Malware Finances and Operations: a Data-Driven Study of the Value Chain for Infections and Compromised Access

Juha Nurmi
Tampere University
Tampere, Finland
juha.nurmi@tuni.fi
juha@cyberintelligencehouse.com

Mikko Niemelä
Cyber Intelligence House
Singapore, Singapore
mikko@cyberintelligencehouse.com

Billy Bob Brumley
Tampere University
Tampere, Finland
billy.brumley@tuni.fi

ABSTRACT

We investigate the criminal market dynamics of infostealer malware and publish three evidence datasets on malware infections and trade. We justify the value chain between illicit enterprises using the datasets, compare the prices and added value, and use the value chain to identify the most effective countermeasures.

We begin by examining infostealer malware victim logs shared by actors on hacking forums, and extract victim information and mask sensitive data to protect privacy. We find access to these same victims for sale at Genesis Market. This technically sophisticated marketplace provides its own browser to access victim's online accounts. We collect a second dataset and discover that 91% of prices fall between 1–20 US dollars, with a median of 5 US dollars.

Database Market sells access to compromised online accounts. We produce yet another dataset, finding 91% of prices fall between 1–30 US dollars, with a median of 7 US dollars.

CCS CONCEPTS

• **Security and privacy** → **Malware and its mitigation**; *Multifactor authentication*; *Authorization*; **Web application security**; **Economics of security and privacy**.

KEYWORDS

Anti-virus Software; Computer Hacking; Malicious Program Traits; Intrusion Detection

1 INTRODUCTION

Malicious software, or malware, is a persistent executable on a victim system controlled by an attacker [24]. Infostealer malware steals personal information from the victim [26]. This typically includes (but is not limited to) system data, login credentials, browser cookies, payment card information, and cryptocurrency wallets [10]. Existing literature classifies malware in a tree-like structure, and there are distinct families of infostealer malware [13, 21, 34, 46].

The five most popular infostealers are Taurus, Raccoon, Azorult, Vidar, and RedLine [19]. We also found active Kpot and Arkei malware family victims, and all of them are available for sale [22]. After purchasing malware control software, the adversary convinces victims to download malicious code and become infected. Infostealers communicate with a Command & Control (C&C) Server, which remotely controls the victim's computer [22, 46].

Protection software cannot detect all intrusions by malware [4, 28], because code obfuscation methods hide the infection [7]. Predator the Thief, for example, is an infostealer available on hacking forums, and is constantly updated to include a variety of mechanisms to evade detection [15]. Updated stealth mechanisms make detection difficult for anti-virus software [14, 23].

Unethical software developers provide infostealer control software via Malware-as-a-Service (MaaS) [1], which is a form of Cybercrime-as-a-Service (CaaS). In 2022, prices range from 75 US dollars per week, to 125–200 US dollars per month, to 1,000 US dollars for an unlimited lifetime subscription for improved versions of RedLine [1, 44]. Subscribing threat actors gain access to an administration panel enabling them to customise the malware, retrieve stolen data, and create new malware builds [1].

Malicious actors infect victims with infostealer malware using (most frequently) phishing emails, cracked and pirated software, game cheating packages, browser extensions, and cryptocurrency-related software [10, 20]. In 2022, for example, one infostealer spread via spam emails that included an Excel spreadsheet containing macros that downloaded a persistent executable [20].

The malware network owner profits from victim data by gaining access to the compromised services or by selling the victim data to other criminals who profit from it [21, 46]. A study from 2016 indicates, "Cybercriminals steal access credentials to webmail accounts and then misuse them for their own profit, release them publicly, or sell them on the underground market" [33]. For example, between January 2021 and March 2022, one particular marketplace sold 4,368,909 units of victim data [19].

Research questions. We propose the following research questions related to the previously described actors in the malware ecosystem.

RQ1. How do criminal enterprises sell malware, victims, and online account access?

RQ2. How are the prices determined for compromised devices?

RQ3. Compared to access to a victim's device, what is the cost of online account access?

Contributions. These actors form illicit markets, and we study them (in part) using economics nomenclature: "variables that can be observed in, or derived directly from, markets (for example, prices of publicly traded securities and interest rates)" (A926 IFRS 17 Insurance Contracts manual[1]). Our interests resemble those of economics literature: "The interest in market factors is threefold: (1) to identify the factors that influence a product's or service's demand, (2) determine the relationship between the factor and the product or service, and (3) forecast that market factor for future years." [48, p. 52].

Contribution 1: Release data describing the value chain. For our research, we collect, curate, and model three datasets, which we publish under the CC BY 4.0 licence. We represent each victim using a single JavaScript Object Notation (JSON) data file. Data sources provide sets of victim JSON data files from which we extract the essential information and omit Personally Identifiable Information (PII).

MalwareInfectionSet: Underground hacking forums for malware logs. RaidForums[2] operated as a forum for unethical hackers from 2015 until 2022, when the FBI seized the domain. The website was distributing data breaches and malware tools. We use RaidForums as a starting point to track criminal profiles, and then add Telegram and Discord channels based on their posts. This enables us to follow the content criminals share in a broader fashion.

We discover (and, to the best of our knowledge, document scientifically for the first time) that malware networks appear to dump their data collections online. We speculate this could possibly be due to one or more of the following factors. (i) They exploited the victims and hid their tracks by attracting others to exploit the victims on a large scale as well. For example, for at least a decade on pastebin.com, cybercriminals have released publicly accessible lists to compromised targets [27]. (ii) By offering free example data, they showcase their skills and specialisation. For example, spammer botnets distribute free samples and display their capabilities in this manner [33, 42]. (iii) A rival adversary leaked information to harm the operations of another cybercriminal; see [2] for an example case from 2019.

In this study, we utilise 245 malware log dumps from 2019 and 2020 originating from 14 malware networks. MalwareInfectionSet contains 1.8 million victim files, with a dataset size of 15 GB.

Contribution 2: Sale of compromised devices. To answer RQ1, we demonstrate how Infostealer malware networks sell access to infected victims.

VictimAccessSet: Genesis Market for compromised access. Genesis Market[3] is a cybercriminal marketplace specialising in the sale of web browser data from malware victims (see "Buying Bad Bots Wholesale: The Genesis Market" report in 2021[4]). The marketplace operates within the anonymous Tor network but also through the clear web; the `genesis.market` domain was accessible until April 2023, before the FBI seized it.

Genesis Market focuses on user-friendliness and continuous supply of compromised data. Marketplace listings include everything necessary to gain access to the victim's online accounts, including passwords and usernames, and also detailed information which provides a clone of the victim's browser session. Indeed, Genesis Market simplifies the import of compromised victim authentication data into a web browser session.

To utilise the stolen data in a user-friendly way, Genesis Market offers both a web browser plugin and a dedicated web browser: the Genesis Security Plugin, a standard browser plugin, or the Genesium Browser, a Chromium-based browser built specifically for Genesis Market users, featuring the Genesis Security Plugin pre-installed. Once a customer purchases a victim's complete access information, browser technology enables the customer to access the victim's stolen online profiles and authenticated sessions.

Contribution 3: Victim prices. To answer RQ2, we measure the prices on Genesis Market and how compromised device prices are determined.

To build our dataset VictimAccessSet, we crawled the website between April 2019 and May 2022. We use the collected web pages offering the resources for sale to investigate market factors – in particular those that influence the sale price. VictimAccessSet contains 0.5 million victim files, with a dataset size of 3.5 GB.

AccountAccessSet: Database Market. Database Market[5] specialises in selling credentials to compromised online accounts. The marketplace operates inside the anonymous Tor network. Vendors offer their goods for sale, and customers can purchase them with Bitcoins. The marketplace sells online accounts, such as PayPal and Spotify, as well as private datasets, such as driver's licence photographs and tax forms.

To build our dataset AccountAccessSet, we crawled the website between November 2021 and June 2022. AccountAccessSet contains 33,896 victim files, with a dataset size of 400 MB.

Contribution 4: Account prices. To answer RQ3, we measure the prices on Database Market and illustrate an increase in value: an attacker with access to a single victim can sell access to individual accounts at a higher price.

Contribution 5: Value chain. The contributions reveal a broader perspective: we draw the value chain for the malware economy from our findings and related prices in the literature.

