Examining the trends and operations of modern Dark-Web marketplaces

Víctor Labrador
University Carlos III of Madrid
100440595@alumnos.uc3m.es

Sergio Pastrana
University Carlos III of Madrid
Sergio.Pastrana@uc3m.es

***Abstract*—Currently, the Dark Web is one key platform for the online trading of illegal products and services. Analysing the `.onion` sites hosting marketplaces is of interest for law enforcement and security researchers. This paper presents a study on 123k listings obtained from 6 different Dark Web markets. While most of current works leverage existing datasets, these are outdated and might not contain new products, e.g., those related to the 2020 COVID pandemic. Thus, we build a custom focused crawler to collect the data. Being able to conduct analyses on current data is of considerable importance as these marketplaces continue to change and grow, both in terms of products offered and users. Also, there are several anti-crawling mechanisms being improved, making this task more difficult and, consequently, reducing the amount of data obtained in recent years on these marketplaces. We conduct a data analysis evaluating multiple characteristics regarding the products, sellers, and markets. These characteristics include, among others, the number of sales, existing categories in the markets, the origin of the products and the sellers. Our study sheds light on the products and services being offered in these markets nowadays. Moreover, we have conducted a case study on one particular productive and dynamic drug market, i.e., *Cannazon*. Our initial goal was to understand its evolution over time, analyzing the variation of products in stock and their price longitudinally. We realized, though, that during the period of study the market suffered a DDoS attack which damaged its reputation and affected users' trust on it, which was a potential reason which lead to the subsequent closure of the market by its operators. Consequently, our study provides insights regarding the last days of operation of such a productive market, and showcases the effectiveness of a potential intervention approach by means of disrupting the service and fostering mistrust.**

***Index Terms*—Crawler, Web Scraping, Dark Web, Drugs, Dark Marketplaces**

1. Introduction

Nowadays, most of the content on the Internet, estimated at 95% [1] in 2020 and known as the Deep Web, is not indexed by search engines such as Google. Part of these sites are hosted in The Onion Router (Tor) network, which is intended to provide anonymity both to clients navigating the network, and also to service providers by means of `.onion` sites. These sites can be easily accessed through a special browser, which inherently creates a circuit (an anonymous networking path between a source and a destination). This, together with other mechanisms such as encryption, allows users to navigate the so-called Dark Web anonymously, making it hard for third parties to track the network traffic and to locate the actual hosting of `.onion` sites. However, to access these sites, it is necessary to know the URL beforehand, or to use specialized services such as the the Hidden Wiki [2] or custom search engines [3].

Due to anonymity provided [4], the Dark Web has attracted miscreants to carry out different illegal activities. One of the most widespread of such activities is the trading of illicit products, such as drugs, hacking material, weapons or child pornography [5]. The trading in the Dark Web usually occurs in dedicated black markets, some of them specialized in certain products, while other having a general purpose [6]. In order to sale and buy these products, sellers and buyers need to have an account the marketplace. It is indeed possible to create one on each of them, since it is in their interest to increase the number of users and, therefore, potential customers to offer and buy the products.

These marketplaces provide sellers with a platform that preserves their identity and facilitates communication with their customers. Some of the most famous marketplaces by the end of 2021 are: *ToRReZ Market*, *Dark0de*, *DarkFox Market* or *ASAP Market*. A common pattern of all these markets is that the commonly focus on the selling of drugs [7], and also digital goods (e.g., porn, credit cards, pirate (fake) software, hacking material, etc.). Moreover, with the 2020 pandemic, the volume of trading in such marketplaces has increased, possibly due to the lockdown effect [8]. Besides these two categories, other categories which might pose higher risks to society are also prevalent. These are harder to reach and detect, and include the trading of military weapons, hitmen, child pornography, human trafficking or terrorist movements [9]. It is thus important to understand what these activities are, and how they evolve in order to intervene on them.

Unlike the traditional web, it is difficult to index and compile all the existing information in the Dark Web. Given the ample offer of products[1] for sale with a relatively viable access for any user, it is of vital importance to know as much as possible all the information

1. In this paper, we use the term listing and product indifferently since we do not analyse the presence of duplicates.

related to these markets: users, sellers, products, amount of sales, origin of shipments, etc. For this purpose, focused crawlers are used [10], [11], a type of software designed and developed to navigate through the different websites (in this case marketplaces) obtaining and processing all the information (scraping) that exists on these pages. This allows to get offline snapshots of the content and make a subsequent analysis, which allows to know as much as possible about these transactions considered illegal with the intention of reducing or eliminating their activity.

