activities Detected by EMAS (extended version)

Hardware Tampering Activity. This is related to direct disc access activities, where a process tries to access data directly on the disc or disc root, as depicted in *Figure 89*. Such access is typically denied from utilisation by user processes since a malware can write its code to any places on the hard disc without hosting OS awareness.

Process	Started	C:\Documents and Settings\admin\Local Settings\Temp\ab7d31db100a8cf944914761b4e9cbfe.exe *Parentname*: C:\WINDOWS\explorer.exe *Command Line*: "C:\DOCUME~1\admin\LOCALS~1\Temp\ab7d31db100a8cf944914761b4e9cbfe.exe" MD5: ab7d31db100a8cf944914761b4e9cbfe SHA1: d72e71342e85a3628e88ec9a0536749a51cf89f2
File	Open	C:
File	Failed	C:\WINDOWS\system32\config\system
File	Close	C:
Malicious Alert	Hardware Tampering Activity	*Message*: Direct disk access
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\SystemInfo-vc100-mt.dll
File	Failed	C:\WINDOWS\system32\SystemInfo-vc100-mt.dll
File	Failed	C:\WINDOWS\system\SystemInfo-vc100-mt.dll
File	Failed	C:\WINDOWS\SystemInfo-vc100-mt.dll
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\SystemInfo-vc100-mt.dll

Figure 90: Direct disc access observed in analysis of 54545dc3868032ab9eac2cb95bdc1227.exe on MS Windows XP SP3

Generic DLL Load Activity. Dynamically linked libraries (DLL) are used by Microsoft Windows to provide access to common API function calls. By default, most of the necessary OS functionality is implemented in such libraries, which are supplied with MS Windows. *Figure 90* shows a routine that was performed before loading the DLL. First, a specific temporary directory was created to administrator a local directory. Then, '_shfoldr.dll' was placed in the folder to be loaded at a later date. Notably, the loaded DLL is similar in name to 'shfoldedr.dll,' which is Microsoft Shell Folder Service that is responsible for shared folder functionality in Windows.

File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\is-6AU5B.tmp
Folder	Created	C:\Documents and Settings\admin\Local Settings\Temp\is-6AU5B.tmp
Folder	Created	C:\Documents and Settings\admin\Local Settings\Temp\is-6AU5B.tmp_isetup
File	Created	C:\Documents and Settings\admin\Local Settings\Temp\is-6AU5B.tmp_isetup_shfoldr.dll
File	Close	C:\Documents and Settings\admin\Local Settings\Temp\is-6AU5B.tmp_isetup_shfoldr.dll MD5: 92dc6ef532fbb4a5c3201469a5b5eb63 SHA1: 3e89ff837147c16b4e41c30d6c796374e0b8e62c
DLL Loaded		*Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\is-AO937.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *DLL Path*: C:\Documents and Settings\admin\Local Settings\Temp\is-6AU5B.tmp_isetup_shfoldr.dll MD5: 92dc6ef532fbb4a5c3201469a5b5eb63 SHA1: 3e89ff837147c16b4e41c30d6c796374e0b8e62c
Malicious Alert	Generic Dll Load Activity	*Message*: DLL loaded
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\is-AO937.tmp\shfolder.dll
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\is-AO937.tmp\shfolder.dll
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\is-AO937.tmp\shfolder.dll
New Dialog Popup		*Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\is-AO937.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp
API Call		*API Name*: ShellExecuteA *Address*: 0x004413ae *Params*: [0x0, NULL, C:\DOCUME~1\admin\LOCALS~1\Temp\83e90785a659ccc2673ed0982cdc1fbf.exe, /SILENT, NULL, 5] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\is-AO937.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *DLL Name*: Shell32.dll
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\is-AO937.tmp\netapi32.dll

Figure 91: Loaded DLL activity found in analysis of 83e90785a659ccc2673ed0982cdc1fbf.exe using MS Windows XP SP3

Network Activity. Network traffic is one of the most common indicators of malicious activities. With modern ubiquitous connectivity, malware developers almost always use network communications to

send commands and retrieve information from infected computers. *Figure 91* shows that there was an attempt to communicate to digicert.com, a private certification authority that starts with a DNS query to resolve a domain name to an IP address.

Network	Dns Query	*Protocol Type:* udp *Qtype:* Host Address *Hostname:* cacerts.digicert.com *Imagepath:* c:\Users\Administrator\AppData\Local\Temp\624e2bab14c48d0c84ee265125811169.exe	2832		
Malicious Alert	Network Activity	*Message:* Network outbound communication attempted			
Network	Dns Query Answer	*Protocol Type:* udp *IP Address:* 199.16.199.2 *Hostname:* cacerts.digicert.com *Imagepath:* c:\Users\Administrator\AppData\Local\Temp\624e2bab14c48d0c84ee265125811169.exe	2832		
API Call		*API Name:* GetSystemTime *Address:* 0x000007fef37cd8cf *Params:* [1badf390] *Imagepath:* C:\Users\Administrator\AppData\Local\Temp\624e2bab14c48d0c84ee265125811169.exe *DLL Name:* kernel32.dll	2832		
Network	Http Request	*Protocol Type:* tcp *Destination Port:* 80 *IP Address:* 199.16.199.2 *Imagepath:* c:\Users\Administrator\AppData\Local\Temp\624e2bab14c48d0c84ee265125811169.exe	2832		

Figure 92: Outgoing network communication to digicert.com detected while executing 624e2bab14c48d0c84ee265125811169.exe on MS Windows 7 x64

More specifically, the analysed software is communicated to the certification authority through Microsoft Crypto API and the implementation of cryptographic algorithms that are available to developers. Such API was used by malware previously to ensure secret communication between the malicious or otherwise unwanted software and service providers[155]. In fact, the initial traffic was sent to Certificate Revocation List (CRL) to validate whether the certificate was legitimate and should not be revoked. Furthermore, Online Certificate Status Protocol (OCSP) was employed to check the validity of the certificate. Finally, 'Error: unsupported certificate purpose' in EMAS reports, which means that the certificate is used for the wrong purpose, such as server certificates for client-side or in the opposite way.

Bot Communication Details:
Event: 1848261

Server DNS Name: *crl3.digicert.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command		User-Agent	Host	Connection	Pragma
GET	/DigiCertAssuredIDRootCA.crl HTTP/1.1		Microsoft-CryptoAPI/6.1	crl3.digicert.com	Keep-Alive	
	Others	Accept: */*				
GET	/sha2-assured-cs-g1.crl HTTP/1.1		Microsoft-CryptoAPI/6.1	crl3.digicert.com	Keep-Alive	
	Others	Accept: */*				

Server DNS Name: *crl4.digicert.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command		User-Agent	Host	Connection	Pragma
GET	/DigiCertAssuredIDRootCA.crl HTTP/1.1		Microsoft-CryptoAPI/6.1	crl4.digicert.com	Keep-Alive	
	Others	Accept: */*				
GET	/sha2-assured-cs-g1.crl HTTP/1.1		Microsoft-CryptoAPI/6.1	crl4.digicert.com	Keep-Alive	
	Others	Accept: */*				

Server DNS Name: *cacerts.digicert.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command		User-Agent	Host	Connection	Pragma
GET	/DigiCertAssuredIDRootCA.crt HTTP/1.1		Microsoft-CryptoAPI/6.1	cacerts.digicert.com	Keep-Alive	
	Others	Accept: */*				

Server DNS Name: *ocsp.digicert.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command		User-Agent	Host	Connection	Pragma
GET	/MFEwTzBNMEswSTAJBgUrDgMCGgUABBT3xL4LQLXDRDM9P665TW442vrsUQQUReuir%2FSSy4IxLVGLp6chnfNtyA8CEAQJGBtf1btmdVNDtW%2BVUAg%3D HTTP/1.1		Microsoft-CryptoAPI/6.1	ocsp.digicert.com	Keep-Alive	
	Others	Accept: */*				
GET	/MFEwTzBNMEswSTAJBgUrDgMCGgUABBSnR4FoxLLkl7vkvsUIFlZt%2BlGH3gQUWsS5eyoKo6XqcQPAYPkt9mV1DlgCEAztJjptShv7XNqq865y%2BKw%3D HTTP/1.1		Microsoft-CryptoAPI/6.1	ocsp.digicert.com	Keep-Alive	
	Others	Accept: */*				

Figure 93: Several queries are used to check the validity of the certificates by employing Microsoft Crypto API during launch of 624e2bab14c48d0c84ee265125811169.exe

[155] Bisson, D., 'Dyre Developer Helped Create TrickBot Malware, Say Researchers', *The State of Security*, 17 October 2016; retrieved from https://www.tripwire.com/state-of-security/latest-security-news/dyre-developer-helped-create-trickbot-malware-say-researchers/.