Structure. Section 2 discusses previous malware-related work in the context of CaaS, including economic aspects. Section 3 presents our methods, and Section 4 shares the three datasets and provides a step-by-step tutorial to allow anyone to reproduce our findings. Consequently, Section 5 provides our primary contributions: we investigate the prices and the value chain between illicit enterprises. Section 6 establishes coverage metrics to measure the quality and extensiveness of our datasets. Section 7 discusses ethical and privacy considerations we account for in our work. We conclude our research in Section 8.

---

[1]https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/ifrs-17-insurance-contracts.pdf
[2]https://raidforums.com/
[3]https://genesis7zoveavupiiwnrycmaq6uro3kn5h2be3el7wdnbjti2ln2wid.onion/
[4]https://www.netacea.com/wp-content/uploads/2021/04/Genesis_market_report_2021.pdf

[5]http://database6e2t4yvdsrbw3qq6votzyfzspaso7sjga2tchx6tov23nsid.onion/

2 RELATED WORK

Cybercrime economics. In prior research, the cybercrime value chain has been established in *Framing Dependencies Introduced by Underground Commoditization* [43], however this value chain is justified with circumstantial evidence and the data are not interconnected, as the research points out: "the future of black market research lays in developing a better understanding of the tenuous connections between the actors in this space … we note that research into many of these profit centers is sorely lacking: we rely on industry and government estimates of profit in the absence of methodologically sound measurements … research contributions must shift from analysis based off anecdotal evidence to thorough investigations of the ecosystem". In our study, we address this gap in the literature and we establish robust connections between the malware networks and the illicit markets.

Already in 2011, Stone-Gross et al. [42] identify financial variables and the specialisation of cybercriminals. The authors describe how botnet operators sell their services, while other vendors only sell email address lists: "spam-as-a-service can be purchased for approximately $100–$500 per million emails sent" and "Rates for one million email addresses range from $25 to $50".

Broadhurst et al. [8] study MaaS offerings on the illicit Dream Market from 2017 to 2018. During this period, they calculate that 47 AUD is the average price for a single set of compromised credentials.

Huang et al. [25] propose a conceptual model describing CaaS. The authors identify an impressive 24 families of CaaS, including pricing models and estimated costs. Our dataset MalwareInfectionSet falls under their categorisation Personal Profile as a Service (PPaaS), for which they provide a representative price of 4–20 US dollars per record in 2018 [25, Tbl. 1].

Meland and Sindre [29] study 11 dark web illicit markets offering CaaS. They explicitly omit compromised credentials from their study, since their focus is on the sale of offensive services [29, Sect. 3].

Bhalerao et al. [6] study CaaS using machine learning techniques to perform supply chain detection. They validate their techniques using the same underground forum as Vu et al. [47].

Lastly, Campobasso and Allodi [11] crawl the Russian cybercrime market that provides user impersonation profiles and estimate a daily trade volume of up to 700 profiles, equating to daily sales of up to 4,000 US dollars.

Security dataset classification. Zheng et al. [49] study security-related datasets to provide a taxonomy, as well as to understand the scholarly ramifications of releasing said datasets. In their classification, our datasets falls under the "Attacker-Related" category, and furthermore the "cybercrime activities" subcategory, since they contain "information on the infrastructure and operations used by malicious actors to perpetrate attacks" [49, Sect. 3]. While the authors give other examples from this class including data collected from illicit marketplaces inside the Tor network [12, 17, 18, 32], ours is the first study dedicated to the malware ecosystem.

Gathering cybercrime data. Motoyama et al. [31] study data from six underground forums dedicated to the sale of illicit goods and services. They focus on social network aspects, including connectivity, reputation, and community policies. Furthermore, they study

these marketplaces to determine how the different social network aspects affect market performance. One of these forums (Carders, German speaking) specialised in stolen credit card numbers, yet interestingly roughly 3% (334/9923) of the sale listings were victim logs [31, Tbl. 5]. The original data source for their study was leaked SQL databases [31, Sect. 3], sharing conceptual similarities with the original data source for our MalwareInfectionSet.

Schäfer et al. [39] build a tool called BlackWidow to monitor illicit forums. The tool is also capable of machine learning-based information extraction. The authors apply their tool to seven different forums, subsequently analysing the cross-forum social network aspects.

Pastrana et al. [36] propose CrimeBot, a framework for scalable data collection from illicit online forums. The authors apply CrimeBot to produce their CrimeBB dataset [36, Sect. 4], containing 48 million posts and their associated metadata. CrimeBB continues to grow, now approaching 100 million posts in 2022 [37].

Cybercrime data creates new research. As CrimeBB grows, it increasingly fosters academic studies for those with access to the data through a legal agreement with the University of Cambridge.[6] The following CrimeBB usage examples are relevant to our research and motivate us to publish datasets for the benefit of other researchers.
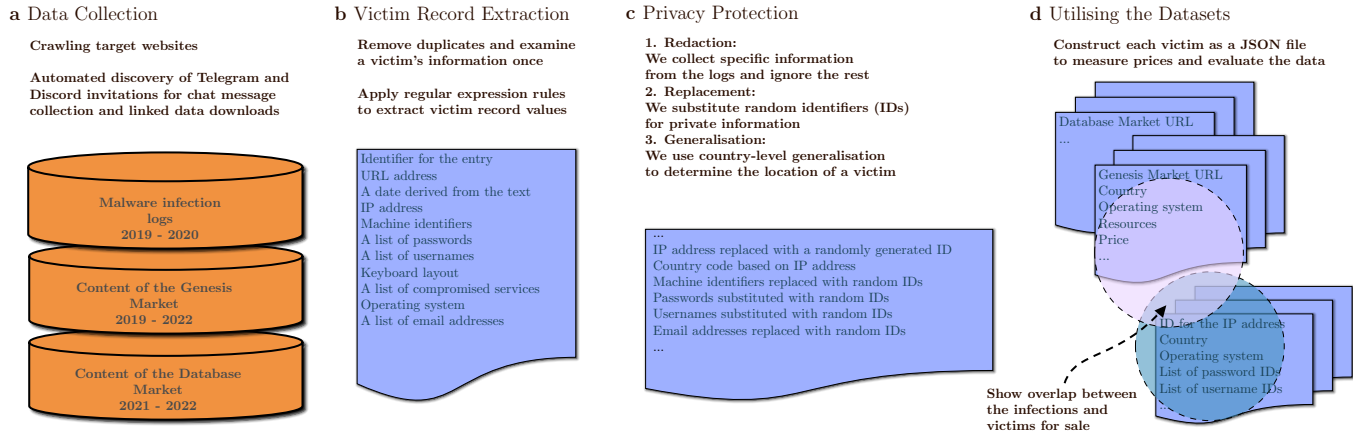
Pastrana et al. [35] utilise CrimeBB data to characterise the key individuals that are likely to engage in major cybercriminal activities, aiding law enforcement. In a similar vein with a focus on individuals, Pete et al. [38] analyse social network aspects of six dark web forums and use graph-theoretic similarities to identify properties in individual users that lead to these actors taking different forum roles. In a CrimeBB followup study, Collier et al. [16] investigate cybercrime infrastructure work and show that botnet administration tasks are a burden and a weak point in the cybercrime ecosystem, suitable for law enforcement targeting. Vu et al. [47] analyse the reputation properties in one particular illicit marketplace, and Bermudez-Villalva and Stringhini [5] compare dark web vs. deep web markets.

Lastly, and most relevant to our study, Akyazi et al. [3] use the CrimeBB data to analyse various CaaS offerings on one particular illicit marketplace. The vast majority of these offerings were related to botnets, reputation boosting, and traffic flooding.

3 METHODS

Figure 1 provides a summary of our automated malware ecosystem monitoring: web crawling, chat message collection, and linked resource downloads. We process the raw data into victim record files while protecting the user's privacy. The privacy protection measures replace the same input with the same ID. For instance, throughout the dataset, the same IP address has the same ID. Similarly, if there are passwords for services but the passwords are identical, the replacement IDs are identical. We curated together datasets to measure malware infections, the sale of access to victims, and the reselling of online accounts (see Section 4 on how to download the datasets and scripts and replicate the results).

