Busting up Monopoly: Methods for modern darknet marketplace forensics

By:
Daniel Dolejška, Michal Koutenský, Vladimír Veselý, Jan Pluskal

From the proceedings of
The Digital Forensic Research Conference
DFRWS APAC 2023
Oct 17-20, 2023

Forensic Science International: Digital Investigation 46 (2023) 301604

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

DFRWS 2023 APAC - Proceedings of the Third Annual DFRWS APAC

Busting up Monopoly: Methods for modern darknet marketplace forensics

Daniel Dolejška *, Michal Koutenský, Vladimír Veselý, Jan Pluskal

Faculty of Information Technology, Brno University of Technology, Božetěchova 1/2, Brno, 612 00, Czech Republic

ARTICLE INFO

Keywords:
Dark market
Illegal trade
Market analytics
Web scraping
Crypto
Cryptocurrency
Dark marketplace
Cryptomarket
Tor
Overlay networks

ABSTRACT

Darknet marketplaces represent the most delinquent evolution step in distributing illicit goods such as drugs, steroids, firearms, warez, or leaked personal information. On the one hand, law enforcement agencies try to catch vendors, buyers, and operators of darknet marketplaces. On the other hand, the criminals mentioned above constantly stretch the limits of overlay networks, applied cryptography, and cryptocurrency pseudonymity. This paper intends to provide relevant and up-to-date (for the year 2022) information about potential ways to deal with darknet marketplaces from the perspective of investigators. The paper outlines methods (based on periodic web scraping) that may help sworn officers to gather evidence about darknet marketplace (ab)users. The potential is demonstrated in a real-life case study of the Monopoly Market. For instance, suggested approaches seem capable: monitoring the demography and activities of darknet marketplace users, estimating the number of procurements and their value, and correlating user identities with their cryptocurrency addresses. The paper also provides an applicability analysis of proposed methods on the subset of currently trending darknet marketplaces.

1. Introduction

Silk Road's success paved a new era, where darknet marketplaces are used as meeting points for vendors and buyers on the darknet. Darknet markets established a proven e-commerce platform where people can trade illicit commodities such as drugs, firearms, counterfeit currencies, and cyber arms. Fraud shops (lighter subsets of darknet marketplaces) offer strictly "less" problematic commodities such as forged documents, stolen personal information, credit cards, and warez.

There are numerous explanations for the "darknet marketplace". Instead of a definition, we outline a list of components and IT concepts necessary for darknet marketplace operation. Dark marketplace, a.k.a. darknet market (hereafter DNM), is a website operated within an overlay network (e.g., Tor, which offers an anonymized connection between client and server using a hidden service). DNM buyers, vendors, and operators employ end-to-end encryption for communication, which prevents any eavesdropping. DNM users conduct payments with cryptocurrencies because they provide a nearly instant, decentralized, and trustless transfer of value.

In our study, we have focused on Monopoly Market, which we chose to represent the state-of-the-art DNM (see Section 1.3). We have monitored this market using periodical web-scraping to collect.

1.1. Problem statement

The state-of-the-art DNMs have learned from their forebears and are adopting a wallet-less approach while supporting a direct deal (not escrows) modus operandi. This state renders methods, which detect DNM transactions directly on blockchains (Hiramoto and Tsuchiya, 2020; Tsuchiya and Hiramoto, 2021), no longer possible.

In our previous publication (Dolejška et al., 2022), we have introduced methods allowing web scraping with many finer-grained snapshotting intervals. Each item is snapshot roughly every 15 min, which allows for novel investigation methods to be examined.

Although the amount of DNMs currently available diminishes, LEA still needs to keep an open eye and deal with this threat to public safety with due rigour. Available sources, such as (Chainalysis Inc., 2021, 2022), show that the revenue that DNMs render is constantly increasing.

1.2. Target audience

This paper is primarily intended for law enforcement representatives and cybersecurity researchers. However, writing a paper describing the methodology and methods to help law enforcement agencies (LEA) get the upper hand over criminals might be tricky since disclosing suggested techniques may obsolete them in the future. The information contained in this paper is based on a DNM which already ceased to exist. Therefore,

* Corresponding author.
E-mail addresses: dolejska@fit.vut.cz (D. Dolejška), koutenmi@fit.vut.cz (M. Koutenský), veselyv@fit.vut.cz (V. Veselý), ipluskal@fit.vut.cz (J. Pluskal).

https://doi.org/10.1016/j.fsidi.2023.301604

Available online 13 October 2023

we believe publishing this paper will not endanger any ongoing investigation. On the contrary, we encourage any sworn officer or cybersecurity researcher to contact us so that we can share our intel and potentially join forces.

1.3. Selection of Monopoly Market

Monopoly Market was one of the attractive places (Darknetlive, 2021) to buy illicit goods such as drugs in 2021. Apart from its attractiveness for the end-users (buyers and vendors), we chose this market because of the following parameters that appealed to us.

account-less access to the website, product listing, and shopping, which increases anonymity for end-users, limits user traceability, which increases vulnerability and makes the site prone to web scraping (see Section 3.1);
wallet-less eliminates risks of potential exit scams, which appeals to end-users but also mandates that any transaction needs to be a direct deal between buyer and vendor, which makes the transaction visible on the blockchain and allows correlation (see Section 3.5)
only single-product orders are supported by the market, which allows a more straightforward correlation of purchase and blockchain transactions (see Section 3.5);
simple CAPTCHA that allows automated resolution, in our case, by paid third-party service[1] (see Section 4.2);
link distribution network (LDN) publication of the most up-to-date marketplace address eliminates manual intervention in case that market address changes over time (see Section 4.2);
obfuscation-free blockchain used for money transfers allows for correlation of purchase and blockchain transactions (see Section 3.5);

In our previous publication (Dolejška et al., 2022), we demonstrated how to conduct fine-grained periodical scraping of DNM, in our case, Monopoly Market, resulting in a data set containing any deal procured on this market with roughly 15 min accuracy. This data set is freely available to any researcher or LEA officer per request.

1.4. Paper structure

The paper is structured as follows. Section 2 provides a state-of-the-art overview and the background of Monopoly Market operations. Sections 3 and 4 form the core of our recent research work involving the analysis of DNMs. Section 3 describes details on periodic web scraping of specific data from DNMs, which is an enabler for subsequent analysis providing interesting evidence about vendors, buyers and procured commodities. Section 4 practically demonstrates Section 3 consequences on a previously published data set collected during the nearly year-long monitoring of Monopoly Market. Section 5 provides an executive summary and mentions our next research steps.

