An Analysis of Pay-per-Install Economics Using Entity Graphs

Platon Kotzias
IMDEA Software Institute &
Universidad Politécnica de
Madrid, Spain
platon.kotzias@imdea.org

Juan Caballero
IMDEA Software Institute
Madrid, Spain
juan.caballero@imdea.org

ABSTRACT

Potentially unwanted programs (PUP) are a category of undesirable software which includes adware and rogueware. PUP is often distributed through commercial pay-per-install (PPI) services. In this work we perform what we believe is the first analysis of the economics of commercial PPI services. To enable the economic analysis, we propose a novel attribution approach using *entity graphs* that capture the network of companies and persons behind a PUP *operation*, e.g., a commercial PPI service or a set of PUP.

We analyze 3 Spain-based operations. Each operation runs a commercial PPI service, develops PUP, and manages download portals. For each operation, we collect financial statements submitted by the companies and audit reports when available. This data allows us to analyze not only the operation revenues, but also their profits (and losses), which can widely differ from revenues depending on operational costs.

Our analysis answers 6 main questions. (1) How profitable are the commercial PPI services and the operations running them? We measure that the three operations have a total revenue of 202.5M €, net income (i.e., profits) of 23M €, and EBITDA of 24.7M €. Overall, expenses are high and margins low. (2) What are the revenue sources of the operations? The largest source of revenue is the PPI service, which provides up to 90% of an operation's revenue. But, we also observe the operations to draw revenue from advertising, download portals, PUP, and streaming services. (3) How has the PPI business evolved over time? Peak revenue and net income happened in 2013 and there was a sharp decrease starting mid-2014 when different vendors deployed new defenses that significantly impacted the PPI market, which did not recover afterwards. (4) How many companies are involved in an operation? We find that each operation runs from 15 up to 32 companies, but most of them are shell companies. (5) How many persons are involved in an operation? We find that a small number of 1–6 persons manage each operation. (6) How long have the operations been active? Operations start as early as 2003, but the PPI services do not operate until 2010–2011.

1. INTRODUCTION

Potentially unwanted programs (PUP) are a category of undesirable software that includes adware performing ad-injection, ad-replacement, pop-ups, and pop-unders, as well as rogue software (i.e., rogueware) that pushes users through scary warnings to buy licences of the rogueware, despite its limited functionality. PUP's undesirable behaviors prompt user complaints and have led security vendors to flag PUP in ways similar to malware. PUP prominence has quickly increased over the last years. Thomas et al. [54] showed that ad-injectors, a popular type of PUP that injects advertisements into user's Web surfing, affect 5% of unique daily IP addresses accessing Google. They also measured that Google's Safe Browsing generates 60 million warnings related to PUP - three times that of malware [55]. And, Kotzias et al. measured that over 54% of the 3.9 M real hosts they examined had PUP installed [39] and that PUP dominates so-called malware feeds [14] .

PUP is often distributed through commercial pay-per-install (PPI) services [39, 55]. A commercial PPI service acts as an intermediary between advertisers that want to distribute their programs and affiliates that own programs (typically freeware) that users want to install. To monetize installations of its free program, an affiliate bundles the free program with a downloader from a PPI service, which it distributes to users looking for the free program. Affiliates are paid by the PPI service $2.00–$0.01 per installation, depending on the geographic location of the user. During the installation process of the free program, users are prompted with offers to also install programs from the PPI advertisers. Advertisers pay the PPI service for successful installs of their advertised programs. Commercial PPI services are often used (or abused) by PUP publishers to advertise their programs and play an important role in PUP distribution [39,55]. Undesirable programs from advertisers, commercial PPI downloaders, and affiliate programs bundled with the PPI downloader are all typically flagged as PUP by AV vendors. PPI services also exist for distributing malware [25], but we call those *underground PPI services* to differentiate them from the commercial PPI services that PUP uses (which for short we simply call PPI services in this paper). Prices paid to affiliates by underground PPI services range $0.18–$0.01 per install [25], showing that malware distribution can be an order of magnitude cheaper than PUP distribution for the most demanded countries. Prior work has shown that both types of PPI services are largely disjoint [39, 55].

This work has two goals. Our main goal is to *measure the economics* of the commercial PPI services used to distribute PUP, e.g., how profitable they are. Understanding their economic evolution over time is an essential step for evaluating the effect of deployed defenses [58]. To measure their economics, we first need to perform *attribution*, i.e., identify the entities behind them.

A fundamental difference between PUP and malware is that PUP is often published by companies, and companies also run the commercial PPI services used to distribute PUP. In contrast, malware publishers are cybercriminals with hidden identities and use underground PPI services also run by cybercriminals. Since companies are behind PUP and the PPI services PUP uses, attribution is potentially easier (compared to malware) because those companies may be required to publish information about them, their business activities, and the people that manage them. For example, legislation may require companies to register in a national company register, to submit financial reports, and in some cases to be the subject of external audits. Although PUP attribution is arguably easier than malware attribution, it is still challenging because behind PUP and commercial PPI services there are often networks of companies [14] and those companies are created, dissolved, and renamed over time. It is also challenging because company information widely varies among countries, comes from different sources, is often incomplete, and only available as text documents (e.g., PDF).

In this work we call *operation* the network of persons and companies that operate a commercial PPI service. To perform attribution of an operation we propose *entity graphs*. Nodes in an entity graph represent companies and persons. An edge from a person to a company indicates that the person is part of the company's management. Company nodes are annotated with corporate information such as creation date, address, country where registered, fiscal identification number, list of names used over time, and type of economic activity. An entity graph enables structured attribution by tracking the business relationships among persons and companies in an operation. Our approach to build an entity graph takes as input an initial list of companies, possibly only one, known to belong to an operation. It uses company registers for obtaining company information, identifying the persons managing the company, and finding other companies also managed by those persons. This approach discovers new companies in the operation, not present in the input set of companies, thus expanding the operation's coverage.

Once we have an entity graph for an operation, we obtain financial and audit reports for the identified companies, and use them to analyze the operation's economics. We analyze the revenue, net income (i.e., profits and losses), and EBITDA. We also examine the number of employees and, when available, expenses and revenue split by source. We focus on the 2013–2015 time period. As far as we know, this is the first work that looks at the economics of PUP operations, and in particular of the commercial PPI services used to distribute PUP. Prior work has analyzed the economics of diverse malicious activities, e.g., [42, 47, 52, 57]. A key difference is that those papers analyze revenue data obtained from data leaks or estimated through external measurements. However, high revenues do not necessarily mean high profits, since the operational expenses can also be high. In contrast, by using company financial statements, we have not only revenue data, but also profits and losses, and in some cases expenses split by category. Thus, we can truly analyze how profitable the operations running commercial PPI services are.

While our approach is generic, in practice it is challenging to obtain the data for building entity graphs for the reasons exposed above. Therefore, in this work we focus on operations based in Spain because for this country we are able to collect, in a semi-automated fashion, the company information for building the entity graphs and analyzing the operations' economics. In particular, we analyze three Spain-based operations. All three operations run a PPI service for Windows programs, but are also involved in other parts of the PUP ecosystem such as publishing their own PUP (e.g., system cleaning utilities) and managing freeware download portals.

Our analysis addresses the following 6 main questions:

1. *How profitable are commercial PPI services and the operations behind them?* We measure that the three operations have a total revenue of 202.5M €, net income of 23M €, and EBITDA of 24.7M €. The most profitable operation has revenue of 92.2M € and net income of 11M € obtained in 2013–2015. Most of the revenue of each operation comes from a small subset of companies. There is a large gap between revenue and net income in all operations, indicating large expenses and low margins.

2. *What are the revenue sources?* The largest source of revenue for all three operations is the PPI service, which provides up to 90% of an operation's revenue. But, we also observe the operations to draw revenue from other sources such as advertising, download portals, PUP products they develop, and video streaming services.

3. *How has the PPI business evolved?* Peak revenue and net income happened in 2013. We observe a sharp decrease on both revenue and income for all three operations starting mid-2014, leading to all three operations to have losses in 2015. We conclude that improved PUP defenses deployed by different vendors in mid-2014 [18, 19, 23] significantly impacted the PPI market, which did not recover afterwards.

4. *How many companies are involved in an operation?* We find that each operation runs from 15 up to 32 companies, but most of them are *shell companies* that have no employees, no revenue, share address with other companies, are often created in batches, and have no website. We observe those shell companies being used to obtain code signing certificates from certification authorities, later used to sign the distributed executables. While all three operations are based in Spain, two of them also use 2–5 companies registered abroad, namely in Israel and the state of Delaware in the US, a known tax haven [48].

5. *How many persons run an operation?* We find that a small number of 1–6 persons manages the large number of companies in each operation. One of the operations is run by a single person that manages 21 companies.

6. *How long have they been in operation?* The lifetime of each operation is 7–13 years, with companies created as early as 2003. However, the companies that run the PPI service in each operation were created in 2010–2011. Prior to 2010, the revenue came from other activities such as PUP licenses and download portals.

Our contributions are the following:

- We perform the first economic analysis of PUP operations and specifically of commercial PPI services used to distribute PUP. We acquire financial and audit reports for the companies involved and use them to analyze the revenue, net income, and EBITDA. When available, we also analyze expenses and sources of revenue.

- We propose a novel approach to perform PUP attribution using entity graphs. Nodes in an entity graph are companies or persons and edge from a person to a company indicates the person holds a management position in the company.

- We generate the entity graphs for three Spain-based operations, each running a commercial PPI service and involved in other related activities. The entity graphs comprise of 15–32 companies and 1–6 persons.

2. OVERVIEW

In this section we describe privacy and legal considerations (Section 2.1), introduce the operations analyzed (Section 2.2), define the entity graph (Section 2.3), and present the input company lists (Section 2.4).

Operation	PPI	DP	PUP
OP1	✓	16	1
OP2	✓	1	4
OP3	✓	2	1

Table 1: Whether each operation runs a PPI service, download portals, and PUP software.

2.1 Privacy & Legal Considerations

Our main goal is to analyze the economics of operations running commercial PPI services. For this, we build entity graphs for three Spain-based PUP operations. At no point we aim to point the finger to these particular operations or the people behind them. They have been chosen simply because they are Spain-based and thus we can obtain the needed data for the analysis. Any other operation could have been analyzed if their country of origin makes available the needed data.

Our ethics advisory board has mandated that we anonymize the operations to prevent putting the spotlight on the people running these three operations, and to avoid time-consuming legal actions. Specifically, we anonymize the names of the operations, as well as the names of the companies and persons involved in each operation. For the rest of the paper we refer to the three operations as: OP1, OP2, and OP3. And, we refer to specific companies using the operation and a company identifier, e.g., OP1.C02.

We strive to achieve a balance between the privacy of the persons behind the operations and the value of the information provided. Our anonymization is best-effort since all the data analyzed is public and accessible freely or by paying small fees. Thus, the raw sources could be analyzed independently of our results. We understand that providing information about the operations' rankings (Section 2.2) reduces the anonymity set for the operations, but we believe that it is not much additional information given the availability of the raw data. Furthermore, we believe that the rankings are fundamental for readers to understand how representative the three operations are and thus the extent of our results.

We note that the anonymization process does not affect our analysis since it is performed a posteriori. We also believe that it does not significantly impact the presentation of our results, while helping protect the privacy of the persons behind the operations.

We also note that providing a definition of what behaviors make a program PUP (or malware) exceeds the scope of this paper. Instead, to determine if a sample is PUP, malware, or benign we use a previously proposed approach that examines PUP-related keywords that appear in the labels output by AV engines during scanning of suspicious samples [14]. In a nutshell, we rely on AV vendors to identify PUP samples, and use the digital signatures in those samples (when available) to identify the companies in charge of the PUP.