---

[6]https://www.cambridgecybercrime.uk/process.html

a Data Collection

Crawling target websites

Automated discovery of Telegram and Discord invitations for chat message collection and linked data downloads

Malware infection logs
2019 - 2020

Content of the Genesis Market
2019 - 2022

Content of the Database Market
2021 - 2022

b Victim Record Extraction

Remove duplicates and examine a victim's information once

Apply regular expression rules to extract victim record values

Identifier for the entry
URL address
A date derived from the text
IP address
Machine identifiers
A list of passwords
A list of usernames
Keyboard layout
A list of compromised services
Operating system
A list of email addresses

c Privacy Protection

1. Redaction:
We collect specific information from the logs and ignore the rest
2. Replacement:
We substitute random identifiers (IDs) for private information
3. Generalisation:
We use country-level generalisation to determine the location of a victim

...
IP address replaced with a randomly generated ID
Country code based on IP address
Machine identifiers replaced with random IDs
Passwords substituted with random IDs
Usernames substituted with random IDs
Email addresses replaced with random IDs
...

d Utilising the Datasets

Construct each victim as a JSON file to measure prices and evaluate the data

Database Market URL
...

Genesis Market URL
Country
Operating system
Resources
Price
...

ID for the IP address
Country
Operating system
List of password IDs
List of username IDs

Show overlap between the infections and victims for sale

Figure 1: (a) Our crawlers authenticate on RaidForums, Genesis Market, and Database Market for data collection. Our bots follow messages by joining the Discord and Telegram channels mentioned on RaidForums. They download linked resources, including malware infection log files. (b) To extract victim records from the data sources, we use regular expressions. (c) We apply redaction, replacement, and generalisation to omit PII. (d) Finally, we create JSON victim record sets for pricing measurements, demonstrate overlap between infections and for-sale victims, and evaluate our data.

3.1 Data extraction for MalwareInfectionSet

Our web crawler authenticates with RaidForums in order to gather pages and extract Discord and Telegram invitations. Then, our bots join channels on Discord and Telegram to monitor messages. The bots visit the shared links and download resources, including log files containing malware (if this is not feasible, the resources are reported for manual human involvement).

We utilise 245 malware log dumps from 2019 and 2020 originating from AZORult, KPOT, Raccoon, Qulab, Masad, Arkei, Oski, Taurus, Redline, Mystery, Ficker, Predator, Ducky, and Vidar malware networks. The total size of these original malware log files is 30 GB. Figure 2 is an example of a victim record file.

The malware logs include system information, login credentials to online services, user information, credit card numbers, screenshots of victim desktops, operating system, keyboard languages, IP address, geolocation, hardware specifications, process list of active software, browser information and cookies, cryptocurrency wallets, victim files, automatically filled site forms, gaming software configurations, chat service configurations, clipboard data, installed software, browsing and download history, and mail, VPN, RDP, and FTP account information.

We structured and cleaned the data, aggregating the following fields from the logs to build our dataset MalwareInfectionSet: (1) Identifier for the entry; (2) A date derived from the logs; (3) IP address replaced with a randomly generated ID; (4) Machine identifiers replaced with random IDs; (5) A list of passwords we substituted with random IDs; (6) Total number of one-of-a-kind passwords; (7) A list of usernames we substituted with random IDs; (8) Total number of distinct usernames; (9) All domain names in the log, including browser history; (10) Keyboard layout; (11) A list of compromised services; (12) Count of compromised services; (13) Country code which is primarily based on IP address; (14) Operating system; (15) A list of email addresses replaced with random IDs;

```
1  {"id": 12345,
2   "timestamp" : "2018-10-07",
3   "device_ip_addr" : "46609a9930",
4   "machineids" : ["c8ce18fc03", "7c299b41c2"],
5   "passwords" : ["61c321014", "c7dc63658e", "c7dc63658e"],
6   "passwords_count" : 2,
7   "usernames" : ["5006ff6e3", "dcbefb2b82", "dcbefb2b82"],
8   "usernames_count" : 2,
9   "domain_names" : ["battle.net", "gmail.com"],
10  "keyboard" : "uk-UA",
11  "services" : ["https://gmail.com", "https://battle.net"],
12  "services_count" : 2,
13  "country" : "UA",
14  "os" : "Windows 10 Pro [x64]",
15  "email_addresses" : ["dcbefb2b82"],
16  "email_addresses_count" : 1,
17  "antivirus": ["mcafee virusscan", "microsoft security"]}
```

Figure 2: An example JSON file from MalwareInfectionSet depicting victim data.

(16) Total number of unique email addresses; (17) Possible endpoint protections.

We remove email addresses, IP addresses, and payment card numbers from raw text fields, such as those used to list compromised services. This means that service count (number of distinct services) may exceed the total number of services listed. In that case, we remove services which contain private information from the list.

3.2 Data extraction for VictimAccessSet

We crawled the Genesis Market between April 2019 and May 2022, collecting the web pages offering the resources for sale. Figure 3 is an example of a victim information page.

Since the victim page is continuously updated, we use the oldest page as the reference entry (or entry of record) in VictimAccessSet. As such, Figure 4 is an example of a file. This record corresponds
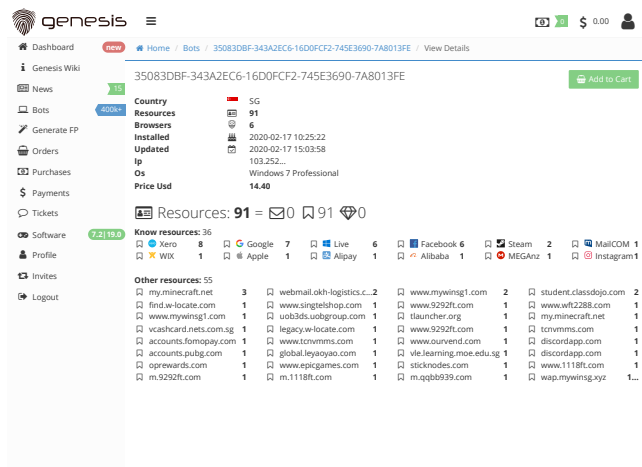
Figure 3: The infostealer malware adds compromised data to the Genesis Market for sale.

```
1  {"bot": "35083DBF-343A2EC6-16D0FCF2-745E3690-7A8013FE",
2    "url":
   ↪    "https://genesis.market/client/bots/view?id=245092557",
3    "text": "35083DBF- ... Add to Cart __ Reserve __ Buy\n\n",
4    "domain_names": ["m.qqbb939.com", "uob3ds.uobgroup.com",
   ↪    "wft2288.com", "...", "m.9292ft.com", "m.1118ft.com"],
5    "country": "SG",
6    "resources": 91,
7    "installed": "2020-02-17",
8    "updated": "2020-02-17",
9    "ip": "103.252",
10   "os": "Windows 7 Professional",
11   "price": 15.4}
```

Figure 4: An example JSON file from VictimAccessSet which represents one victim.

to the page in Figure 3, however, due to different capture dates, the price has increased from 14.40 to 15.50 US dollars.

We consolidated the following fields to build our dataset VictimAccessSet: (1) Bot identifier from the webpage; (2) URL address; (3) Text of the webpage; (4) All domain names on the webpage; (5) Country code; (6) Total number of resources; (7) The malware installation date; (8) Date of victim information update; (9) Partial IP address; (10) Operating system; and (11) Price of the access.

3.3 Data extraction for AccountAccessSet

We crawled the website between November 2021 and June 2022, collecting the web pages offering the credentials for sale. Figure 5 is an example of a vendor advertising compromised credentials for sale on Database Market.

Figure 6 is an example of a file, where we populated the following fields: (1) URL address of the webpage; (2) Text of the webpage; and (3) Data collection date.

Because vendors can style advertisements in any way they like, it is not possible to accurately collect price and sale item information across the full AccountAccessSet in an automated way. We manually inspected a random sample of 599 out of 33,896 pages

Figure 5: Compromised data for sale on Database market.

```
1  {"url": "http://database6...sid.onion/products/1573",
2    "text": "DATABASE DATABASE\n\n__ Shop ...",
3    "timestamp": "2022-05-31"}
```

Figure 6: An example JSON file from AccountAccessSet which represents one page.

on Database Market. We carefully verified that our chosen subset contains individual accounts with price information and no other products for sale, such as merged accounts, company databases, or stolen software. Then, we created an extraction script for these 599 pages to evaluate the prices.

3.4 Relationship between malware infections and victims for sale on Genesis Market

To illustrate the value chain, we compare the malware infections with the records for sale on Genesis Market by cross-referencing MalwareInfectionSet with VictimAccessSet. We are identifying unique users based on service and country data, then comparing their presence in both datasets.

