Spams meet Cryptocurrencies: Sextortion in the Bitcoin Ecosystem

Masarah Paquet-Clouston
GoSecure
mcpc@gosecure.net

Matteo Romiti
Austrian Institute of Technology
Matteo.Romiti@ait.ac.at

Bernhard Haslhofer
Austrian Institute of Technology
Bernhard.Haslhofer@ait.ac.at

Thomas Charvat
Excello
tc@excello.cz

ABSTRACT

In the past year, a new spamming scheme has emerged: sexual extortion messages requiring payments in the cryptocurrency Bitcoin, also known as *sextortion*. This scheme represents a first integration of the use of cryptocurrencies by members of the spamming industry. Using a dataset of 4,340,736 sextortion spams, this research aims at understanding such new amalgamation by uncovering spammers' operations. To do so, a simple, yet effective method for projecting Bitcoin addresses mentioned in sextortion spams onto transaction graph abstractions is computed over the entire Bitcoin blockchain. This allows us to track and investigate monetary flows between involved actors and gain insights into the financial structure of sextortion campaigns. We find that sextortion spammers are somewhat sophisticated, following pricing strategies and benefiting from cost reductions as their operations cut the upper-tail of the spamming supply chain. We discover that one single entity is likely controlling the financial backbone of the majority of the sextortion campaigns and that the 11-month operation studied yielded a lower-bound revenue between $1,300,620 and $1,352,266. We conclude that sextortion spamming is a lucrative business and spammers will likely continue to send bulk emails that try to extort money through cryptocurrencies.

1 INTRODUCTION

In the past year, a new spamming scheme has emerged: sexual extortion messages requiring payments in the cryptocurrency Bitcoin, known as *sextortion* spam. The threat is about the disclosure of compromising images and/or videos to the recipient's contacts if an amount of money is not sent to a specific Bitcoin address. This sextortion spam scheme mixes a well-known existing online problem, spams, [16, 17, 29, 35, 36], while leveraging *cryptocurrencies* as a new channel for illicit financial transactions. It represents a first integration of the use of cryptocurrencies by members of the spamming industry. Such spamming scheme is much simpler than

conventional spamming that requires, for example, links to affiliate marketing [29] for the sale of goods such as pharmaceutical products [16, 17] or counterfeit goods [36]. Indeed, sextortion spams cut the upper-tail of the spamming supply chain: there are no product shipments or hosting of illicit websites involved.

Since 2018, sextortion spams have been distributed in a dozen languages, most likely with the use of the Necurs botnet [13, 32]. Shultz [32] already conducted a primary analysis of the sextortion spammers potential revenues by inspecting 58,611 Bitcoin addresses found in two sextortion-related spam campaigns lasting 60 days. He concluded that a total of 83 addresses received approximately $146,280. However, the author only summed incoming payments without applying more advanced methods for tracing monetary flows in the Bitcoin transaction graph, which have already been applied for studying other forms of cryptocurrency-related cybercrime, such as ransomware [12, 28]. Shultz's [32] findings provide a first overview of a much bigger puzzle, which has yet to be uncovered and understood. Moreover, the simplicity behind sextortion spams — or any spams taking advantage of cryptocurrencies — may influence many to leverage this strategy, developing new threats and/or offering different products or services to convince spam recipients to send money to a cryptocurrency address. Uncovering the potential returns of this scheme, as well as the level of sophistication behind these campaigns, may indicate the extent to which this illicit scheme will be prevalent in the near future.

This research builds on unique and innovative methods to understand this new spamming scheme, providing insights into how sextortion spammers operate and the extent to which they are successful. To do so, we dive into a dataset of 4,340,736 emails related to sextortion, identify different threat campaigns and evaluate spammers' pricing strategy. Then, we extract Bitcoin addresses and investigate transactions related to sextortion in the Bitcoin ecosystem. Using the open-source *GraphSense Cryptocurrency Analytics Platform*[1], we project them onto transaction graphs by applying the well-known multiple-input clustering heuristic [21, 30]. We estimate spammers' potential revenue through different techniques and analyze patterns in the monetary flows. Summarizing, our contributions are as follows:

(1) We develop a simple and efficient *email bucketing heuristic* to syntactically group 4,340,736 emails into 96 sextortion-related buckets, from which we manually derive 35 specific

[1]https://graphsense.info/

campaigns, allowing us to overcome spammers' obfuscation techniques of changing words and sentences.

(2) We find that sextortion spammers modify the amounts asked in spams, based on the language in which the spam is sent and the campaign in which it is included.

(3) From analyzing the 245 sextortion addresses active in the Bitcoin blockchain, which we make available for research reproducibility[2], we find that one single entity is likely controlling the financial backbone of the majority of the sextortion campaigns studied.

(4) Filtering out unrelated transactions, we conclude that this sextortion scheme generated a lower-bound revenue between $1,300,620 and $1,352,266 for an 11-month period.

(5) We find that the holding period of sextortion payments is, on average, 5.5 days and we track, to the best of our capacity, monetary flows from spammers to known entities.

Our results show that sextortion spammers are somewhat sophisticated, developing pricing strategies before sending spams, yet they are not tracking whether the developed strategies worked, as Bitcoin addresses are reused among spams. We also find that the lower-bound revenues of sextortion spams are similar in magnitude compared to traditional spams [16, 17]. However, by cutting the upper-tail of the spamming supply chain, sextortion spammers make near $122,933 per month while facing little costs. These results suggest that sextortion spams, or any spamming scheme using cryptocurrencies to attempt at influencing spam recipients to blindly send money, are not likely to disappear, due to the lucrative returns. The anti-spam community should thus focus on actively flagging these emails, preventing them to reach users' inbox. We make our research reproducible by publishing the 12,533 Bitcoin addresses found in the 4,340,736 sextortion spams[2].

In the following paper, we start by introducing the necessary background on sextortion spams, summarizing what is currently known about their characteristics and the profitability of the spamming industry, and providing an overview of known methods for estimating illicit revenues using Bitcoin. Then, in Section 3, we describe the simple and efficient *email bucketing heuristic* we developed to group emails in buckets and subsequently campaigns, as well as the computations done on known Bitcoin addresses to expand the dataset and infer additional information from the Bitcoin ecosystem. We present results about the spammers' pricing strategy, the reuse of Bitcoin addresses, the evolution of spamming revenues as well as an analysis of the money flows in the subsequent sections. We finish with a discussion and a conclusion summarizing the main results of this study.

2 BACKGROUND AND RELATED WORK

Since sextortion is a specific spamming scheme, we first summarize what is currently known about spam and the industry behind it in general, before we present known facts about spam sextortion. Then, we focus on profitability of spam and summarize known methods for measuring illicit cryptocurrency revenues.

2.1 Spam and the Industry behind it

Spam refers to the sending of in-mass "messages from economic agents who do not have previous relationship with the customer and who do not offer opt-out provisions" [29, p.96] or, put more simply, "unsolicited bulk emails" [1]. Spam was first attributed to unwanted emails, but now refers to other kinds of undesirable messages, such as unsolicited text messages [3], calls [37], social media posts [3, 33] and fake reviews [11].

The spamming industry, referring to all economic agents involved in the spamming business, has been thriving since the 1980's with the creation of the Simple Mail Transfer Protocol (SMTP) sender push technology [29]. Since its inception, this industry has been closely related to affiliate marketing [29], advertising various products and services, such as pharmaceutical products [16, 17], counterfeit goods, online dating and search engine optimization services [36].

Stringhini et al. [36] suggested that the lower-tail of the spamming supply chain is segmented, with each participant specializing in one activity. They also point out that traditional spammers buy emails lists, rent botnets and use them to send spam. Target email lists are bought from email harvesters, which are individuals or groups of individuals crawling the web and compiling large lists of emails for sale. Then, botnets would be rented by spammers to send their unsolicited emails [14, 17, 35]. Botnets enable spam to be sent in bulk from a vast pool of infected hosts with geographic diversity, especially if the infected host is located in a Dynamic Host Protocol (DHCP). In such case, the IP address of the infected host may be constantly changing, thus hardening the flagging process by the anti-spam community [29].