2.2 PUP Operations Analyzed

All three operations run a commercial PPI service during our analysis period. The three PPI services have been ranked by prior work among the top 15 commercial PPI services by user installation base, and have been estimated to affect a few millions of users in total [39]. Other prior work ranks two of these operations among the Top 10 PUP operations by number of signed samples and the other in the Top 30 [14]. Thus, while we do not know exactly what fraction of the commercial PPI market the three operations represent, we do know that they play a significant role, i.e., they run some of the largest commercial PPI services and affect a large number of users. Thus, we believe that the insights gained on the commercial PPI market from these operations are representative of the ecosystem.

Attributes		Objects		Datasets		
Attribute	**Type**	**Object**	**Type**	**BE**	**HP**	**IF**
Person name	str	Node	Person	✓	✗	✗
Fiscal ID	str	Node	Comp.	✓	✗	✓
Company names	str list	Node	Comp.	✓	✗	✗
Company type	str	Node	Comp.	✓	✗	✗
Economic activity	str	Node	Comp.	✓	✗	✓
Employees	int	Node	Comp.	✗	✗	✓
Telephone number	str	Node	Comp.	✓	✗	✓
Address	str	Node	Comp.	✓	✗	✓
City	str	Node	Comp.	✓	✗	✓
Country	str	Node	Comp.	✓	✗	✓
Creation date	date	Node	Comp.	✓	✗	✓
Dissolution date	date	Node	Comp.	✓	✗	✓
Last modification date	date	Node	Comp.	✓	✗	✗
Capital	float	Node	Comp.	✓	✗	✓
Earnings	float	Node	Comp.	✗	✗	✓
Revenue	float	Node	Comp.	✗	✗	✓
EBITDA	float	Node	Comp.	✗	✗	✓
Certificates	str list	Node	Comp.	✗	✓	✗
Revoked certificates	str list	Node	Comp.	✗	✓	✗
Active roles	str list	Edge	Roles	✓	✗	✗
Inactive roles	str list	Edge	Roles	✓	✗	✗

Table 2: Attributes used in the entity graph, the objects holding the attribute, and the datasets used to obtain their information. The datasets are described in Section 3 and correspond to BORME (BE), HerdProtect (HP), and Infocif (IF).

In addition to running a PPI service, the operations have been involved in other related activities. Specifically, all operations have developed at least one PUP product and have managed at least one download portal to assist in the distribution of their PPI downloaders and PUP products.

Table 1 summarizes the number of download portals and PUP products we have identified. These numbers are only a lower bound since we may have missed other software products and download portals. OP1 develops a download manager used to offer advertiser programs to users that install it. They also manage a large number of download portals that offer freeware bundled with their PPI downloader. Of their 16 download portals, 9 are blocked by SafeBrowsing as unsafe. OP2 develops several rogueware, namely system cleaning utilities and media players, and operated until 2015 a download portal. The audit report for OP3.C18, the company that runs the PPI service in OP3, has a nice description of how the company operates. Translated to English, it states: "The company obtains its revenue predominantly from offering to users visiting their download portals third-party applications from which they receive a payment for each installation or a share of the revenues the application generates. In the first case, the revenue is accounted for when the application is installed by the final user; in the second case, the revenue is accounted for as it is confirmed by our clients". Thus, advertisers can opt for a pay-per-install or a revenue sharing model. The download portals are used by the PPI service to attract users looking for freeware to have them install the PPI downloader.

2.3 Entity Graph

The entity graph is an undirected graph where a node is either a person or a company. An edge from a person to a company means that the person holds, or held in the past, a directive position in the company. A person may have (or have had) multiple positions in a company (e.g., administrator and treasurer). Companies are uniquely identified by their fiscal identification number because they may change names over time. The left part of Table 2 summarizes the node and edge attributes of the entity graph. For each attribute it shows the attribute type (i.e., string, integer, float, list, boolean), the type of object where stored (i.e., node or edge), and the node type (i.e., person, company, roles). We detail node and edge attributes below.

Node attributes. Persons have only one attribute, the name of the person. Companies have generic, economic, and code signing certificate attributes. Generic attributes include the list of company names, company type (e.g., limited liability), economic activity, number of employees, fiscal identification number, telephone number, address, city, country, creation date, dissolution date (if any), and the date of the last modification of the company data. Economic information attributes include parts of the annual balance for all available years. These attributes are initial capital, revenue, net income, and EBITDA (Earnings Before Interest, Taxes, Depreciation, and Amortization). At last, there are two code signing certificate attributes: the list of known code signing certificates issued to the company and the list of those certificates that have been revoked. If the company has not been used for obtaining a code signing certificate both attributes are empty.

Edge attributes. Edges have two attributes: the list of active roles (if any) that the person currently has in the company, and the list of past roles (if any) that the person had in the company.

2.4 Input Company List

Our approach takes as input an initial list of companies that are part of a PUP operation. This initial list can be obtained from different sources such as the contact information and privacy policies of PUP websites, Whois registration data for PUP domains, or the digital signatures of PUP samples. In this work, we obtain the list of initial companies from prior work that clustered PUP samples into operations using information from their digital signatures [14].

The intuition of using the digital signatures from PUP samples to identify companies is that PUP publishers are constantly looking for ways to make their programs look benign in order to convince the user to install them and to avoid detection. One such way is code signing, where the software is distributed with a digital signature which, if valid, certifies the integrity of the software and the identity of the publisher. Signed programs look more benign and may be assigned higher reputation by security products. In Windows, properly signed programs avoid scary warnings when a user executes them and are assigned higher reputation when downloaded through Internet Explorer [45]. Furthermore, kernel-mode code is required to be signed. To sign Windows programs, publishers need to obtain a valid *code signing certificate* from a Certification Authority (CA), which requires providing the publisher's identity to the CA and paying a fee ($500–$60 for 1-year certificates). While not all PUP samples are signed [54], prior work has shown that properly signed samples detected by AV engines are predominantly PUP [14], as identity validation by CAs poses an important barrier for malware.

The input lists obtained from Malsign [14] comprises of 15 companies for OP1, 9 companies for OP2, and 29 companies for OP3. Despite the large number of input companies, our approach still discovers 15 previously unknown companies, as well as the people managing the companies. Furthermore, we have also evaluated our approach by using an input list with a single company for each operation. With this reduced input list, the produced entity graphs still contain all companies registered in Spain present in the entity graphs obtained with the larger input lists.

3. DATASETS

We leverage a variety of datasets for this work. We use company registers for obtaining company information and for establishing the persons managing a company; audit and business reports for obtaining company financial data; a dataset of signed PUP executables for identifying the initial list of companies in each operation; the website of a security company for determining which companies have been used for obtaining code signing certificates; a malware repository to measure the prevalence of samples from each operation over time; and certificate transparency logs to identify websites belonging to the operations.

Company registers. Company registers collect information about companies in the jurisdiction they operate under. Each country has its own norms regarding the existence of such register, whether the register is centralized or distributed (e.g., to its regions), what type of information it collects on companies, and how publicly available the data is. Countries may provide public access to some of the data collected by their company registers, for example Germany [8], Israel [12], Spain [3], and United States [6].

In Spain, there exist 52 regional registers and a central register called *Registro Mercantil Central* [16]. The regional registers collect the information on the companies in their region. The central register is in charge of providing access to the information collected by the regional registers since January 1st, 1990. The central register has a publication, called *Boletín Oficial del Registro Mercantil* (BORME), which provides free public access to much of the collected company information. Every day, BORME publishes a PDF document with all changes that occurred in the regional registers. Among others, BORME reports the following company events: creation, dissolution, changes in the type of economic activity identified by a CNAE[1] code (e.g., 6312 - Web portals), changes in administrators, capital increases and reductions, balances, company name changes, corporate split-ups, and adsorptions.

Not all company data collected by the regional registers appears in BORME. In particular, Spanish law requires all companies to submit an annual financial report to their regional register. These financial reports are not included in BORME, but can be acquired from the central or regional registers for a fee. The financial report contains a balance sheet, an income statement, a statement of changes in equity, a profit and loss statement, and a statement of cash flows. Depending on specific criteria such as company size and net turnover, a company may be required to submit only an abbreviated version of the above documents [36].

To access the information from BORME, we leverage LibreBORME [13], an open source project that provides a public API to query BORME data published since 2009. LibreBORME parses the PDF files published every day and stores their data in a central database. It provides a clean interface, but only to a subset of the data from BORME. By querying LibreBORME with a company name, we can obtain: the fiscal identification number (NIF) that uniquely identifies a company in Spain, its creation date, the last modification date of the company's data, and the names of the persons with management positions (e.g., administrator, secretary, liquidator). By querying LibreBORME with a person name, we can obtain the list of companies in which the person holds management positions.

Audit reports. Countries may require companies satisfying certain criteria (e.g., income or number of employees above some threshold) to have periodic audits of their financial reports by external certified professional accountant (CPA) firms, i.e., auditors. The auditors examine and validate the financial reports of the company and perform a detailed analysis of its business activities, as well as a comparison with the previous year. Audit reports often contain details not included in other financial reports. For example, they may detail the sources of income for a company (e.g., company

[1] *Clasificación Nacional de Actividades Económicas*

products), the type of expenses, the reasons behind big changes on revenue or net income, and an analysis of the business risks.

In Spain, companies need to perform an audit if they fulfill two of the three following requirements for two consecutive years: (a) revenue over 5.7M €, (b) total assets over 2.8M €, (c) an average number of employees higher than 50 [17]. Audit reports can be acquired from business portals (see below). We use the Infocif [11] Web portal to obtain 5 audit reports for the three companies that fulfill the above requirements: OP1.C02, OP2.C08, OP3.C18. Those three companies correspond to the largest company for each operation. Each audit report covers two years, the year being audited and the previous year for comparison. If a company is part of a corporate group, the audit report also contains financial information for the other group companies.

For OP1.C02 and OP2.C08, we obtain the audit reports for 2014 and 2015, which cover 2013–15. For OP3.C18, only the 2014 audit report is available, which covers 2013–14. Each audit report costs 10 € and is performed by one of the following CPA firms: PricewaterhouseCoopers [15], Deloitte [5], Audalia Laes Nexia [2], or AFP Audit & Consulting [1]. The audit reports have varying degree of detail. example, the PricewaterhouseCoopers and Deloitte reports contain a revenue split by income source, but the ones from Audalia Laes Nexia and AFP Audit & Consulting do not.

Business reports. Many online services offer company reports comprising of corporate and financial data. Such reports are typically compiled by aggregating data from multiple sources including financial statements from company registers, public audit reports, financial reports published by the company, mandatory statements from listed companies, and Internet data such as news clips. These reports can be acquired by paying a fee. They are typically not as detailed as audit reports and their information can widely vary among services. One advantage is that their aggregation may cover longer time periods (e.g., up to 3 years), which is useful since we focus on the 2013–2015 period.

We use the Infocif service [11] to acquire reports for all companies in our entity graphs. Infocif reports contain financial data such as revenue, net income, EBITDA, number of employees, and an abbreviated version of the profit and loss statement. They also contain corporate information such as company address, phone number, CNAE code, and the company's website. Corporate information, except the company's website, comes from BORME, but is not accessible through LibreBORME.

Overall, we were able to acquire reports from Infocif for 59, out of 68 companies, with a total cost of 215 €. For the remaining 9 companies we could not obtain reports because 8 are registered outside of Spain and the other one is created in 2016, and thus had not yet filed their first financial statement when we ordered the reports. From the 59 acquired reports, 5 are empty. An empty report indicates that a company has not submitted their financial report to the company register, or that the report is pending approval (or digitalization) by the company register.