For instance, the victim information for sale on Genesis Market, present in VictimAccessSet (ID=245092557, 245092557.json) and depicted in Figure 4, has details matching a record from MalwareInfectionSet (ID=487238, 0487238.json). This Singaporean victim has the following rare compromised services: m.qqbb939.com, either.io, wft2288.com, m.9292ft.com, m.1118ft.com, uob3ds.uobgroup-.com, legacy.w-locate.com, and tcnvmms.com.

The matching is as follows: (i) select the victims that have at least five completely unique services based on the domain names; (ii) calculate the intersection of three or more of these unique domain names between the datasets; and (iii) check that the countries are identical.

The matching criteria identify users who are unique based on the services they use (they use uncommon services that other users do not) and who are present in both datasets. As a result, the tight matching criterion finds only the most unique victims in both datasets, not all of the victims in both datasets.

4 DATASETS TO REPLICATE OUR RESULTS

Using our well-documented scripts, it is straightforward to reproduce our findings. You can download our datasets, Python scripts, and tutorials on the Zenodo service for Open Science[7]. It takes an estimated 1 hour of human time and 3 hours of computing time to duplicate our key findings from MalwareInfectionSet (malwarevictims.tar.gz, 1.7 GB); around one hour with VictimAccessSet (genesisvictims.tar.gz, 335 MB); and minutes to replicate the price calculations using AccountAccessSet (database.tar.gz, 7 MB).

5 RESULTS

RQ1: How do criminal enterprises sell malware, victims, and online account access?

We discovered information-stealing malware logs from victims of the Taurus, Raccoon, Azorult, Vidar, RedLine, Kpot, and Arkei families. These MaaS controlling software are for sale on underground markets with prices ranging from 75 to 200 US dollars per week and 125 to 200 US dollars per month. After acquiring an MaaS subscription, the adversary convinces victims to download malicious code, causing them to become infected.

5.1 Malware targets Windows

The infostealers target users of Microsoft's Windows operating system. Windows 10, Windows 7, and Windows 8 are the main infected operating systems according to MalwareInfectionSet and VictimAccessSet. Victims in VictimAccessSet employ one of the following Windows operating systems: 76% (365,676/480,994) use Windows 10, 7% (32,989/480,994) use Windows 8, and 16% (77,170/480,994) use Windows 7. We extracted MalwareInfectionSet from heterogeneous text data logs in which the operating system information is sometimes absent, incorrect, or described in multiple ways; nonetheless, if we search os:NT* or os:Windows*, we obtain 62% (1,115,263/1,809,384) of the infection victims.

Stealth mechanisms are making malware difficult to detect for anti-virus software. According to MalwareInfectionSet, some users have installed popular anti-virus software: Avast, Mcafee, Symantec, Microsoft Security Essentials, Bitdefender, Microsoft/Windows Defender, F-secure anti-virus, Zonealarm anti-virus, and Kaspersky anti-virus. The infection occurred, regardless.
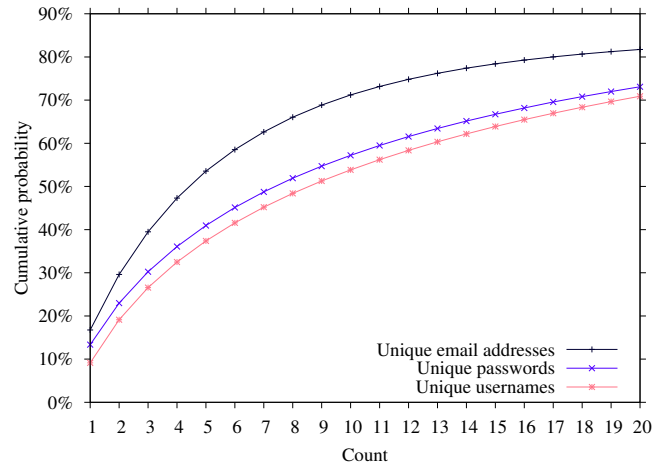
5.2 Credentials and password reuse

Password reuse is a common problem for security: We study credential reuse with MalwareInfectionSet, and find that 50% of victims have 1–8 unique passwords.

Furthermore, Figure 7 also shows that the majority of victims have few distinct usernames. In total, 9% of victims have a single unique username across services, while 10% have two. The majority of victims have between 1–9 usernames and between 1–5 unique email addresses.

If we perform linear regression on the number of unique usernames and unique passwords across all victims, we discover that a typical user has 1.12 times as many unique usernames as unique passwords:

$$Usernames = 1.12 * Passwords + 1.78$$

Figure 7: Cumulation of unique passwords, usernames, and email addresses for each victim in MalwareInfectionSet.

Investigating the correlation between the number of unique credentials and compromised services similarly with linear regression, we obtain:

$$EmailAddresses = 0.11 * Services + 3.67$$

$$Usernames = 0.36 * Services + 6.19$$

$$Passwords = 0.31 * Services + 4.43$$

This means that a typical user creates a new unique password and username for every third service.

5.3 Commonly compromised services

The top 10 services that are most frequently compromised in MalwareInfectionSet are: google.com, facebook.com, live.com, roblox.com, twitter.com, discord.com, twitch.tv, epicgames.com, instagram.com, and paypal.com.

Because each of these is a well-known web service (c.f. Similarweb's "Most Visited Websites Ranking Analysis"[8]), the results are consistent with expectations.
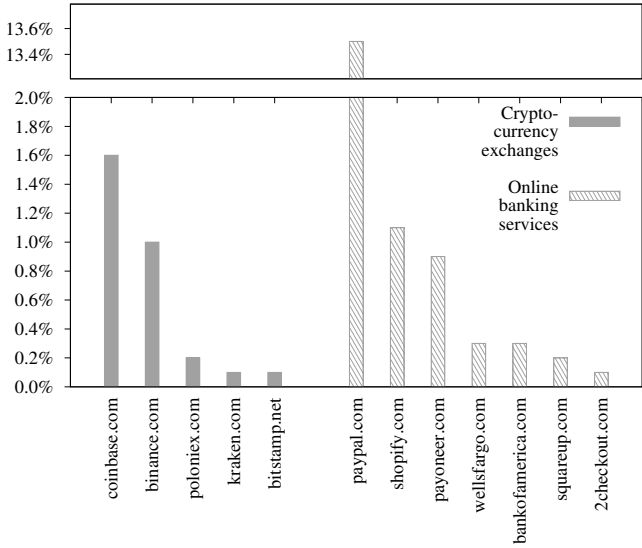
5.4 Compromised financial services

The victims used the financial services listed in Figure 8, consisting of online banking services and cryptocurrency exchanges. In total, 13.5% of the victims in MalwareInfectionSet have their PayPal online payment account compromised.

5.5 Victim records for sale

Adversaries may attempt to profit from stolen credentials, e.g., by the owner of the bot network using them, or the adversary selling the access information forward. We now provide evidence of the latter with our matching criteria described in Section 3 (methods).

We first automatically extract and identify unique domain names (compromised services) from VictimAccessSet, and select records

---

[7]https://doi.org/10.5281/zenodo.8047204

[8]https://www.similarweb.com/top-websites/

Figure 8: Financial services in MalwareInfectionSet: cryptocurrency exchanges and online banking. 13.5% of the victims have their PayPal online payment account compromised.

Figure 9: Price of access to compromised devices (VictimAccessSet) and accounts (AccountAccessSet). In the former, 91% fall between 1 and 20 US dollars, with a median of 5 US dollars. In the latter, 91% fall between 1 and 30 US dollars, with a median of 7 US dollars.

that contain at least five of these distinct domain names. This selection yields 2,116 records from VictimAccessSet. We then search for matches from MalwareInfectionSet.

As an example of our selection criteria, consider 1502766.json from MalwareInfectionSet and 1814801285.json from VictimAccessSet. Here, countries (IT and IT) match, and both records contain a total of 11 distinct services: gravel.ltd, bitminer.biz, shamining.com, coincity.in, navminer.biz, luckfarm.fun, ethereum-master.com, aristabank.com, flovin.ltd, getdoge.io, and xcryptos.io.

As a result, we discover that 5% (106/2,116) of the malware-infected victims are present in both datasets. Please note that this method only provides evidence that infostealer malware victims are available for purchase on Genesis Market; the exact percentage must be higher than this. This unequivocally establishes a link in the supply chain: Infostealer malware infections are a supplier for cybercriminal marketplaces that sell access to victim data.

### 5.6 Access to a malware victim: ask prices

RQ2: How are the prices determined for compromised devices?