Another type of network activity that was discovered includes a simple HTTP query to an external host (possibly affiliated with an attacker). This can either be requested to download additional payload or to submit statistics and information about the user who installs the software.

Network	Dns Query	*Protocol Type*: udp *Qtype*: Host Address *Hostname*: www.heydown.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\d0e96f86c1e2f9943e200afa9c1a4fd7.exe
Malicious Alert	Network Activity	*Message*: Network outbound communication attempted
Network	Dns Query Answer	*Protocol Type*: udp *IP Address*: 199.16.199.2 *Hostname*: www.heydown.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\d0e96f86c1e2f9943e200afa9c1a4fd7.exe
API Call		*API Name*: GetSystemDirectoryA *Address*: 0x75999c36 *Params*: [0x36db00, 260] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\d0e96f86c1e2f9943e200afa9c1a4fd7.exe *DLL Name*: kernel32.dll
API Call		*API Name*: Sleep *Address*: 0x75bcd98d *Params*: [60000] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\d0e96f86c1e2f9943e200afa9c1a4fd7.exe *DLL Name*: kernel32.dll
Network	Http Request	*Protocol Type*: tcp *Destination Port*: 80 *IP Address*: 199.16.199.2 *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\d0e96f86c1e2f9943e200afa9c1a4fd7.exe

Figure 94: Outgoing traffic to heydown.com detected while executing d0e96f86c1e2f9943e200afa9c1a4fd7.exe on MS Windows x64

In this case, the request is a simple HTTP GET command that contains md5 hash sum of the file being launched. The website heydown.com claims to be a file-sharing service with up to 5 GB. However, it is a hoax that has a few dummy pages and is probably used to collect statistics from distributed malware. Finally, some of the URLs of these websites redirect users to gamesofpc.com, a website that was used to distribute malware in the very first place.

Bot Communication Details:

Event: 1848417

Server DNS Name: *www.heydown.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command	User-Agent	Host	Connection	Pragma
GET	/gen/1/d0e96f86c1e2f9943e200afa9c1a4fd7.html HTTP/1.1	InetURL/1.0	www.heydown.com	Keep-Alive	

Figure 95: Example of network communication to malicious website that involves sending md5 hash sum of the malware being launched

The next example of network traffic activity involves multiple HTTP requests to the well-known malicious domain *soromomos.com*. It can be seen that there are several requests that appear in the network communication during the execution of the malware sample.

Network	Dns Query	*Protocol Type*: udp *Qtype*: Host Address *Hostname*: rp.soromomos.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe
Malicious Alert	Network Activity	*Message*: Network outbound communication attempted
Network	Dns Query Answer	*Protocol Type*: udp *IP Address*: 199.16.199.4 *Hostname*: rp.soromomos.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe
API Call		*API Name*: SetTimer *Address*: 0x71ec08c8 *Params*: [0x2051e, 0x58, 500, 0x0] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: user32.dll
API Call		*API Name*: Sleep *Address*: 0x02b31f2c *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: kernel32.dll
Network	Http Request	*Protocol Type*: tcp *Destination Port*: 80 *IP Address*: 199.16.199.4 *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe
Network	Http Request	*Protocol Type*: tcp *Destination Port*: 80 *IP Address*: 199.16.199.4 *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe

Figure 96: Outgoing network communication to a malicious website discovered during analysis of f39c99de42f42771b2d4c8ac8e698771.exe using MS Windows 10 x64

Figure 96 shows a summary of the suspicious communication that was discovered and highlighted in the EMAS report. It can be seen that the first communication uses an HTTP POST request to communicate with the server. This is not a common way of communication since the main application area of POST requests are in the forms on the web page that contain multiple fields to be submitted by users.

Bot Communication Details:

Event: 1848533

Server DNS Name: *rp.soromomos.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command	User-Agent	Host	Connection	Pragma
POST	/ HTTP/1.1	Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko	rp.soromomos.com		
	Others	*Accept*: */* *Content-Length*: 1264 *Cache-Control*: no-cache			

Server DNS Name: *info.soromomos.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Raw Command

```
POST /?hofav=0 HTTP/1.1
Accept: */*
Host: info.soromomos.com
User-Agent: Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko
Content-Length: 176
Cache-Control: no-cache

\352^\306S\262\330\267N)Tw?\323G(8\320As(tv\251O\212n_Q\266\256<\312\024UL\302\354R\201\234,\005\177
  \260:>~\306\276\204\037\211\337\252\333\345\036L\232w\247\\H\310\206\377|\265\226\021\330\243Y\37
  3\033\270,0\373s\223\214\312\305\030U1\362\267\242!Cb\233\266$!\005\272\201S\251\345W\326>\271\23
  7\001q\336\351\321\223\226\245I\321\365\277k\247\313\003\036\350\322\020::\266\251\314\235_\343G(
  p\311\331\3272\025DO\274\300\016
<fG\031{\320\3673W\036\212\233\\P\0374\367rG2\312m\003\272+\262\004POST /?tuvi=1 HTTP/1.1
Accept: */*
Host: info.soromomos.com
User-Agent: Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; rv:11.0) like Gecko
Content-Length: 176
Cache-Control: no-cache
```

Figure 97: Summary of the network communication

It becomes more interesting when the communication statistics are collected. As can be seen, a pattern in HTTP POST requests is going to specific URLs. Furthermore, some patterns, and obviously a program, use an algorithm to generate a 3-5 letter URL variable name with corresponding value [0,1,2]. Other files from this website generate similar behaviour with multiple POST requests.

Severity	Summary	Group	Protocol	Count
Warning	Connection reset (RST)	Sequence	TCP	49
Note	The acknowledgment number field is nonzero while the AC...	Protocol	TCP	7
Chat	Connection establish request (SYN): server port 443	Sequence	TCP	39
Chat	Connection establish acknowledge (SYN+ACK): server port ...	Sequence	TCP	39
Chat	M-SEARCH * HTTP/1.1\r\n	Sequence	SSDP	3
Chat	Connection establish request (SYN): server port 80	Sequence	TCP	32
Chat	Connection establish acknowledge (SYN+ACK): server port 80	Sequence	TCP	32
Chat	POST / HTTP/1.1\r\n	Sequence	HTTP	21
Chat	POST /?hofav=0 HTTP/1.1\r\n	Sequence	HTTP	2
Chat	POST /?tuvi=1 HTTP/1.1\r\n	Sequence	HTTP	2
Chat	Connection finish (FIN)	Sequence	TCP	45
Chat	POST /?per=2 HTTP/1.1\r\n	Sequence	HTTP	1
Chat	POST /?wawet=0 HTTP/1.1\r\n	Sequence	HTTP	2
Chat	POST /?jekelat=1 HTTP/1.1\r\n	Sequence	HTTP	2
Chat	POST /?voto=2 HTTP/1.1\r\n	Sequence	HTTP	2

Figure 98: Expert information provided as a summary of abovementioned network communication analysed by Wireshark