Malsign. The Malsign dataset consists of 142 K signed PUP (and a few malware) samples, as well as their clustering into operations/families [14]. The clustering results are based on statically extracted features from the samples with a focus on features from the Windows Authenticode signature [43]. These features include: the leaf certificate hash, leaf certificate fields (i.e., public key, subject common name and location), the executable's hash in the signature (i.e., Authentihash), file metadata (i.e., publisher, description, internal name, original name, product name, copyright, and trademarks), and the PEhash [59]. From the Malsign clustering results, we select the three operations based in Spain. For each of these three operations, we use the clustering results to extract an initial list of companies, which comes from the subject common names (CN) in the certificates of samples in the cluster, after being normalized to remove duplicates. Overall, Malsign contains 15 companies for OP1, 9 for OP2, and 29 for OP3.

HerdProtect. HerdProtect [9] is a security company that provides a host-based defense against PUP and malware. Their website provides detailed threat information including certificate information from PUP samples they observe in their users' hosts. For each company in an entity graph, we leverage HerdProtect's website to query for all code signing certificates they have observed issued to the company. For each certificate found, we collect the Subject, the Issuer (i.e., certification authority), the validity period, and the certificate's serial number. The data from HerdProtect enables us to identify which companies have been used to obtain code signing certificates from CAs. While we could use Malsign for this step, HerdProtect's coverage is larger, containing many certificates not included in Malsign.

VirusShare. We collect 27.7 M hashes of malware and PUP executables from the VirusShare repository [21]. We query those hashes to VirusTotal [22] (VT) to get their detection labels by multiple AV engines, as well as the timestamp when they were first submitted to VT. We use the AV labels as input to AVClass [50], a malware labeling tool that outputs for each sample the most likely family name and a confidence factor based on the agreement across engines. We use the AVClass results to identify samples that belong to the three operations and the VT first seen timestamp to measure the fraction of samples of each operation in VirusShare over time.

Certificate transparency logs. We analyze 38.3 M HTTPS certificates from Google's Certificate Transparency logs [4] to check if companies in entity graphs have a website.

The right part of Table 2 summarizes which dataset is used to extract each attribute in the entity graphs.

4. BUILDING ENTITY GRAPHS

This section describes how an entity graph is built given as input an initial list of companies known to belong to a PUP operation. This process may identify additional companies, not present in the initial company list, which are also part of the operation. In addition, it identifies the persons managing the companies in the operation. Building an entity graph comprises of four steps: building an initial graph, collecting certificate data, trimming the initial graph, and acquiring financial data. The first three steps we have been able to automate, so that they are reusable for other Spanish PUP operations. Acquiring the financial data and incorporating it into the entity graph is a manual process.

Building an initial graph. To build the initial graph we leverage LibreBORME to obtain information about companies and the persons managing them. For each company in the input list, our approach first queries LibreBORME to obtain its data (fiscal identification number, creation date, last modification date) as well as the names and positions of the people managing the company. If the company is found, a node is added for it in the initial graph. In addition, one node for each person managing the company is also added, as well as edges from each person node to the company node. If the company is not found, for example because it is registered in a country other than Spain, a node is still added for the company, but no person nodes or edges are added. Thus, foreign companies introduce disconnected components into the graph. Then, for each person found in the previous step, LibreBORME is queried again to obtain any other companies where the person has managing positions. For each additional company identified,

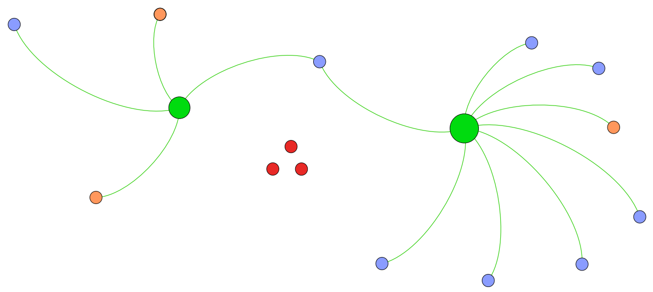


Figure 1: Mock example of an entity graph.

not yet present in the initial graph, the process recurses and the two steps above are repeated to obtain the new company's data and possibly identify new persons managing the new company. The process recurses until no new companies or persons are found.

During the recursion, if a person is found to hold managing positions in an unusually large number of companies (i.e., more than 100), the person is not added to the entity graph. This rule prevents lawyers to appear in the entity graph. Such lawyers are used by one operation to perform the initial registration of a company, after which the company is transferred to the real managers that are part of the operation. If we included the lawyers, we would also include the companies they register for other clients, which often number in the hundreds and are unrelated to the operation.

Collecting certificate data. For each company in the initial graph, our approach queries HerdProtect using the company name to obtain the list of code signing certificates issued to the company that HerdProtect has observed being used to sign PUP executables. For each certificate, we check the revocation status using the OCSP protocol and certificate revocation lists (CRLs). HerdProtect does not provide us with the raw certificate, but rather with its metadata. To query the revocation status we use the certificate's serial number and the CA that issued the certificate, both obtained from HerdProtect.

Trimming the initial graph. The initial graph goes through a trimming process to remove persons and companies unrelated to the PUP operation. The trimming applies two heuristic rules in sequence. The first rule aims at removing persons unrelated to the PUP operation, which did not satisfy the lawyer rule while building the initial graph. Specifically, this rule removes persons that only have positions in companies for which no certificates have been collected from HerdProtect. After removing those persons, any (Spanish) companies with no managing persons are also trimmed, since the reason they were initially added was the person no longer considered part of the operation. The second rule aims at trimming companies managed by people that are part of the operation, but that are used for purposes different than the PUP operation. This is important because persons involved in the PUP operation may also be involved in other unrelated activities such as real estate or finance. Specifically, this rule removes companies for which their business area is unrelated to information technology and which have not been used for obtaining code signing certificates.

Acquiring financial data. The trimmed graph corresponds to the entity graph for the PUP operation. For each company in the entity graph, we try to acquire a business report (and an audit report if applicable) from Infocif to obtain its financial data, as well as additional corporate information (e.g., address and phone number) not available through LibreBORME.

Visualization. Figure 1 shows a mock example of an entity graph visualized using Gephi 0.9.1 [20]. We differentiate nodes using colors. Green nodes are persons and we use three colors for company nodes: orange, purple, and red. Orange companies have no code signing certificate; purple companies have at least one certificate; and red companies are not registered in Spain.

5. PUP ENTITY GRAPHS

This section describes the entity graphs produced for the three operations. We first compare the three entity graphs and then analyze each operation in more detail in its own subsection.

Table 3 summarizes the entity graphs. The table is split in three parts. The leftmost part shows the number of company nodes, person nodes, and edges in the entity graph. The numbers in parentheses indicate how many companies are new in the entity graph, i.e., they were not present in the initial list of companies used as input to build the entity graph. The middle part contains company data: the number of distinct addresses for the companies, the number of companies that had at least one name change since their creation, and the number of countries that these companies are registered in (a value larger than one means non-Spanish companies appear in the entity graph). The rightmost part summarizes the code signing certificates: the number of distinct companies used for obtaining certificates, the number of certificates issued to those companies, the number of CAs that issued those certificates, and the number of revoked certificates. The numbers in parentheses indicate certificates issued to new companies not in the initial list.

The entity graphs show that OP3 is the largest operation with 32 companies and 6 persons managing them. All operations have a large number of companies, ranging from 15 for OP2 up to 32 for OP3. But only a handful of people manage those companies, from 1 in OP1 up to 6 persons in OP3. In the case of OP1, a single person is the sole manager for 21 companies.

The number of new companies (in brackets) show that our approach to build entity graphs enables discovering additional companies that were not present in the initial list of companies used as input. Specifically, we discover 15 previously unknown companies: 6 in OP1, 6 in OP2, and 3 in OP3. Thus, in addition of capturing the relationships between companies and their managers, the entity graphs amplify the coverage for all three operations. This amplification happens despite the initial list of companies in each operation, obtained from Malsign, being fairly large. We expect that for other operations the initial list of companies may come from less complete sources and be much smaller, perhaps even a single company. In fact, we have also tested building the entity graphs for all three operations starting with only the main company in the initial list. The produced entity graphs are identical to the ones in Table 3 except in that companies not registered in Spain are not identified since they do not appear in BORME.

For all three operations the number of distinct street addresses is much lower than the number of companies. This indicates that multiple companies share the same address. For example, OP1 uses 7 addresses for all 21 companies. This points towards some of the companies not having real activity. We confirm this in Section 6 by examining the number of employees and other financial data.

Not shown in Table 3 is that both OP1 and OP2 often create multiple companies on the same day. For example, three OP2 companies (OP2.C12, OP2.C13, OP2.C14) were created on the same day in April 2014. Such batch registrations are often performed by lawyers that later transfer the companies to the real managers.

Renaming companies is a common behavior. All operations have companies that have changed names since their creation. Renaming a company is cheaper than creating a new company. The new

	Graph			Companies			Certificates			
Operation	Companies	Persons	Edges	Addresses	Renames	Countries	Companies	Certs	CAs	Revoked
OP1	21 (6)	1	21	7	3	1	18	48 (22)	5	2
OP2	15 (6)	3	20	8	5	3	12	85 (51)	8	24
OP3	32 (3)	6	55	15	3	2	14	54 (26)	7	16

Table 3: Summary of entity graphs for the three operations.

company name can be used to obtain new code signing certificates, e.g., if a CA does not verify that the fiscal identity number of the requesting company matches a company with a certificate already issued. Even when a CA already issued a certificate for the company under a different name, it is logical that the managers may want to update their certificate after a company name change, making it difficult to deny the request. In OP2, five companies have been renamed and two of the new names have been used for obtaining new certificates. In each of the other two operations, three companies have been renamed and one new name has been used to obtain a new certificate.

It is also not uncommon for the operations to set up companies in multiple countries. OP2 and OP3 have companies registered outside of Spain, specifically in Israel and the United States. Interestingly, 6 companies (1 in OP2 and 5 in OP3) are registered in the US state of Delaware, a known tax haven [48].

Certificates. The number of code signing certificates issued to all three operations is four times larger than the number of companies across the three operations. This indicates that companies are used to obtain multiple code signing certificates (four on average). For OP2 this ratio goes up to 7 certificates per company. Certificates may be issued to the same company by different CAs. We also observe multiple certificates for the same company from the same CA using slight variations in the company name, e.g., *FakeComp SL* and *Fake Comp S.L.* All operations obtain certificates from multiple CAs (from 5 to 8). Focusing on a small number of CAs reduces the effort for obtaining the information required for the identity validation process. All three operations have certificates revoked, but the revocation ratio is quite low ranging from 4% for OP1 to 29% for OP3. The number of revoked certificates is especially low considering the lifetime of the operations. For example, OP1 had only 2 certificates revoked over 7 years.

Coverage. While we have shown that entity graphs amplify coverage, an important question is how much coverage do they achieve? We evaluate the coverage by comparing the companies in the entity graphs with the companies listed in the audit reports as members of a corporate group. The audit reports identify 21 OP1 companies, 11 OP2 companies, and 6 OP3 companies. Overall, the audit reports identify 39 companies, compared with 68 companies in our entity graphs. Thus, the entity graphs have significantly higher coverage. Of all the companies identified in the audit reports, only three OP3 companies are missing in the entity graphs. One of those three companies is registered in Ireland and it was not identified because there are no certificates in Malsign for that company. A manual check for the other two missed companies reveals that they are Spanish companies, but LibreBorme has not properly parsed the person that created the company. Thus, our approach did not identify the companies. On the other hand, there are 29 companies in the entity graphs that do not appear in the audit reports. One company in OP1 was created in 2016, i.e., after the audit reports were issued. The other 28 companies are not listed in the audit reports as being part of the corporate group, however the entity graphs reveals that they are connected to the operations. In fact, many of them have been issued certificates used to sign PUP samples of the operation. This indicates that solely relying on the information reported by the company in the audit reports is not enough to understand the full scope of the operation.