In this work, we present a crawler that connects to the Dark Web marketplaces and scrapes the information, obtaining the data for all the listings (products or services being offered) and the users registered in that marketplace. We use the crawler to conduct an analysis on 6 Dark Web markets. In summary, our main contributions are:

1) We describe a crawler capable of collecting data from Dark Web marketplaces. This crawler deals with modern anti-scraping techniques.
2) We collect more than 123k listings from 6 different markets, and conduct an analysis the data. We confirm that the most prevalent category of products relate to drugs. We also analyse the geography of the sales (i.e., origin and destination of the offered products) and also the vendors.
3) Finally, with an initial attempt of conducting a longitudinal analysis, we select one productive and dynamic market to conduct periodic crawls and get different snapshots over time, in order to understand its evolution. Concretely, we choose *Cannazon* market, due to its large volume of sales [8]. During the period of study, the market suffered a DDoS attack, and then it was suddenly closed. We provide insights on the latest moments of the market, and analyse the relationship of the different variations in the stock and prices of the products with the attack and the closure of the site.

Overall, we believe that our paper provides new insights on the cybercrime operations carried out in the Dark web, as well as potential intervention approaches for law enforcement officers. To foster research on the area, and to allow for reproducibility, we open source the code for the crawler.[2]

2. Background and related work

Darknet markets are online marketplaces hosted on the Dark Web which are usually used for selling illegal goods and services, such as drugs, pornography, digital fraud, or weapons, in a way that transactions are carried out through cryptocurrencies. The analysis of these markets within the Dark Web represents an interesting, yet challenging problem within the field of cybersecurity due to the popularity acquired by these services in recent years with the aim of obtaining illegal products or services [12]. The problem arises due to the anonymity offered by Tor's hidden services which make it difficult for law enforcement to track down all types of criminals.

Different types of analysis has been proposed to find cybercrime activities in the DarkWeb. For example, by exploring its structure [13]. Also, by finding out where the servers are located or tracing users of terrorist forums [14]. In this way, more information about these individuals can be obtained, to deter and stop these activities.

As previously mentioned, the Dark Web consists of a small percentage of the Deep Web, so in order to analyze these markets, it is also important to know in what context it is located. For this purpose, more modern techniques are emerging, providing crawlers with the ability to automatically discover search forms [15] or even crawlers supported by artificial intelligence (e.g., for learning the structure of the site being crawled [16], [17]), focused on obtaining a picture of the real size of the Deep Web and all existing URLs, navigating from one to another.

However, obtaining this structure does not provide information on all the products and services offered on the Dark Web marketplaces and, although there are currently a large number of manuals on how to scrape "normal" websites accessible from any browser anonymously through the Tor browser, there is little information on scraping onion websites belonging exclusively to the Dark Web.

In recent years, considerable progress has been made in the study of the Dark Web, following the first successful takedown of an anonymous marketplace known as Silk Road in October 2013. But this spurred this ecosystem of marketplaces to re-emerge stronger and in greater numbers, with at least 16 different marketplaces reported in 2015 [18]. Thanks to these studies, collections of up to 1.6 TB containing information on 89 marketplaces and more than 37 related forums have been obtained and are now public [19]. Despite the popularity of this dataset in the research community [20] [21], the contained information is outdated. In addition, due to the advancement of these markets and technology in recent years, it has become more difficult to collect this information because it is less accessible, more security measures have been added such as *CAPTCHAs* that are more difficult to resolve and other similar mechanisms that make crawling more difficult [22].

In fact, most of the crawlers found related to the Dark Web focus on obtaining its structure by removing all the links present on a page and accessing recursively until no more new links are found to browse, in order to obtain new hidden Dark Web sites [23] or even companies that offer customized services to obtain this data [24], but there are hardly any recent crawlers focused on obtaining the existing information in each market with the aim of analyzing this data publicly.

In November 2020, OWASP presented a new project called TorBot [25], focused, like the tools mentioned above, on obtaining the largest possible number of URLs by analyzing the relevance of each of them and finally returning the most relevant of them all. It has other features such as obtaining e-mails. However, the storage of the links obtained in a database is not yet implemented,

2. https://github.com/vicviclablab/darkmarkets_crawler

which is a key property required for this work.