Furthermore, by looking at the specific HTTP POST requests, we can see, inter alia, the following information (set of variables with corresponding values) that is travelling in plain text mode from 10.0.0.68.49715 to an external IP address, 199.16.199.4.80. In this regard, the information represents

different characteristics of the victim's OS that is collected by malicious or otherwise unwanted software. For this, we used both Bro and tshark network monitoring tools, which can be successfully applied to analyse dumped network traffic against any malicious patterns[156].

cpuid_pid=178BFBFD00000623

winID=f266b8414e7865cecc84c1c3c2741494

HostParamsMS=124

mdl_ttl=79

mdl_codes=

mdl_names=

mdl_dbver=3

prc_codes=

prc_names=

prc_dbver=4

KernelVer=7.42.3.6671

IRVER=7.42

BRW=Microsoft.MicrosoftEdge_8wekyb3d8bbwe!MicrosoftEdge

IEVer=11.0.10240.16384

BRW_CERT=Microsoft Corporation

CarrierName=

CHNL=

PadTotal=6138

PadSize=0

PadVer=4

c_ver=1.0.8.50650

_makeDate=20170726023411826

_makerVer=3.19.23.6722

pe_cert=../Certificates/OS201608243761BestStandard220617.pfx

_isDbg=0

iHostVer=5.19

bHostVer=8.01

hostBuild=6671

svnRev=78721

svnPath=\Downloaders\roms43\roms43\trunk\Release_OS201608243761BestStandard220617_stub.ini

[156] Valenzuela, I., 'Identifying Malware Traffic with Bro and the Collective Intelligence Framework (CIF)', *Open Security Research*, 11 March 2014; retrieved from http://blog.opensecurityresearch.com/2014/03/identifying-malware-traffic-with-bro.html.

tmpDirSts=1

scr_MonCnt=1

scr_HSzMM=271

scr_VSzMM=203

scr_HRes=1024

scr_VRes=768

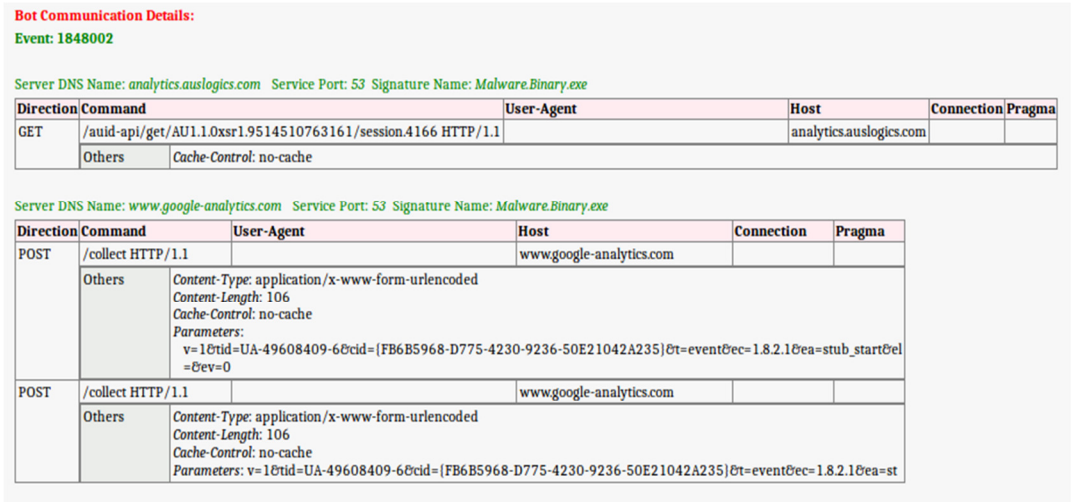
scr_dpi=96

si.dpcto=0

si.iaoi=1

si.

Another example of network traffic activity is when a legitimate service can be used to track user's activities. In *Figure 98*, an example of HTTP POST requests is shown, which was generated by malware during the EMAS analysis. These queries are attributed to Google Analytics, a web analytics platform that collects statistics from web pages based on geo-specific indicators of a user's activity, which are shared with the website owner.



Bot Communication Details:

Event: 1848002

Server DNS Name: *analytics.auslogics.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command	User-Agent	Host	Connection	Pragma
GET	/auid-api/get/AU1.1.0xsr1.9514510763161/session.4166 HTTP/1.1		analytics.auslogics.com		
	Others	*Cache-Control*: no-cache			

Server DNS Name: *www.google-analytics.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command	User-Agent	Host	Connection	Pragma
POST	/collect HTTP/1.1		www.google-analytics.com		
	Others	*Content-Type*: application/x-www-form-urlencoded *Content-Length*: 106 *Cache-Control*: no-cache *Parameters*: v=1&tid=UA-49608409-6&cid={FB6B5968-D775-4230-9236-50E21042A235}&t=event&ec=1.8.2.1&ea=stub_start&el=&ev=0			
POST	/collect HTTP/1.1		www.google-analytics.com		
	Others	*Content-Type*: application/x-www-form-urlencoded *Content-Length*: 106 *Cache-Control*: no-cache *Parameters*: v=1&tid=UA-49608409-6&cid={FB6B5968-D775-4230-9236-50E21042A235}&t=event&ec=1.8.2.1&ea=st			

Figure 99: Bot communication details of 8d89e96947ba51f4924245d1fa77f3ca.exe under analysis on MS Windows XP SP3

The query that was sent to the Google Analytics server is represented below. Among all the parameters, the 'ea' parameter in Google Analytics API implies the custom category of the event[157], which can also be used by the malware developer to gather specific information.

v=1
tid=UA-49608409-6
cid={FB6B5968-D775-4230-9236-50E21042A235}
t=event

[157] Google Analytics, 'Measurement Protocol Parameter Reference', https://developers.google.com/analytics/devguides/collection/protocol/v1/parameters.

ec=1.8.2.1
ea=stub_start
el=
ev=0

As we do not have access to users' information in Google API defined by the token 'tid,' it is hard to say anything about the owner of this account. However, it is clear that HTTP POST requests are being used as legitimate traffic to represent details of the user who installs malicious or otherwise unwanted software.

Process Based Anomaly. *Figure 99* shows how the process creates a duplicate process named explorer.exe. This might be done to create the same inheritable Windows process handle with different access permissions. In addition, this might be a point where the Windows process is used to inject malicious code into the duplicated handle of the original process.

Folder	Created	C:\Users\admin\AppData\Local\Programs	3256
Process	Opened	*Source:* C:\Users\admin\AppData\Local\Temp\is-8J5EV.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *Target:* C:\Windows\explorer.exe	3256 2008
Malicious Alert	Process Based Anomaly	*Message:* Duplicate handle acquired on Windows process	
Process	Opened	*Source:* C:\Users\admin\AppData\Local\Temp\is-8J5EV.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *Target:* C:\Windows\explorer.exe	3256 2008
Process	Opened	*Source:* C:\Users\admin\AppData\Local\Temp\is-8J5EV.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *Target:* C:\Windows\explorer.exe	3256 2008
API Call		*API Name:* GetSystemDirectoryW *Address:* 0x75403d91 *Params:* [0x0, 0] *Imagepath:* C:\Users\admin\AppData\Local\Temp\is-8J5EV.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *DL L Name:* kernel32.dll	3256
API Call		*API Name:* GetSystemDirectoryW *Address:* 0x75403db8 *Params:* [0x68e660, 20] *Imagepath:* C:\Users\admin\AppData\Local\Temp\is-8J5EV.tmp\83e90785a659ccc2673ed0982cdc1fbf.tmp *DL L Name:* kernel32.dll	3256

Figure 100: Example of duplicated process handle while executing 83e90785a659ccc2673ed0982cdc1fbf.exe using Microsoft Windows 7 SP1

Application Crash Activity. This is a general class of activities that can appear during execution of software, also followed by a specific error exception if this was predefined by developers. As can be seen in *Figure 100*, the application crashed with a 'status_access_violation' error type, which could mean that either the process tried to access memory space that it was not supposed to access or it was operating with data outside the permitted area.