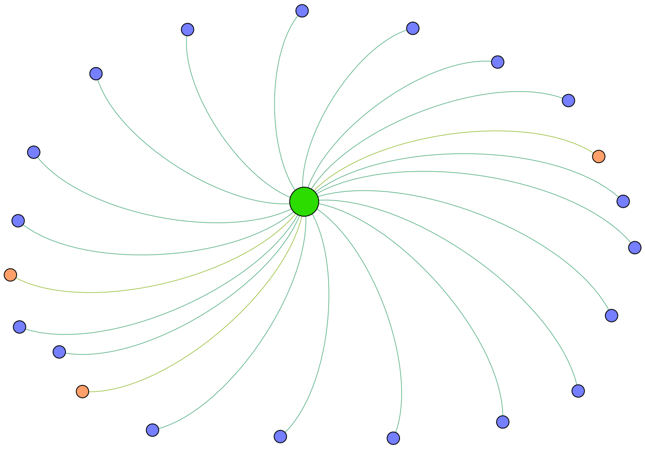


Figure 2: Anonymized OP1 entity graph. Green nodes represent persons, orange nodes companies without code signing certificates, and purple nodes companies with certificates.

Graph building rules. The building of the entity graphs excluded 5 persons (3 in OP1 and 2 in OP3) that manage an unusually large number of companies. They correspond to lawyers that, if included, would add hundreds of unrelated companies to the entity graphs. Additionally, the two trimming rules applied to the initial entity graph removed 50 companies and 12 persons across the three operations (4 companies and 2 persons in OP1, 2 companies and 3 persons in OP2, and 44 companies and 7 persons in OP3).

5.1 OP1 Analysis

Figure 2 shows the OP1 entity graph. It's the simplest entity graph among the three PUP operations with one person controlling the 21 companies in the operation. In contrast with the other two entity graphs, it does not contain any disconnected nodes indicating that all companies are registered in Spain.

Figure 3 presents a timeline of the 21 OP1 companies. The length of each line represents the lifetime of a company from creation to dissolution (or January 2017 if still active). A circle marks the issuing date of the first certificate for a company (if any). A star marks a date when a company was renamed (if any). The timeline shows that OP1 has existed for seven years, with the first company (*OP1.C00*) being created on March 2009. The company that runs the PPI service (OP1.C02, the one for which we have audit reports) was created in June 2010. For the first five years, at least one new company was created each year. In 2014, the rate of company creation increases significantly, with 14 companies created in the span of one year. These recent registrations often happen in batches with multiple companies being created simultaneously on the same date. The high rate of company registrations in 2014, and thus of certificates used, may indicate increased pressure by security vendors during that time period. Since January 2015, only one new company has been created. This may be due to the recently observed

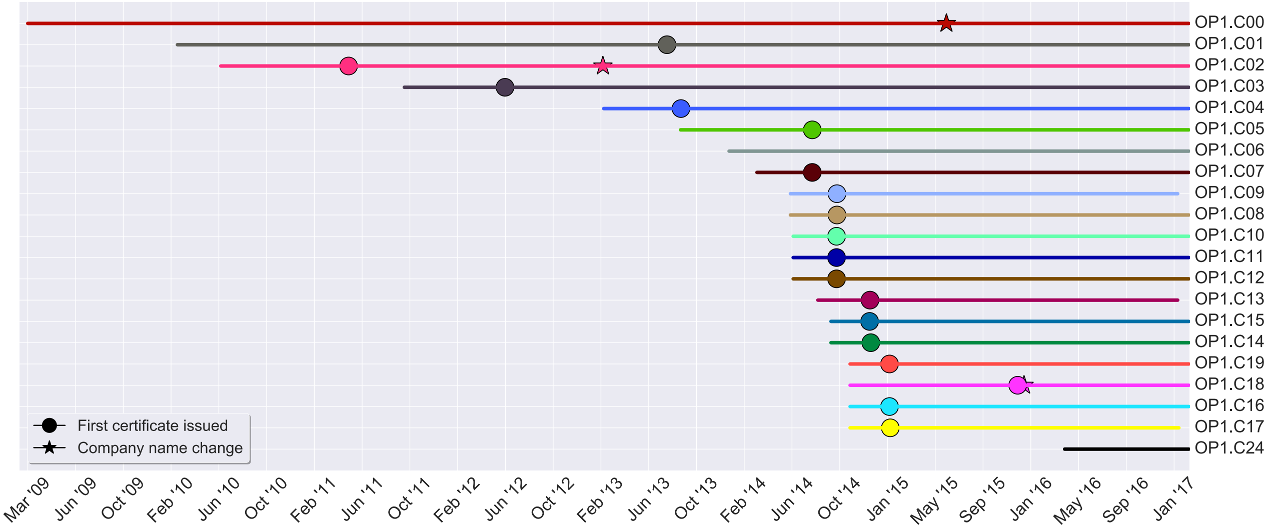


Figure 3: Each line represents the lifetime of an OP1 company. Circles mark the date of the first issued certificate of a company (if any) and stars mark the date of the company name change (if any).

shift of focus in OP1 towards other activities such as distribution of mobile applications.

OP1 follows a unique company registration pattern, not present in the other operations. Initially, an employee of a law firm creates the companies, but after a few months the company is transferred to the real manager that appears in the entity graph. When created, all companies mark the type of activity (i.e., CNAE code) as real estate. But, when ownership is transferred, the type of activity is modified to be development of web portals, which is one of OP1's activities. Similar to the company creation, changes of ownership occur in batches. For example, five companies (OP1.C08, OP1.C09, OP1.C10, OP1.C11, and OP1.C12) changed ownership on the same day in September 2014.

The first OP1 certificate was issued in 2011 (for OP1.C02). Of the 21 companies, 18 have been used to obtain certificates. We observe that certificates are often issued in batches using multiple CAs to request certificates on the same day. For example, on the same day in September 2014, 5 certificates were issued to 4 companies (OP1.C08, OP1.C09, OP1.C10, OP1.C12). We also observe that, especially since 2014, the code signing certificates are issued close to the registration date of the company. We measure how fast a company is used for obtaining a certificate by measuring the difference in days between the company creation date and the issuing date of the first certificate for the company. The median delay is 117 days with the shortest being 97 days (OP1.C15) and the longest 1,230 days (OP1.C01). Instead, if we measure the delay starting from the date the company is transferred to the real manager, the median time drops to 26.5 days, with the fastest being 14 days after the company transfer. This indicates that once the real manager is in charge of the company, within a month he obtains a code signing certificate for the company. This may indicate that obtaining such certificate is one of the main reasons for the company creation.

5.2 OP2 Analysis

Figure 4 shows the OP2 entity graph, which comprises of 3 persons and 15 companies. One company (OP2.C08) connects the 3 persons. Beyond OP2.C08, each person manages a quite independent set of companies, except for OP2.C17 which is managed by two persons. There are 3 disconnected nodes, which correspond to companies registered outside of Spain: two in Israel, and another in Delaware, US.

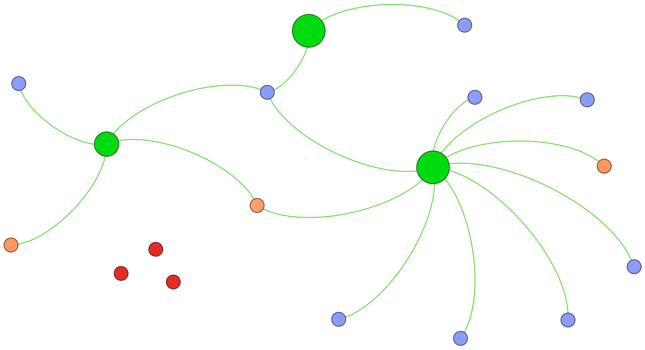


Figure 4: Anonymized OP2 entity graph. Nodes are colored similar to Figure 2.

Interestingly, all the companies in the entity graph appear in Malsign. However, in Malsign the companies were split among two different clusters. Even if we only used one of the Malsign clusters as input for the entity graph creation, our approach identifies that the companies in the other Malsign cluster (not used as input) also belong to the operation.

Figure 5 shows the company timeline for OP2. The operation has been active for 9 years, with the first company (OP2.C05) being created on December 2007. The company in the operation that runs the PPI service (OP2.C08, the one for which we have audit reports) is created in February 2011. Similar to OP1 company creation happens in batches, but in contrast with OP1, lawyers are not used by OP2. For example, on the same day in April 2014 three companies are created (OP2.C12, OP2.C13, OP2.C14). Interestingly, OP2.C09 and OP2.C10 are also created on the same day, although they are managed by different persons in the operation. We see a spike on company registrations between September 2013 and May 2014, which largely coincides with the spike in OP1. Similar to OP1, we do not observe any new companies created in the second half of 2014 and throughout 2015.

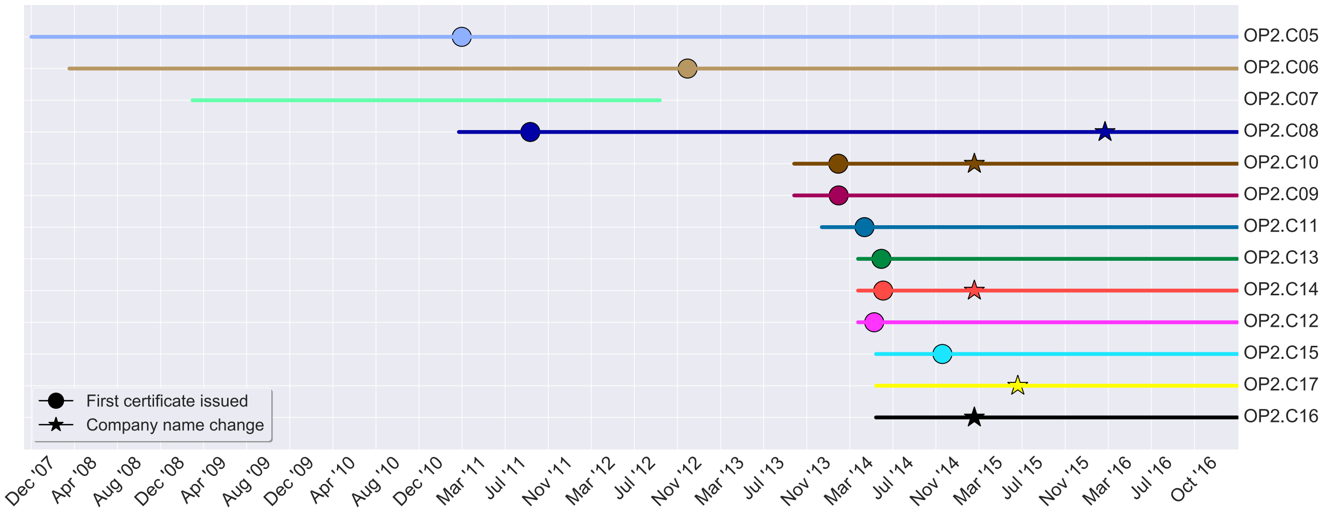


Figure 5: Each line represents the lifetime of an OP2 company. Circles mark the date of the first issued certificate of a company (if any) and stars mark the date of the company name change (if any).