Genesis Market sells access to victim information, with stolen credentials from victims serving as the primary pricing factor. Addressing RQ2, we discover that the number of compromised services seems to dictate the ask price. Using linear regression to determine the price-to-services correlation, we obtain:

$$Price = 0.05 * Services + 6.26$$

Interestingly, there is no effect of country on pricing. We determined this by creating separate linear regression models for each nation, then comparing them.

Finally, we examine price distribution to conclude our study of RQ2. Rounding the prices to the nearest dollar, the range of prices is from 1 to 350 US dollars, and prices over 125 US dollars each have
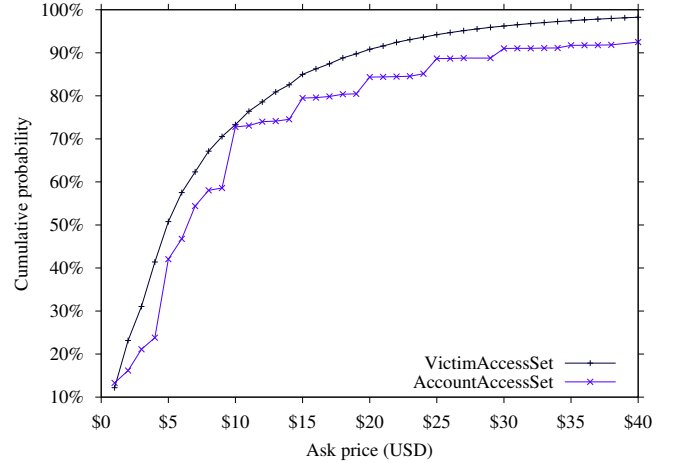
fewer than 10 victims. Figure 9 illustrates these ask prices, where 91% fall between 1 and 20 US dollars, with a median of 5 US dollars.

5.7 Access to an online account: ask prices

RQ3: Compared to access to a victim's device, what is the cost of online account access?

AccountAccessSet reveals the prices of compromised online account access, where we chose a subset of 599 account advertisement pages. We manually confirmed these contain individual accounts with price information, and no other items for sale, such as accounts combined into a single item, organisation databases, or stolen software.

Figure 9 illustrates these ask prices, where 91% fall between 1 and 30 US dollars, with a median of 7 US dollars. This data provides quantitative estimates for the costs of individual online accounts, addressing RQ3.

In summary, after purchasing access to a victim's device for 1–20 US dollars, a cybercriminal is able to verify working online accounts of the victim and resell the harvested online accounts for 1–30 US dollars per account.
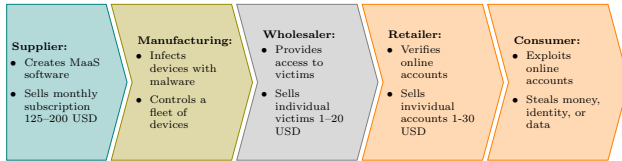
5.8 Malware supply chain

Malware suppliers, infections, victim sales, and online account sales form the value chain in Figure 10.

Malware victim and account access are both perishable goods, so the manufacturer must sell them before they grow stale. To sell victims, the manufacturer needs economies of scale (roughly, an economics concept formalising cost advantages) and a streamlined production process. In Table 1, we display the known costs and projected earnings for the operations.

Cause interruption. Victim wholesale is precisely what Genesis Market offers. Wholesale is the single weakest link in the chain:

Table 1: Prices and profitability in the malware economy value chain. Many goods, such as licences, have fixed costs and zero per-unit production costs at scale. The average number of compromised services for victims in the Genesis Market is 48.

Actor	Price (US dollars)	Pricing model	Monthly profit is revenues minus total costs
Malware-as-a-Service	125–200 / monthly license	Monthly subscription	$\sum Pay\text{-}per\text{-}subscriptions$
Spam-as-a-Service	100–500 / million sent emails	Pay-per-million	$\sum Pay\text{-}per\text{-}million\text{-}emails$
Email Target List Supplier	25–50 / million addresses	Pay-per-million	$\sum Pay\text{-}per\text{-}million\text{-}addresses$
Infection Manufacturing	Bulk sale of victims	Pay-per-victim-access	$\sum (Pay\text{-}per\text{-}victim\text{-}access - comission)$
Access Wholesaler	1–20 / victim access	Commission pricing (?)	$\sum Pay\text{-}per\text{-}commission$
Account Retailer	1–30 / stolen record	Pay-per-access	$\sum (Pay\text{-}per\text{-}account\text{-}access - Pay\text{-}per\text{-}victim\text{-}access/48)$
Criminal Consumer	Exploits individual accounts	Per-victim-revenue	$\sum (Account\text{-}revenue - Pay\text{-}per\text{-}account\text{-}access)$

Figure 10: Malware value chain: malware suppliers, infections, victim sales, and online account sales.

As access to victim devices is a single point of failure through one marketplace, law enforcement should focus effort there.

6 LIMITATIONS AND QUALITY METRICS

6.1 MalwareInfectionSet: representativeness

Here we study the geographic distribution of entries in MalwareInfectionSet. The World Bank publishes open access to world statistics[9], including estimated number of Internet users. This statistic allows us to calculate the proportion of victims per million Internet users (International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database, CSV file)[10]. We compared the number of victims to that estimate for each country. Figure 11 illustrates the country distribution across this dataset, which ranks countries by the number of victims per million Internet users.

Although the distribution of victims varies between countries, nearly every country has victims. The number of victims may vary due to the country- or language-specific nature of some infection sources. For example, Niger, France, and Australia have a disproportionately high number of cases in comparison to their Internet user populations.
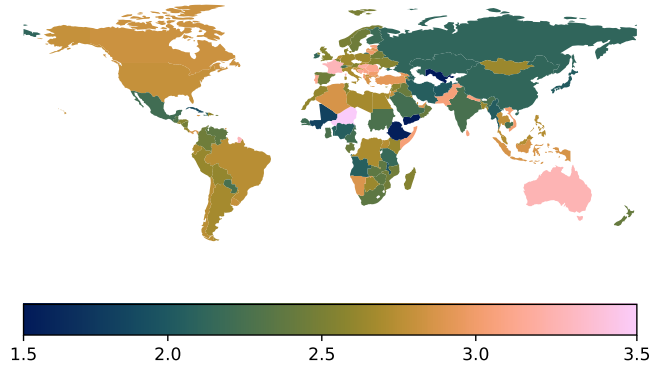
MalwareInfectionSet contains compromised credentials for 424 unique onion domains from 293 victims, i.e. addresses from Tor-Browser for use within the Tor anonymity network. We opened onion domains to view them: Most of them are offline, but the online onion services appear to be marketplaces, based on the landing page. Typically, marketplaces on Tor sell illegal substances and goods [9].

The Tor Network has an estimated 3 million concurrent users, based on daily data collection of the Tor Project for 2022[11]. This
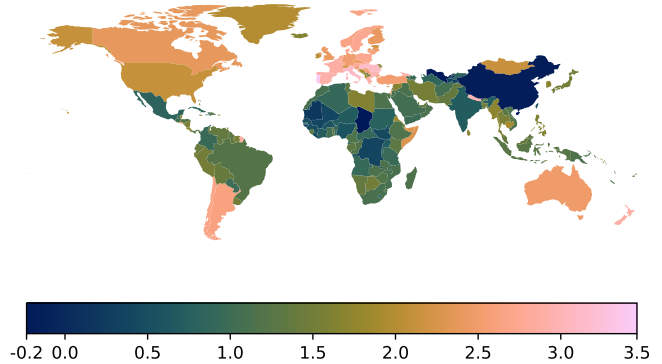

Figure 11: Countries in MalwareInfectionSet by the number of victims per million Internet users, using a base-10 logarithmic scale.

Figure 12: Genesis Market sells access to victims. Countries in VictimAccessSet by the number of victims per million Internet users, using a base-10 logarithmic scale.

equates to 0.075% of the total 4 billion Internet users[12]. We now compare this percentage to what our dataset indicates. Using MalwareInfectionSet, we are aware of the percentage of victims who authenticate on onion websites (the total number of victims using Tor is greater than this), which is 0.016% (293/1.8 million). The proportions have the same order of magnitude, and MalwareInfectionSet also represents anticipated Tor users.

---

[9]https://doi.org/10.1596/978-1-4648-0382-6_open_data

[10]https://doi.org/10.5281/zenodo.3969472

[11]https://metrics.torproject.org/userstats-relay-country.html

[12]https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?most_recent_value_desc=true

According to a Google study in conjunction with Harris Poll in 2018, 13% of US adults use a single password for all of their accounts[13]. We conducted a similar study using MalwareInfectionSet, and Figure 7 depicts the results. We observe that 13% of victims use a single password for all services, strikingly consistent with the aforementioned study.