1.5. Contribution

This paper aims to elaborate on periodic web scraping as a tool enabling numerous different methods of assessing users' activities on DNMs (five of which are presented in the following sections). The methods are then evaluated on a data set of open-source intelligence gathered during the nearly one-year-long monitoring of the Monopoly Market. Results reveal interesting findings about DNM vendors, buyers, and operators, which should support the suitability of the proposed methods. The capabilities to apply these methods are thoroughly analyzed on a subset of currently active DNMs.

2. State of the art

The darknet has been an exciting cybersecurity research topic for more than ten years due to its built-in nature, offering a haven for various prohibited activities. This trend has increased with the introduction of cryptocurrencies and DNMs as e-commerce platforms for trading illegal goods. This section provides a broader context for this paper's research and outcomes. The section starts with a general overview of other relevant scientific works. Then, it continues with economic insight and trends into DNM businesses. This section is concluded with publicly available information about the history of Monopoly Market operation.

2.1. Relevant research

For several years, criminal activity centred around the trade of illicit goods has been primarily done through Dark Markets (Kermitsis et al., 2021). These marketplaces are only accessible through what is known as the dark web.

Websites (accessible using standard tools and technologies and indexable by search engines) constitute the Surface Web. Although this is the part of the average Internet users are most familiar with, it makes up only a fraction of the total Internet traffic.

Most traffic belongs to the Deep Web, which covers resources locked behind authentication or APIs and is thus inaccessible without additional tooling or knowledge. It is estimated that over 90% of total Internet traffic is of this type (Chertoff, 2017).

The dark web is characterized by using overlay networks such as Tor or I2P, which provide some amount of anonymity for all actors involved. These technologies were initially intended for privacy-conscious users or professions where anonymity might be a matter of security, such as journalists or military personnel (Chertoff, 2017). However, the features they provide make them an attractive option for individuals engaging in illegal activities online.

To conduct monetary transactions, DNMs do not use fiat currencies, such as the euro or dollar, but cryptocurrencies. Cryptocurrencies are digital currencies that are decentralized, pseudonymous, and based on cryptographic principles. These principles allow actors to exchange value without the need to trust an intermediary to process the transaction or disclose their identity to any parties involved and are resistant to blocking and censorship (Spagnoletti et al., 2021). These properties make them a natural fit as a payment system for DNM, where anonymity and lack of trust are expected.

Bitcoin was the first cryptocurrency based on blockchain technology, created in 2008 by Satoshi Nakamoto. To this day, Bitcoin remains one of the most popular cryptocurrencies, with a market valuation of more than 26B USD.[2] While many Dark Markets still accept Bitcoin, many are adopting newer currencies with more advanced privacy features, such as Monero (van Saberhagen, 2013). Compared to Bitcoin, Monero uses various additional mechanisms which allow the sender and the receiver, as well as the amount transferred, to be hidden in the transaction. This trend can be seen in our analysis presented in Table 1 as well as recent reports (Chainalysis Inc., 2022).

DNMs saw their rise in popularity with Silk Road, a drug-focused market operating from 2011 to 2013. Silk Road was shut down by LEAs in late 2013, and its administrator, Ross Ulbricht, was arrested. This case was thoroughly analyzed by Christin (2013). Until then, trade was primarily based around Silk Road; market fragmentation followed, with many new markets trying to fill the gap left by Silk Road.

In the field of dark web scraping, namely DNM web scraping, rigorous research has been conducted. Christin (2013) web-scraped data from Silk Road with a 14-h period focusing on detailed listings and vendor ratings followed by a detailed analysis of the Silk Road case.

[1] Services such as https://2captcha.com or https://anti-captcha.com [Accessed 12th October 2022].

[2] https://coinmarketcap.com/[Accessed 12th October 2022].

Table 1
A table shows the observed features of various popular DNMs. A green check means a feature is present; a red cross that a feature is not present. A yellow tilde in the *Product Sales Counter* represents that only the total vendor sales counter is available (i.e., no numbers per product listing). A gray question mark is used for features we were unable to determine, mainly due to DNMs being unavailable for prolonged time periods during our survey. **AB**: AlphaBay; **ARC**: Archetyp; **ARES**: Ares; **ASAP**: ASAP Market; **BHM**: Bohemia; **DFM**: DarkFox Market; **CPH**: Cypher Marketplace; **OMG**: OMG!OMG! **MM**: Monopoly Market; **NMS**: Nemesis; **RYL**: Royal Market; **T2D**: Tor2Door; **VC**: Vice City.

Feature	AB	ARC	ARES	ASAP	BHM	DFM	CPH	MM	NMS	RYL	T2D	VC
Vendor Profile Page	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vendor Activity Status	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product Sales Counter	✓	✓	~	~	✓	✓	~	✓	~	✓	✓	~
Product Origin	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product Reviews	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Buyer Profile Page	✗	?	✗	✗	?	~	✗	✓	✗	✗	?	✗
Monero	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bitcoin	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Account-less Orders	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
Wallet-less Payment	✗	?	✓	✗	?	✗	✓	✓	✓	✓	?	✗
Multisig Support	✗	?	✗	✗	?	✓	✗	✓	✗	✗	?	✗
Escrow Support	✓	?	✓	✓	?	✓	✓	✓	✓	✓	?	✓
Direct Deal Support	✓	?	✓	✓	?	✓	✓	✓	✓	✓	?	✓
Controlled Substances	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Counterfeits	✓	?	✓	✓	?	✓	✓	✗	✓	✓	?	✓
Digital	✓	?	✓	✓	?	✓	✓	✗	✓	✓	?	✓
Fraud	✓	?	✓	✓	?	✓	✗	✗	✓	✓	?	✓
Weapons	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Others	✓	?	✗	✗	?	✓	✓	✗	✓	✓	?	✓

Soska and Christin (2015) updated Christin (2013) collection with data from additional 16 Dark Markets in the following two years. In their last data set, Tai et al. (2019) numbered 12 markets around 2016–2018. Van Buskirk et al. (2016), or Van Buskirk et al. (2017) have performed long-term web scraping attempts, usually of multiple marketplaces, but with a considerable period that spans hours or days. LEAs have used collected data sets in the past as an essential source for a criminal investigation of drug-related illicit activities EMCDDA (2017). It would seem that the most extensive data set available (Darknet Market Archives (DNM)) collected by Branwen et al. (2015) sourced data from 89 markets, 37 darknet forums 1.6 TB in size. The oldest records span 2011–2013; main, manually scraped data are from 2013 to 2015 obtained in daily up to weekly intervals. Another recent effort is the AZSecure data set, where Du et al. (2018) focused on cyber threats. Therefore, the data set contains additional sources like IRC and forums communication.