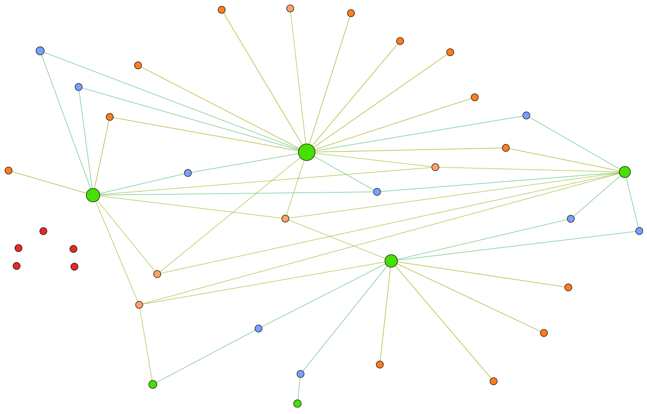


Figure 6: Anonymized OP3 entity graph. Nodes are colored similar to Figure 2.

Op.	Period	Employees	Revenue	Income	EBITDA
OP1	2012-15	$\leq$ 40 (3)	81.8 M	8.2 M	7.3 M
OP2	2013-15	$\leq$ 66 (4)	92.2 M	11.0 M	12.3 M
OP3	2008-14	$\leq$ 65 (8)	28.5 M	3.8 M	5.1 M
Total	**2008-15**	**$\leq$ 171 (15)**	**202.5 M**	**23.0 M**	**24.7 M**

Table 4: Summary of financial data for all operations. Revenue, net income, and EBITDA are provided in Euros.

Overall, OP2 has been issued 85 certificates, the largest number among the three operations. The first certificate for OP2 was issued in March 2011 (for OP2.C05) and it was revoked one day later by Comodo. The second certificate was issued in September 2011 for the main company (OP2.C08). Since then, 83 other certificates have been issued for this operation with a peak in 2014 with 46 issued certificates. This operation also requests certificates in batches, but each certificate in a batch is issued on a separate, but consecutive, day. The median time between the creation of a company and the issuing of the first certificate is 123 days, with the shortest being 45 days (OP2.C12) and the longest 4.7 years (OP2.C06).

Similar to the other operations, multiple OP2 companies are registered on the same address of the same city. Interestingly, three companies (OP2.C18, OP2.C05, OP2.C17) are registered by different people in different years, but all three on the same address.

5.3 OP3 Analysis

Figure 6 represents the OP3 entity graph. It is the largest and most complex entity graph with 32 companies and 6 persons. Of the 6 persons, 4 are connected to at least 4 companies and the other two are less central, being only connected to 1 or 2 companies. The company with the highest degree is OP3.C09, which connected 4 persons in the graph, but was dissolved in March 2014. The company that operates the PPI service (and for which we have an audit report) is OP3.C18, which connects 3 of the persons. The entity graph has five disconnected nodes for companies registered in Delaware, US.

The timeline in Figure 7 shows that OP3 has operated for 13 years, making it the longest-lived operation. The first OP3 company (OP3.C00) was created in July 2003 and was soon followed by three other companies (OP3.C01, OP3.C02, OP3.C03). Since then, 1–3 new companies were created every year. The company running the PPI service was created on June 2011. In 2012, 4 new companies are created and 8 new companies are added in 2013. The last company was created in May 2014. Compared to the other operations, OP3 companies are more spread over the years and the spike occurs earlier, in the second half of 2012 and the first half of 2013. We do not observe batch company registrations in OP3.

Overall, 54 certificates have been issued for OP3 companies. The first certificate was issued in 2004 for OP3.C04 and the highest number is 22 certificates issued during 2013. Each company is used for obtaining 3.8 certificates on average. The highest use is for OP3.C18, which has been issued 12 certificates. The median time between the creation of a company and the issuing of the first certificate is 236 days with the fastest being 13 days (OP3.C33) and the longest 9.5 years (OP3.C00). Thus, OP3 is the slowest among the three operations in using new companies to obtain certificates.

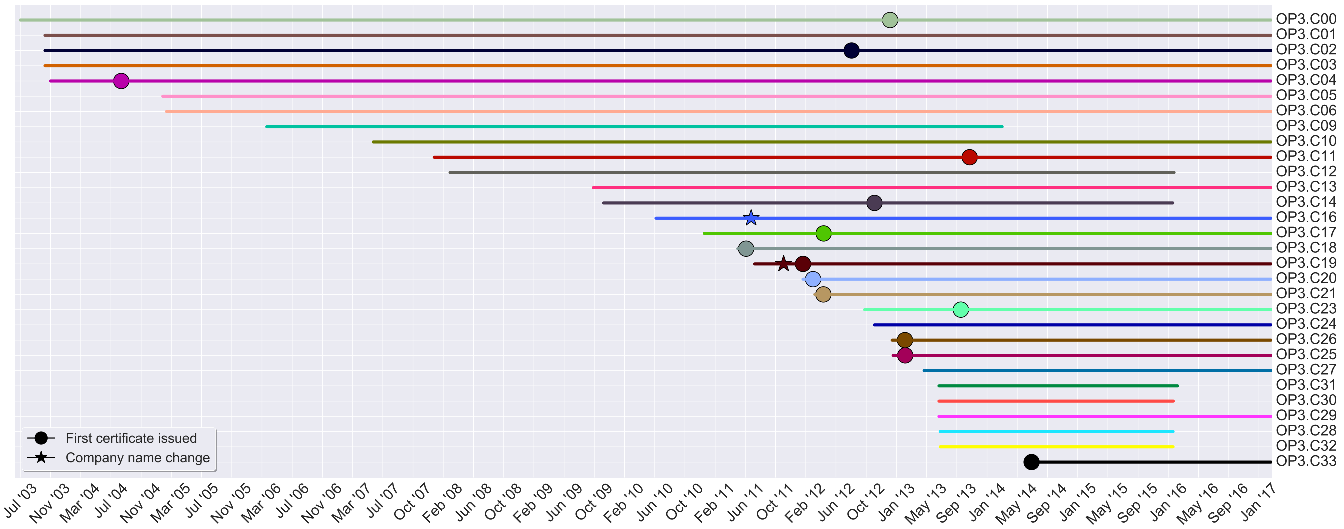


Figure 7: Each line represents the lifetime of an OP3 company. Circles mark the date of the first issued certificate of a company (if any) and stars mark the date of the company name change (if any).

6. PUP ECONOMICS

In this section we analyze the financial data obtained from the business and audit reports. We first provide a summary of the three operations and then detail each operation in its own subsection.

Table 4 summarizes the financial data. For each operation, it shows the period covered by the financial data, an upper bound on the number of employees across all companies in the operation (and the number of companies reporting at least one employee in brackets), and the total revenue, net income, and EBITDA across the whole period and for all companies in the operation. All currency values are in Euros.

The number of employees in Table 4 is the sum of the maximum number of employees reported by each company across the period. It is an upper bound because there could be overlaps between employees in different companies and because not all employees may have been contracted at the same time. The numbers in brackets show that the majority of the companies in all three operations have no employees. In each operation, there is one company that provides almost all employees and a few companies with a very low number of employees. For example, of the 21 OP1 companies only 3 have reported any employees, and of those two have reported only one employee, with the main company in the operation providing the other 38 employees.

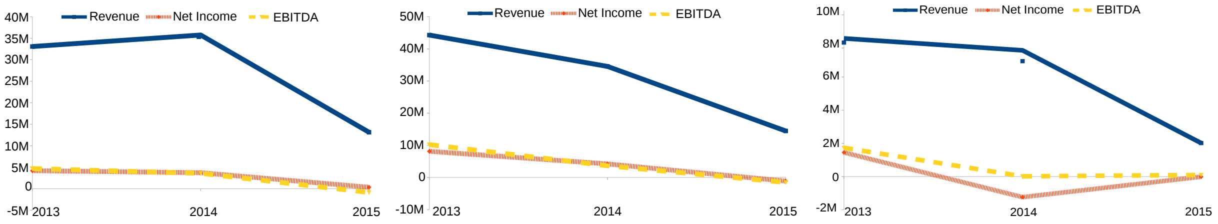
In total, the three operations have revenue of 202.5M €, net income of 23M €, and EBITDA of 24.7M €. These amounts are lower bounds since we do not have financial data for every year for each company. We provide the period covered for each individual company in Tables 6–10. OP2 is the most profitable operation with net income of 11M € and EBITDA of 12.3M €. OP1 ranks second with profits of 8.2M €and OP3 third with profits of 3.8M €. Thus, the number of companies in the operation does not directly influence its financial data as OP2 has the fewest companies, followed by OP1 and OP3. In each operation, there is a small subset of companies that brings the most revenue. For example, OP1 has 26 companies but five of them are responsible for 93% of the total revenue and 98% of the total net income. Similarly, 3 OP2 companies and 5 OP3 companies are responsible for 99% and 95% of the total revenue of those operations. We examine the individual companies of each operation in Sections 6.1–6.3.

Category	OP1.C02	OP2.C08	OP3.C18
Personnel	205K (<1%)	5.7M (6%)	2.7M (28%)
Advertising	-	64.9M (72%)	2.9M (30%)
Supplies	39.3M (75%)	-	-
Other	5.2M (10%)	8.0M (9%)	2.6M (27%)

Table 5: Expenses of the 3 audited companies for 2013–15. Percentages are calculated over the company's yearly revenue.

Expenses. The large difference between revenue and net income in all operations indicates large expenses. The expenses data comes from the financial reports submitted by the companies. While many categories exist, the declared expenses are typically under one of three categories: personnel, supplies, and a generic other costs. The personnel expenses for the whole operation are highest for OP3 reaching 17% (5M) of the total revenue of the operation, followed by OP1 with 9% (7.5M) and OP2 with 6% (6M). Unfortunately, the other two categories are too generic to understand the nature of the expenses. For the three companies required to have their financial statements audited by a CPA firm, i.e., the ones running the PPI services, the audit reports provide more detail into the generic other costs category. That generic category may include, among others, advertisement, office rentals, maintenance, insurance fees, banking fees, taxes, and provisions for losses. Table 5 summarizes the expenses declared in the audit reports by the three audited companies. Although the main business activity of all three companies is to operate a PPI service, the declared expenses differ significantly. OP1.C02 declares that 75% of the revenue is spent in (unspecified) supplies. However, some parts of the audit report label these supplies as external services provided to the company. For OP2.C08 and OP3.C18, the largest expenses are in advertising, which correspond to 72% of all revenue for OP2.C08 and 30% for OP3.C18. We suspect that the supplies expenses in OP1.C02 and the advertising expenses in OP2.C08 and OP3.C18 include the payments to the PPI affiliates. Overall, the data indicates commercial PPI services have high expenses and low margins.

Evolution over time. Figure 8 shows how the revenue, net income, and EBITDA of each operation has evolved in the period 2013–2015. The figures illustrate the large gap between revenue and net



(a) OP1 (b) OP2 (c) OP3

Figure 8: Economic data for the three PUP operations for the period 2013-15.

