EUIPO
EUROPEAN UNION
INTELLECTUAL PROPERTY OFFICE

www.euipo.europa.eu

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

## EXECUTIVE SUMMARY


September 2018

# 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.

# 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.

# 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 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

[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.

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.

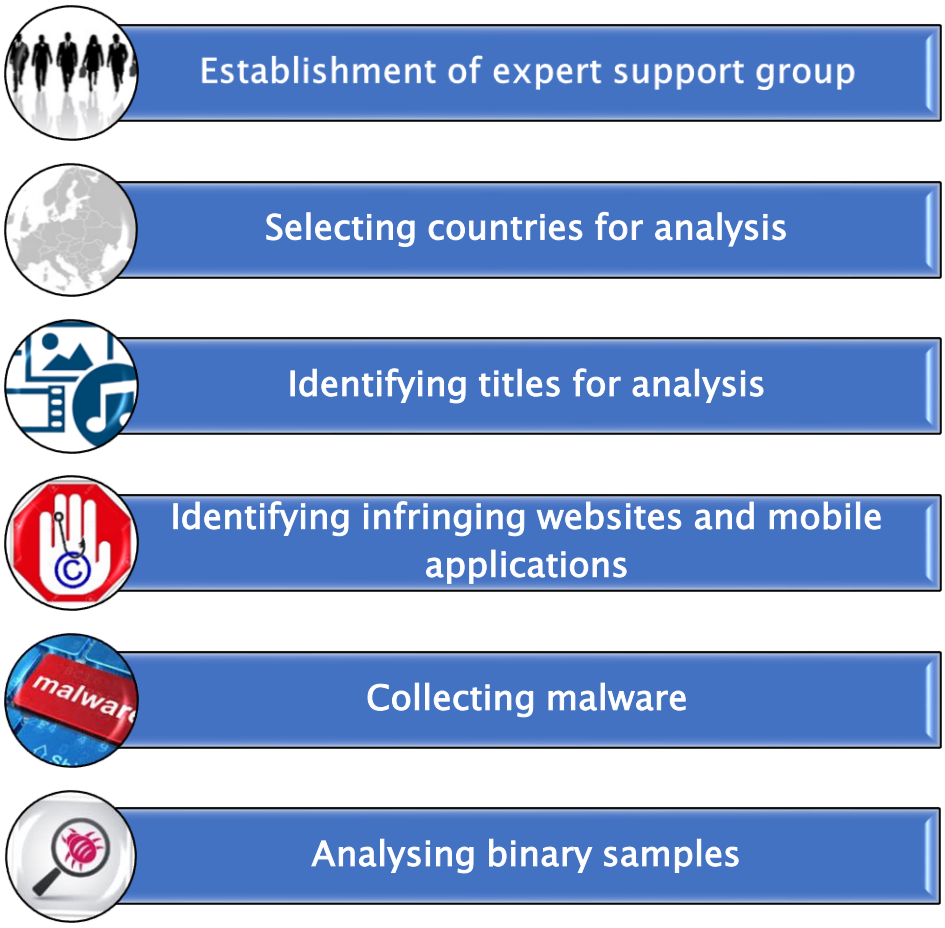
5. Phase V consisted of collection of malware and PUPs in addition to mobile applications on the identified websites, to be examined at a later stage for proper categorisation. The data acquisition phase included two rounds of malware collection and analysis performed during the summer of 2017. The first 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.

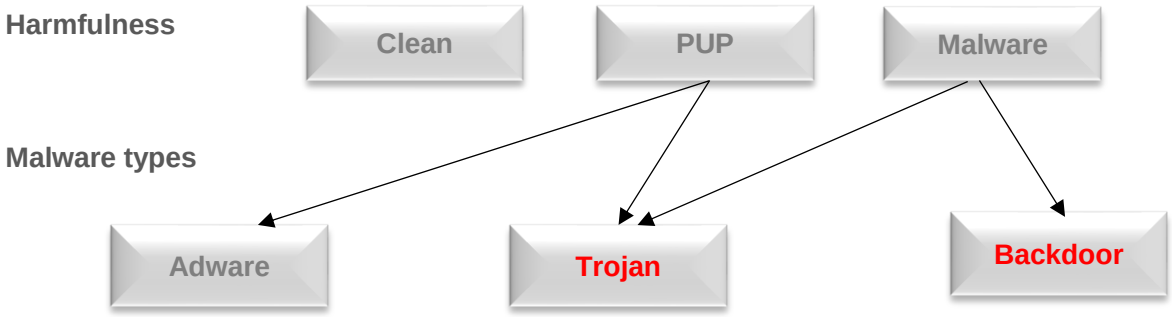
| | **Round 1** | **Round 2** |
|---|---|---|
| **Date** | 28 July 2017 | 10 August 2017 |
| **Discovered websites across 10 EU countries** | 5 240 | 5 606 |
| **Unique websites** | 1 054 | 1 057 |
| **Relevant files** | 617 | 3 665[2] |
| **Size of relevant files, GB** | 47 | 167 |
| **Delivered to 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. 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.

[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.




**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.


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

## EXECUTIVE SUMMARY

September 2018