In the paper Dolejška et al. (2022), we show that it is possible to automatically web scrape a well-protected DNM with a smaller period that spans 15 min on average. We have used a programmatic approach to automate and scale the web scraping process to overcome limitations encountered by our predecessors (such as CAPTCHAs, rate-limiting, etc.) to compose finer-grained data sets allowing for advanced investigation method applications. We introduce such methods in Section 3.

2.2. Evolution and trends

Overlay networks (e.g., Tor, I2P), Pretty Good Privacy (PGP), and cryptocurrencies (e.g., Bitcoin, Monero) are the building blocks of DNMs. These technologies offer (a) anonymous network connection (between the user and DNM); (b) secure end-to-end encrypted message exchange (among DNM users); and (c) decentralized, trustless transfer of value (from buyer towards vendor or DNM operator). It all started with Silk Road DNM (which revolutionized the application of technologies mentioned above into web services) in February 2011. Silk Road was shut down in 2013, and immediately other copycat projects took its market share. Since then, the continuous proliferation of similar websites has kept the DNM ecosystem alive and blooming.

There are four different endings for any DNM listed according to (EMCDDA, 2018).

- seizure by LEAs (e.g., Farmer's Market, Silk Road, AlphaBay;
- the hack caused by other parties (e.g., Flo Market, Onionshop, Havana/Absolem);
- scam conducted by operators (e.g., Sheep Marketplace, Atlantis, Deepbay);
- voluntarily exit by operators (e.g., Agora, Acropolis, Dream Market, White House Market).

Being an operator of a DNM is a profitable yet precarious business (considering associated criminal charges such as drug distribution, money laundering, and participation in organized criminal groups). Despite that, online means of communication have taken a more predominant role as the source of drug distribution with the outbreak of the COVID-19 pandemic (EMCDDA, 2020). DNMs have gained importance because COVID-19 disrupted usual supply chains (e.g., street dealers, parties, trusted suppliers) since the in-person contact between people was limited (EMCDDA and Europol, 2020). Nevertheless, a recent report (EMCDDA, 2022) shows that the estimated revenues of DNMs dropped to 30k EUR per day by the end of 2021 compared with 1M EUR per day during 2020. The estimation is based on just a subset of DNMs in the ecosystem. Some of them were affected by long downtime periods, scams, voluntary exits, and also successful LEA operations — including the seizure of the Hydra market (Europol, 2021), the largest market since 2015 with approx. 2400 vendors serving nearly 0.5M users.

Let us focus on the valuation of DNMs according to blockchain analytics published in (Chainalysis Inc., 2022). In 2022, overall DNM revenues reached up to 2.1B USD. From this number, 1.8B USD was generated by DNMs. The remaining 300M USD were trades conducted on fraud shops (i.e., markets with "less" problematic goods like credit cards, stolen logins, and exploit kits). The amount of 1.8B USD also includes 110M USD, representing value sent directly between DNM buyers and vendors (i.e., direct deals without DNM as a middleman providing escrow service). Direct deals' total value and count are rapidly increasing (nearly tripling from 2019 to 2022), which empowers our research and effort to conduct further investigations of wallet-less DNMs. According to (Chainalysis Inc., 2021) and (Chainalysis Inc.,

2022), the number of DNMs is declining (both reports show graphs with downward trends but with mismatching numbers across the years, probably due to the different methodologies of categorization of what is/was DNM at the time of the report).

Hopefully, some portion of cryptocurrency assets entering and leaving dark markets will be easier to trace with the impending and more thorough adoption of the "FATF Traveler Rule" guidance (Financial Action Task Force, 2021) or "Markets in Crypto-Assets" bill (European Commission, 2020). In compliance with these acts, every cryptocurrency transaction above a significant monetary threshold must have originators and beneficiaries identified if handled by regulated virtual asset service providers. This should help cryptocurrency exchanges and banks identify risky transactions and (not only DNM) users faster, even before any LEA interventions.

2.3. Background of Monopoly Market

Monopoly Market[3] was launched in August 2019 (Darknetlive, 2019). It gained a lot of popularity in late 2021 after the White House Market (WHM) announced its retirement. The WHM was one the largest (45, 000 advertised product listings (Barratt et al., 2022)) and most popular (326, 570 active user accounts (Darknetlive, 2021)) DNMs. Monopoly Market was recommended as its successor, namely because it supported a *true wallet-less, direct deal* schema (Darknetlive, 2021). Another market recommended as a viable alternative was Versus market, because it *enforced multisig* transactions, thus minimizing the chance of exit scams. Monopoly Market went dark on 28th December 2021. It has been later revealed that the marketplace has been shut down as a part of international dark marketplace seizure operation SpecTor (Europol, 2023).

Monopoly Market had several notable features: (a) it supported both Monero (XMR) and Bitcoin (BTC); (b) it has used a completely wallet-less monetization system; (c) and offered direct deal transactions as well as a multi-sig escrow. Usage of direct deals increases confidence that the market does not end with an exit scam thus, customers do not lose funds and are fully in control of the payment. In contrast, usage of escrow demands that the buyer pays to the "trusted" deposit held by the market operator, which prolongs payment to the vendor to a later date. With the direct deal adaptation, Monopoly Market charged commissions on a monthly basis.

This market was focused on selling narcotics, steroids, stimulants, and prescription drugs. A very small percentage of sold items did not fall under one of these categories. While the proportional representation of these categories changed over time, the best-selling category was cannabis (Dolejška et al., 2022).

3. Methods, datapoints and use cases

DNMs come and go. Sometimes out of the operator's free will and sometimes with a little help from LEAs. When a marketplace is closed, it also takes all the data and publicly accessible information to its grave. All the information provided by the marketplace website can be used as digital evidence. Collecting and periodically archiving such content before it disappears brings certain benefits. From helping paint a bigger picture (categories, counts, popularity, offer, demand, trends) to sometimes providing direct evidence and aiding the conviction. It allows post-mortem analysis, development timeline visualization, and the ability to "go back in time" in order to find things that would not otherwise exist anymore.

The following subsections present various data and metadata collection and analysis methods. There is a wide variety of metadata properties which can be gathered from a market. Each of the proposed methods focuses on a different subset. Similarly, they target distinct entities which engage in or support activities on the marketplace (i.e., service operators, vendors, buyers). Furthermore, the proposed methods include real-life use case examples. They showcase which (meta)data are gathered and processed on a selection of popular and sizeable marketplaces (at the time of writing this paper).