High Cpu		*Imagepath:* C:\Users\Administrator\AppData\Local\Temp\5175ea1cecb14da7c521cb1943fc178d.exe
ProcessTelemetryReport		*Imagepath:* C:\Users\Administrator\AppData\Local\Temp\5175ea1cecb14da7c521cb1943fc178d.exe
High Cpu		*Imagepath:* C:\Users\Administrator\AppData\Local\Temp\5175ea1cecb14da7c521cb1943fc178d.exe
Appexception		*Exception Faulting Address:* 0x34 *Exception Code:* 0xc0000005 *Exception Level:* SECOND_CHANCE *Exception Type:* STATUS_ACCESS_VIOLATION *Instruction Address:* 0x05eb995e *Description:* N/A *Imagepath:* C:\Users\Administrator\AppData\Local\Temp\5175ea1cecb14da7c521cb1943fc178d.exe *Call Stack:* Frame No. | Instruction Addr. | Module Name | Symbol Name | SD
Malicious Alert	Application Crash Activity	*Message:* Application crash detected
Process	Terminated	C:\Users\Administrator\AppData\Local\Temp\5175ea1cecb14da7c521cb1943fc178d.exe *Parentname:* C:\Windows\explorer.exe *Command Line:* N/A
High Cpu		*Imagepath:* C:\Users\Administrator\AppData\Local\Temp\5175ea1cecb14da7c521cb1943fc178d.exe

Figure 101: Application was putting a high load on the CPU with subsequent crashing for file 5175ea1cecb14da7c521cb1943fc1.exe launched under MS Windows 7 x64

Install Activity. The process of installing software on Microsoft Windows OS includes several stages, among those creating files, registry entries, establishing links to files, etc. The most popular script language/system that is used to create such installers is NSIS[158] (*Nullsoft scriptable install system*). In

[158] Nullsoft scriptable install system, http://nsis.sourceforge.net/.

Figure 101, it can be seen that a folder is first created with the corresponding DLL file. Furthermore, the 'System.dll' library is loaded by the process. Even though the files pretend to be system files, it is hardly possible that this is a benign component.

Folder	Created	C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp
File	Failed	C:\Users\ADMINI~1\AppData\Local\Temp\nsf4937.tmp\SYSTEM.DLL
File	Created	C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp\System.dll
Malicious Alert	Install Activity	*Message*: NSIS Install Activity
File	Close	C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp\System.dll
File	Close	C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp\System.dll MD5: 959ea64598b9a3e494c00e8fa793be7e SHA1: 40f284a3b92c2f04b1038def79579d4b3d066ee0
DLL Loaded		*Imagepath*: C:\Users\Administrator\AppData\Local\Temp\edabc5d017281cf973587185ceb56307.exe *DLL Path*: C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp\System.dll MD5: 959ea64598b9a3e494c00e8fa793be7e SHA1: 40f284a3b92c2f04b1038def79579d4b3d066ee0
Malicious Alert	Generic Dll Load Activity	*Message*: DLL loaded
File	Failed	C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp\System.dll
API Call		*API Name*: CLSIDFromString *Address*: 0x10002a04 *Params*: [""] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\edabc5d017281cf973587185ceb56307.exe *DLL Name*: Ole32.dll
File	Failed	C:\Users\Administrator\AppData\Local\Temp\nsf4937.tmp\System.dll

Figure 102: Installer activity detected during execution of edabc5d017281cf973587185ceb56307.exe on MS Windows 10 x64

Kernel-Level Activity. A driver provides access to the hardware layer from the OS, meaning that it needs corresponding kernel-level access. Therefore, this may indicate rootkit activity if the process attempts to load a driver.

File	Rename	*Old Name*: C:\Users\Administrator\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations\0AE3PGXIGTWOESFG5.OML.temp *New Name*: C:\Users\Administrator\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations\873c1e62f5d6da3f.customDestinations-ms *Imagepath*: C:\Program Files (x86)\Mozilla Firefox\firefox.exe
File	Delete	C:\Users\Administrator\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations\873c1e62f5d6da3f.customDestinations-ms~RF698853.TMP
Malicious Alert	Kernal Level Activity	*Message*: Process is loading a driver

Figure 103: Kernel activity for 8b3ccf367c2b033ca560b37e83f47875.dll on MS Windows 7 x64

DGA Activity. A domain-generation algorithm (DGA) is a specific set of methods designed to generate domain names and is known only by the attacker to obfuscate network communication and guarantee persistence of the malware in the network. An example below demonstrates EMAS detection for suspicious third-level domain names.

Network	Dns Query	*Protocol Type*: udp *Qtype*: Host Address *Hostname*: crl4.digicert.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\61330d8acbb7800e49e63bd411ef20ab.exe
Network	Dns Query Answer	*Protocol Type*: udp *IP Address*: 199.16.199.5 *Hostname*: crl4.digicert.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\61330d8acbb7800e49e63bd411ef20ab.exe
API Call		*API Name*: GetSystemTime *Address*: 0x000007fef328d8cf *Params*: [1c8af070] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\61330d8acbb7800e49e63bd411ef20ab.exe *DLL Name*: kernel32.dll
Network	Http Request	*Protocol Type*: tcp *Destination Port*: 80 *IP Address*: 199.16.199.5 *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\61330d8acbb7800e49e63bd411ef20ab.exe
File	Failed	C:\Users\Administrator\AppData\LocalLow\Microsoft\CryptnetUrlCache\MetaData\616AD1AB067CFD351D6C0EF6F3E12F40
Network	Dns Query	*Protocol Type*: udp *Qtype*: Host Address *Hostname*: crl3.digicert.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\61330d8acbb7800e49e63bd411ef20ab.exe
Malicious Alert	DGA Activity	*Message*: Suspicious Network Activity
Network	Dns Query Answer	*Protocol Type*: udp *IP Address*: 199.16.199.6 *Hostname*: crl3.digicert.com *Imagepath*: c:\Users\Administrator\AppData\Local\Temp\61330d8acbb7800e49e63bd411ef20ab.exe

Figure 104: Example of DGA traffic for 61330d8acbb7800e49e63bd411ef20ab.exe detected during execution on MS Windows 7 x64

High Repeated Sleep Calls. Sleep function is used in malicious or otherwise unwanted software to delay the execution of the payload or to wait until some conditions are met. This means that if malicious payload execution is delayed, then it becomes extremely difficult to detect such activities using behavioural indicators.

API Call		*API Name*: GetLocalTime *Address*: 0x02b0bfe5 *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: kernel32.dll
API Call		*API Name*: Sleep *Address*: 0x02b31f2c *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: kernel32.dll
API Call		*API Name*: Sleep *Address*: 0x02b31f2c *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: kernel32.dll
Malicious Alert	High Repeated Sleep Calls	*Message*: High repeated sleep calls
API Call		*API Name*: GetSystemTime *Address*: 0x701678fa *Params*: [0x19ec0c] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: kernel32.dll
API Call		*API Name*: GetSystemTime *Address*: 0x701678fa *Params*: [0x19e19c] *Imagepath*: C:\Users\Administrator\AppData\Local\Temp\f39c99de42f42771b2d4c8ac8e698771.exe *DLL Name*: kernel32.dll

Figure 105: Multiple calls of sleep functions done by f39c99de42f42771b2d4c8ac8e698771.exe during execution on Microsoft Windows 10 x64

Keylogging Activity. This activity captures any keys that are being pressed on a keyboard by a user in order to steal sensitive information, including logins and passwords. It can be seen that the application is called SetWindowsHookEx() function, which is designed to set up a specifically designed hook into the execution hook chain to monitor a specific set of events.