While Stringhini et al. [36] suggested a segmented industry at the lower-tail, Kanich et al. [16] investigated the Storm spam botnet and concluded that the spamming industry would be more vertically integrated, with the harvesters, the botnet owners and the spammers coming from the same organization.

Then, a study conducted by Levenchenko et al. [19] showed that, at the upper tail of the traditional spamming supply chain, several operations need to be performed. The authors discussed, more precisely, spam campaigns related to pharmaceutical and software products and explained that sending spam is only a small part of the whole operation. Once in a virtual inbox, spammers depend on the propensity of email recipients to click on the URL displayed in the message and for the URL to remain available. Although this process seems simple at first, Levenchenko et al. [19] explain that spammers need to orchestrate and plan an elaborated operation to avoid URL and domain blacklisting, as well as website take downs by Internet Service Providers (ISPs) and registrars. Such an operation is quite sophisticated and requires URL redirection strategies, dealing with bullet-proof hosters and even sometimes using fast flux domain name system (DNS). Moreover, spammers must maintain credible websites, provide reliable payment processes and find banks that will accept to receive payments, regardless of the illicit activity being conducted. Lastly, they may have to acquire the product and deliver it to the customer and/or develop the service. Since the supply chain for the spamming of counterfeit products and software services is quite complex and requires sophisticated operations,

---

[2]https://github.com/MatteoRomiti/Sextortion_Spam_Bitcoin

spammers can also resort to enterprises with lenient values to conduct business with, sending spams on their behalf [29].

2.2 Spam Sextortion

With the rise of Bitcoin as a preferred currency for specific illicit activities, such as ransomware [12, 28], some spammers have returned to spamming through emails while exploiting a new strategy: sexual extortion that requires payments in Bitcoin. The scheme, known as *spam sextortion*, is simple: it aims at extorting money from spam recipients through the threat that compromising photos or videos will be sent to the recipients' contacts if the amount asked in Bitcoin is not paid. Such spamming scheme represents a new approach that leverages Bitcoin and is cost-saving: it cuts the spamming supply chain to its bare minimal as there is no need to develop sophisticated techniques to monetize the scheme.

Indeed, the upper-tail of the supply chain is butchered when considering sextortion spam campaigns: the spam is built on a false threat that if believed by the spam recipient, will generate an amount of money to the spammers. There are no URL redirections and bulletproof hosting to deal with, credible websites to maintain, or business partners to find. The cost of running a sextortion spam campaign is thus largely reduced compared to traditional spams, which may explain why, in the past year, such spams have been prevalent. To this date, sextortion spams have been discussed in mid-2018 in a short article from IBM X-Force [13] as well as a security blog written by Schultz (2018) [32]. We review both articles below.

The first short article [13] presents examples of spam sextortion and hypothesizes that these spams have been distributed since September 11, 2018, by the Necurs botnet, a well-known spamming botnet [18]. According to the blog, the amounts asked varied between $250 and $550 and the threat was found to be distributed in seven languages: English, German, French, Italian, Japanese, Korean and Arabic. The second article, written by Schultz [32] in 2018, investigated two sextortion campaigns (a total of 233,336 emails) that were received by SpamCop, a reporting spam service [3], between August 30 and October 26, 2018. The author found additional languages: Czech, Spanish, Norwegian, Swedish and Finnish, illustrating that the sextortion spamming scheme is inevitably global. Schultz also found that Bitcoin addresses displayed in spams are reused throughout sextortion messages: the spammers do not even bother to generate a unique address per message. This is further supported in the previous article [13] which stated that one Bitcoin address was reused in 3 million messages. Schultz [32], moreover, found overlaps of Bitcoin addresses in sextortion spam emails and a spam where a Russian girl promised to send an explicit video in exchange for $40 in Bitcoins.

These findings suggest that there might be a common financial backbone behind the spam campaigns. However, a more detailed picture on the structure of that backbone and revenues made by sextortion spammers are still missing. But, since spammers clearly tend to reuse addresses, we can infer and show later in this paper that such a picture can be drawn using state-of-the-art methods for tracing and tracking monetary flows in Bitcoin.

2.3 The Potential Profitability of Spams

Traditional spamming is a lucrative business: in 2008, Kanich et al. [16] infiltrated a botnet for 26 days and found that 350 million pharmaceutical spams resulted in 28 sales with a total revenue of $2,731.88, which corresponds to a conversion rate of 0.00001%. The authors claim that, when extrapolating their findings to additional bots, the revenue could go up to $9,500 per day and thus $3.5 million per year. Later, the same authors conducted a second study [17] in which they investigated multiple spamming campaigns and found an even more profitable picture: spammers would earn, on average, $50 per purchase and the revenues per spamming campaign would range between $400,000 and $1,000,000 per month per affiliate program. In both cases, however, such revenues do not include the costs of additional activities related to traditional spams, such as bullet-proof domain hosting or shipping.

Botnet operators are, as discussed above, key actors in the industry and Stone-Gross et al. [35] estimated their potential profit. They estimated the costs of maintaining the Cutwail botnet infrastructure at $1,500 to $15,000 on a recurring basis (with 121,336 unique IP addresses online per day) and concluded that botnet owners must have made between $1.7 to $4.2 million in profit in the past approximate year. According to this same study, spam-as-a-service can be purchased at a price of $100-500 per million of spams or botnets can be rented at a price of $10,000 per month for 100 million of spam emails.

Rao and Reiley [29] identified three tactics for estimating spamming revenues sent through botnets: monitor a botnet and a market for spams, hijack a botnet and estimate the number of sales made, or evaluate the number of completed orders on a website related to spam by looking at the sequential order IDs in the URL.

In our work, we follow a new tactic, one that can be added to Rao and Reiley's [29] typology. The tactic aims at estimating illicit sextortion spam revenues by analyzing associated Bitcoin transactions, which are openly accessible on a ledger, also known as the blockchain.

2.4 Estimating Illicit Bitcoin Revenues

While Bitcoin's share of the cryptocurrency market is shrinking, it still remains the predominant cryptocurrency encountered in cybercrime investigations [4]. Hence, tracking and tracing monetary flows has become a de facto standard forensic method for cybercrime investigations and is now supported by a number of commercial (e.g., Chainalysis[4], CipherTrace[5]) and open-source tools (e.g. BlockSci [15]) and analytics platforms (e.g., GraphSense[6]).

These tools rely on three main complementary techniques: first, they apply various clustering heuristics to group multiple addresses into maximal subsets (clusters) that can be likely assigned to the same real-world actor. Multiple-input clustering [25, 30], for instance, leverages the idea that each Bitcoin address has a private key that is required to sign a transaction and spend the money associated with it. Thus, if two addresses are used as inputs in a transaction, the private keys of the two addresses must be used to

---

[3]https://www.spamcop.net/

[4]https://www.chainalysis.com/
[5]https://ciphertrace.com/
[6]https://graphsense.info

sign the transaction and, subsequently, be accepted by the peer-to-peer network and added to the public ledger. These two addresses can thus be attributed to the same real-world entity, known as a *cluster*. A previous study argued that the multiple-input clustering heuristics are surprisingly effective due to the reuse of addresses and report that super clusters covering thousands of addresses typically represent gambling sites, exchanges or large organizations [8]. Moreover, Nick (2015) [26] tested the heuristics on 37,000 wallets and found that it successfully guessed, on average, 68.9% of all addresses belonging to a wallet. Next to that heuristics, other studies have proposed to form clusters based on change addresses (c.f., [21]) or other behavioral patterns [22].

The second technique is to create network abstractions from the underlying blockchain in which nodes represent addresses or real-world actors by their set of addresses (clusters) and edges represent transactions carried out between addresses or clusters [9, 30, 34]. This provides the data structure required for tracking and tracing monetary flows within a cryptocurrency ecosystem. Finally, address nodes in those networks are enriched with so-called *attribution tags*, which are the key for de-anonymizing actors within an ecosystem; and therefore, the most valuable asset in cryptocurrency forensics tasks. An attribution tag associates a single address with some real-world actor, such as the name of an exchange. The strength lies in the combination of these techniques: a tag attributed to a single address belonging to a larger cluster can easily de-anonymize hundreds of thousands cryptocurrency addresses [6].