In any case, in spite of being found to a lesser extent, in the last two years crawlers have been developed that are more focused on obtaining information on products, categories, users, countries, etc. [26] [27], which is essential if a statistical analysis of these illegal activities is to be carried out. But it is important to note how fast these markets are advancing, with a crawler that obtained a total of 6387 products from a market in 2019 [28] [29], and now in 2021 there are several markets with between thirty and forty thousand products. Additionally, recent works have studied the effect of the COVID'20 pandemic in the Dark Web activities. [30] [31] [32].

While it was relatively easy to scrape a Dark Web marketplace in 2015, today the mechanisms to prevent this have been modernised, making the task considerably more difficult. Because of this, there are hardly any crawlers today that focus on the content of a Dark Web marketplace, but rather on the links present and the relationships between websites as a search engine might do. We thus develop for this work a custom tool that allows to analyse in detail the products and services offered on the black markets of the Dark Web.

3. Methodology

In order to gather all the market information mentioned in Section 2, it is necessary to make a connection just as the Tor browser does, creating a circuit of nodes, using at least one node that serves as an exit node in this circuit detailed in Subsection 3.1. For this purpose, we develop a custom software, that relies on the popular Selenium library for web crawling and scraping, and stores all the content in a relational database (SQLite).

3.1. Software architecture and crawler operation

The proposed system implemented is composed by four components:

- *Connection Manager*, responsible for making the connection to access `.onion` sites through a set of parameters and a proxy with the Firefox browser. It allows to choose an user agent and an exit node, that can be changed at each execution to reduce the chances of detection by the markets. The driver is then obtained through the Selenium library which allows scraping the content from each market.
- *Data collector*. The system modulates the data collection tasks to dedicated functions tailored to each of the different markets, i.e., which adapts to their particular structures. These functions are basically two for each market:
  - Function in charge of obtaining all the links of the products browsing along all the existing categories in the market and saving each of these URLs in a file.
  - Function in charge of accessing all the URLs in the previously mentioned file with a logic to delete the ones that have already been processed and thus keep the pending ones in the file (in order to have a record of which ones have already been processed). Subsequently, it accesses all the fields of each product and stores them in the database.
- *Database Manager*. The system includes a relational database, where the information obtained by the previous component is stored. This module allows to manage the DB connections and to conduct the necessary queries for inserting and selecting the data.

The details on the working procedure of the crawler are shown in Figure 1. As it can be observed, it receives three different input parameters, i.e., the name of the marketplace, the port through which the circuit will be created to access Tor, and the name of the file where the URLs of all the products of that marketplace will be stored.

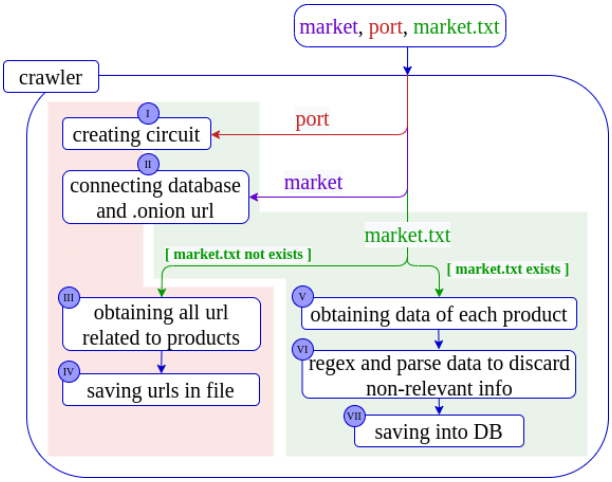


Figure 1: Structure and workflow of the crawler.

The working flow is separated into two main flows: collecting the URLs of all the products to save them in a file (when the file does not exist) and obtaining the information of each product in that file (when the file already exists). The first and second steps are common to both processes, centered on the creation of a circuit (Fig. 1, Step I), which uses the previously mentioned port parameter to, via the Stem library, connect using the Tor control protocol. In the next step (Fig. 1, Step II), the database belonging to each market is linked (creating it if it does not exist) and the driver is obtained with the Selenium library, which is connected through Firefox, being necessary to modify some preferences such as *network.dns.blockDotOnion* to *False* or *network.proxy.sock_remote_dns* to *False*, among others, to be able to access the .onion sites, together with the changes of Step I.