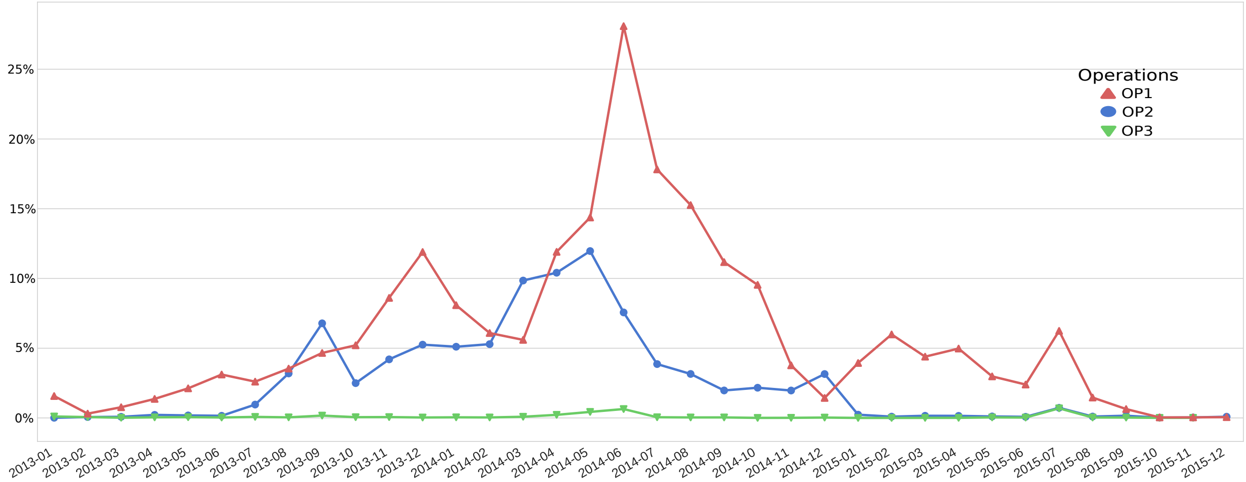


Figure 9: Percentage of samples in VirusShare that belong to each of the three operations for the period 2013–2015, over the total number of samples collected by VirusShare in that month. The two largest operations show growth until Summer 2014 where the number of samples sharply declines and does not later recover.

income (or EBITDA). In fact, OP1 and OP2 have losses in 2015, despite revenue of over 10M each. OP3 has losses in both 2014 and 2015, although the EBITDA is slightly positive in both years. Overall, the trend is that in 2014 revenue stabilized with respect to 2013 with OP1 showing a small increase, OP3 a small decrease, and OP2 a larger decrease. Then, in 2015 all 3 operations have a steep decrease in revenue.

We know that on June 2014 Symantec announced that their AV engines would start to flag PUP [18], that Microsoft enabled stricter PUP detection rules on July 1st 2014 [23], and that on August 2014 Google introduced policies against PUP in SafeBrowsing [19]. Since we only have yearly financial data it is possible that those events are responsible for the drop in revenue and profits, which does not clearly manifest in Figure 8 in 2014 because the PPI market was still growing during the first half of 2014.

We further investigate this assumption using the VirusShare repository. Figure 9 shows the fraction of samples in VirusShare that belong to each of the three operations over the period 2013–15 (using the AVClass family classification and VirusTotal first seen timestamp). For the two largest operations (i.e., OP1, OP2) we observe growth until May–June 2014, followed by a steep drop in June–August 2014. A drop on the PPI samples observed in the wild is an indication that fewer programs are distributed through PPI services, which in turn indicates less revenue for the PPI services.

While correlation in time does not necessarily mean causality, we also observe that the 2014 audit report of OP3.C18 identifies as a main business risk the changes in the policies of both Google and Microsoft, which it states can significantly affect the installations of their customers products. The audit also mentions that their R&D department plans to recover from the losses by developing better techniques that can address the installation demands of their clients.

Thus, we conclude that improved PUP defenses deployed by different vendors in mid-2014 significantly impacted the PPI market, which did not recover afterwards.

Web presence. We check if the companies have a website using three sources. First, the business reports may include the company's official website. Second, we query search engines using the company names. Third, we search for domains belonging to the companies in the certificate transparency logs. From the business reports and search engines we identify that 8 of the 68 companies have an official website. From the CT logs, we identify 31 HTTPS certificates for 6 companies containing 42 domains[2]. From these 42 domains, only four have a website, in each case describing a product rather than a company, e.g., a registry cleaner offered by OP2.C11. The results show that only a minority of the companies have a Web presence. This is surprising since according to the declared type of activity they provide Internet services like website development or online marketing. The lack of Web presence, in addition to the lack of employees and the minimal business activity, indicates that most of the companies are shell companies.

[2] Some certificates contain additional domains in the Subject Alternative Name extension.

Company	Creation	Diss.	Cert.	Web	Period	Emp.	Revenue €	Net Inc. €	EBITDA €
OP1.C00	03/09	-	✗	✓	2013-15	20-38	11.9 M	2.2 M	-923 K
OP1.C01	03/10	-	✓	✗	2013-15	1	8.6 M	285.5 K	399.3 K
OP1.C02	06/10	-	✓	✓	2012-15	0-1	52.6 M	6.0 M	7.7 M
OP1.C03	10/11	-	✓	✗	2014-15	0	2.1 M	237.0 K	371.1 K
OP1.C04	02/13	-	✓	✗	2014-15	0	2.0 M	96.7 K	193.6 K
OP1.C05	08/13	-	✓	✗	2014-15	0	3.1 M	106.4 K	220.2 K
OP1.C06	12/13	-	✗	✗	2014-15	0	33.7 K	5.2 K	7.2 K
OP1.C07	03/14	-	✓	✗	2014-15	0	384.1 K	-14.6 K	-6.2 K
OP1.C08	05/14	-	✓	✗	2014-15	0	648.4 K	-485.2 K	-643.9 K
OP1.C09	05/14	01/17	✓	✗	2014-15	0		-16.5 K	
OP1.C10	06/14	-	✓	✗	2014-15	0	277.5 K	-57	8.1 K
OP1.C11	06/14	-	✓	✗	2014-15	0		-2.7 K	
OP1.C12	06/14	-	✓	✗	2014-15	0	59.9 K	-8.5 K	-8.9 K
OP1.C13	08/14	01/17	✓	✗	2014-15	0	26.6 K	-1.5 K	-1.9 K
OP1.C14	09/14	-	✓	✗	2014-15	0	538	-1.3 K	-1.8 K
OP1.C15	09/14	-	✓	✗	2014-15	0	85.8 K	3.2 K	2.1 K
OP1.C16	10/14	-	✓	✗	2015	0	2.3 K	-1.5 K	-2.1 K
OP1.C17	10/14	01/17	✓	✗	2015	0		-4.3 K	
OP1.C18	10/14	-	✓	✓	2015	0		-2.5 K	
OP1.C19	10/14	-	✓	✗	2015	0		-185.6 K	
OP1.C24	04/16	-	✗	✗					
Total					**2012-15**	**0-38**	**81.8 M**	**8.2 M**	**7.3 M**

Table 6: OP1 financial data.

6.1 OP1 Economics

Table 6 summarizes the financial data obtained from the business reports of each OP1 company. The left part of the table contains general company data: the creation and dissolution (if any) dates, whether the company has been issued at least one code signing certificate, and whether the company has any website (corporate or product). The right part of the table shows financial data: the period covered in the business report, the number of employees reported, the revenue, the net income, and the EBITDA. From the 21 companies, 5 companies do not report any revenue, and one was created in 2016 and thus had not filed any financial report.

The lead company in the operation is OP1.C02 with 52.6M in revenue and 6M in net income. These correspond to 64% of the total revenue and 73% of the net income of the whole operation. In general, the companies created before 2014 show significant activity, while companies created in 2014–2015 have little business activity and mostly report losses. The three largest companies by revenue are the ones that report some employees, but the top company (OP1.C02) reports a single employee.

The content of OP1's download portals shows that different companies in the operation run the portals. The companies behind the download portals are the older ones, created before May 2014, and thus the ones that have largest activity.

Audit reports. We have acquired the 2014 and 2015 audit reports for OP1.C02, which were performed by Audalia Laes Nexia [2] and AFP Audit & Consulting [1] respectively. Unfortunately, these audit reports do not include as much information as the audit reports for the other two operations. In particular, they do not detail how much specific products and services are contributing to the bottom line. The audit report identifies 21 companies in the corporate group. OP1.C02 reports transactions of 1.9M with 11 of the companies in the corporate group, which are typically services provided to the auditee by those other companies.

6.2 OP2 Economics

Table 8 shows the financial data for the 12 OP2 companies registered in Spain, with the same structure as Table 6. Of those 12 companies, two report no revenue.

Category	2013	2014	2015	Total
PPI	39.4M (91%)	34M (99%)	10M (76%)	**83.4M**
Mobile Adv.	2.0M (5%)	34K (<1%)	-	**2.0M**
Down. Portal	1.4M (3%)	-	-	**1.4M**
Stream. Portal	-	-	3.1M (24%)	**3.1M**
Rogueware	-	4K (<1%)	41K (<1%)	**45.1K**
Other	0.5M (1%)	80K (<1%)	-	**0.6M**

Table 7: OP2.C08 revenue split. Percentages are calculated over the company's yearly revenue.

The lead company in the operation is OP2.C08 with 90.6M in revenue and 11.3M in net income for the period 2013–2015. These correspond to 98% of the total revenue and 100% of the net income of the operation. Similar to OP1, the companies that report employees are the ones with most business activity, save for OP2.C07 that does not report any revenue.

Audit reports. We have acquired the 2014 and 2015 audit report of OP2.C08, both performed by PricewaterhouseCoopers [15]. The two audit reports cover the 2013–2015 period. The audit reports contain a detailed revenue split, summarized in Table 7. The revenue split reveals that the main source of revenue is the PPI service, which generates revenue of 83.4M for the period 2013–15. This is 92% of the total revenue of OP2.C08, and 90% of the total revenue of the operation in that period. The next largest revenue source is a video streaming service launched in 2015. The video streaming service targets the US and offers a free 5-day unlimited content trial that automatically renews to $59.95 per month when the trial ends (compared to Netflix $9 monthly fee for a similar service). Other relevant sources of income are mobile advertising and a download portal that generated 5% and 3% of the 2013 revenue, respectively, but did not generate significant revenue in 2014–2015.

The audit reports also split the revenue of OP2.C08 by geographical area. From the 90.6M of revenue, 2.5% (2.3M) comes from Spain, 10% (9.1M) from other countries in the European Union, and 87.5% (79.2M) from the rest of the world. Thus, most business comes from outside Spain and is produced in US dollars. We believe this split represents where the advertisers using the PPI service to promote their programs originate from.

Company	Creation	Diss.	Cert.	Web	Period	Emp.	Revenue €	Net Inc. €	EBITDA €
OP2.C05	12/07	-	✓	✓	2013-14	0-1	1.3 M	86.1 K	154.0 K
OP2.C07	03/09	10/12	✗	✓	2009-10	0-2		-200.4 K	
OP2.C08	03/11	-	✓	✓	2013-15	58	90.6 M	11.3 M	12.2 M
OP2.C09	10/13	-	✓	✗	2014-15	0	88.5 K	3.0 K	4.3 K
OP2.C10	10/13	-	✓	✗	2014-15	0	209 K	5.9 K	14.4 K
OP2.C11	12/13	-	✓	✓	2014-15	0	41.8 K	-2.3 K	-2.7 K
OP2.C12	04/14	-	✓	✗	2014-15	0	4.9 K	-460	-579
OP2.C13	04/14	-	✓	✓	2014-15	0	4.7 K	-471	-624
OP2.C14	04/14	-	✓	✗	2014-15	0	1.8 K	-887	-1.2 K
OP2.C15	05/14	-	✓	✗	2014-15	0	1.7 K	679	-930
OP2.C16	05/14	-	✗	✗	2014-15	0-5	1.1 M	-69.4 K	-55.5 K
OP2.C17	05/14	-	✗	✗	2015	0		-91.6 K	
Total					**2013-15**	**0-58**	**92.2 M**	**11.0 M**	**12.3 M**

Table 8: OP2 financial data.