The effectiveness of the multiple-input clustering heuristics can, however, provide misleading results if transactions were involved in Bitcoin mixing, either through centralized mixing services (also called tumblers) or through CoinJoin transactions. Centralized mixing services act as middlemen and amalgamate different users' funds before creating new transactions in order to confuse the trail back to the funds' original source (c.f. [24]). A CoinJoin is a feature often supported by wallet providers or dedicated platforms, which allows several users to combine their funds as input of a single Bitcoin transaction without sharing private keys [23]. Therefore, CoinJoins must be taken into account when analyzing cryptocurrency flows and a number of detection heuristics are already available in tools like BlockSci [15]. Furthermore, previous studies have shown that mixing services are typically not used in the first step, when victim pay ransom amounts to some payment address controlled by the attacker [28]. This suggests that mixing services are used at later stages before cybercriminals want to cash out and turn tainted funds into clean, spendable fiat currencies.

Above tools and techniques have so far been applied in two studies: Paquet-Clouston et al. [28] investigated modern ransomware strains that encrypt a user's computer files and asks for a ransom, usually in Bitcoin, for the user to recover the encrypted files. They collected a set of Bitcoin seed addresses related to ransomware and expanded that set using the multiple-input heuristics. This allowed them to trace ransomware monetary transactions related to 35 ransomware families and to estimate that the lower-bound revenue of these 35 ransomware families, from 2013 to mid-2015, was near $12.7 million. Concurrently, Huang et al. [12] extracted Bitcoin addresses from malware files related to ransomware and followed a similar methodology. They ended up estimating that the market for ransomware, over a two-year period, was worth near

16 million of US dollars. These two studies are among the first to provide reliable and reproducible measurement on the revenues of a specific cybercrime: ransomware.

In this paper, we will build and expand on the methodologies presented in those studies and investigate over four million sextortion emails. This should provide insight into the money flow and the potential profitability of sextortion campaigns. Furthermore, this study represents a first analysis of the conjuncture between the spamming industry and cryptocurrencies, a mingle that will likely continue in the next years.

3 DATA AND METHODS

In the following section, we describe the dataset, including over 4.2 million spams, and the different methods used to filter and cluster these spams in different campaigns. We also provide summary statistics on the information extracted from each spam.

3.1 Raw Email Acquisition

Potentially relevant emails for this research were collected by setting up a tailored spam filter in Virusfree's email security solution between 2018-10-02 until 2019-02-13. It filtered all emails containing the terms *BTC* or *Bitcoin* in singular or plural form AND were classified as Spam by Virusfree's spam filter AND originated from an IP that is known by Virusfree as being a member of some botnet. Thus, none of the emails in our dataset reached the recipient. Personally Identifiable Information (PII) in emails, such as recipients' email addresses, were masked before being handed over to us. In total, the raw email dataset contained 4,360,575 emails. These emails cannot be published for privacy reasons.

3.2 Email Bucketing

To cluster similar spams together, we had to group emails with textual similarity into buckets and identify a representative template for each bucket. In short, we had to find sets of emails with a relatively large intersection (high textual similarity). We tackled this problem by developing a simple *email bucketing heuristic* that allowed us to efficiently compute sets of syntactically similar emails and retrieve a representative email template for each bucket. Our heuristic is based on two observations we made upon manual inspection of random email samples: first, attackers introduce intentional spelling or grammar mistakes or language variations, such as synonyms, across the entire email in order to avoid spam filters. Second, the end of the email contained the most important part of the message, where the attacker explained to the victims what to do. Based on these two observations, we defined our heuristic as follows:

Definition 3.1. Let E be the set of emails in our raw dataset and B the set of baskets representing similar sextortion emails.

(1) For each $e \in E$ find a basket $b \in B \subset E$ with the representative email having the same l last words. If there is such a basket, add e to basket b; otherwise, create a new basket with e_t as representative template email.

(2) Compute the Jaccard similarity[7] between the l last words in all pairs of template emails e_t representing a basket and

---

[7]The Jaccard similarity between two sets S and T is defined as $J_{S, T} = |S \cap T| / |S \cup T|$
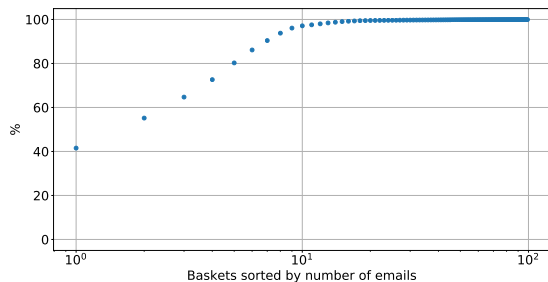
Figure 1: Distribution of emails among the first 100 baskets.

merge two baskets if they have a similarity higher than a threshold t.

In order to maximize Jaccard similarities within baskets, we tuned parameters t and l by running experiments on 10% of the sample emails with all possible pairs (t, l) where $t \in \{0.2, 0.3, 0.4, 0.5\}$ and $l \in \{20, 30, 40, 50, 60\}$. We found that parameters $l = 50$ and $t = 0.3$ yielded maximum similarity with most buckets achieving values greater than 0.8. We considered this a satisfactory result since spammers change textual contents in order to avoid spam filters.

Overall, our email bucketing heuristic produced 1,810 buckets with syntactically similar emails, each one having a representative template email. Figure 1 shows the cumulative distribution of the number of emails in the first 100 baskets ordered by size and shows that 96.1% of all emails fall into the top 10 buckets.

3.3 Identifying Buckets Related to Sextortion

However, at this point, it still remained unclear whether a bucket contained emails related to sextortion or other emails mentioning terms related to Bitcoin. Therefore, we investigated each template representing a bucket and annotated it as being related to *sextortion* or *other*. Since many emails were non-English, we used the Python langdetect[8] module for automatically detecting the email's language. In case of non-English templates, we used Google Translate[9] to translate their bodies and kept an English translation alongside the original template in our dataset.

This annotation procedure was carried out twice by two independent annotators on all 1,810 representative email templates. Disagreement on 12 syntactically corrupted email templates was resolved by jointly investigating them in more detail and classifying them as being sextortion. Finally, we ended up with a manually curated corpus of 4,342,380 sextortion emails filed into 96 buckets representing 99.6% of the total number of emails shared with us.

3.4 Categorizing Buckets by Campaign

Then, we went through each of the 96 templates and found that some of them should be grouped together. Yet, at this point, a manual investigation was necessary and could not be automated. Human interpretation was required because, although the sentence

structure and the wording differed among specific templates, the meaning of the message and the nature of the threat were the same and could only be grasped intellectually. This can be illustrated by considering the two following templates:

Template No.0 (translated from Czech): *Hi! As you may have noticed, I sent you an email from your account. This means I have full access to your account. Your account ([...@....com]) password: [...] I've been looking at you for several months. The fact is that you have been infected with malicious software via the adult website you have visited.*

Template No.4 (translated from Czech): *Hello! As you may have noticed, I sent you an email from your account. This means that I have full access to your account. I've been watching you for a few months now. The fact is that you were infected with malware through an adult site that you visited.*

In template No.0, the attacker has the email and the password of the victim, whereas in template No.4, the attacker doesn't. Yet, the threat that follows is the same with different wording and sentence structure. The buckets representing template No.0 and No.4 should thus be merged, as they encompass the same campaign.

To ensure that our manual investigation was sound, we performed three iterations: going three times manually through each template and grouping them only if specific features around the threat were the same, even though sentence structures and wording differed. Based on this template analysis, a total of 19 campaigns, encompassing at least two buckets or more, were found. These 19 campaigns include 80 of the 96 buckets. The 16 remaining buckets that are not grouped still represent a campaign by themselves. Overall, 35 distinctive sextortion campaigns were found and they represent our final dataset.

Table 1 presents the specific features of 15 campaigns that were detected in the investigation above. It shows, for example, that Campaign D represents a merged group of templates that mentions the presence of an international hacking group while Campaign G represents another that states that the attacker comes from the DarkNet. Campaign F, on the other hand, details two options to the email recipient: not to pay and live in shame or pay and live safely, while Campaign L mentions the exact *Common Vulnerability Exposure (CVE)* used to infect the device.