In summary, it is fair to say MalwareInfectionSet represents global Internet users.

6.2 MalwareInfectionSet: accuracy

We compare MalwareInfectionSet to other data sources, including the number of active Facebook users. Given that there are 2.9 billion monthly active users, Facebook estimates that 60% of the world's active Internet users log in each month in 2021[14]. In MalwareInfectionSet, 46% (829,630/1,809,384) of victims had their Facebook accounts compromised (matching in services:*.facebook.com).

Multi-victim entries. It is possible that the infected device is a shared computer with dozens of users in a library, school, or business. In MalwareInfectionSet, 3.3% of entries (60,163) have over ten distinct email addresses, over 50 passwords, and over 100 compromised services. As such, multiple victims are almost certainly represented by these particular entries. However, in 96.7% of cases it appears as though one infected device equates to one human victim in MalwareInfectionSet.

Not all timestamps are accurate. The primary reason is that text extractors are interpreting the victim's information incorrectly. For instance, we noticed that a victim has a computer named 2011-01-1, and because this information is near the actual time field, the extractor chose it. The second reason is that the timestamp appears to be incorrect (perhaps it uses the victim's clock time).

Empty victim entries. 0.6% of entries (11,157) contain no detailed victim information. This indicates that the extractors did not discover any timestamps, IP addresses, domain names, or email addresses. We retain these empty entries in MalwareInfectionSet because they represent victims that had no meaningful data stolen.

Anti-virus information is only indicative. Although the anti-virus field contains snippets of potential endpoint protection software, it does not always show if the device is running the most recent protection. This field is also populated if the victim is merely a customer of a security company (e.g., credentials for my.kaspersky.com). Consequently, we do not attempt to quantify the effectiveness of endpoint protection with our dataset.

Malware also stores incorrect credentials for services. In the raw data, there are numerous credentials for the same service where the username (which can be an email address) and password vary slightly. When victims enter their credentials incorrectly, malware obliviously stores these incorrect credentials for services. For instance, it is common for victims to have usernames and email addresses like name@gmail.com and name@gmail.co.

6.3 VictimAccessSet: representativeness

Here we study the geographic distribution of entries in VictimAccessSet. Figure 12 illustrates the country distribution across this dataset, which ranks countries by the number of victims per million Internet users (i.e., analogous to Figure 11, but for VictimAccessSet).

According to Figure 12, nearly every country is in VictimAccessSet. Although the distribution of victims varies between countries, nearly every country has victims. The number of victims may vary due to the country- or language-specific nature of some infection sources.

For example, Portugal, Hungary, Romania, Italy, and Bulgaria have a disproportionately high number of cases in comparison to their Internet user populations. In contrast, Uzbekistan (post-Soviet state), China, and Tajikistan (post-Soviet state) have a disproportionately low number of cases.

Certain countries do not appear in VictimAccessSet at all: For example, the lack of Russian victims is clear from Figure 12. This indicates that Genesis Market operates from Russia, where cybercriminals are actually allowed to operate freely as long as they do not target Russian citizens. Furthermore, this explains the two post-Soviet state outliers previously mentioned, as well as the following post-Soviet states that are absent from VictimAccessSet: Ukraine, Belarus, and Kazakhstan. Lastly, North Korea is also absent from VictimAccessSet.

This marketplace sells access to devices in all major countries except Russia (and several post-Soviet states), and indeed the Genesis Market domain resolved to an IP address there (before the FBI seized it in 2023).

Authorities in Russia will not investigate cybercrime unless a Russian company or person files a formal complaint as a victim [40]. The simplest way for these criminals to remain undetected is to ensure that their malware does not infect people in their home country. For instance, DarkSide malware does not spread within the Commonwealth of Independent States (CIS) or former Soviet satellites, and many malware strains do not infect computers with a Cyrillic keyboard [40].

In summary, it is fair to say VictimAccessSet represents global Internet users, sans the noted omissions.

6.4 VictimAccessSet: accuracy

Here we describe a number of validations to demonstrate that VictimAccessSet accurately represents global Internet users (by comparison to anticipated numbers).

Compromised services. In VictimAccessSet, 50% (239,622/480,994) of victims had their Facebook accounts compromised. This corresponds to the expected number of Facebook users in the dataset, as Facebook estimates that 60% of the world's active Internet users log in each month.

Operating systems. Windows 10 has a 72.23% market share among Windows operating systems, according to Statcounter GlobalStats in 2022[15]. We calculate the percentage of Windows 10 users within our VictimAccessSet. According to our data, the proportion of Windows 10 users is directly analogous, at 76% (365,676/480,994).

Web browsers. The Chrome web browser has a market share of 79.9% according to W3Schools' Browser Statistics for 2022[16]. Likewise, 71% (340,933/480,994) of the victims in VictimAccessSet are Chrome users.

7 RESEARCH ETHICS AND PRIVACY

Open Data is an unqualified good, but sometimes it can identify individuals. For example, Tran and Scholtes [45] suggest that avoiding the collection of certain types of data can protect against civil rights violations, and countries with a history of data misuse have resolved not to collect data that associates an individual with a vulnerable group.

We obtained the original information for our research from public sources, even though some sources are difficult to locate and access. Bear in mind that at least some of the original logs have been available online for three years. Worse yet, as we demonstrated in Section 5, this information is partially available for sale. As a result, an adversary can locate the original information and identify the victims from the original data, as their real names are typically included in the available data logs.

Our research results use anonymised victim data, which contains no personal information. The original data sources for VictimAccessSet and AccountAccessSet are anonymous and do not contain personal information, hence require no protection. The original data logs for MalwareInfectionSet contain private and sensitive information. We process MalwareInfectionSet while protecting the anonymity of victims (see Section 3).

Malevolent actors would simply use the original data source, not our datasets. By making our structured datasets publicly available with privacy protections in place, researchers can study malware without accidental exposure to sensitive victim information, nor the security risks of collecting the original data.

8 CONCLUSION

We now summarise the entire infostealer malware ecosystem of actors and their operations. We find that cybercrime is specialising in several roles within the illicit economy and for the first time give a comprehensive illustration of the malware value chain with exact pricing information.

First, unethical software developers create malware. They sell software licenses via the Internet, where the cost of the MaaS model is between 125–200 US dollars per month.

Second, the owner of the malware network infects victims via popular methods, such as email attachments and macros. This attacker may utilise spam-as-a-service, which costs between 100–500 US dollars per million emails sent, and targeted email lists, which costs between 25–50 US dollars per one million email addresses.

Third, once malware infects victims, the network operator sells access to their devices on Genesis Market.

One key new finding is a full understanding of the precise monetary value of victims from a criminal perspective: The price range is between 1–350 US dollars, and the value of a victim's record depends on the number of compromised services. This provides insight into the malware network owner's direct retail value of the

victim's information. Typically, access to a malware victim costs between 1–20 US dollars, with a median of 5 US dollars.

Fourth and finally, a buyer of victim access ultimately profits by gaining access to the victim's online accounts or by reselling individual credentials. Database Market operates inside the anonymous Tor network, and resells access to compromised online accounts. We show how resale prices are typically between 1–30 US dollars, with a median of 7 US dollars.

9 FUTURE WORK

Use datasets for interdisciplinary research. With our datasets, it is possible to investigate completely new research topics outside of security and malware. We provide one example here: studying the effectiveness of Internet censorship. In several nations, including the Philippines, the popular adult website Pornhub is blocked[17]. Although using the service without creating an account is free, additional paid features are available after logging in. We investigate this using MalwareInfectionSet, and provide a Python script for this with the published data. 0.47% (133/27,978) of victims in the Philippines have a premium account on Pornhub although the service is censored. Comparatively, 1% (1,861/1,864,61) of victims in the United States and 0.2% (108/56902) of victims in Turkey have premium Pornhub accounts. As a way to use these datasets further, we just showed that Internet filtering in the Philippines does not prevent many people from using a popular adult website.