3.1. Product Data Retention

This method is focused on the retention of product listings and their later usage as direct or at least corroborative evidence against the vendor. Data related to this method has almost universally the highest occurrence on the marketplaces and is the easiest to gather — the information about advertised products and the corresponding vendors. One such example of product listing is shown in Fig. 1. The available data properties vary from market to market but usually consist of the following.

- product name or title (which should be distinguishable or unique for the whole DNM),
- product showcase (photos),
- product descriptions,
- product price,
- buyers' comments and reviews,
- vendor's profile/biography,
- other product metadata (created date of ad creation, category, country of origin, a list of countries for product delivery, sales count),
- other vendor profile metadata (joined date, country of shipping, avatar, PGP keys).

Product photos may be valuable digital evidence because they are often taken from the vendor's home or production surroundings. Such photos may contain additional information about the vendors (e.g., handwriting samples) and their environment (e.g., same table or easily recognizable furniture).

3.1.1. Use case — home search

Police arrive at the suspect's house with a search warrant. No drugs are found at the scene; however, a carpet with a pattern and a table with various scratch marks matching some of the photos published on the marketplace website are found there. The person's involvement is apparent, and this knowledge can now be used as leverage during an interrogation.

3.1.2. Use case — looking back

Having the data archived, we can go back to them whenever we want an additional content analysis. That may come in handy, especially

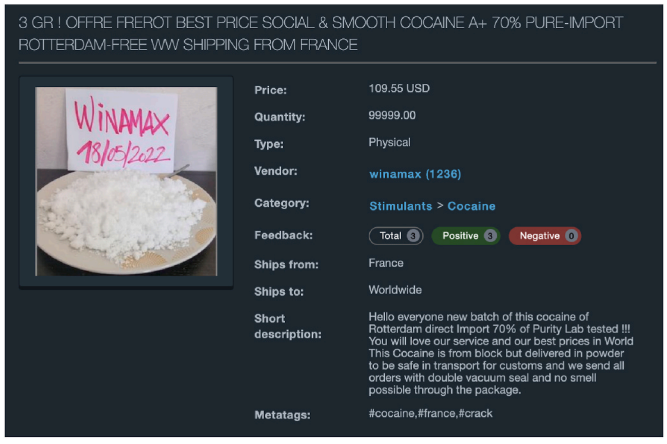

Fig. 1. Product listing detail from ASAP Market.

[3] Review available at https://darknetlive.com/markets/monopoly-market/[Accessed 12th October 2022].

when a new vendor or product appears somewhere (on a different marketplace). Looking back, we can determine whether the vendor or product in question already appeared previously (on another existing marketplace or even the ones that do not exist anymore or are temporarily out of service).

3.1.3. Use case — general market analysis

New studies and analyses can be conducted on available products, vendors, and marketplace features, potentially considering new (perhaps previously ignored) factors and features. The number of available/sold products (controlled substances, fraudulent documents, etc.) can be compared between marketplaces empirically, even after the marketplaces cease to exist.

3.2. Operator data retention

This method targets DNM globally and people in charge (i.e., operators, moderators, admins) who are not necessarily participating as vendors or buyers.

Any information about the DNM service operators is essential and can prove to be highly valuable. Data sources providing such information can be scattered all over the marketplace website or even other related sites. For example, e-commerce-based marketplaces tend to have a separate forum to ask questions, solve disputes, announce news or just chat. Two such instances can be seen in Figs. 2 and 3. The images show admin and moderator activity on the forums at certain times, providing us with information about their activity window. As for the data points, these include but are not limited to the following.

- available server metadata (software versions, used technology stack, active libraries, significant HTTP headers),
- times of service maintenance and update outages,
- general website (un)availability periods,
- any content created by DNM operators (forum comments/answers, tutorials, guides, public announcements, PGP keys, signed messages, canaries, and their update period), and the corresponding timestamps,
- periods of detectable activity of moderators, administrators and operators (hand-made content — comments, posts, replies, announcements, service updates, online status changes).

A structured data set of these data points could be helpful in setting up a timeline of the operator's actions and the marketplace website service. The data points mentioned above may not be too significant by

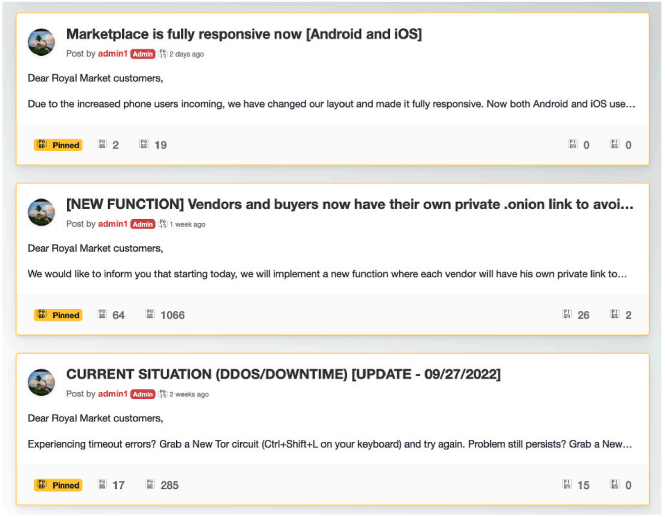

Fig. 2. Forum posts from ASAP Market.

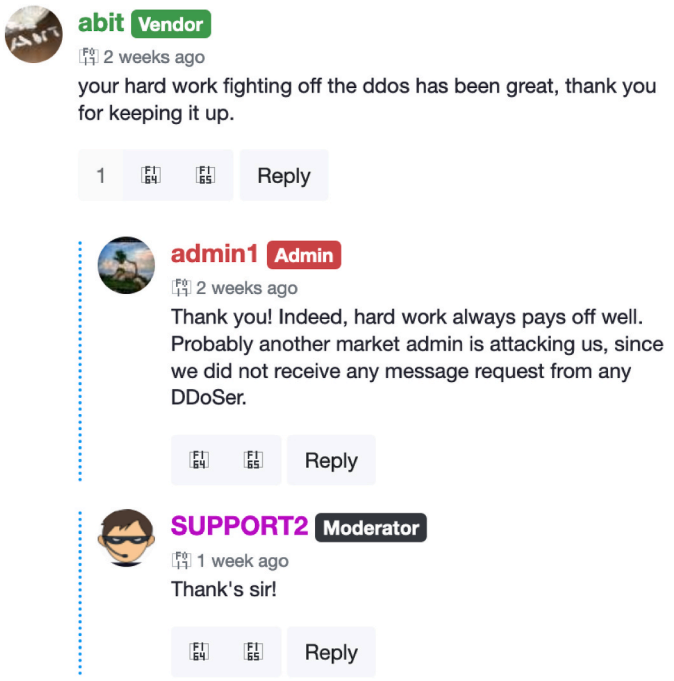

Fig. 3. Forum comments from ASAP Market.

themselves, even though they offer additional insight into the operations of the marketplace. However, correlating these dark web data points with other sources makes them significantly more attractive. Other sources may include real-life events, open-source intelligence (OSINT), human intelligence, lawful interception (LI), network intelligence, and LEA surveillance information.