After having identified buckets related to sextortion and having categorized them by campaign, we extracted the following data points from each email in those buckets: the amount asked[10] normalized in USD, the time when an email was sent from its `Date` field in the header, the user's alleged password (or phone number) mentioned in the email, and the Bitcoin address the victim is supposed to send money to. We define all bitcoin addresses found in the spams as *sextortion payment address*. Table 2 provides summary statistics on the information extracted from the top 15 baskets representing 99.0% of all emails in our dataset. We also identified four campaigns (not shown in Tables 1 and 2) that did not involve hacked email accounts and passwords, but phone numbers.

---

[8]https://pypi.org/project/langdetect/
[9]https://translate.google.com

[10]Amounts were asked in the following currencies: USD, EUR, GBP, NOK

Table 1: Distinctive features of 15 sextortion spam campaigns

Campaign	Key Distinctive Features for each Campaign
A	A Trojan virus exploited the device and an anti-virus can't detect it as a "Trojan Driver" updates the signature every 4 hours
B	RDP was exploited with a keylogger and the device was infected while watching compromising videos
C	The primary goal was to lock the device and ask for money, but a sextortion ransom was asked instead
D	The attacker is a member of an international hacking group and knows all of the recipient's secrets
E	A RAT software was installed on the device and discovered compromising videos
F	A virus infected the device via porn websites and two options are presented: pay or not pay
G	The attacker is known on the DarkNet and hacked the recipient's mailbox a long time ago
H	The attacker is from China and the email is a last warning
I	The recipient's password was intercepted via a porn website and a malicious software now intercepts the password every time it is changed
J	A Trojan with remote access was installed on the device and the device was additionally hacked through porn websites
K	The nationality of the attacker is mentioned as an excuse for the poor language structure of the spam
L	The attacker claims to be a spyware software developer and mentions the CVE used to compromise the device
M	The email recipient's cellphone was compromised and all pictures and videos backed up by the attacker
N	The email is presented as a serious last warning
O	A Trojan infected the device and downloaded all of the recipient's private information, the camera also captured "soloing"

Table 2: Summary statistics on data points extracted from top 15 baskets.

Basket	Campaign	Language	#Emails	Range [$]	Currency	Time Period	#Passwords (Unique)	Jaccard Mean (Var)
0	A	cs	1,811,062	280 | 290	EUR	2018-12-02|2018-12-23	387,943 (283,828)	0.98 (0.0)
1	E	cs	594,904	250 | 550	USD	2018-09-06|2018-11-12	544,448 (388,809)	0.96 (0.0)
2	B	en	415,925	300 | 7,000	USD	2018-09-22|2019-01-24	414,785 (297,514)	0.91 (0.0)
3	C	en	345,969	695 | 899	USD	2018-10-29|2019-02-12	303,910 (234,354)	0.92 (0.0)
4	A	en	331,742	500 | 1,000	USD,EUR	2018-11-30|2019-02-13	0 (0)	0.82 (0.01)
5	G	cs	256,252	200 | 290	USD	2018-09-27|2018-11-06	256,252 (189,173)	0.98 (0.0)
6	D	cs	186,723	250 | 250	USD	2018-09-03|2018-11-24	0 (0)	1.0 (0.0)
7	F	en	146,709	700 | 750	USD	2018-12-20|2019-02-15	111,114 (97,485)	0.94 (0.0)
8	L	en	100,395	700 | 999	USD	2018-10-14|2019-01-15	54,883 (49,721)	0.86 (0.01)
9	C	en	44,790	500 | 1,000	USD	2018-09-09|2018-12-27	44,454 (26,920)	0.93 (0.0)
10	A	de	19,850	336 | 784	EUR	2018-12-07|2019-01-21	0 (0)	0.99 (0.0)
11	E	ko	19,683	400 | 500	USD	2018-09-14|2018-12-26	0 (0)	1.0 (0.0)
12	G	en	18,314	200 | 2,000	USD	2018-09-18|2019-02-13	0 (0)	0.79 (0.03)
13	C	de	13,732	336 | 504	EUR	2018-11-24|2019-01-22	0 (0)	0.99 (0.0)
14	F	en	12,476	500 | 7,000	USD	2018-09-13|2019-01-15	12,213 (1,472)	0.45 (0.0)

3.5 Finding Sextortion Payment Addresses

From the 96 considered sextortion-related buckets, we extracted 12,533 Bitcoin *seed addresses*, but found that only 245 of them received funds. We consider these 245 as being our set of *seed addresses*.

Next, we used the *GraphSense Cryptocurrency Analytics Platform*[11] to expand these seed addresses to maximal address subsets (clusters) computed over the entire Bitcoin blockchain from its inception until April 30th, 2019 (block 573,989). GraphSense applies the well-known [5, 31] multiple-input clustering heuristics for that purpose. The underlying intuition is that if two addresses (i.e., *A* and *B*) are used as inputs in the same transaction while one of these addresses along with another address (i.e., *B* and *C*) are used as inputs in another transaction, then the three addresses (*A*, *B* and *C*) must somehow be controlled by the same entity [21], who conducted both transactions and therefore possesses the private keys corresponding to all three addresses. This heuristic can fail when

CoinJoin transactions [23] are taken into account, because they combine payments from different spenders that do not necessarily represent one single entity. Being aware of this problem, we filtered these transactions out using detection heuristics similar to those found in the tool BlockSci [15] before applying the multiple-input heuristics. Also, as discussed in Section 2.4, we assume that victims do not directly pay into other centralized mixers.

We found that 49 seed addresses could not be associated with clusters, while the remaining 196 appear in 52 different clusters. Table 3 reports the number of addresses as well as the first transaction date of the top 10 clusters sorted by amounts received in US dollars. We noticed that two clusters, cluster 0 and 1, differed significantly from the others in terms of the number of addresses, the total amount received and the date of the first transaction. Based on attribution tags retrieved from *walletexplorer.com*, we were able to identify one of these clusters (0) as being related to the exchange *Luno.com*. The real-world identity of the other cluster is unknown, but from its characteristics, we concluded that it represents, most likely, an exchange or another large real-world entity. From these

---

[11]https://graphsense.info

Table 3: Summary statistics of top 10 clusters by amounts received.

Cluster	#Seed Addresses	#Addresses	Amount Received [$]	First Tx
0	1	398,777	2,155,478,411	2013-03-22
1	3	564,595	2,110,090,570	2016-03-21
2	16	52	240,221	2018-09-25
3	15	25	200,146	2019-01-07
4	13	22	178,245	2018-10-30
5	6	19	108,652	2018-10-25
6	13	37	103,660	2018-09-12
7	8	29	77,062	2018-09-10
8	3	11	50,106	2019-01-24
9	11	47	36,661	2018-10-15

observations, we decided to disregard the addresses included in these two clusters, keeping only their seed addresses.

Overall, when disregarding the two large address clusters, we expanded our dataset to 485 address. We assume that the addresses included in this *expanded dataset* represent sextortion payment addresses as well.

3.6 Filtering Sextortion Transactions

With 485 sextortion payment addresses at our disposal, we tried to identify Bitcoin transactions representing payment from email recipients to spammers. Since sextortion payment addresses could also be used for other purposes, or other kinds of spamming schemes as discussed in [32], we approached this problem heuristically and subsequently employed three filtering methods with varying degrees of restrictiveness on all incoming transactions. This allowed us to gradually restrict the set of possible payments to a very narrow, lower-bound estimate.

Inspired by previous work that estimated the ransomware market, while considering potential collectors (addresses used to aggregate money) in the dataset [28], we developed a first filter, dubbed *CollectorFilter*. This filter removes transactions with at least one sextortion address among the inputs and one in the outputs, which indicates money flowing among the group of spammers, and not a victim's payment.

The second filtering method, named *RangeFilter* selects only the transactions where sextortion addresses received payments within the range found in spams, or more precisely, of $[(1 - p) * \min S, (1 + p) * \max S]$, where $p = 0.1$ is a tolerance factor taking into account possible exchange rates oscillations between the email time and the payment time, and S is the set of all the ransom amounts we found in the emails.

The third filter, dubbed *Moving-MoneyFilter* removes transactions with only one output, also known as *exact-amount payment* [2]. Such filter assumes that these transactions represent a sender using the *send maximum amount* feature of a wallet to transfer funds from one address to other, both addresses belonging to the mentioned sender.