API Call		*API Name*: GetComputerNameExW *Address*: 0x77927048 *Params*: [0, 0x236ed5c, 0x236ed58] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\7fe2fdbacf6b4563cf895a19c0375059.exe *DLL Name*: kernel32.dll
Regkey	Queryvalue	\REGISTRY\MACHINE\SYSTEM\ControlSet001\Control\ComputerName\ActiveComputerName\"ComputerName"
API Call		*API Name*: GetComputerNameExW *Address*: 0x779270ab *Params*: [3, 0x236ed5c, 0x236ed58] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\7fe2fdbacf6b4563cf895a19c0375059.exe *DLL Name*: kernel32.dll
API Call		*API Name*: SetWindowsHookExW *Address*: 0x02536828 *Params*: [7, 0x259c591, 0x0, 3744] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\7fe2fdbacf6b4563cf895a19c0375059.exe *DLL Name*: user32.dll
Malicious Alert	Keylogging Activity	*Message*: High-level keyboard hook registered
Malicious Alert	Keylogging Activity	*Message*: Thread specific hook registered
API Call		*API Name*: SetWindowsHookExW *Address*: 0x02536839 *Params*: [2, 0x250a3c0, 0x0, 3744] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\7fe2fdbacf6b4563cf895a19c0375059.exe *DLL Name*: user32.dll

Figure 106: Example of keyboard hook being registered by 7fe2fdbacf6b4563cf895a19c0375059.exe while executing on MS Windows XP SP3

Generic Anomalous Activity. This class of events in the EMAS report describes general anomalies that can be suspicious and related to malicious activities. In *Figure 106*, the suspicious activity that was detected when a process made an attempt to launch explorer is shown. Moreover, from the sequence of events, it can be seen that the software first deleted a default browser value in the registry. Then, third-party software, DefaultBrwoserFinder.exe, was launched, whose probable intention was to find a default browser used by the OS. These actions can be explained by an attempt to replace the default browser or modify settings, such as a home page in the default browser, so a user is redirected there each time the browser is launched.

Folder	Open	C:\Users\Administrator\AppData\Roaming\Microsoft\Windows\Cookies
File	Failed	C:\Users\Administrator\AppData\Local\Microsoft\Internet Explorer\DOMSTORE
File	Failed	C:\Users\Administrator\AppData\LocalLow\Microsoft\Internet Explorer\DOMSTORE
Regkey	Deleteval	\REGISTRY\MACHINE\SOFTWARE\Wow6432Node\TweakBit\Driver Updater\1.x\Settings\"General.DefWebBrowser"
Process	Terminated	C:\Users\Administrator\AppData\Local\Temp\is-CII47.tmp\DefaultBrowserFinder.exe *Parentname:* C:\Users\Administrator\AppData\Local\Temp\is-T0NV9.tmp\e735319ff70ebb722f1949bec8519bdd.tmp *Command Line:* N/A
Malicious Alert	Generic Anomalous Activity	*Message:* Process Opening explorer

Figure 107: Default browser modification in e735319ff70ebb722f1949bec8519bdd.exe launched on MS Windows 7 x64

Another anomalous activity is when software launched a Processs32First()function to retrieve information about a first process in the process list. This can be done for a number of reasons in combination with the Sleep() function, such as the need to have some instance of an arbitrary program running to check whether the previously invoked program is actually running.

API Call		*API Name:* Sleep *Address:* 0x01ed1cef *Params:* [100] *Imagepath:* C:\Users\admin\AppData\Local\Temp\d8b5eeb2ecd229c8869f32eb925ce23a.exe *DLL Name:* kernel32.dll
API Call		*API Name:* Process32First *Address:* 0x01e6f4e4 *Params:* [0x314, 0x3b4f5d0] *Imagepath:* C:\Users\admin\AppData\Local\Temp\d8b5eeb2ecd229c8869f32eb925ce23a.exe *DLL Name:* kernel32.dll
Malicious Alert	Generic Anomalous Activity	*Message:* Enumerating running processes
API Call		*API Name:* VerifyVersionInfoA *Address:* 0x01ec2e0a *Params:* [5, 1, 0, 0, 2, 0, 0, 0, 51, 800000000001b01b] *Imagepath:* C:\Users\admin\AppData\Local\Temp\d8b5eeb2ecd229c8869f32eb925ce23a.exe *DLL Name:* kernel32.dll
API Call		*API Name:* Sleep *Address:* 0x01f1032b *Params:* [500] *Imagepath:* C:\Users\admin\AppData\Local\Temp\d8b5eeb2ecd229c8869f32eb925ce23a.exe *DLL Name:* kernel32.dll

Figure 108: Looking into the process list on the computer in d8b5eeb2ecd229c8869f32eb925ce23a.exe launched on MS Windows 7 SP1

Rootkit Activity. This is an activity that malware can show while attempting to get unauthorised access through employing different approaches, including injections in DLL for further advancing on a victim's machine. EMAS did not provide any further details for such activity, except that a hidden process in user mode was created.

Hiddenproc	User mode	*Imagepath:* C:\Windows\SysWOW64\WerFault.exe	10732
Malicious Alert	Rootkit Activity	*Message:* Rootkit behavior observed	
DLL Loaded		*Imagepath:* C:\Windows\SysWOW64\rundll32.exe *DLL Path:* C:\Users\Administrator\AppData\Local\Temp\bc83108b18756547013ed443b8cdb31b.dll MD5: bc83108b18756547013ed443b8cdb31b SHA1: 79bcaad3714433e01c7f153b05b781f8d7cb318d	9012
DLL Loaded		*Imagepath:* C:\Windows\SysWOW64\rundll32.exe *DLL Path:* C:\Users\Administrator\AppData\Local\Temp\bc83108b18756547013ed443b8cdb31b.dll MD5: bc83108b18756547013ed443b8cdb31b SHA1: 79bcaad3714433e01c7f153b05b781f8d7cb318d	6680
Regkey	Queryvalue	\REGISTRY\MACHINE\SYSTEM\ControlSet001\Control\ComputerName\ActiveComputerName\"ComputerName"	6900
DLL Loaded		*Imagepath:* C:\Windows\SysWOW64\rundll32.exe *DLL Path:* C:\Users\Administrator\AppData\Local\Temp\bc83108b18756547013ed443b8cdb31b.dll MD5: bc83108b18756547013ed443b8cdb31b SHA1: 79bcaad3714433e01c7f153b05b781f8d7cb318d	6900

Figure 109: Suspicious rootkit behaviour for bc83108b18756547013ed443b8cdb31b.dll analysed on MS Windows 10 x64

Adding CA Certificate Activity. A digital certificate can be added by malicious or otherwise unwanted software to ensure successful operation on a computer. As *Figure 109* demonstrates, the software adds its own third-party root certificate authority into the Windows registry. Under Microsoft Windows, this implies that any non-Microsoft certificates can be placed there. By adding such certificates, malware can initialise any kind of protected communication. Moreover, malware has added at least a dozen such certificates in the system.