3.2.1. Use case — operator surveillance

Person(s) of interest, in this case, possible DNM operator(s), are being monitored, and a log of what they do is created. These data could be what this person was physically doing, where they were, and who they were with. The log could also include their online presence at given times (i.e., what they were sharing on social media). Such a log provides the perfect data source for cross-correlation with data about operator actions/service events on the dark web.

3.3. User activity metadata

This method is similar to the previous method, but the targets are primarily users (either vendors or buyers) with different observable activities.

One aspect greatly benefiting from scraping a marketplace frequently enough is being able to derive activity metadata of vendors in addition to gathering product listing contents and account information.

In this scenario, we are focusing not on the contents of the scraping (such as the product being offered or the listed price) but on establishing time frames for events that have occurred. If, for example, we observe a new product listing in the most recent scrape, we can conclude that this listing has been created in the period between this scrape and the previous one. It follows that the higher the frequency of scraping, the better accuracy we have for trying to timestamp such events.

Utilizing such sources of information effectively relies on two conditions: (a) the time frames are small enough to provide helpful activity indicators; and (b) we have enough data samples to establish patterns. A single data point in a month-long time window is not as valuable as five-day-long windows in the same week, which is less valuable than a hundred 15-min windows for several months.

In all cases, the base information we get is that the user has been

active during this period. That implies several other things, such as the user being awake, being at a computer, accessing the internet, etc. If we have a more significant number of events observed, we can use them to figure out more information about the user, such as (a) the user's activity — regular or seemingly random; (b) the user's time zone (based on the observed event timestamps); and (c) coincidence of events and their patterns with other observed activities gathered by OSINT or LI.

The metadata sources we can use are not limited to new product listings. Any content change we observe, such as a counter of sold products, the number of buyers' ratings received, or an update of the last seen timestamp on the user profile page, might provide useful information. What kind of information we might be able to infer from observation metadata differs on a case-by-case (source-by-source) basis.

3.3.1. *Use case — targeted surveillance*

Similarly to the previous method and its use case, we can use collected (meta)data to presume the active presence of the user on DNM. For instance, vendor-specific detected events (such as the listing of new products, as depicted in Fig. 6) can be used to assess a vendor's online presence. The same information can be obtained from the user profile page, which usually also contains an indication of when the user was last logged in to DNM (see Fig. 4).

3.4. *Procurement tracking*

Another way to utilize publicly shared data about products and orders on DNMs is procurement tracking (order fulfilment). Let us look at Figs. 5 and 6 for website view examples that can be used as appropriate data sources. Product reviews, as shown in Fig. 5, have significant importance on darknet marketplaces, and they provide us with the following.

- believable confirmation of fulfilled order,
- timestamp of the confirmation,
- additional information about the order — state of packaging, contents, speed, and potentially more.

A subset of these data points can sometimes be obtained from a single page, as shown in Fig. 6. The figure shows an "activity feed" page that contains all significant events on the marketplace in the near past, such as new product reviews from buyers, newly published products, or user bans with reasons. Such a page on the marketplace website makes web scraping simple and the acquisition of such an exciting piece of information very straightforward.

When user reviews cannot be used (i.e., the marketplace only shows aggregate results), or we simply do not wish to use them, there are alternatives to that approach. The sales counter feature itself can be used to determine a successful product purchase at a specific time. Leveraging automated periodic scraping and content change detection (counter increment in this case), we are able to detect distinct product purchases. Same as the review-based approach; however, heavily dependable on

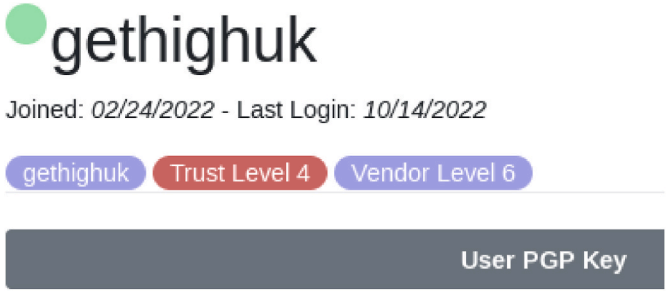

Fig. 4. A random user profile from AlphaBay with join and last login timestamps.

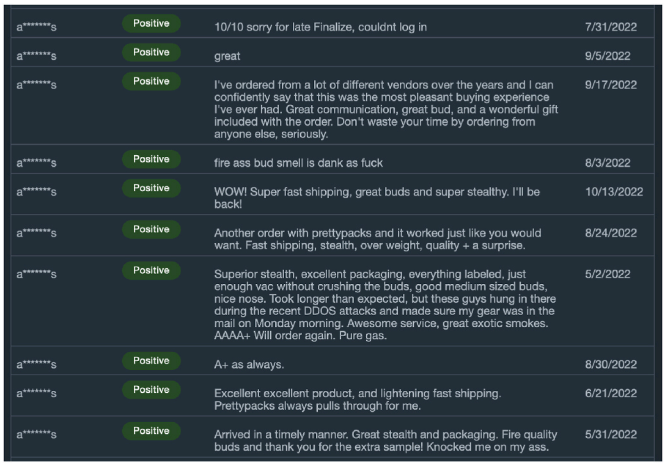

Fig. 5. Product reviews from ASAP Market.