4 ANALYSIS

After having collected Bitcoin addresses and transactions related to sextortion campaigns, we start our analysis by analyzing spammer's pricing strategies. Then we examine reuse of Bitcoin addresses

across campaigns, analyze the accumulation of revenues longitudinally, and provide some deeper insights into the monetary flows involved in sextortion spam.

4.1 Spammers' Pricing Strategies

Sextortion spams in our dataset have various threat structures, known as campaigns, and are written in different languages. We hypothesized that sextortion spammers could follow different pricing strategies, based on the language in which the spam is sent and/or the content of the campaigns, as well as whether the email contains personal information, such as the recipient's password or phone number.

First, to systematically assess whether spammers consider language in their pricing strategy, we tested the following *null hypothesis* between language pairs: *the mean amount asked in a certain language does not statistically differ from other languages*. We computed pairwise t-tests, comparing ransom amounts asked based on the language in which the spam was sent. For each language, we conducted normality tests on amounts asked as discussed in [20]. If, on average, the random samples tested followed a normally shaped distribution, we kept the samples to compute pairwise t-tests. If not, we concluded that the samples were too small in size for the t-tests to be robust to non-normality. We thus disregarded them in the analysis below to avoid doing a Type 1 error (wrongly rejecting a null hypothesis) when comparing means of amounts asked. Moreover, since variances among spams grouped per language and per campaign showed heteroscedasticity, we applied Satterthwaite correction on sample variances by computing the Welch-Satterthwaite t-test, a test that is designed for unequal variances [27, 38]. Lastly, to adjust the p-values for multiple comparisons we used the conservative Bonferroni correction.

The pairwise t-tests revealed that the means of amounts asked per language are significantly different for most of the pairwise comparisons. To discuss these results, Figure 2 presents the mean amounts asked per language, along with the standard deviations in blue and the 1.96 standard mean errors in red. Overlaps of standard mean error bars indicate that the difference between means is zero.

If we look at pairwise comparisons over all campaigns and languages in Figure 2a, we find that the means of amounts asked in English spams are significantly higher than the means of amounts in all other languages, except Slovenian. We can also observe that the means of amounts asked in German, French, and Dutch are not significantly different. In the lower range of mean payments asked, we find Polish, Italian, and Spanish, which do not significantly differ within their group, but do significantly differ from other languages demanding higher amounts. Japanese, Slovakian and Norwegian spam samples did not pass the normality tests discussed above. Overall, spammers ask higher amounts, on average, for spams sent in English, Slovenian and Korean and lower amounts, on average, for spams sent in Polish, Italian and Spanish.

Next, in order to assess if this finding was robust within one single campaign, we looked at amounts asked per language in Campaign A, a campaign that contains six languages and more than two million spams. Based on Figure 2b and pairwise t-tests between each language in Campaign A, we find that the means of amounts asked in English are significantly higher than all other
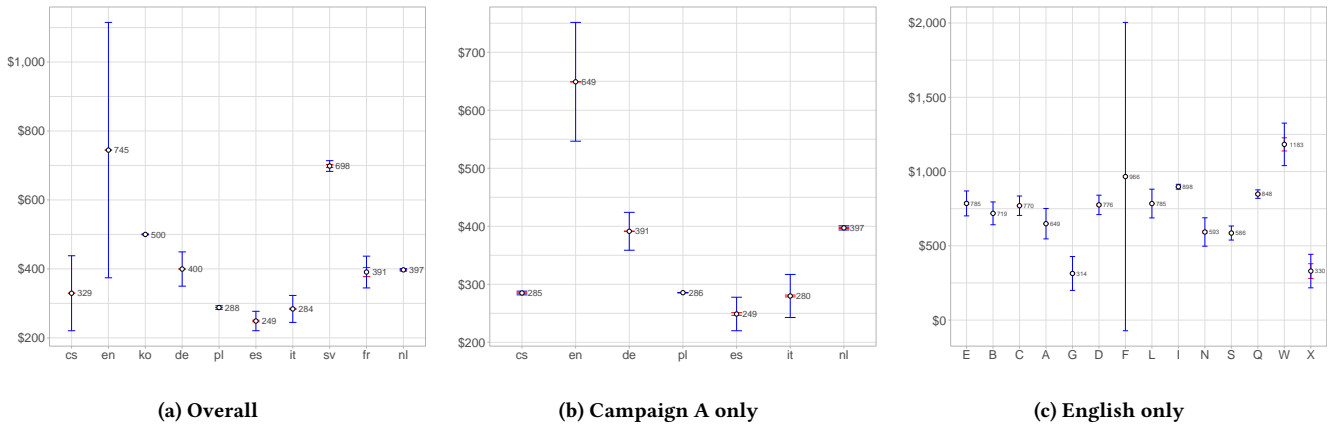
| (a) Overall | (b) Campaign A only | (c) English only |

Figure 2: Mean amounts asked per Language and per Campaign

languages. The means of amounts asked in Dutch are also significantly higher than all other remaining languages. The means of amounts asked in Czech spams are significantly higher than Spanish and Italian spams, but are not significantly different with Polish spams. Overall, we find similar results, with recipients of English spams facing higher amounts asked than recipients of spams in all other languages. Also within Campaign A, Spanish spams asked the cheapest amounts, on average.

We looked if amounts asked differed per campaign. To avoid grasping a language effect, we selected only English spams, a language encompassing 1,427,311 spams and 31 campaigns. Then, we selected the 14 campaigns that passed the normality test mentioned above. As shown in Figure 2c and based on pairwise t-tests, we find that there are significant differences in the means of amounts asked depending on the campaign. For example, the means of amounts asked in Campaign G are significantly lower than all other campaigns except Campaign X. Amount asked in Campaign W are also significantly higher than amounts asked in all other campaigns. Campaign F also shows that amounts asked in English vary greatly. Overall, we find that there are significant differences in the amounts asked per campaign.

Lastly, we found no significant differences between spams sent with and without the recipient's password or phone number, even when controlling for the language in which the spam was sent and the campaign in which it was classified. Spammers do not ask for higher amounts when they have additional information from the recipient. In summary, from these findings, we can conclude that sextortion spammers consider language in their pricing strategy within and across campaigns. They also ask for different amounts depending on the campaigns.

4.2 Passwords and Public Data Breaches

Schultz [32] hypothesized that the emails and passwords used by sextortion spammers come from publicly available data breaches. To explore this hypothesis, we selected all unique clear text passwords (as email addresses were masked for privacy purposes) from the spamming dataset. We then created a list of clear text passwords that were seen only once in the dataset, to avoid passwords such as *12345*,

and that comprised at least four characters, to avoid passwords such as *!!*.

This yielded a list of 632,688 unique passwords. We fetched 21 publicly available data breaches (e.g., MySpace, Linkedin, Twitter, Yahoo, 665KK), encompassing nearly 800 million passwords. Due to computational limits, we randomly selected 25% of the passwords included in the list above and investigated whether they were contained in one of these breaches. A total of 80,944 passwords were found to be a match. In short, about 51% of our randomly selected password samples used by sextortion spammers were passwords found in previous clear text data breaches. These findings strongly support the idea that sextortion spammers use publicly available data breaches to populate the content of their spams.
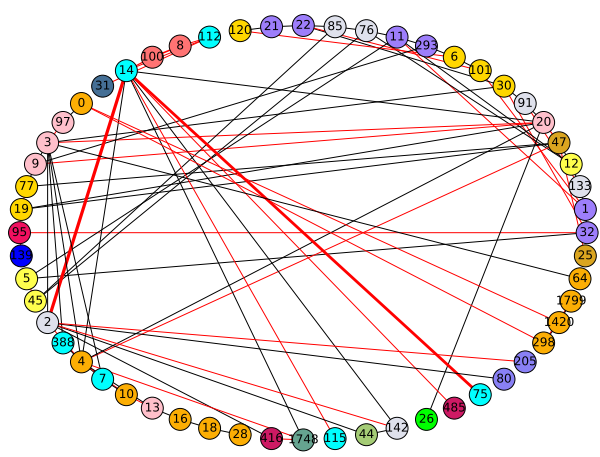
4.3 Reuse of Bitcoin Addresses across Campaigns

When investigating Bitcoin addresses extracted from the buckets listed in Table 2, we noticed that addresses and associated clusters are reused across buckets and sextortion campaigns, confirming similar observations made by researchers previously [13, 32]. More specifically, we observed that 50 clusters appear in at least two distinct baskets, and from this, we hypothesized that the financial revenue streams of a large majority of sextortion campaigns are controlled by a single real-world entity despite their semantic and language differences.