From here, the program flow that occurs when the file with the URLs of the products does not exist (Fig. 1,

red part, left side) is based on the connection to the corresponding marketplace. This is done in such a way that the URL is checked to see which marketplace it belongs to, and whether the scraping function is done. If so, the flow is derived to the specific function of that marketplace and that is where it starts navigating between the product pages with the aim of getting all the existing product URLs at that moment (Fig. 1, Step III). In case a *CAPTCHA* appears or a login is required, the program stops and waits for the *CAPTCHA* to be resolved manually, which then allows the crawler to proceed automatically for the remainder tasks. This human involvement limits the operation of the crawler, but the design is flexible enough in such a way that an automatic *CAPTCHA* solver could be integrated[3]. The frequency with which these *CAPTCHAs* appeared varied slightly from one market to another and did not follow an specific pattern.

Once the market pages have been accessed, the URLs of all the listings are stored in a separate file, passed as a parameter (Fig. 1, Step IV). This prevents missing URLs even in the event of a crash during the crawling process. The URLs for a market are then visited and the products are scraped (Fig. 1, Step V) to collect all the desired information. This logic prevents to lose the information already obtained in case the connection fails. Indeed, once a product is obtained, it is saved in the database (Fig. 1, Step VII) before accessing another one and, in addition, the URL line of the file is removed so as to have a record of the products already viewed. For the vast majority of data obtained from each product, it is necessary to process the information using regular expressions (Fig. 1, Step VI) in order to discard any information that is not relevant for subsequent analysis.

Due to the size of some markets, the crawler works in parallel by means of a script that separated all the product URLs into several files and launched as many tasks as there were files. This way, the required time to obtain the products can be reduced. However, to avoid flooding the target servers with multiple requests (see ethical discussions below) we limit to 6 the maximum number of parallel processes. In addition, a delay of 10 seconds was also added to each requests, to avoid saturating the server and, at the same time, to avoid being recognized as a bot and thus reducing the chances of being challenged by a CAPTCHA. As a result of all the previous steps, a database with all the products for each of the markets is ready for data analysis.

Ideally, once the crawler finishes the first phase, it automatically starts reading the list of URLs and starts working on obtaining the information to avoid products being removed during this waiting time. For marketplaces with high volume of products (e.g., Dark0de), it might takes several hours to obtain all the URLs. In such cases, when finishing with all the existing URLs and starting with the second phase, there may be products that no longer exist if they are removed. These cases are contemplated in the structure of the crawler so that, if a URL in the list is no longer accessible (certain fields do not appear in the scraping such as price, description, etc.), this product is removed from the list and is not taken into account. This is a limitation since the crawler access all products in sequence, which could be overcome by setting parallel crawlers for the two phases. This, however, would increase the load to the server, which might result in the banning of the account and increases the load on the Tor network.

3. We note that the different markets implement various *CAPTCHAs* challenges, which might require different tools. For example, in one of them it was necessary to move a figure to its place, in another one it was required to select the time of an image of a clock or even enter a code that was displayed and hidden letter by letter

3.2. Ethical issues

During our study, we developed crawlers to access various `.onion` sites hosted in the Tor Network. To prevent overloading, and also with the aim at mimicking human behaviour, we have established due time delays during the crawling activities. Also, when various crawlers were run in parallel, they were limited at 6 per time, which is a reasonable amount that could be expected by a human user.

The methodology was designed following ethical standards for academic research in computer science [33], [34]. We also account for particularities of dark market research [35]. We did not compromised into the servers, nor interacted with the community. In our experimentation, the CAPTCHA were filled manually, though the methodology allows for the inclusion of automatic solvers if the research can be conducted under proper legal and ethical agreements. The data collected has been publicly released by their owners. Most of these data refers to product information, including description, prices, and shipping locations. The only potential personal data collected are the usernames of the vendors, which are expected to be pseudonyms since they are operating in anonymous markets. We have not attempted to deanonymize such users. The dataset is kept in a database stored in our servers with appropriate security mechanisms, including encryption and restricted access control.

4. Experimental evaluation

This section first presents the steps taken to obtain standardised data to work with and how this data has been processed. Secondly, the analysis of the data is presented, both of the products and of the sellers as well as the markets in which they are found.