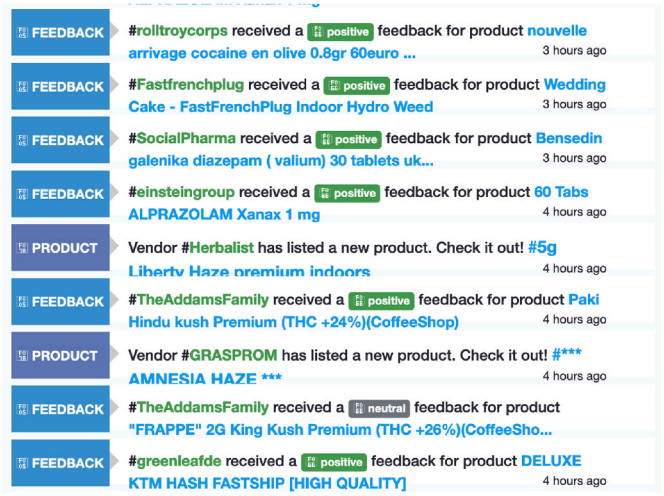

Fig. 6. Marketplace-wide activity feed from Royal Market.

the actual scraping period length. Too long a period results in an imprecise purchase window definition and possibly window merging — a situation when multiple purchases occur/are detected in a single given period.

3.4.1. *Use case — earnings estimation*

When a successful order confirmation from a buyer can be detected and attributed to a specific product (and hence also a vendor), it only takes a value to be assigned to the detected purchase to start estimating earnings. When assigning the lowest possible price of a product to each detected purchase, we can arrive at a reliable absolute minimum revenue estimate. Furthermore, we can generate reasonably valid revenue estimates by applying a reasonable Gaussian distribution model to the prices assigned to the detected purchases.

3.5. *Cryptocurrency transaction correlation*

This method will inform how to use the sales count and product price as correlators for finding the corresponding transaction in the blockchain.

Among the data collected by the *Product Data Retention* method (see Section 3.1 above) is the sales counter, which represents how many trades have already been done via the DNM. On many DNMs, the sales counter is increased by +1 immediately after the purchase, which sends a positive signal to other potential buyers about the product. That

strengthens the vendor's reputation without waiting for the review, which is optional and makes sense after the trade is (un)successfully concluded. If we sample the value of the sales counter for a single product with a reasonable frequency, we can detect the number of purchases between two rounds of web scraping. If the sampling is fast (i.e., it takes just minutes for the web scraping engine to visit the same product page and collect data), then we can detect the purchase more accurately concerning its time of occurrence. This detection cannot be used if the sales counter is unavailable on the product webpage. Moreover, some DNMs show sales counter in the form of aggregated value (e.g., less than 100 sales) rather than accurate one (e.g., 42 sales).

Product price is usually in Bitcoin or Monero. It can also be displayed in fiat currency (e.g., USD, EUR), which is converted to its cryptocurrency equivalent during order placement. Fiat conversion is done according to the exchange rate provided dynamically/statically by the DNM operator.

Once we detect the product purchase via the sales counter within a certain period, we can search for blocks in the corresponding time range and filter out these transactions that match the estimated product price. By conducting this correlation, we obtain a list of transactions where.

- originators, i.e., input address(es) may belong to buyers of the product;
- beneficiaries, i.e., output address(es) may belong to either vendor (in the case of direct deals) or operator (in the case of DNMs offering escrow services).

It is essential to mention that this use case is a heuristic. The list of transaction candidates can have numerous items, where all except one are false positives. The accuracy of this heuristic depends on (a) the value of the purchase price and its uniqueness concerning simultaneous transactions recorded in the blockchain; and (b) the number of blocks in which the transaction is looked up — the more extended time window yields more blocks and potential transaction candidates.

3.5.1. Use case — blockchain identification of vendor and buyer

Let's assume that the agent has detected a single new purchase of cocaine sold by the investigated vendor. The purchase price of 2 g was 171 USD, and money was transferred via an equivalent amount in Bitcoins. The period between two web scraping rounds is 15 min. Therefore, the investigator is searching in two Bitcoin blocks for any transaction having an output address value of 0.00290497 BTC. This heuristic generates three candidate transactions, with average deviance from the targeted price 0.0000000042 BTC. One of these candidate transactions includes the buyer's addresses as inputs and the vendor's address as one of the outputs.

4. Evaluation

This section focuses on the more practical aspects of the proposed approaches. Section 4.1 demonstrates the actual usage of previously presented methods on a real dark marketplace (Monopoly Market) during 2021. Followed by Section 4.2 showing the availability of features described in Section 3 on various popular DNMs at the time of writing this paper.

4.1. Demonstration on Monopoly Market

Monopoly Market (MM) featured the sales counter as well as customer reviews. An example listing from MM can be seen in Fig. 7. Visualizations are mainly based on the data set we previously published (Dolejška et al., 2022). A few results acquired by the previously described methods are shown in the following subsections. Each subsection describes the results from different contexts and methods.

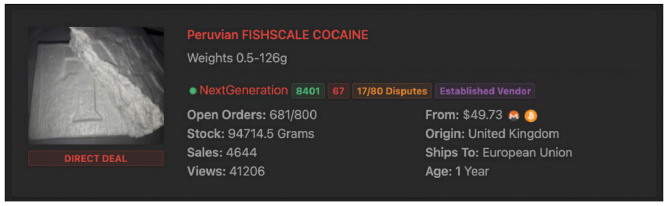

Fig. 7. Monopoly Market product listing. (Dolejška et al., 2022)

4.1.1. Vendor and Product Data Retention

As described in Section 3.1, vendor and product data can be used to analyse activity on the monitored marketplace. Advertised products and active vendor profiles were periodically archived in order to view their changes over time. Fig. 8 shows a number of active vendor profiles on MM over time. Such a metric can tell us about the marketplace ecosystem and how it is developing over time — if new vendors are coming in or if they are leaving. If the metric is available across multiple marketplaces, we can even monitor how the vendors are expanding their business to new places.

A number of products advertised by the active vendors is shown in Fig. 9. The figure only shows a number of products; however, the data set also contains all the necessary details. Knowing which products were available on which marketplaces, at which period and from which vendor also helps in tracking their activities. A cross-analysis between marketplaces is also possible, showing an expansion of new products between marketplaces and vendors over time.

4.1.2. Procurement tracking

We were primarily focused on the sales counter and detection of its change — using techniques described in Section 3.4.

Using the gathered data, we were able to construct a purchase heat map. It can be seen in Fig. 10 and it shows a number of distinct product purchases over time in 15-min windows. The map includes all the products from all the vendors on the market; however, the data set allows both product and vendor-based filtering.