We evaluated our hypothesis by creating an undirected graph abstraction from the 96 buckets that were identified as being related to sextortion campaigns. Nodes in the graph represent buckets while edges between nodes indicate shared clusters or shared seed addresses among two buckets. Thus, an edge between two nodes indicates that access to funds collected from sextortion victims is controlled by the same real-world entity, holding the private keys of some addresses included in both nodes.

When analyzing that graph abstraction, we found that the overall graph consists of 4 disconnected components, with one giant component covering 56 out of 96 nodes, which include 99.4% of the emails and 72% of the sextortion addresses. Two other disconnected components have only three nodes while the last component has

Figure 3: Relationship among sextortion buckets, campaigns, bitcoin addresses and clusters.

Figure 4: Sextortion's cumulative revenues combining different filters and single payments to sextortion addresses.

two nodes. All the remaining nodes are disconnected. Figure 3 represents the giant component and shows the financial connections between buckets and campaigns identified by overlapping Bitcoin clusters or addresses. Red edges indicate that the connected buckets share at least one Bitcoin address. Black edges denote intersections in the set of addresses (clusters) associated with two connected bucket nodes. The strength of an edge indicates the number of shared entities (either clusters or addresses) and node colors represent campaigns.

In summary, our analytics strongly indicates that a large majority of the financial revenue streams generated by sextortion spam can essentially be attributed to a single real-world entity. We note that this does not necessarily imply that all sextortion campaigns are designed and carried out by a single person. It is more likely that botnet infrastructures, which can be rented by spammers, are nowadays designed with financial features that handle the creation of Bitcoin addresses, and potentially handle the financial revenue streams within the Bitcoin ecosystem. However, more research on the matter is needed.
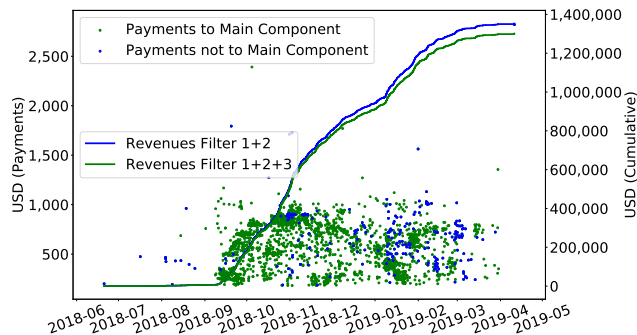
4.4 Sextortion Revenues

Next, in order to create reliable and reproducible statistics on the profitability of sextortion spam, we estimated the profitability of sextortion spams based on the emails contained in our dataset. Since our observation period covers five months (2018-10-02 until 2019-02-13), we also wanted to understand how the revenue stream evolved over time.

To gain insight into the overall profitability, we first quantified revenues with various filter settings: we applied the *CollectorFilter* (1) and the *RangeFilter* (2) presented previously on all incoming transactions of the expanded dataset, providing a reasonable revenue estimation. Additionally, we applied the *Moving-MoneyFilter* (3). However, we considered the estimate including the third filter as a *conservative one*. Indeed, the *Moving-MoneyFilter* could disregard potential transactions related to sextortion, as victims could send money to a sextortion address without providing a change address.

Table 4 shows the number of payments, as well as the total amounts received by the sextortion addresses based on the different filtering criteria presented above. It shows that the overall revenue of sextortion spams lies between $1,300,620 and $1,352,266.

Figure 4 illustrates all the individual payments to sextortion addresses and the cumulative sum over time. Although we collected spams from September 2018 to March 2019, Figure 4 shows that we captured payments before and after this period. The revenue estimations thus include payments between June 2018 and April 2019. Figure 4 also shows that sextortion addresses related to the one single component, discussed in the previous section, accounts for 88.3% of the overall sextortion revenues. Moreover, from September 2018 to March 2019, we find that sextortion addresses receive an almost steady flow of payments, with slightly higher revenues in October, November and January with $336,979, $286,462 and $250,189 respectively. The average monthly revenue being $122,933.

4.5 Holding Periods and Monetary Flows

To better understand when and how victims' coins change hands, we investigated the holding periods of received payments and subsequent monetary flows to other addresses or clusters.

A holding period is the time elapsed between the moment a sextortion address received coins from a victim and the moment these coins are spent. We found that holding periods vary between few minutes and few months, with an average holding time of 5.5 days and a median of 3.8 days, as shown in Figure 5.

Then, we investigated monetary flows that start from a victim's address (depth 0) and forward amounts to other addresses through several hops. Technically, this corresponds to traversing a transaction graph, in which nodes represent addresses and edges represent transactions. In order to limit computational efforts, we stopped
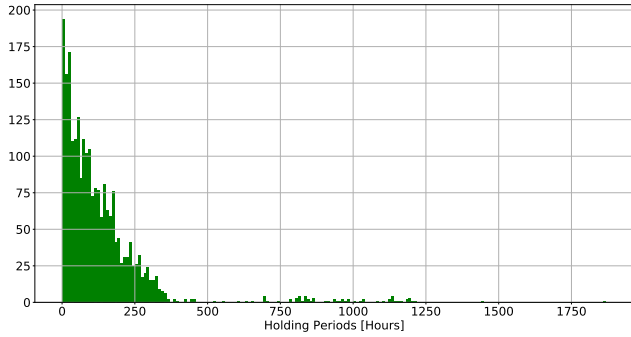
Figure 5: Histogram of holding periods considering the amounts received by sextortion addresses. Each bin has a width of 10 hours.

Table 5: Top 15 clusters by BTC received from victims at depth 2.

Cluster	Total Spent [$]	First Tx	#Txs Out	BTC
466805936	113,082	2019-01-09	1	19.324660
374981799	52,789,437	2017-12-27	179	18.418278
363386240	2,110,090,569	2016-03-21	377,671	10.377687
452279937	16,690	2018-11-19	1	3.781298
426830314	36,454	2018-07-23	20	3.513247
477564438	17,636	2019-01-20	1	3.476584
329777716	154,186,706	2013-04-16	226,156	2.977402
484209796	488,989,966	2018-08-04	159,198	2.959614
498338351	27,513,490,714	2014-09-04	3,150,371	2.945446
310058166	6,016,682	2016-10-28	633	2.180812
490354283	1,344,877,122	2018-10-04	13,230	1.764936
469936395	11,736	2019-01-15	1	1.654278
468575380	33,950	2019-01-09	22	1.544722
496671952	3,421,775	2019-01-11	129	1.517212
386226499	9,883,607	2016-12-07	565	1.420000

traversing when we hit an address representing a known exchange or when we encountered more than 100 nodes at a certain depth.

For the analysis, we only considered transactions that have been flagged as sextortion payments, based on the *Collector-* and the *Range* filters, yielding 2,252 sextortion payments related to 33 addresses and 48 clusters. We did not consider the *Moving-Money* filter because it might exclude payments where the victim does not care about the change.

We found that all the coins sent from the victims to the sextortion addresses have been moved to other wallets. Within two hops, we already found entities representing known cryptocurrency exchanges, such as Huobi.com, CoinPayments.net, Cubits.com, and again Luno.com, thanks to the attribution tags retrieved from *walletexplorer.com*. These known entities, however, only received 3.8181 BTC, representing about 0.135% of the total estimated sextortion revenues. Such finding would indicate that spammers do not necessarily cash out within the first hops. However, since our attribution dataset is limited to *walletexplorer.com*, we cannot conclude with certainty this assumption, as untagged clusters may represent exchanges or money-laundering services that we are not aware of.