Category	2013	2014	Total
PPI	5.6M (87%)	2.3M (68%)	**7.9M**
Advertising	0.4M (6%)	0.3M (10%)	**0.7M**
Software	44K (<1%)	-	**44K**
Other	0.4M (7%)	0.7M (22%)	**1.1M**

Table 9: OP3.C18 revenue split. Percentages are calculated over the company's yearly revenue.

Finally, the audit report shows transactions of 10.7M with 9 other OP2 companies. The largest transactions are performed with OP2.C20, which is registered in Israel. No transactions are reported with the other Israel-based company or with the company registered in Delaware, US.

6.3 OP3 Economics

Table 10 shows the financial data for the 27 OP3 companies registered in Spain, with the same structure as Tables 6–8. Of those 27 companies, 5 have no business reports and 10 report no revenue.

The largest company by revenue is OP3.C09 with 10.6M in 2008–2009, but the central company in the operation is OP3.C18, which has most employees and runs the PPI service. OP3.C18 has 9.7M in revenue and 443K in net income during 2013–2014, which represent 34% of the total revenue and 11% of the net income. Compared to the other operations, the revenue of OP3 is more diversified and less reliant on the revenue of the PPI service. Once again, the companies that report employees are the most active ones.

Audit report. We have acquired the 2014 audit report for OP3.C18, performed by Deloitte [5]. There is no audit report for 2015. The revenue split in the audit report is summarized in Table 9. The table shows that most revenue (7.9M) comes from the PPI service, which represents 80% of the revenue for OP3.C18 and 28% of the revenue of the whole operation. Advertising provides an additional 7% of the revenue, while software revenue is minimal (44K in 2013). Of the 9.7M revenue of OP3.C18, 8% (824K) comes from Spain, 20% (1.9M) from other countries in the European Union, and 72% (7M) from the rest of the world. Similar to OP2, most revenue comes from outside Spain. This geographical split likely represents where advertisers using the PPI service for distribution come from. The company has transactions of 1.3M with 6 other OP3 companies. No transactions are reported with the 5 companies registered in Delaware, US.

7. DISCUSSION

This section discusses different aspects of the operations and limitations of our approach.

Defenses. The economic analysis of malicious and undesirable operations has two main applications: evaluating the deployment of defenses and proposing new defenses [58]. Our results are useful towards the first goal by demonstrating the impact on the PPI market of PUP defenses deployed in mid-2014 by different vendors. A possible defense using entity graphs would be that once the persons behind an operation are identified using an entity graph, company registers could be periodically queried to find new companies created by those persons and put them in a watchlist. Those watchlists could be used by CAs for identifying certificate requests from PUP operations.

Shell companies. Our analysis shows that the three operations employ a large number of companies, but most of them have no employees, use the address of other companies, report no revenue, and have no Web presence. While we cannot be certain of the purpose of such shell companies, we do observe them being used to obtain code signing certificates that are later used to sign PUP samples.

False positives. Our approach to generate entity graphs went through successive iterations to define the trimming steps to avoid including unrelated persons and operations. The resulting entity graphs have undergone extensive manual curation by the authors to verify that no unrelated entities are included. While our approach to build entity graphs can be applied to automatically produce entity graphs for other (Spain-based) operations, given the high cost of wrong attribution, we recommend the final entity graphs are manually curated, as we did, to guarantee that no unrelated entities are included.

Financial data trustworthiness. Our economic analysis is based on the yearly financial statements filed by the companies, and a few audit reports by CPA firms. A limitation of this approach is that it is possible for companies to falsify their results in financial statements, e.g., for fiscal reasons [7,10]. However, such manipulation constitutes a fraud in countries like Spain that mandate yearly financial statements. Verifying the accuracy of financial data is a complex task that requires full access to the finances of a company and is outside the scope of this work.

Certificates. We observe operations using 48–85 code signing certificates, but those are rarely revoked by CAs. This raises the question of why large numbers of certificates are needed. We believe that certificate changes help evading detection by security products. Specifically, it is common for AV engines to include detection signatures that focus on the certificate information, e.g., the signature may correspond to the subset of the certificate's Subject field that

Company	Creation	Diss.	Cert.	Web	Period	Emp.	Revenue €	Net Inc. €	EBITDA €
OP3.C00	07/03	-	✓	✗	2013-14	0		-107.4 K	
OP3.C01	10/03	-	✗	✗	2013-14	0	6.2 K	141.1 K	
OP3.C02	10/03	-	✓	✗	2013-14	1	156.3 K	22.3 K	-48.3 K
OP3.C03	10/03	-	✗	✗	2013-14	0		124.9 K	
OP3.C04	11/03	-	✓	✗	2014-15	0		-17.0 K	
OP3.C05	02/05	-	✗	✗	2009-10	0-5	224.0 K	923.1 K	-475.8 K
OP3.C06	02/05	-	✗	✓					
OP3.C09	03/06	03/14	✗	✓	2008-09	0	10.6 M	2.6 M	3.8 M
OP3.C10	05/07	-	✗	✓	2014-15	12	3.8 M	345.4 K	335.6 K
OP3.C11	01/08	-	✓	✗					
OP3.C12	03/08	02/16	✗	✗	2013-14	0		-17.8 K	
OP3.C13	10/09	-	✗	✗					
OP3.C14	11/09	02/16	✓	✗	2013-14	0		-6.6 K	
OP3.C16	06/10	-	✗	✗	2013-14	0		-10.3 K	
OP3.C17	12/10	-	✓	✓					
OP3.C18	05/11	-	✓	✓	2013-14	31	9.7 M	443.6 K	1.7 M
OP3.C19	07/11	-	✓	✗	2013-14	6	3.3 M	82.0 K	144.3 K
OP3.C23	09/12	-	✓	✓					
OP3.C24	11/12	-	✗	✗	2013-14	0-1	405.6 K	-60.2 K	-50.1 K
OP3.C27	05/13	-	✗	✗	2013-14	0	28.2 K	52.4 K	-52.4 K
OP3.C28	07/13	02/16	✗	✗	2013	0		-1.2 K	
OP3.C29	07/13	-	✗	✗	2013-14	0	68.8 K	-2.7 K	-2.7 K
OP3.C30	07/13	02/16	✗	✗	2013	0		-1.2 K	
OP3.C31	07/13	02/16	✗	✗	2013-14	0	47.0 K	-5.7 K	-5.7 K
OP3.C32	07/13	02/16	✗	✗	2013	0		-1.2 K	
OP3.C34		-	✗	✓	2014-15	1	12.9 K	-729.3 K	-345.9 K
OP3.C36		-	✗	✗	2013-14	0-8	191.2 K	56.8 K	126.2 K
Total					**2008-14**	**0-31**	**28.5 M**	**3.8 M**	**5.1 M**

Table 10: OP3 financial data.

captures the company name. Re-signing a program with a clean certificate for another company bypasses those signatures. We have performed experiments (not detailed in the paper) demonstrating that by removing the certificate chain from a detected PUP sample, the number of AVs detecting the sample reduces significantly. In addition, new companies can be used to reset the reputation for PUP programs. For example, changing the name of a program and its publisher (i.e., certificate) makes it difficult for a user to check if a suspicious program has already been reported by other users as undesirable.

8. RELATED WORK

A large amount of research has used graphs and graph analysis to investigate criminal and malicious activities [26, 27, 29, 30, 34, 37, 38, 51, 60]. The graphs in these approaches use as nodes persons, group of persons, companies, or resources (e.g., telephone numbers, bank accounts). The edges capture widely different relationships such as telephone calls, bank transfers, or resource ownership. In this work we propose the use of entity graphs to perform the first analysis of the economics of PUP operations, and in particular of the commercial PPI services used to distribute PUP. We do not claim novelty of the entity graphs themselves, since they are similar to the graphs used in prior research. On the other hand, the novelty relies on the detailed description of the data sources and the approach used for building the entity graphs, as well as the economic analysis they enable.

PUP. Early work on PUP focused on its deceptive methods. In 2005–2007, Edelman studied deceptive installations by spyware and other unwanted software [31]. Good et al. [32] studied the influence of the form and content of End User Licence Agreements (EULAs) on user's software installation decisions. They discovered that users have limited understanding of EULAs and often regret their installation decisions once informed of the contents of those. Good et al. [33] analyzed user behavior during the installation process of spyware and showed that a short notice before the installation, significantly reduced the number of spyware installations. In 2012, Pickard and Miladinov [49] studied a rogue anti-malware program concluding that while not malicious, it only detected 0.3% of the malware and its main purpose was convincing the user to pay the license.

Research on PUP has recently revived with a number of papers examining PUP prevalence and its distribution through commercial PPI services. Thomas et al. [54] measured that ad-injectors, a type of PUP that modifies browser sessions to inject advertisements, affect 5% of unique daily IP addresses accessing Google. Kotzias et al. [14] studied abuse in Windows Authenticode by analyzing 356K samples from malware feeds. They found that PUP has quickly increased in so-called malware feeds since 2010, that the vast majority of properly signed samples are PUP, and that PUP publishers use high file and certificate polymorphism to evade security tools and CA defenses such as identity validation and revocation. In a separate work, Kotzias et al. [39] used AV telemetry of 3.9 M real hosts for analyzing PUP prevalence and its distribution through commercial PPI services. They found PUP installed in 54% of the hosts and identified 23 commercial PPI services that distribute over a quarter of all the PUP in their 2013–14 dataset. In simultaneous and independent work, Thomas et al. [55] analyzed the advertiser software distributed to US hosts by 4 commercial PPI services. They used SafeBrowsing data to measure that PPI services drive over 60 million download events every week in the second half of 2015, nearly three times that of malware. Nelms et al. [46] analyzed web-based advertisements that use social engineering to deceive users to download PUP. They found that most programs distributed this way are bundles of free software with PUP.

In contrast to these papers, our work analyzes PUP economics and in particular the economics of commercial PPI services.

Malware economics. Prior work has measured the revenue of different malicious activities. McCoy et al. [42] analyzed the leaked databases of three pharmaceutical affiliate programs, finding that they had a total revenue of $170M during a 4-year period between 2005–2010. Stone-Gross et al. [52] analyzed three fake antivirus programs with a combined revenue of $130M in 2008–2010. Thomas et al. [57] measured that the 10-month revenue of 27 merchants of fraudulent Twitter accounts reach $127-459K. Pearce et al. [47] measured that the ZeroAccess botnet had earnings of $2.7M per month in 2013. Liao et al. [41] performed a one year study in 2013–2014 on the Bitcoin addresses used by CryptoLocker ransomware and made a lower-bound measurement of revenue of 1.128 BTC (i.e., $310K) per year. In our work, we measure a combined revenue across the three operations of 195M€ in the three-year period of 2013–2015 (202.5M€ throughout all the analysis period). A key difference with these works is that we have access not only to the revenue, but also to the net income of the operations. Since high revenue does not imply high profit, our data enables to truly examine how profitable commercial PPI services are.

Prior work has also studied other aspects of malware economics. Zhen et al. [40] proposed an economic model for understanding the effective rental size and the optimal botnet size that can maximize the profits of botnet masters. Cormac and Dinei [35] analyzed IRC underground markets finding that these markets are a very low-value channel for exchanging goods. Anderson et al. [24] performed a systematic study of the losses caused by various types of cybercrime. To the best of our knowledge these works have not studied the economics of PUP and commercial PPI services.