Using the data from the same technique as before, we can take a look at high-level trends between product categories or even competing

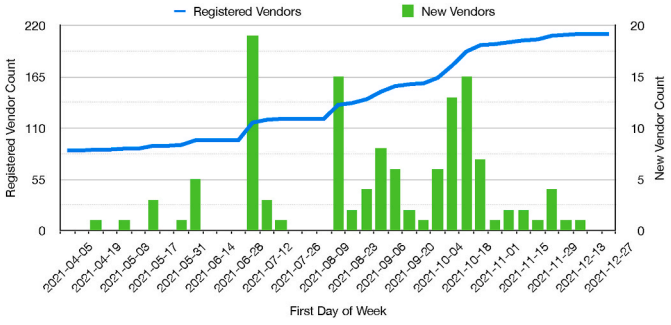

Fig. 8. Active vendor count over time. (Dolejška et al., 2022)

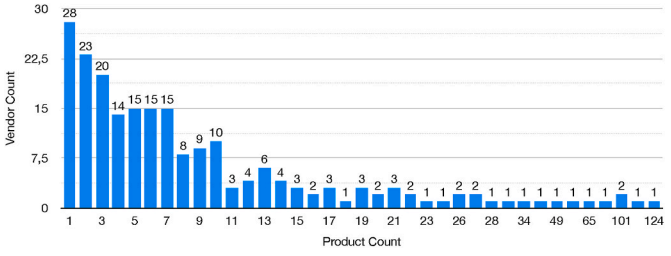

Fig. 9. Counts of products advertised by vendors on Monopoly Market. (Dolejška et al., 2022)

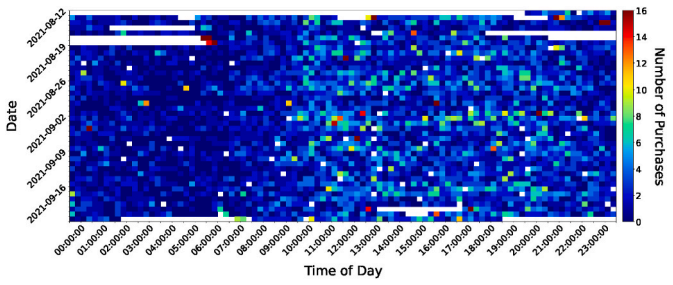

Fig. 10. Count of product sales over time. (Dolejška et al., 2022)

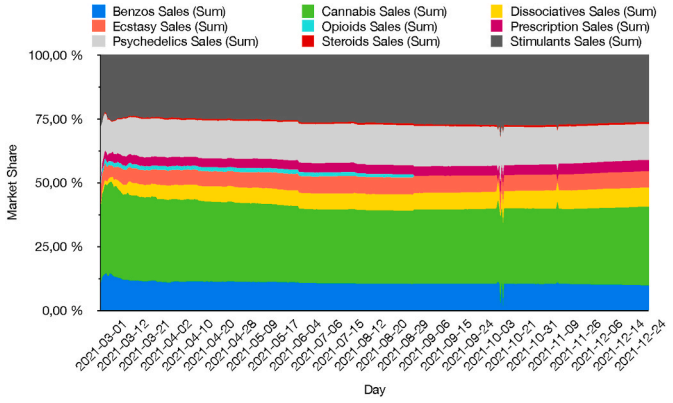

Fig. 11. Percentage of total sales per product category over time. (Dolejška et al., 2022)

products in the same category over time. The category trends are shown in Fig. 11.

4.2. Survey of applicability

We have surveyed various features (properties) of DNMs currently available. This was done to determine the applicability of our proposed methods on DNMs other than Monopoly Market (which our data set is based on).

The features we focused on can be grouped into three categories. The first are features that dictate what kind of information is present on the DNM and therefore can be scraped and archived. Methods that use data and/or metadata from the market are dependent on these features. The second group focuses on how the order and payment processes are carried out. This is useful for methods that try to correlate market activity with cryptocurrency (blockchain) activity. Lastly, we catalogued the types of goods advertised and sold on the market.

We have attempted to access and survey as many DNMs available on popular onion link catalogues[4] as we could. This is a time-consuming process. In our experience, accessing these DNMs over the Tor network is generally slow. Additionally, most DNMs use some kind of DDOS protection which requires solving captchas to gain access, as well as registering an account. The ephemeral nature of DNMs further complicates the process. During our survey, we observed a number of DNMs going offline and becoming inaccessible for several days.

The survey was done manually by accessing the DNMs using the Tor browser. After registering an account, we looked at the available market/product/vendor pages as well as support articles that might contain information about what the particular DNM offers to its users. Our findings can be seen in Table 1.

[4] https://darknetlive.com, https://darkcatalog.com, https://tor.taxi, https://dark.fail.

5. Conclusion

DNMs are subjects of active cybersecurity research supported not only by LEAs but generally by the social need in countries where drug distribution and abuse are considered unlawful.

In this paper, we assess the current state-of-the-art of DNM ecosystem, focusing on web scraping and subsequent (meta)data analysis as crucial investigation techniques. We provided a thorough summary of research papers and, most importantly, data sets containing (meta)data from the long-term monitoring of DNMs (including our data set from a nearly one-year-long observation of Monopoly Market).

We elaborated on trends and evolutionary steps DNMs have been experiencing during the last couple of years. Despite the declining number of active DNMs, total revenue increases yearly. Buyers tend to trust DNMs more as a working platform for the distribution of illicit goods. Nevertheless, due to the successful LEA operations, hacks, and exit scams, DNMs adopt rapidly direct deals as a primary way of exchanging money between vendors and buyers.

This paper's core contribution describes five methods that leverage various periodically collected data from DNMs. Using our data set Dolejška et al. (2022), we have demonstrated these five methods in action. They are able to provide helpful intel about: (a) vendors and evidence of their activity and portfolio; (b) products, including heuristics for the detection of purchases; (c) overall trends and demography for the whole DNM. Last but not least, we have checked the applicability of these methods on numerous other currently trending darknet marketplaces, specifically in the Tor overlay network.

Among the plans for our future work is.

- to further speed improvements of periodic web scraping on our monitoring infrastructure;
- to use other data sets as inputs for the five methods mentioned above and analyze results;
- to focus on heuristic-based DNM purchase correlation with blockchain transactions, with the aim to enhance this heuristic with more blockchain intelligence.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This paper and associated research work was supported by the national project "BAZAR: Building better and safer darknet without walletless marketplaces" (with code VJ01030004) funded by the Ministry of the Interior of the Czech Republic in 2021 and 2022. The authors would like to thank and support a nameless group of LEA investigators who are making our world a safer place every day.