As a first explanatory exercise, we present in Table 5 all unknown clusters found at depth 2, meaning the clusters that received money from known sextortion addresses. As shown in this Table, some clusters spent millions of dollars, were active years before sextortion spams were first seen, and completed thousands of transactions. Considering these features, such clusters could represent exchanges or money-laundering services. As an extra exercise, we computed the proportion of sextortion revenues that were moved to clusters that have a transaction that is older than June 2018. Such criterion -June 2018- represents the month in which we observed a first sextortion payment in our dataset. Clusters being active before June 2018 could represent potential cash outs, since they are older than the spammers' observed activities. We find a total of 48.218340 BTC, 17% of spammers' revenue going through these clusters, representing potential cash outs. However, this analysis is arbitrary and further research is needed to determine what features of a cluster (e.g. total amount spent, number of addresses, activity period) can hint an individual that the cluster in which the coins are flowing, although untagged, is a potential cash out service.

5 DISCUSSION

In this section, we first discuss our findings about the limited sophistication of sextortion spammers. We then provide estimates of spammers' revenues and profits and lay down the limits of this research.

5.1 The Somewhat Sophistication of Sextortion Spam Operations

This study finds that sextortion spammers developed a large array of threat strategies to convince spam recipients to pay the amount asked: a total of 80 different campaigns were found. The amount asked also differed based on the campaign in which a spam was included into. However, we couldn't find a precise reason as to why a campaign would display higher average amounts compared to other campaigns. Identified features, such as Campaign W stating that both the device and the spam recipient's phone were exploited, or campaign G, mentioning that the attacker comes from the Dark-Net, were insufficient to find why some campaigns would be asking for higher amounts compared to others.

We also found that sextortion spammers adjust the price of spams to the language in which the spam is sent. Our results showed that, on average, English ($745), Slovenian ($698) and Korean ($500) are charged higher amounts than Polish ($288), Italian ($284) and Spanish ($249) recipients. This is in line with other related results: Hernandez et al. [10] found that ransomware authors charged their victims differently depending on many variables, including language. Another study found malware-as-a-service was sold at a higher price on English websites compared to Russian ones [7]. Thus, our study finds that, similar to other illicit online activities, spammers also conduct price discrimination based on language. However, explanations as to why the language of a spam recipient changes the amount asked by sextortion spammers are yet to be discovered. There are many potential arguments, such as testing different threat strategies or subjectively deciding that a specific

language-speaking population should be charged more. However, such analysis is beyond the scope of this paper.

We also found that spammers do not charge higher prices for spams that include passwords and/or phone numbers. This is surprising considering that these additional elements could make the threat more convincing. Moreover, spammers do not seem to track down which threat strategy is more efficient or which language-speaking population pays more, as Bitcoin addresses are being reused among spams, regardless of the campaign or the language in which the spam is sent. Such findings indicate two tendencies: spammers do develop strategies related to the spam content, but they do not track afterwards the success of their operations (which campaign paid more, for example). Their sophistication is thus, limited.

Potentially, the fact that spammers do not track payments could be related to the idea that botnet infrastructures are nowadays designed with financial features that handle the creation of Bitcoin addresses, and possibly handle the financial revenue streams within the Bitcoin ecosystem. The sophistication would thus be limited to pre-spamming activities.

5.2 Revenues and Costs of Sextortion Spams

Our research results indicate that one single entity is likely controlling the financial backbone of the majority of sextortion campaigns investigated in this study. Moreover, depending on the applied filter, sextortion spams generated a lower-bound revenue between $1,300,620 and $1,352,266 for an 11-month operation, or an average of $ 122,933 per month, between June 2018 and April 2019 when payments were captured.

Traditional spam was found to be as lucrative, with findings on revenues ranging from 3.5 million per year [16] to between $400,000 to 1 million a month [17], depending on the study. However, as discussed above, sextortion spams cut the upper-tail of the supply chain compared to the traditional spamming industry [19]. The costs of sending sextortion spams are much lower than that the costs of sending traditional spams related affiliate marketing [29] or the sale of pharmaceutical products [16, 17] or counterfeit goods [36], as no third-party relationships are needed, nor product shipments or hosting of illicit websites, for example.

Moreover, a large proportion of the passwords found in the dataset were related to available data leaks and pastes. Sextortion spammers are thus likely to use emails found online, through public data breaches, rather than buy email lists from harvesters, as argued in previous research [14, 17, 35]. The price of harvesting these emails would thus be quite limited.

At another level, all evidences point to the idea that sextortion spammers are using botnets to disseminate their spams, just like traditional spammers. Indeed, two previous articles on sextortion [13, 32] argued that this worldwide campaign was sent by the Necurs botnet. This assumption was based on the geolocations of the IP addresses sending the spams, which were similar to the Necurs's infected hosts. Our sextortion dataset was also created partly through a filter that selected emails sent from IP addresses that were known to be part of botnets, supporting this hypothesis as well.

Moreover, according to Stone-Gross et al. [35], spam-as-a-service can be purchased at a price of $100-500 per million of spams or $10,000 per month for 100 million of spams. Kanich et al. [16] also found that 350 million of spams soliciting pharmaceutical products led to 28 sales. This means that 12.5 million of spams would need to be sent for one sale to be conducted. Hypothesizing that the lower-tail of the spamming industry is segmented [36] and that sextortion spammers operated within these eleven months by renting a botnet at a price of $10,000 per month for 100 million of spams [16], a practice that was found in several previous research [14, 17, 35], sextortion spammers would have paid a total of $110,000 to botnet owners. This means that they would have made a lower-bound profit of $1,190,620 and $ 1,242,266 for eleven months. Such estimate, however, does not include, other costs, such as the costs of crafting the spams or of cashing out through the Bitcoin ecosystem. Yet, this potential profit magnitude shows that sextortion spamming is a lucrative business that is likely to continue, especially considering that steady flows of payments every month were observed in our dataset.

5.3 Limits and Future Research

This research faces some limits that can drive forthcoming studies. First, the dataset studied is based on emails that never reach the intended recipients. The lower-bound estimates are derived from the reuse of Bitcoin addresses by spammers in their sextortion spams. Potentially, the reported numbers underestimate the magnitude of the spammers' revenues. Thus, possibly, our dataset represents only a small proportion of a global campaign. The multiple-input heuristic allowed us to find many addresses related to sextortion, but we cannot confirm that the expanded dataset is holistic. Future research could reproduce our methodology with additional addresses. Upcoming studies could also look into the potential conversion rate of sextortion spams, a measure that we could not find due to our limited dataset.

Second, the address clustering method we applied in our study is a best effort approach based on behavioral patterns such as reuse of addresses across transactions. We cannot preclude that it unifies entities that have no association in the real world, or that it misses to unify entities. Our method also relies on the availability and reliability of attribution tags, which are currently harvested from public sources (mainly *walletexplorer.com*) without a verification process in-between. Future investigations that aim at understanding the real-world actors involved in sextortion campaigns and associated monetary flows clearly have to provide safeguards for ensuring the evidential value of such kind of analysis, by recording, for example, the provenance of attribution tags. Moreover, the tracking of money flows is limited by these attribution tags. Upcoming research should look into determining clusters' features that can detect a potential cash-out service, without the use of the attribution tag, allowing the tracking to stop once such a cluster is found.

The expanded dataset may also include transactions that are not related to sextortion spams. Indeed, Schultz [32] did find one example of an address being reused in sextortion spam and in a scam where a Russian girl would send a video in exchange of a small Bitcoin amount. Such spams were not found in our dataset nor was the example provided by Schultz (2018) found in the campaigns

studied. Yet, we cannot exclude the possibility that the revenues estimated include other kinds of spamming schemes that aim at influencing spam recipients to send money to addresses. Subsequent studies should aim at understanding other schemes developed by spammers to take advantage of cryptocurrencies and extract money from spams recipients, an amalgamation that is unlikely to go away.

6 CONCLUSION

This study shows that spammers have upped their game: mixing the cryptocurrency bitcoin with spams by trying to extort money from spam recipients. This is done through a false threat: compromising videos or photos will be sent to the recipient's contacts if an amount of money is not sent to a specific bitcoin addresses. We find that sextortion spammers are somewhat sophisticated, following specific pricing strategies, based on language and threat structures. Yet, our findings also indicate that they do not track whether their strategies are successful, hence the idea of "somewhat". Moreover, we conclude that sextortion spammers cut the upper-tail of the spamming supply chain, reducing the spamming cost. Tracking their financial transactions in the Bitcoin ecosystem, we find that these spammers made a lower-bound revenue between $1,300,620 and $1,352,266 for an 11-month operation.