Also related are analysis of different malicious ecosystems. Caballero et al. [25] showed that miscreants can distribute their malware through underground PPI services by paying $100-$180 for a thousand unique installs in the most demanded regions. Motoyama et al. [44] analyzed CAPTCHA solving services with a cost of $1 per thousand CAPTCHAs. Thomas et al. [56] found that Google phone verified accounts are sold for $85-$500 per thousand. Twitter accounts are also offered from merchants at $1-$20 per thousand [57]. Stringhini et al. [53] showed that Twitter followers are offered for $20-$100 per thousand and promoted tweets for $10 per thousand. De Cristofaro et al. [28] showed that Facebook likes can be bought for $15–$70 for worldwide users and $60–$190 for US users. Recently, Thomas et al. [58] developed a taxonomy of profit centers and support centers for reasoning about the flow of capital and their dependencies within the black market.

9. CONCLUSION

We have performed what we believe is the first analysis of the economics of commercial PPI services. To enable the economic analysis, we have proposed a novel attribution approach using entity graphs. We have generated entity graphs for 3 Spain-based operations, Each operation runs a commercial PPI service, develops PUP, and manages download portals. For each company in a entity graph, we collect financial statements and audit reports when available.

Our economic analysis has shown that the three operations have a total revenue of 202.5M €, net income of 23M €, and EBITDA of 24.7M €. Operation expenses are high and margins low. The largest source of revenue for each operation is the PPI service, which provides up to 90% of an operation's revenue. But, we also observe the operations to draw revenue from advertising, download portals, PUP, and streaming services. The operations start as early as 2003, but the PPI services do not operate until 2010–2011. Peak revenue and net income happened in 2013. There was a sharp decrease in 2015 leading to losses that year. Each operation runs from 15 up to 32 companies, but most of them are shell companies. Those companies are managed by a small number of 1–6 persons.

10. ACKNOWLEDGMENTS

We are grateful to the developers of LibreBORME for making it open source. We thank Alfonso Allen-Perkins for the fruitful discussions on LibreBORME and the Spanish central company register. We thank George Kapetanios and Joseph Kapetanios for their insights on the analysis of economic and audit reports.

This research was partially supported by the Regional Government of Madrid through the N-GREENS Software-CM project S2013/ICE-2731 and by the Spanish Government through the DEDETIS grant TIN2015-7013-R. All opinions, findings and conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the sponsors.

11. REFERENCES

[1] AFP Audit and Consulting. http://www.afpaudit.com/.

[2] Audalia Laes Nexia. http://www.audalialaesnexia.com/en/.

[3] Boletín Oficial del Registro Mercantil (BORME). https://www.boe.es/diario_borme/.

[4] Certificate Transparency. https://www.certificate-transparency.org/.

[5] Deloitte. https://www2.deloitte.com/global/en.html.

[6] Electronic Data Gathering, Analysis, and Retrieval (EDGAR). https://www.sec.gov/edgar/searchedgar/companysearch.html.

[7] Financial Statement Manipulation An ever-present problem for investors. http://www.investopedia.com/articles/fundamental-analysis/financial-statement-manipulation.asp.

[8] Handelsregister - Commercial Register of Germany. https://www.handelsregister.de/rp_web/welcome.do.

[9] Herdprotect. http://www.herdprotect.com/index.aspx.

[10] Here's why you cannot trust a company's financials. http://fortune.com/2015/10/19/auditors-financial-reports/.

[11] Infocif - Informes de empresas. http://www.infocif.es/.

[12] Israeli Corporations Authority. havarot.justice.gov.ilIIs.

[13] LibreBorme. https://libreborme.net/.

[14] Malsign Project. http://www.malsign.org/.

[15] PricewaterhouseCoopers. https://www.pwc.com/.

[16] Registro Mercantil Central. http://www.rmc.es/.

[17] Requisitos para auditoria obligatoria. http://bcsconsultoresasociados.blogspot.com.es/2012/01/requisitos-para-auditoria-obligatoria.html.

[18] Symantec security products will soon detect and remove Potentially Unwanted Programs (PUPs). http://botcrawl.com/symantec-security-products-will-soon-detect-and-remove-potentially-unwanted-programs-pups/.

[19] That's not the download you're looking for... https://security.googleblog.com/2014/08/thats-not-download-youre-looking-for.html.

[20] The Open Graph Viz Platform. https://gephi.org/.

[21] VirusShare. http://virusshare.com/.

[22] VirusTotal. http://www.virustotal.com/.

[23] Adware: A new approach, April 2017. https://blogs.technet.microsoft.com/mmpc/2014/04/02/adware-a-new-approach/.

[24] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. Van Eeten, M. Levi, T. Moore, and S. Savage. Measuring the Cost of Cybercrime. In *Workshop on Economics of Information Security*. 2012.

[25] J. Caballero, C. Grier, C. Kreibich, and V. Paxson. Measuring Pay-per-Install: The Commoditization of Malware Distribution. In *USENIX Security Symposium*, 2011.

[26] W. Coady. Automated Link Analysis - Artificial Intelligence-Based Tool for Investigators. *Police Chief*, 52(9):22–23, 1985.

[27] R. H. Davis. Social Network Analysis: An Aid in Conspiracy Investigations. *FBI Law Enforcement Bulletin*, 50:11, 1981.

[28] E. De Cristofaro, A. Friedman, G. Jourjon, M. A. Kaafar, and M. Z. Shafiq. Paying for Likes?: Understanding Facebook Like Fraud using Honeypots. In *ACM Internet Measurement Conference*, 2014.

[29] W. Didimo, G. Liotta, and F. Montecchiani. Network Visualization for Financial Crime Detection. *Journal of Visual Languages & Computing*, 25(4):433–451, 2014.

[30] W. Eberle, L. B. Holder, and J. Graves. Using a Graph-Based Approach for Discovering Cybercrime. In *FLAIRS Conference*, 2010.

[31] B. Edelman. Spyware Installation Methods. http://www.benedelman.org/spyware/installations/.

[32] N. Good, R. Dhamija, J. Grossklags, D. Thaw, S. Aronowitz, D. Mulligan, and J. Konstan. Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware. In *ACM Symposium on Usable Privacy and Security*, 2005.

[33] N. S. Good, J. Grossklags, D. K. Mulligan, and J. A. Konstan. Noticing Notice: A Large-Scale Experiment on the Timing of Software License Agreements. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, 2007.

[34] W. R. Harper and D. H. Harris. The Application of Link Analysis to Police Intelligence. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 17(2):157–164, 1975.

[35] C. Herley and D. Florêncio. Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy. In *Workshop on Economics of Information Security*. 2009.

[36] C. Horwath. Country by Country Financial Reporting and Auditing Framework Spain, 2014. https://www.crowehorwath.net/uploadedfiles/crowe-horwath-global/services/audit/financial_reporting_frameworks/financial%20reporting%20-%20spain%20may%202014.pdf.

[37] J. Howlett. Analytical Investigative Techniques: Tools for Complex Criminal Investigations. *Police Chief*, 47(12):42–45, 1980.

[38] C. Jedrzejek, J. Bak, and M. Falkowski. Graph Mining for Detection of a Large Class of Financial Crimes. In *International Conference on Conceptual Structures*, 2009.

[39] P. Kotzias, L. Bilge, and J. Caballero. Measuring PUP Prevalence and PUP Distribution through Pay-Per-Install Services. In *USENIX Security Symposium*, August 2016.

[40] Z. Li, Q. Liao, and A. Striegel. Botnet Economics: Uncertainty Matters. In *Workshop on Economics of Information Security*. 2008.

[41] K. Liao, Z. Zhao, A. Doupé, and G.-J. Ahn. Behind Closed Doors: Measurement and Analysis of CryptoLocker Ransoms in Bitcoin. In *APWG Symposium on Electronic Crime Research*, 2016.

[42] D. McCoy, A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. M. Voelker, S. Savage, and K. Levchenko. Pharmaleaks: Understanding the Business of Online Pharmaceutical Affiliate Programs. In *USENIX Security Symposium*, 2012.

[43] Microsoft. Windows authenticode portable executable signature format, Mar. 21 2008. http://download.microsoft.com/download/9/c/5/9c5b2167-8017-4bae-9fde-d599bac8184a/Authenticode_PE.docx.

[44] M. Motoyama, K. Levchenko, C. Kanich, D. McCoy, G. M. Voelker, and S. Savage. Re: CAPTCHAs-Understanding CAPTCHA-Solving Services in an Economic Context. In *USENIX Security Symposium*, 2010.

[45] MSDN. "Stranger Danger" - Introducing SmartScreen Application Reputation. http://blogs.msdn.com/b/ie/archive/2010/10/13/stranger-danger-introducing-smartscreen-application-reputation.aspx.

[46] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad. Towards Measuring and Mitigating Social Engineering Malware Download Attacks. In *USENIX Security Symposium*, 2016.

[47] P. Pearce, V. Dave, C. Grier, K. Levchenko, S. Guha, D. McCoy, V. Paxson, S. Savage, and G. M. Voelker. Characterizing Large-Scale Click Fraud in Zeroaccess. In *ACM SIGSAC Conference on Computer and Communications Security*, 2014.

[48] R. Phillips. What Makes Delaware an Onshore Tax Haven, December 2015. http://www.taxjusticeblog.org/archive/2015/12/what_makes_delaware_an_onshore.php.

[49] C. Pickard and S. Miladinov. Rogue Software: Protection Against Potentially Unwanted Applications. In *IEEE International Conference on Malicious and Unwanted Software*, 2012.

[50] M. Sebastián, R. Rivera, P. Kotzias, and J. Caballero. AVClass: A Tool for Massive Malware Labeling. In *International Symposium on Research in Attacks, Intrusions and Defenses*, 2016.

[51] M. K. Sparrow. The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects. *Social networks*, 13(3):251–274, 1991.

[52] B. Stone-Gross, R. Abman, R. A. Kemmerer, C. Kruegel, D. G. Steigerwald, and G. Vigna. The Underground Economy of Fake Antivirus Software. In *Workshop on Economics of Information Security*. 2011.

[53] G. Stringhini, M. Egele, C. Kruegel, and G. Vigna. Poultry Markets: On the Underground Economy of Twitter Followers. In *ACM Workshop on Online Social Networks*, 2012.

[54] K. Thomas, E. Bursztein, C. Grier, G. Ho, N. Jagpal, A. Kapravelos, D. McCoy, A. Nappa, V. Paxson, P. Pearce, N. Provos, and M. A. Rajab. Ad Injection at Scale: Assessing

Deceptive Advertisement Modifications. In *IEEE Symposium on Security and Privacy*, 2015.

[55] K. Thomas, J. A. E. Crespo, R. Rastil, J.-M. Picodi, L. Ballard, M. A. Rajab, N. Provos, E. Bursztein, and D. Mccoy. Investigating Commercial Pay-Per-Install and the Distribution of Unwanted Software. In *USENIX Security Symposium*, 2016.

[56] K. Thomas, D. Iatskiv, E. Bursztein, T. Pietraszek, C. Grier, and D. McCoy. Dialing Back Abuse on Phone Verified Accounts. In *ACM SIGSAC Conference on Computer and Communications Security*, 2014.

[57] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson. Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. In *USENIX Security Symposium*, 2013.

[58] K. Thomas, D. Yuxing, H. David, W. Elie, B. C. Grier, T. J. Holt, C. Kruegel, D. Mccoy, S. Savage, and G. Vigna. Framing Dependencies Introduced by Underground Commoditization. In *Workshop on Economics of Information Security*, 2015.

[59] G. Wicherski. PEHash: A novel approach to fast malware clusteringi. In *USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 2009.

[60] J. J. Xu and H. Chen. Fighting Organized Crimes: Using Shortest-Path Algorithms to Identify Associations in Criminal Networks. *Decision Support Systems*, 38(3):473–487, 2004.