References

Barratt, M.J., Lamy, F.R., Engel, L., Davies, E., Puljevic, C., Ferris, J.A., Winstock, A.R., 2022. Exploring Televend, an Innovative Combination of Cryptomarket and Messaging App Technologies for Trading Prohibited Drugs. Drug and Alcohol Dependence 231. Publisher: Elsevier, 109243.

Branwen, G., Christin, N., Décary-Hétu, D., Andersen, R.M., , StExo, Presidente, E., Anonymous, Lau, D., Sohhlz, D.K., Cakic, V., Buskirk, V., Whom, McKenna, M., Goode, S., 2015. Dark net market archives, 2011-2015. https://www.gwern.net/DNM-archives.

Chainalysis Inc., 2021. The 2021 Crypto crime report. https://go.chainalysis.com/rs/503-FAP-074/images/Chainalysis-Crypto-Crime-2021.pdf.

Chainalysis Inc., 2022. The 2022 Crypto crime report. https://go.chainalysis.com/rs/503-FAP-074/images/Crypto-Crime-Report-2022.pdf.

Chertoff, M., 2017. A public policy perspective of the Dark Web. Journal of Cyber Policy 2, 26–38. https://doi.org/10.1080/23738871.2017.1298643, 10.1080/

23738871.2017.1298643 10.1080/23738871.2017.1298643. publisher: Routledge _eprint.
Christin, N., 2013. Traveling the Silk Road: a measurement analysis of a large anonymous online marketplace. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 213–224.
Darknetlive, 2019. Market announcement: monopoly is open for business. https://darknetlive.com/post/market-announcement-monopoly-is-open-for-business/.
Darknetlive, 2021. PSA: white house market is retiring. https://darknetlive.com/post/psa-white-house-market-is-retiring/.
Dolejška, D., Veselý, V., Pluskal, J., Koutenský, M., 2022. Shedding light on Monopoly: temporal analysis of drug trades. In: Digital Forensics and Cyber Crime. Springer Cham, Boston, MA, United States. https://doi.org/10.1007/978-3-031-36574-4_9, 10.1007/978-3-031-36574-4_9.
Du, P.Y., Zhang, N., Ebrahimi, M., Samtani, S., Lazarine, B., Arnold, N., Dunn, R., Suntwal, S., Angeles, G., Schweitzer, R., Chen, H., 2018. Identifying, collecting, and presenting hacker community data: forums, IRC, carding shops, and DNMs. In: 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 70–75.
EMCDDA, 2018. Darknet markets ecosystem: lifetimes and reasons for closure of over 100 global darknet markets offering drugs. sorted by date. https://www.emcdda.europa.eu/system/files/publications/8347/Darknet2018_posterFINAL.pdf.
EMCDDA, 2020. Special report - COVID-19 and drugs: drug supply via darknet markets. https://www.emcdda.europa.eu/system/files/publications/13042/EMCDDA-report_COVID19-darknet-final.pdf.
EMCDDA, 2017. Drugs and the Darknet: Perspectives for Enforcement. Research and Policy.
EMCDDA, 2022. European Drug Report 2022: Trends and Developments. Publications office of the European Union, Luxemburg.
EMCDDA, Europol, 2020. EU Drug Markets: Impact of COVID-19. Publications Office, LU. https://data.europa.eu/doi/10.2810/19284.
European Commission, 2020. Proposal for a regulation of the EUROPEAN parliament and of the council on markets in crypto-assets, and amending directive (eu). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020PC0593.
Europol, 2021. DarkMarket: world's largest illegal dark web marketplace taken down. https://www.europol.europa.eu/media-press/newsroom/news/darkmarket-worlds-largest-illegal-dark-web-marketplace-taken-down.
Europol, 2023. 288 dark web vendors arrested in major marketplace seizure. https://www.europol.europa.eu/media-press/newsroom/news/288-dark-web-vendors-arrested-in-major-marketplace-seizure.
Financial Action Task Force, 2021. Updated guidance for a risk-based approach to virtual assets and virtual asset service providers. https://www.fatf-gafi.org/media/fatf/documents/recommendations/Updated-Guidance-VA-VASP.pdf.
Hiramoto, N., Tsuchiya, Y., 2020. Measuring dark web+ marketplaces via Bitcoin transactions: from birth to independence. Forensic Sci. Int.: Digit. Invest. 35, 301086. https://www.sciencedirect.com/science/article/pii/S2666281720303887, 10.1016/j.fsidi.2020.301086.
Kermitsis, E., Kavallieros, D., Myttas, D., Lissaris, E., Giataganas, G., 2021. Dark web markets. In: Akhgar, B., Gercke, M., Vrochidis, S., Gibson, H. (Eds.), Dark Web Investigation. Springer International Publishing, Cham. Security Informatics and Law Enforcement, pp. 85–118. https://doi.org/10.1007/978-3-030-55343-2_4, 10.1007/978-3-030-55343-2_4.
Soska, K., Christin, N., 2015. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: 24th USENIX Security Symposium (USENIX Security 15), pp. 33–48.
Spagnoletti, P., Ceci, F., Bygstad, B., 2021. Online black-markets: an investigation of a digital infrastructure in the dark. Information systems frontiers. https://doi.org/10.1007/s10796-021-10187-9, 10.1007/s10796-021-10187-9.
Tai, X.H., Soska, K., Christin, N., 2019. Adversarial matching of dark net market vendor accounts. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, New York, NY, USA, pp. 1871–1880. https://doi.org/10.1145/3292500.3330763, 10.1145/3292500.3330763.
Tsuchiya, Y., Hiramoto, N., 2021. Dark web in the dark: investigating when transactions take place on cryptomarkets. Forensic Sci. Int.: Digit. Invest. 36, 301093. https://www.sciencedirect.com/science/article/pii/S2666281720303954, 10.1016/j.fsidi.2020.301093.
Van Buskirk, J., Naicker, S., Roxburgh, A., Bruno, R., Burns, L., 2016. Who Sells what? Country Specific Differences in Substance Availability on the Agora Cryptomarket. International Journal of Drug Policy, vol. 35. Publisher: Elsevier, pp. 16–23.
Van Buskirk, J., Bruno, R., Dobbins, T., Breen, C., Burns, L., Naicker, S., Roxburgh, A., 2017. The Recovery of Online Drug Markets Following Law Enforcement and Other Disruptions. Drug and Alcohol Dependence, vol. 173. Publisher: Elsevier, pp. 159–162.
van Saberhagen, N., 2013. CryptoNote v 2.0. https://www.getmonero.org/ru/resources/research-lab/pubs/whitepaper_annotated.pdf.