Regkey	Deleteval	\REGISTRY\MACHINE\SOFTWARE\Microsoft\SystemCertificates\ROOT\Certificates\"357E6DF727AA1C553B0057 38C 261E1D46F9C4ACF"
Regkey	Setval	\REGISTRY\MACHINE\SOFTWARE\Microsoft\SystemCertificates\ROOT\Certificates\357E6DF727AA1C553B00573 8C2 61E1D46F9C4ACF\"Blob" = 04 00 00 00 01 00 00 00 10 00 00 00 98 cf 24 6c d0 67 9a 45 84 92 87 c5 4 3 85 d9 a3 14 00 00 00 01 00 00 00 14 00 00 00 cb 82 fa 3b 2a 5d 21 b9 35 98 06 61 49 73 c5 61 87 d9 0f 7a 03 00 00 00 01 00 00 00 14 00 00 00 35 7e 6d f7 27 aa 1c 55 3b 00 57 38 c2 61 e1 d4 6f 9c 4a cf 19 00 00 00 01 00 00 00 10 00 00 00 bf fb 58 ab 01 b8 00 f1 fe ab 74 1a 23 63 10 30 20 0 0 00 00 01 00 00 00 70 03 00 00 30 82 03 6c 30 82 02 54 a0 03 02 01 02 02 09 00 b4 5f ac 4e d6 f3 cb 79 30 0d 06 09 2a 86 48 86 f7 0d 01 01 0b 05 00 30 43 31 0b 30 09 06 03 55 04 06 13 02 55 53 31 1a 30 18 06 03 55 04 0a 13 11 47 6c 6f 62 61 6c 53 69 67 6e 65 72 20 49 6e 63 2e 31 18 30 16 0 6 03 55 04 03 13 0f 47 6c 6f 62 61 6c 53 69 67 6e 65 72 20 43 41 30 1e 17 0d 31 36 30 36 31 37 30 30 34 33 34 30 5a 17 0d 32 36 30 36 31 35 30 30 34 33 34 30 5a 30 43 31 0b 30 09 06 03 55 04 06 13 02 55 53 31 1a 30 18 06 03 55 04 0a 13 11 47 6c 6f 62 61 6c 53 69 67 6e 65 72 20 49 6e 63 2e 3 1 18 30 16 06 03 55 04 03 13 0f 47 6c 6f 62 61 6c 53 69 67 6e 65 72 20 43 41 30 82 01 22 30 0d 06 09 2a 86 48 86 f7 0d 01 01 01 05 00 03 82 01 0f 00 30 82 01 0a 02 82 01 01 00 a4 b5 41 fa 28 69 90 6d 9a 21 14 b9 4e 9e 61 91 f2 e8 5a 40 23 e5 a3 92 72 ca 4d 7f 56 17 c8 29 17 db 99 3a 7b c7 f e 6c 3a ff e5 ed 9d 7c e0 1f 29 c1 bc 17 21 8f 73 d9 06 4a af c3 ba 38 cd f7 cd 47 5f c3 be d7 6f 01 fd 44 44 b7 ba 38 90 b0 e9 b1 d3 83 12 7b 78 62 ea b7 16 65 6b ea fd 0a b7 64 dd 11 5b f1 12 8e 4e 72 9b a5 29 d9 19 54 c4 be 5c 92 5c 19 03 4e 79 ce 77 2b 7c 8d 7d 7e 6e f7 ff 48 fb c9 bc c 4 ae d9 ef 2d 51 0f 78 e6 28 d0 01 e6 6d 05 6f 30 31 3a 0b fa 2e fa 29 da f1 a7 c0 a3 57 3d 6f 9f 1e d5 3d 37 37 f6 8f 71 f8 f7 88 62 fe f4 ba 3b d4 83 03 7e f8 e3 95 af 18 a4 c7 5d de 84 bb 4d a4 f2 c8 de ac 1b 9f ee c4 54 20 a0 6a 3d 31 b7 d3 7c 65 68 ec d6 47 3f 73 f2 80 1e 28 e6 28 47 e d 35 9e 88 55 53 d4 ec 5b 07 6a 36 e1 87 4a ea 14 7e b4 92 4b 5c 47 bd 02 03 01 00 01 a3 63 30 61 30 1d 06 03 55 1d 0e 04 16 04 14 cb 82 fa 3b 2a 5d 21 b9 35 98 06 61 49 73 c5 61 87 d9 0f 7a 30 1f 06 03 55 1d 23 04 18 30 16 80 14 cb 82 fa 3b 2a 5d 21 b9 35 98 06 61 49 73 c5 61 87 d9 0f 7a 3 0 0f 06 03 55 1d 13 01 01 ff 04 05 30 03 01 01 ff 30 0e 06 03 55 1d 0f 01 01 ff 04 04 03 02 01 86 30 0d 06 09 2a 86 48 86 f7 0d 01 01 0b 05 00 03 82 01 01 00 31 cc 48 58 47 8b fe 22 d5 20 13 3f f2 67 de 9c 2b c8 4f 5f cb a1 4b ce 03 5e 43 23 54 dc 05 c7 fe 3a 83 96 e6 bb 0e 0d f9 66 52 e5 d b 88 94 3c e4 88 c4 b3 f3 fa d6 8c 95 af af 38 42 96 6c 49 8b 79 80 7c f9 6d 5c 77 5c da e5 f2 56 b2 9e 19 04 dc 57 a2 92 d2 7d 91 17 f7 c5 35 64 70 80 aa bf 01 6a b8 50 f4 0c 26 22 53 83 e5 67 69 b4 b8 9a 4c 77 bc 5e aa 41 77 cf d9 3f f6 74 99 4a fa 71 01 76 33 89 54 e1 7c b4 5c fa bf c2 4 1 6e a8 5a d7 de 77 52 75 0e 1f 55 67 60 1a ee 09 70 56 7c 66 96 1e 69 5a 41 3f 63 5e 7b 7a a5 f7 7d 7f b1 8e da 4d ce e6 4c b4 91 82 74 3a 97 2c 3f ee 44 3f 5c 92 57 3b bc 33 4e aa 9d 33 6c d9 ec ec f8 cc 4a c7 1d cf a0 f7 56 f9 40 43 49 07 d3 57 df a3 9a de ee 87 91 bd 99 93 94 5c 4a 56 8 e 7d e7 85 f7 81 d9 7c cc 2b eb ea b1 3c bd 21 18 27
Malicious Alert	Adding CA Certificate Activity	*Message:* Adding CA Certificate
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\Microsoft\SystemCertificates\AuthRoot\Certificates
Regkey	Deleteval	\REGISTRY\MACHINE\SOFTWARE\Microsoft\SystemCertificates\AuthRoot\Certificates\"4EFCED9C6BDD0C985CA 3C 7D253063C5BE6FC620C"
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\Microsoft\SystemCertificates\AuthRoot\Certificates\4EFCED9C6BDD0C985CA3 C7 D253063C5BE6FC620C

Figure 110: Example of the content of digital certificate entered into registry value for the system certificates by edabc5d017281cf973587185ceb56307.exe during execution on MS Windows XP SP3

Process Cloned. It might be that the malware clones itself for a variety of purposes by creating another file on the disc that might look legitimate. As can be seen in *Figure 110*, the malware first created a directory and placed the executable there. Furthermore, it created and modified a Zone.Identifier file, which is normally created when a file is being downloaded from other sources. In this case, Windows might display a warning message based on the Zone.Indentifier value. However, malware developers can modify it to eliminate the appearance of such alerts. Furthermore, timestamps had been modified and a file copied, also adding a registry entry to rename it during the next boot. Later, the new file had been launched, meaning that the process created a new copy of itself that might look less suspicious.

Folder	Created	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp	604		
File	Failed	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe	604		
File	Created	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe	604		
File	Close	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe	604		
File	Open	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe	604		
File	Created	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe:Zone.Identifier	604		
File	Date Change	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe:Zone.Identifier	604		26
File	CopyTimestamp	*Source:* C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe:Zone.Identifier MD5: fbccf14d504b7b2dbcb5a5bda75bd93b SHA1: d59fc84cdd5217c6cf74785703655f78da6b582b	604		26
File	Date Change	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe	604		67600
File	CopyFile	*Source:* C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe MD5: b92fdca08753528a148317864f99ab6f SHA1: 57f3edc316ea126ad8301eb0025b8c38f5458131	604		67600
Malicious Alert	Process Cloned	*Message:* Process clones and starts itself			
Regkey	Setval	\REGISTRY\MACHINE\SYSTEM\ControlSet001\Control\SESSION MANAGER\"PendingFileRenameOperations" = \??\C :\Users\ADMINI~1\AppData\Local\Temp\~nsuA.tmp\Un_A.exe\0\0\0	604		
Process	Started	C:\Users\Administrator\AppData\Local\Temp\~nsuA.tmp\Un_A.exe *Parentname:* C:\Users\Administrator\AppData\Local\Temp\b92fdca08753528a148317864f99ab6f.exe *Command Line:* "C:\Users\ADMINI~1\AppData\Local\Temp\~nsuA.tmp\Un_A.exe" _?=C:\Users\ADMINI~1\AppData\Local\Temp\ MD5: b92fdca08753528a148317864f99ab6f SHA1: 57f3edc316ea126ad8301eb0025b8c38f5458131	1608	604	67600
File	Failed	C:\Users\ADMINI~1\AppData\Local\Temp\~nsuA.tmp\UI\SWDRM.DLL	604		
File	Failed	C:\Users\ADMINI~1\AppData\Local\Temp\~nsuA.tmp\UN_B.EXE	604		

Figure 111: Example of cloned process by b92fdca08753528a148317864f99ab6f.exe while running on MS Windows 7 x64

Suspicious Directory. Under normal conditions, user programs must work with files only in some directories. However, there is also a specifically designated folder inside the Microsoft Windows system directory. Such directories are usually protected, as the modification of files in these directories may cause harm. In particular, a host file has been modified, meaning that some of the traffic can be redirected from the victim's computer to another server or machine.