User guidelines and security technologies. We propose a study of novel guidelines and techniques for businesses that should account for the fact that infostealer malware undermines multi-factor authentication (MFA) and endpoint security, thereby rendering the end-user defenceless. If we compare our malware victim data to the National Institute of Standards and Technology's (NIST) guidelines for the prevention and management of malware incidents, we find that infostealer malware undermines these suggestions [30, 41]. According to the guidelines, network-based anti-virus software, host-based anti-virus software, and spyware detection and removal software are the most effective approaches for preventing malicious code [30, p. 77]. A typical victim uses Windows 10, which by default has Microsoft's most recent endpoint defences. Microsoft Office macros remains a common infection method, after decades of security patches against these exploits. A typical victim has 1–8 unique passwords, indicating that the typical user finds it too challenging to use a unique password for each account. A user may employ MFA, but Genesis Market duplicates the victim's authenticated web session, hence weakening MFA protection.

REFERENCES

[1] Lawrence Abrams. 2022. Raccoon Stealer malware suspends operations due to war in Ukraine. https://www.bleepingcomputer.com/news/security/raccoon-stealer-malware-suspends-operations-due-to-war-in-ukraine/

---

[16]https://www.w3schools.com/browsers/default.asp

[17]https://www.cnnphilippines.com/news/2017/01/16/government-blocks-major-porn-websites.html

[2] Ariel Ainhoren. 2019. Massive 'Collection #1' Data Dump: What's In It and How Did it Happen? https://intsights.com/blog/massive-collection-1-data-dump-whats-in-it-and-how-did-it-happen

[3] Ugur Akyazi, Michel van Eeten, and Carlos Hernandez Gañán. 2021. Measuring Cybercrime as a Service (CaaS) Offerings in a Cybercrime Forum. In *21st Annual Workshop on the Economics of Information Security, WEIS 2021, Virtual Event, 28-29 June, 2021.* http://resolver.tudelft.nl/uuid:01cb117a-339c-42e9-9691-8456a12e3947

[4] Stefan Axelsson. 1999. The Base-Rate Fallacy and Its Implications for the Difficulty of Intrusion Detection. In *CCS '99, Proceedings of the 6th ACM Conference on Computer and Communications Security, Singapore, November 1-4, 1999,* Juzar Motiwalla and Gene Tsudik (Eds.). ACM, 1–7. https://doi.org/10.1145/319709.319710

[5] Adrian Bermudez-Villalva and Gianluca Stringhini. 2021. The shady economy: Understanding the difference in trading activity from underground forums in different layers of the Web. In *APWG Symposium on Electronic Crime Research, eCrime 2021, Boston, MA, USA, December 1-3, 2021.* IEEE, 1–10. https://doi.org/10.1109/eCrime54498.2021.9738751

[6] Rasika Bhalerao, Maxwell Aliapoulios, Ilia Shumailov, Sadia Afroz, and Damon McCoy. 2019. Mapping the Underground: Supervised Discovery of Cybercrime Supply Chains. In *2019 APWG Symposium on Electronic Crime Research, eCrime 2019, Pittsburgh, PA, USA, November 13-15, 2019.* IEEE, 1–16. https://doi.org/10.1109/eCrime47957.2019.9037532

[7] Jean-Marie Borello and Ludovic Mé. 2008. Code obfuscation techniques for metamorphic viruses. *J. Comput. Virol.* 4, 3 (2008), 211–220. https://doi.org/10.1007/s11416-008-0084-2

[8] Roderic Broadhurst, David Lord, Donald Maxim, Hannah Woodford-Smith, Corey Johnston, Ho Woon Chung, Samara Carroll, Harshit Trivedi, and Bianca Sabol. 2018. Malware Trends on Darknet Crypto-Markets: Research Review. *SSRN Electronic Journal* (2018). https://doi.org/10.2139/ssrn.3226758

[9] Alexandru Caciuloiu, Pawinee Parnitudom, Mikko Niemelä, Joshua James, Juha Nurmi, and Praphaphorn Tamarpirat. 2020. Darknet cybercrime threats to Southeast Asia. https://www.unodc.org/documents/southeastasiaandpacific//Publications/2021/Darknet_Cybercrime_Threats_to_Southeast_Asia_report.pdf

[10] Michele Campobasso and Luca Allodi. 2020. Impersonation-as-a-Service: Characterizing the Emerging Criminal Infrastructure for User Impersonation at Scale. In *CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020,* Jay Ligatti, Xinming Ou, Jonathan Katz, and Giovanni Vigna (Eds.). ACM, 1665–1680. https://doi.org/10.1145/3372297.3417892

[11] Michele Campobasso and Luca Allodi. 2023. Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale. In *32nd USENIX Security Symposium.* USENIX Association. https://atc.usenix.org/system/files/sec23fall-prepub-383-campobasso.pdf

[12] Nicolas Christin. 2013. Traveling the Silk Road: a measurement analysis of a large anonymous online marketplace. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013,* Daniel Schwabe, Virgílio A. F. Almeida, Hartmut Glaser, Ricardo Baeza-Yates, and Sue B. Moon (Eds.). International World Wide Web Conferences Steering Committee / ACM, 213–224. https://doi.org/10.1145/2488388.2488408

[13] Mihai Christodorescu, Somesh Jha, and Christopher Kruegel. 2007. Mining specifications of malicious behavior. In *Proceedings of the 6th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT International Symposium on Foundations of Software Engineering, 2007, Dubrovnik, Croatia, September 3-7, 2007,* Ivica Crnkovic and Antonia Bertolino (Eds.). ACM, 5–14. https://doi.org/10.1145/1287624.1287628

[14] Mihai Christodorescu, Somesh Jha, Sanjit A. Seshia, Dawn Xiaodong Song, and Randal E. Bryant. 2005. Semantics-Aware Malware Detection. In *2005 IEEE Symposium on Security and Privacy (S&P 2005), 8-11 May 2005, Oakland, CA, USA.* IEEE Computer Society, 32–46. https://doi.org/10.1109/SP.2005.20

[15] Ben Cohen. 2020. Predator the Thief. https://www.cyberark.com/resources/threat-research-blog/predator-the-thief

[16] Benjamin Collier, Richard Clayton, Alice Hutchings, and Daniel Thomas. 2020. Cybercrime is (often) boring: maintaining the infrastructure of cybercrime economies. In *20th Annual Workshop on the Economics of Information Security, WEIS 2020, Virtual Event, 14-15 December, 2020.* https://doi.org/10.17863/CAM.53769

[17] Diana S. Dolliver. 2015. Evaluating drug trafficking on the Tor Network: Silk Road 2, the sequel. *International Journal of Drug Policy* 26, 11 (Nov. 2015), 1113–1123. https://doi.org/10.1016/j.drugpo.2015.01.008

[18] Diana S. Dolliver and Jennifer L. Kenney. 2016. Characteristics of Drug Vendors on the Tor Network: A Cryptomarket Comparison. *Vict. Offenders* 11, 4 (May 2016), 600–620. https://doi.org/10.1080/15564886.2016.1173158

[19] Dov Lerner. 2022. Painless Steal: The Malware of Choice for Initial Access Brokers. https://www.cybersixgill.com/blog/painlessstealinitialaccessbrokers/

[20] Brad Duncan. 2022. Windows MetaStealer Malware. https://isc.sans.edu/diary/Windows+MetaStealer+Malware/28522

[21] Jose Miguel Esparza. 2019. Understanding the credential theft lifecycle. *Computer Fraud & Security* 2019, 2 (2019), 6–9. https://doi.org/10.1016/S1361-3723(19)30018-1

[22] Esullivan. 2020. TAU-TIN - Kpot InfoStealer. https://community.carbonblack.com/t5/Threat-Advisories-Documents/TAU-TIN-Kpot-InfoStealer/ta-p/88517

[23] Matt Fredrikson, Mihai Christodorescu, Jonathon T. Giffin, and Somesh Jha. 2010. A Declarative Framework for Intrusion Analysis. In *Cyber Situational Awareness - Issues and Research,* Sushil Jajodia, Peng Liu, Vipin Swarup, and Cliff Wang (Eds.). Advances in Information Security, Vol. 46. Springer, 179–200. https://doi.org/10.1007/978-1-4419-0140-8_9

[24] André Ricardo Abed Grégio, Vitor Monte Afonso, Dario Simões Fernandes Filho, Paulo Lício de Geus, and Mario Jino. 2015. Toward a Taxonomy of Malware Behaviors. *Comput. J.* 58, 10 (2015), 2758–2777. https://doi.org/10.1093/comjnl/bxv047

[25] Keman Huang, Michael D. Siegel, and Stuart E. Madnick. 2018. Systematically Understanding the Cyber Attack Business: A Survey. *ACM Comput. Surv.* 51, 4 (2018), 70:1–70:36. https://doi.org/10.1145/3199674