We conclude that sextortion spamming is a lucrative business and spammers will continue to send bulk emails that try to extort money through cryptocurrencies. Other schemes, different than sextortion, but involving cryptocurrencies, will likely be developed by spammers, allowing such industry to thrive. The anti-spam community should thus focus on preventing these spams to reach users' inbox by flagging messages that include cryptocurrency addresses and that are coming from IP addresses that are known to be part of botnets.

ACKNOWLEDGMENTS

REFERENCES

[1] Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Sakkis, G., Spyropoulos, C. D., and Stamatopoulos, P. Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach. In *Proceedings of the workshop "Machine Learning and Textual Information Access"* (2000), 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-2000), pp. 1–11.

[2] Bitcoin-Wiki. Privacy. https://en.bitcoin.it/wiki/Privacy, 2019. Retrieved May 17, 2019.

[3] Chen, C., Zhang, J., Chen, X., Xiang, Y., and Zhou, W. 6 million spam tweets: A large ground truth for timely twitter spam detection. In *2015 IEEE international conference on communications (ICC)* (2015), IEEE, pp. 7065–7070.

[4] EUROPOL. Internet organized crime threat assessment report IOCTA. https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment, 2018. Retrieved March 6, 2019.

[5] F. Reid, M. H. An analysis of anonymity in the bitcoin system. In *2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing* (2011).

[6] Filtz, E., Polleres, A., Karl, R., and Haslhofer, B. Evolution of the bitcoin address graph. In *Data Science–Analytics and Applications*. Springer, 2017, pp. 77–82.

[7] Gutmann, P. The commercial malware industry. DEF CON conference, https://www.cs.auckland.ac.nz/~pgut001/pubs/malware_biz.pdf, 2007. Retrieved February 2, 2019.

[8] Harrigan, M., and Fretter, C. The unreasonable effectiveness of address clustering. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)* (2016), IEEE, pp. 368–373.

[9] Haslhofer, B., Karl, R., and Filtz, E. O bitcoin where art thou? insight into large-scale transaction graphs. In *SEMANTiCS (Posters, Demos, SuCCESS)* (2016).

[10] Hernandez-Castro, J., Cartwright, E., and Stepanova, A. Economic analysis of ransomware. Available at SSRN 2937641,, 2017. Retrieved March 6, 2019.

[11] Heydari, A., ali Tavakoli, M., Salim, N., and Heydari, Z. Detection of review spam: A survey. *Expert Systems with Applications 42*, 7 (2015), 3634–3642.

[12] Huang, D. Y., Aliapoulios, M. M., Li, V. G., Invernizzi, L., Bursztein, E., McRoberts, K., Levin, J., Levchenko, K., Snoeren, A. C., and McCoy, D. Tracking ransomware end-to-end. In *2018 IEEE Symposium on Security and Privacy (SP)* (2018), IEEE, pp. 618–631.

[13] IBM-X-Force-Exchange. Necurs delivers language targeted porn scams. IBM, https://exchange.xforce.ibmcloud.com/collection/Necurs-delivers-language-targeted-porn-scams-fdb9d6b7941506807cbe56dd06e142d0, Sept 2018. Retrieved February 7, 2019.

[14] John, J. P., Moshchuk, A., Gribble, S. D., Krishnamurthy, A., et al. Studying spamming botnets using botlab. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)* (2009), vol. 9.

[15] Kalodner, H. A., Goldfeder, S., Chator, A., Möser, M., and Narayanan, A. Blocksci: Design and applications of a blockchain analysis platform. Retrieved March 6, 2019.

[16] Kanich, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G. M., Paxson, V., and Savage, S. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM conference on Computer and communications security* (2008), ACM, pp. 3–14.

[17] Kanich, C., Weaver, N., McCoy, D., Halvorson, T., Kreibich, C., Levchenko, K., Paxson, V., Voelker, G. M., and Savage, S. Show me the money: Characterizing spam-advertised revenue. In *USENIX Security Symposium* (2011), pp. 15–15.

[18] Kessem, L. The necurs botnet: A pandora's box of malicious spam. IBM, https://securityintelligence.com/the-necurs-botnet-a-pandoras-box-of-malicious-spam/, April 2017. Retrieved March 6, 2019.

[19] Levchenko, K., Pitsillidis, A., Chachra, N., Enright, B., Félegyházi, M., Grier, C., Halvorson, T., Kanich, C., Kreibich, C., Liu, H., et al. Click trajectories: End-to-end analysis of the spam value chain. In *2011 ieee symposium on security and privacy* (2011), IEEE, pp. 431–446.

[20] Lumley, T., Diehr, P., Emerson, S., and Chen, L. The importance of the normality assumption in large public health data sets. *Annual review of public health 23*, 1 (2002), 151–169.

[21] Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., and Savage, S. A fistful of bitcoins: Characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference* (2013), ACM, pp. 127–140.

[22] Monaco, J. V. Identifying bitcoin users by transaction behavior. In *Biometric and Surveillance Technology for Human and Activity Identification XII* (2015), vol. 9457, International Society for Optics and Photonics, p. 945704.

[23] Möser, M., and Böhme, R. Join me on a market for anonymity. In *Proceedings of the Workshop on the Economics of Information Security (WEIS)* (University of California at Berkeley, 2016). Retrieved March 7, 2019.

[24] Möser, M., Böhme, R., and Breuker, D. An inquiry into money laundering tools in the bitcoin ecosystem. In *2013 APWG eCrime Researchers Summit* (2013), IEEE, pp. 1–14. Retrieved March 20, 2019.

[25] Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf, 2008. Retrieved January 4, 2019.

[26] Nick, J. D. Data-driven de-anonymization in bitcoin. Master's thesis, ETH-Zürich, 2015. Retrieved March 11, 2019.

[27] Overall, J. E., Atlas, R. S., and Gibson, J. M. Tests that are robust against variance heterogeneity in k× 2 designs with unequal cell frequencies. *Psychological reports 76*, 3 (1995), 1011–1017.

[28] Paquet-Clouston, M., Haslhofer, B., and Dupont, B. Ransomware payments in the bitcoin ecosystem. *Journal of Cybersecurity 5*, 1 (2019), tyz003.

[29] Rao, J. M., and Reiley, D. H. The economics of spam. *Journal of Economic Perspectives 26*, 3 (2012), 87–110.

[30] Reid, F., and Harrigan, M. An analysis of anonymity in the bitcoin system. In *Security and Privacy in Social Networks*. Springer, 2013, pp. 197–223.

[31] Ron, D., and Shamir, A. Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security* (2013), Springer, pp. 6–24. Retrieved February 15, 2019.

[32] Schultz, J. Anatomy of a sextortion scam. Cisco Talos Intelligence, https://blog.talosintelligence.com/2018/10/anatomy-of-sextortion-scam.html, Oct 2018. Retrieved January 17, 2019.

[33] Sedhai, S., and Sun, A. Hspam14: A collection of 14 million tweets for hashtag-oriented spam research. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2015), ACM, pp. 223–232.

[34] Spagnuolo, M., Maggi, F., and Zanero, S. Bitiodine: Extracting intelligence from the bitcoin network. In *International Conference on Financial Cryptography and Data Security* (2014), Springer, pp. 457–468.

[35] Stone-Gross, B., Holz, T., Stringhini, G., and Vigna, G. The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns. *LEET 11* (2011), 4–4.

[36] Stringhini, G., Hohlfeld, O., Kruegel, C., and Vigna, G. The harvester, the botmaster, and the spammer: on the relations between the different actors in the spam landscape. In *Proceedings of the 9th ACM symposium on Information,*

computer and communications security (2014), ACM, pp. 353–364.

[37] Tu, H., Doupé, A., Zhao, Z., and Ahn, G.-J. Sok: Everyone hates robocalls: A survey of techniques against telephone spam. In *2016 IEEE Symposium on Security and Privacy (SP)* (2016), IEEE, pp. 320–338.

[38] Zimmerman, D. W., and Zumbo, B. D. Rank transformations and the power of the student t test and welch t'test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale 47*, 3 (1993), 523.