API Call		*API Name:* GetSystemDirectoryW *Address:* 0x00e92065 *Params:* [0x11af800, 260] *Imagepath:* C:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe *DLL Name:* kernel32.dll
File	Failed	C:\Users\admin\AppData\Local\Temp\DHCPCSVC.DLL
File	Failed	C:\ProgramData\BYTEFENCE\RTOP\HOSTS_BACKUP
File	Failed	C:\ProgramData\BYTEFENCE\RTOP\HOSTS_BACKUP
File	Failed	C:\ProgramData\BYTEFENCE\RTOP\HOSTS_BACKUP
File	Open	C:\Windows\System32\drivers\etc\hosts
File	Modified	C:\Windows\System32\drivers\etc\hosts MD5: 7558204c590726fbb39849ae1a20e20f SHA1: d6ff6a432c3f0f428f9a356e892e3f24bb655976
Malicious Alert	Suspicious Directory	*Message:* File created/tampered/deleted in suspicious location
Mutex		\BaseNamedObjects\unchecky_svc
File	Failed	C:\ProgramData\BYTEFENCE\RTOP\UCLOGFILE.BIN
File	Failed	C:\Users\admin\AppData\Local\Temp\RTOP_BG.EXE
Uac	Service	rtop

Figure 112: Host file has been modified by a8fbdf79f7bff18ac1e55d41ee6a5030.exe during launch on MS Windows 7 SP1

Misc Anomaly. There can be some other suspicious activity during the analysis of malware, such as tampering with network traffic by changing the content of network packets. *Figure 112* below shows an HTTP request that was analysed in EMAS and contains a request to an external server with corresponding details of GET parameters. It might be that this request was tampered with.

Network	Dns Query	*Protocol Type*: udp *Qtype*: Host Address *Hostname*: update.bytefence.com *Imagepath*: c:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe
Malicious Alert	Network Activity	*Message*: Network outbound communication attempted
Network	Dns Query Answer	*Protocol Type*: udp *IP Address*: 199.16.199.2 *Hostname*: update.bytefence.com *Imagepath*: c:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe
API Call		*API Name*: GetSystemTime *Address*: 0x72bcf87c *Params*: [0x105f994] *Imagepath*: C:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe *DLL Name*: kernel32.dll
API Call		*API Name*: SystemTimeToFileTime *Address*: 0x72bcf887 *Params*: [0x105f994, 0x105f9cc] *Imagepath*: C:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe *DLL Name*: kernel32.dll
API Call		*API Name*: GetSystemTime *Address*: 0x00e9aff5 *Params*: [0x11af894] *Imagepath*: C:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe *DLL Name*: kernel32.dll
Network	Http Request	*Protocol Type*: tcp *Destination Port*: 80 *IP Address*: 199.16.199.2 *Imagepath*: c:\Users\admin\AppData\Local\Temp\a8fbdf79f7bff18ac1e55d41ee6a5030.exe
Malicious Alert	Misc Anom	*Message*: Network Tampering Activity
Malicious Alert	Network Tampering Activity	*Message*: Hosts file modified

Figure 113: Anomalies generated by a8fbdf79f7bff18ac1e55d41ee6a5030.exe during launch on MS Windows 7 SP1

Server DNS Name: *update.bytefence.com* Service Port: *53* Signature Name: *Malware.Binary.exe*

Direction	Command	User-Agent	Host	Connection	Pragma
GET	/files/rtop_setup_version?uv=1.0.2&uid=tDtDtBtDtCzBtCtCtDtCyCyD1S1OtC1O&bv= HTTP/1.1	Unchecky/1.0.2	update.bytefence.com	Keep-Alive	

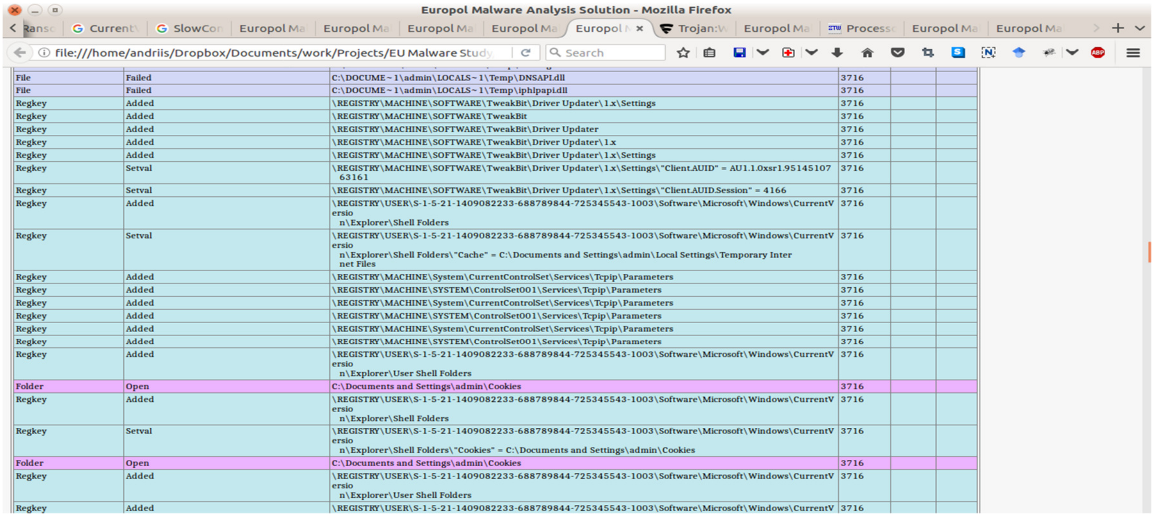
Figure 114: Example of network communications by the abovementioned software

Generic Persistence Activity. Another example of the malicious activities targeting consistent presence on the victim's computer is changing the OS configuration in a way to keep malware running later on. *Figure114* shows that there was a new task created in the Windows task list, which will probably run every time the computer boots.

API Call		*API Name*: CryptAcquireContextA *Address*: 0x767f41f4 *Params*: [NULL, NULL, 12, 4026531840] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\d9a03f672173af04b41f0a0752441199.exe *DLL Name*: advapi32.dll
API Call		*API Name*: CryptAcquireContextW *Address*: 0x767f424a *Params*: [NULL, NULL, 18, 4026531840] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\d9a03f672173af04b41f0a0752441199.exe *DLL Name*: advapi32.dll
File	Created	C:\WINDOWS\Tasks\Start WinZip Driver Updater for TargetBiz@admin(logon).job
Malicious Alert	Generic Persistance Activity	*Message*: System tasks modified
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\dssenh.dll
2 Repeated items skipped		
File	Close	C:\WINDOWS\Tasks\Start WinZip Driver Updater for TargetBiz@admin(logon).job MD5: 5a621e69ae9a2722ac2c650271b5267c SHA1: 7dc70070473db3d88bc4a665b1696d9e76141cdf

Figure 115: Software d9a03f672173af04b41f0a0752441199.exe adds itself to OS tasks on MS Windows XP SP3

Registry Activities. All settings in Microsoft Windows are stored in a hierarchical storage, also called the registry. These include different low-level parameters of the OS, drivers, installed software, user settings, etc. Programs usually make changes or query different keys in this database. It might be challenging to detect clearly malicious activities in those actions. The figure below shows that the registry entries are created by the 'useful' program that is being installed. Later, there are a few registry entries added to the TCP/IP parameters. These usually include different network configuration settings such as DNS servers, etc.