[26] Fabian Kaczmarczyck, Bernhard Grill, Luca Invernizzi, Jennifer Pullman, Cecilia M. Procopiuc, David Tao, Borbala Benko, and Elie Bursztein. 2020. Spotlight: Malware Lead Generation at Scale. In *ACSAC '20: Annual Computer Security Applications Conference, Virtual Event / Austin, TX, USA, 7-11 December, 2020.* ACM, 17–27. https://doi.org/10.1145/3427228.3427273

[27] Leo Kelion. 2012. Pastebin: Running the site where hackers publicise their attacks. https://www.bbc.com/news/technology-17524822

[28] Clemens Kolbitsch, Paolo Milani Comparetti, Christopher Kruegel, Engin Kirda, Xiao-yong Zhou, and XiaoFeng Wang. 2009. Effective and Efficient Malware Detection at the End Host. In *18th USENIX Security Symposium, Montreal, Canada, August 10-14, 2009, Proceedings,* Fabian Monrose (Ed.). USENIX Association, 351–366. http://www.usenix.org/events/sec09/tech/full_papers/kolbitsch.pdf

[29] Per Hå Meland and Guttorm Sindre. 2019. Cyber Attacks for Sale. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI).* IEEE, 54–59. https://doi.org/10.1109/csci49370.2019.00016

[30] Peter Mell, Karen Kent, and Joseph Nusbaum. 2005. *Guide to malware incident prevention and handling.* National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-83

[31] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker. 2011. An analysis of underground forums. In *Proceedings of the 11th ACM SIGCOMM Internet Measurement Conference, IMC '11, Berlin, Germany, November 2-, 2011,* Patrick Thiran and Walter Willinger (Eds.). ACM, 71–80. https://doi.org/10.1145/2068816.2068824

[32] Juha Nurmi, Teemu Kaskela, Jussi Perälä, and Atte Oksanen. 2017. Seller's reputation and capacity on the illicit drug markets: 11-month study on the Finnish version of the Silk Road. *Drug and Alcohol Dependence* 178 (Sept. 2017), 201–207. https://doi.org/10.1016/j.drugalcdep.2017.05.018

[33] Jeremiah Onaolapo, Enrico Mariconti, and Gianluca Stringhini. 2016. What Happens After You Are Pwnd: Understanding the Use of Leaked Webmail Credentials in the Wild. In *Proceedings of the 2016 ACM on Internet Measurement Conference, IMC 2016, Santa Monica, CA, USA, November 14-16, 2016,* Phillipa Gill, John S. Heidemann, John W. Byers, and Ramesh Govindan (Eds.). ACM, 65–79. http://dl.acm.org/citation.cfm?id=2987475

[34] Young Hee Park, Douglas S. Reeves, and Mark Stamp. 2013. Deriving common malware behavior through graph clustering. *Comput. Secur.* 39 (2013), 419–430. https://doi.org/10.1016/j.cose.2013.09.006

[35] Sergio Pastrana, Alice Hutchings, Andrew Caines, and Paula Buttery. 2018. Characterizing Eve: Analysing Cybercrime Actors in a Large Underground Forum. In *Research in Attacks, Intrusions, and Defenses - 21st International Symposium, RAID 2018, Heraklion, Crete, Greece, September 10-12, 2018, Proceedings (Lecture Notes in Computer Science, Vol. 11050),* Michael Bailey, Thorsten Holz, Manolis Stamatogiannakis, and Sotiris Ioannidis (Eds.). Springer, 207–227. https://doi.org/10.1007/978-3-030-00470-5_10

[36] Sergio Pastrana, Daniel R. Thomas, Alice Hutchings, and Richard Clayton. 2018. CrimeBB: Enabling Cybercrime Research on Underground Forums at Scale. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018,* Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 1845–1854. https://doi.org/10.1145/3178876.3186178

[37] Ildiko Pete, Jack Hughes, Andrew Caines, Anh V. Vu, Harshad Gupta, Alice Hutchings, Ross J. Anderson, and Paula Buttery. 2022. PostCog: A tool for interdisciplinary research into underground forums at scale. In *IEEE European Symposium on Security and Privacy, EuroS&P 2022 - Workshops, Genoa, Italy, June 6-10, 2022.* IEEE, 93–104. https://doi.org/10.1109/EuroSPW55150.2022.00016

[38] Ildiko Pete, Jack Hughes, Yi Ting Chua, and Maria Bada. 2020. A Social Network Analysis and Comparison of Six Dark Web Forums. In *IEEE European Symposium on Security and Privacy Workshops, EuroS&P Workshops 2020, Genoa, Italy, September 7-11, 2020.* IEEE, 484–493. https://doi.org/10.1109/EuroSPW51379.2020.00071

[39] Matthias Schäfer, Markus Fuchs, Martin Strohmeier, Markus Engel, Marc Liechti, and Vincent Lenders. 2019. BlackWidow: Monitoring the Dark Web for Cyber Security Information. In *11th International Conference on Cyber Conflict, CyCon 2019, Tallinn, Estonia, May 28-31, 2019*, Tomás Minárik, Siim Alatalu, S. Biondi, M. Signoretti, I. Tolga, and G. Visky (Eds.). IEEE, 1–21. https://doi.org/10.23919/CYCON.2019.8756845

[40] Bruce Schneier. 2021. Adding a Russian Keyboard to Protect against Ransomware. https://www.schneier.com/blog/archives/2021/05/adding-a-russian-keyboard-to-protect-against-ransomware.html

[41] Murugiah Souppaya and Karen Scarfone. 2013. *Guide to malware incident prevention and handling for desktops and laptops*. Vol. 800. National Institute of Standards and Technology. 83 pages. https://doi.org/10.6028/NIST.SP.800-83r1

[42] Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. 2011. The Underground Economy of Spam: A Botmaster's Perspective of Coordinating Large-Scale Spam Campaigns. In *4th USENIX Workshop on Large-Scale Exploits and Emergent Threats, LEET '11, Boston, MA, USA, March 29, 2011*, Christopher Kruegel (Ed.). USENIX Association. https://www.usenix.org/conference/leet11/underground-economy-spam-botmasters-perspective-coordinating-large-scale-spam

[43] Kurt Thomas, Danny Yuxing Huang, David Y. Wang, Elie Bursztein, Chris Grier, Tom Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. 2015. Framing Dependencies Introduced by Underground Commoditization. In *14th Annual Workshop on the Economics of Information Security, WEIS 2015, Delft, The Netherlands, 22-23 June, 2015*. http://www.econinfosec.org/archive/weis2015/papers/WEIS_2015_thomas.pdf

[44] Bill Toulas. 2022. New Meta information stealer distributed in malspam campaign. (2022). https://www.bleepingcomputer.com/news/security/new-meta-information-stealer-distributed-in-malspam-campaign/

[45] Emmie Tran and Ginny Scholtes. 2015. Open data literature review. *Barkeley School of Law, University of California* (2015). https://www.law.berkeley.edu/wp-content/uploads/2015/04/Final_OpenDataLitReview_2015-04-14_1.1.pdf

[46] H. W. J. van Rijn. 2021. *An In-depth Analysis of the AZORult Infostealer Malware Capabilities*. Master's thesis. Eindhoven University of Technology. https://research.tue.nl/en/studentTheses/an-in-depth-analysis-of-the-azorult-infostealer-malware-capabilit

[47] Anh V. Vu, Jack Hughes, Ildiko Pete, Ben Collier, Yi Ting Chua, Ilia Shumailov, and Alice Hutchings. 2020. Turning Up the Dial: the Evolution of a Cybercrime Market Through Set-up, Stable, and Covid-19 Eras. In *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*. ACM, 551–566. https://doi.org/10.1145/3419394.3423636

[48] William Winston, Robert E Stevens, Philip K Sherwood, and John Paul Dunn. 2013. *Market analysis: Assessing your business opportunities*. Routledge. https://archive.org/details/marketanalysisas0000stev

[49] Muwei Zheng, Hannah Robbins, Zimo Chai, Prakash Thapa, and Tyler Moore. 2018. Cybersecurity Research Datasets: Taxonomy and Empirical Analysis. In *11th USENIX Workshop on Cyber Security Experimentation and Test, CSET 2018, Baltimore, MD, USA, August 13, 2018*, Christian S. Collberg and Peter A. H. Peterson (Eds.). USENIX Association. https://www.usenix.org/conference/cset18/presentation/zheng