4.1. Dataset characterization

The operation of the crawler described above has been applied to the following six markets, obtaining six databases with the characteristics shown in the Table 1. We also show the total time required to get a full snapshot of the dataset for each market (we note that some markets are crawled with parallel processes). The period in which these databases were collected was from August to November 2021.

Market	Products	Size	Time elapsed
ASAP*	11,788	61 MB	4h 20 m
Cannazon*	2,535	19 MB	2h 45 m
Dark0de*	46,016	258 MB	61h 50 m
DarkFox*	25,829	109 MB	30h 20 m
Global Dreams	136	76 KB	6 m
ToRReZ*	37,072	162 MB	44h 20 m
TOTAL	123,376	609 MB	143h 41 m

TABLE 1: Characteristics of the databases obtained. * The process for these markets has been parallelised by splitting the product list into several processes.

Once a database has been obtained for each of the markets, and in order to conduct a general analysis, all the data is centralized in a single database. To this end, the six databases have been merged to obtain a total of 123,376 products. For each product we collect the information shown in Table 2, as provided by the sites. We note, however, that not all markets have the same information, causing some fields to have a default *"null"* value. In order to validate that the information stored in the DB is complete, once the crawler has finished we manually check the number of products listed per page and the number of pages that the market has. Then, we compare the total number of products with the ones stored in the DB, confirming that the information of all the products has been correctly obtained in all cases.

timestamp	shipping_from
market	shipping_to
category	seller
subcategory	seller_profile
name_of_product	seller_fingerprint
quantity	seller_rating
price	seller_number_ratings
views_of_product	seller_number_of_sales
product_rating	

TABLE 2: Fields collected for all the products.

4.2. Preprocessing and normalization

In order to be able to make a correct analysis of the data, it is necessary to have certain values in the same format, for example, the price of a product. The problem is that information offered by each market is not the same in all of them, therefore showing the price in different forms, such as "50 USD", "$50", "50 €", "50 EUR", etc. All the values of the column have been taken and only the numerical values have been saved, converting dollars to euros using the exchange rate of the day on which the conversion was made (early December 2021).

In the case of quantities, there was a similar problem, where each market and even each seller represented the quantities in a different way, among which there were some such as "2g", "2gr" or "2G", all three being the same quantity. In addition to, as expected, having very different units such as millilitres, micrograms, milligrams or kilograms, among others. Therefore, to normalize the information, we have converted to grams and stored in a new column "quantity_gr" so as not to lose information. In the case of products that were not sold by weight, such as hacked PayPal accounts or pornography, the decision taken to represent them in the database has been with a "-1" to avoid confusion in the analysis, as it would really be 1 account, but in this case it would be wrong with 1gr. Finally, once all the data have been collected in a database and the data in certain columns have been normalized, it can be analysed.

4.3. Data analysis

Market popularity. The first analysis made is intended to compare the popularity of each one of the markets. This popularity is reflected in the number of sellers willing to offer their products, and subsequently, in the number of products in each market. This can be perceived in Fig. 2 as a percentage. The markets that are currently offering a wider catalog are the following three: *Dark0de*, *ToRReZ* and *DarkFox*. These markets are far away from others focusing on a only product. Examples of this could be *Cannazon* or *Global Dreams*, which only trade with drugs.

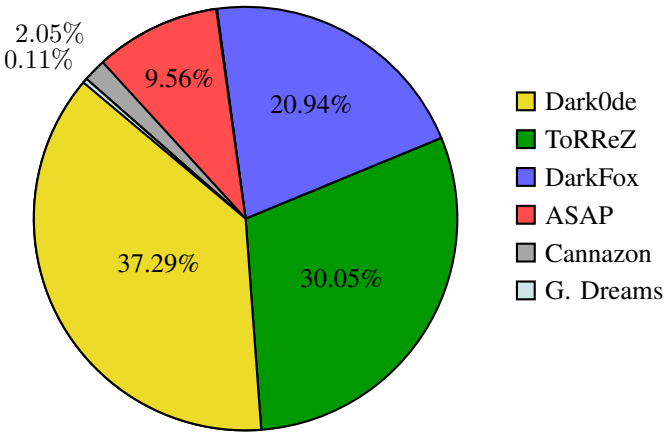


Figure 2: Number of listings offered by markets

Type of products. Regardless of the main target of each of these marketplaces, it is interesting to know which products are most offered by sellers as a whole, without differentiating between marketplaces, in order to find out which type of product or category is currently most offered on the Dark Web. For this purpose, based on the "category" column of the database, all similar categories have been manually grouped according to the name of each category to reduce them to a total of 9 shown in Table 3.