File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\DNSAPI.dll	3716		
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\iphlpapi.dll	3716		
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\TweakBit\Driver Updater\1.x\Settings	3716		
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\TweakBit	3716		
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\TweakBit\Driver Updater	3716		
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\TweakBit\Driver Updater\1.x	3716		
Regkey	Added	\REGISTRY\MACHINE\SOFTWARE\TweakBit\Driver Updater\1.x\Settings	3716		
Regkey	Setval	\REGISTRY\MACHINE\SOFTWARE\TweakBit\Driver Updater\1.x\Settings\"Client.AUID" = AU1.1.0xsr1.95145107 63161	3716		
Regkey	Setval	\REGISTRY\MACHINE\SOFTWARE\TweakBit\Driver Updater\1.x\Settings\"Client.AUID.Session" = 4166	3716		
Regkey	Added	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\Shell Folders	3716		
Regkey	Setval	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\Shell Folders\"Cache" = C:\Documents and Settings\admin\Local Settings\Temporary Inter net Files	3716		
Regkey	Added	\REGISTRY\MACHINE\System\CurrentControlSet\Services\Tcpip\Parameters	3716		
Regkey	Added	\REGISTRY\MACHINE\SYSTEM\ControlSet001\Services\Tcpip\Parameters	3716		
Regkey	Added	\REGISTRY\MACHINE\System\CurrentControlSet\Services\Tcpip\Parameters	3716		
Regkey	Added	\REGISTRY\MACHINE\SYSTEM\ControlSet001\Services\Tcpip\Parameters	3716		
Regkey	Added	\REGISTRY\MACHINE\System\CurrentControlSet\Services\Tcpip\Parameters	3716		
Regkey	Added	\REGISTRY\MACHINE\SYSTEM\ControlSet001\Services\Tcpip\Parameters	3716		
Regkey	Added	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\User Shell Folders	3716		
Folder	Open	C:\Documents and Settings\admin\Cookies	3716		
Regkey	Added	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\Shell Folders	3716		
Regkey	Setval	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\Shell Folders\"Cookies" = C:\Documents and Settings\admin\Cookies	3716		
Folder	Open	C:\Documents and Settings\admin\Cookies	3716		
Regkey	Added	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\User Shell Folders	3716		
Regkey	Added	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentV	3716		

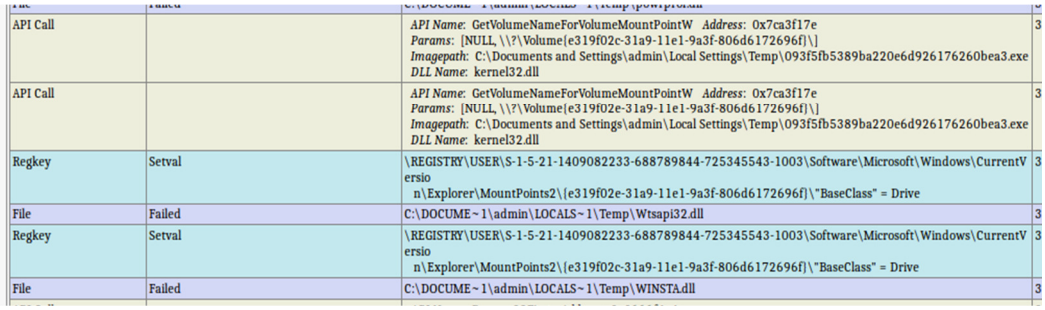
Figure 116: Registry artefacts observed during execution of 8d89e96947ba51f4924245d1fa77f3ca.exe on MS Windows XP SP3

Another example is shown in *Figure 116*, where software queries the system to obtain information about the BIOS manufacturer, its version, CPU name, and product ID of the OS. It is very uncommon to request such details, especially the BIOS-related information. Apparently, this is collected by attackers for statistics and is later transmitted through the network. Otherwise, there is no logical explanation as to why a game installer would need BIOS details.

API Call		*API Name*: GetSystemDirectoryW *Address*: 0x01f3d123 *Imagepath*: C:\Users\admin\AppData\Local\Temp\d242928485fc02b227f11ddcfbf68f3e.exe *DLL Name*: kernel32.dll
4 Repeated items skipped		
API Call		*API Name*: Sleep *Address*: 0x01fa1cef *Imagepath*: C:\Users\admin\AppData\Local\Temp\d242928485fc02b227f11ddcfbf68f3e.exe *DLL Name*: kernel32.dll
Regkey	Queryvalue	\REGISTRY\MACHINE\HARDWARE\DESCRIPTION\System\"SystemBiosVersion"
2 Repeated items skipped		
Regkey	Queryvalue	\REGISTRY\MACHINE\HARDWARE\DESCRIPTION\System\BIOS\"SystemManufacturer"
2 Repeated items skipped		
Regkey	Queryvalue	\REGISTRY\MACHINE\HARDWARE\DESCRIPTION\System\BIOS\"SystemProductName"
Regkey	Queryvalue	\REGISTRY\MACHINE\HARDWARE\DESCRIPTION\System\CentralProcessor\0\"ProcessorNameString"
Regkey	Queryvalue	\REGISTRY\MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\"ProductId"

Figure 117: Registry artefacts found during execution of d242928485fc02b227f11ddcfbf68f3e.exe on MS Windows XP SP3

Finally, the figure below shows how malware uses the Windows API function to retrieve the GUID of a volume or drive mounted in a system twice. The queries have been done with only a difference in the time part of the GUID structure.



API Call		*API Name*: GetVolumeNameForVolumeMountPointW *Address*: 0x7ca3f17e *Params*: [NULL, \\?\Volume{e319f02c-31a9-11e1-9a3f-806d6172696f}\] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\093f5fb5389ba220e6d926176260bea3.exe *DLL Name*: kernel32.dll	3
API Call		*API Name*: GetVolumeNameForVolumeMountPointW *Address*: 0x7ca3f17e *Params*: [NULL, \\?\Volume{e319f02e-31a9-11e1-9a3f-806d6172696f}\] *Imagepath*: C:\Documents and Settings\admin\Local Settings\Temp\093f5fb5389ba220e6d926176260bea3.exe *DLL Name*: kernel32.dll	3
Regkey	Setval	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\MountPoints2\{e319f02e-31a9-11e1-9a3f-806d6172696f}\"BaseClass" = Drive	3
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\Wtsapi32.dll	3
Regkey	Setval	\REGISTRY\USER\S-1-5-21-1409082233-688789844-725345543-1003\Software\Microsoft\Windows\CurrentVersio n\Explorer\MountPoints2\{e319f02c-31a9-11e1-9a3f-806d6172696f}\"BaseClass" = Drive	3
File	Failed	C:\DOCUME~1\admin\LOCALS~1\Temp\WINSTA.dll	3

Figure 118: Software 093f5fb5389ba220e6d926176260bea3.exe successfully queries mounted volume during launch on MS Windows SP3

ISBN 978-92-9156-254-1 doi:10.2814/004056 TB-01-18-336-EN-N

IDENTIFICATION AND ANALYSIS OF MALWARE ON SELECTED SUSPECTED COPYRIGHT INFRINGING WEBSITES

September 2018