As can be seen in Table 3, the most popular category on the Dark Web marketplaces is drugs (mostly cannabis). Within the category of fraud and counterfeiting there are many products related to hacked accounts such as PayPal accounts with balance, credit card or similar. Although it does not appear as a category, pornography is very present in these markets, making up a large percentage of the digital products category. Although it is a very small percentage of the total number of products offered, we also show the category COVID-19 due to being a topic of relevance at the time of this analysis was conducted (end of 2021). We found products such as alleged COVID-19

Category	**Total listings**	Dark0de	ToRReZ	DarkFox	ASAP	Cannazon	G. Dreams
Drugs	**61,062**	23,095	18,476	12,544	4,329	2,482	136
Fraud & Counterfeit	**26,850**	7,243	11,222	5,203	3,182	0	0
Digital Products	**13,820**	8,733	0	3,368	1,719	0	0
Tutorials and e-books	**12,350**	3,297	4,470	3,341	1,242	0	0
Software & malware	**6,484**	2,803	1,555	827	1,299	0	0
Services	**1,921**	87	1,349	485	0	0	0
Jewelry	**598**	598	0	0	0	0	0
COVID-19	**12**	11	0	0	0	1	0
Others	**279**	149	0	61	17	52	0
Total	**123,376**	**46,016**	**37,072**	**25,829**	**11,788**	**2,535**	**136**

TABLE 3: Categories of listings in the Dark Web markets.

medicines and vaccines, or fake COVID-19 vaccination certificates.

Market vendors. With regard to the number of sales, as can be seen in Table 4, there are vendors who, during the few years that these markets have been open, have sold tens of thousands of products. One single seller has made a total of 29,105 sales, all of which are drugs. It is noteworthy that, the vendors with the highest number of sales have been in *Cannazon*, one of the markets with the fewest products on offer. However, in marketplaces with a much larger catalogue, such as *DarkFox*, the number of sales by its vendors is much lower, with its best-selling vendor ranking 90th out of all vendors in all marketplaces. This is similar for the other marketplaces, such as ASAP, whose best-selling seller is ranked 262nd.

Table 4 also shows that the top sellers have drugs as their main product, with hash and cannabis standing out. This can give an idea of what is most commonly bought on Dark Web marketplaces, and although there are more sophisticated products such as jewellery or hacked (supposedly real) bank accounts, it is drugs that predominate. In terms of seller ratings, the sellers with the highest number of sales on the *ToRReZ* marketplace barely reach 30% positive ratings, raising suspicions about the reliability of this marketplace. In the case of *Dark0de* the number of ratings compared to the number of sales is very low, where only 3.50% of the buyers leave a rating in the case of the biggest seller. This is not the case on *Cannazon*, where 87.64% of buyers left a rating for the top seller, which has 97.97% positive ratings, making it a much more reliable marketplace for buyers.

In order to analyse cross-market presence, we analyse the vendors of different markets. From the 2760 total sellers across all marketplaces, 490 of them (17.75%) appear on at least two markets. The way we determine that two sellers are the same person is through their nickname. It may be the case that a seller decides to have different nicknames in each market, but that the same nick is different person in each market is much less likely. We found one seller on all marketplaces (Global Dreams has been excluded as it has no information related to the sellers) with 5765, 293, 136, 79 and 51 sales on *Cannazon*, *ToRReZ*, *Dark0de*, *ASAP* and *DarkFox* respectively. Continuing with this analysis of sellers, the nicks found were very similar to those on the normal website, as is the case of a vendor focused on pornography and present on all marketplaces. If a search on his nickname is done on the normal web using any search engine, all the entries are related to porn sites, and it is highly likely that he is the same person.

As is the case of one seller where 91.16% of his sales are on *Cannazon*, it is also the case with many other sellers such as one vendor in particular who has 10795, 42 and 11 sales on *Cannazon*, *ToRReZ* and *Dark0de* respectively and therefore has 99.51% of his transactions on the first one. This is yet another example of how, with the same seller offering the same products, the sales volume in the *Cannazon* market is much higher than in the other markets, thus making the number of products in the catalogue unrelated to the total number of sales.

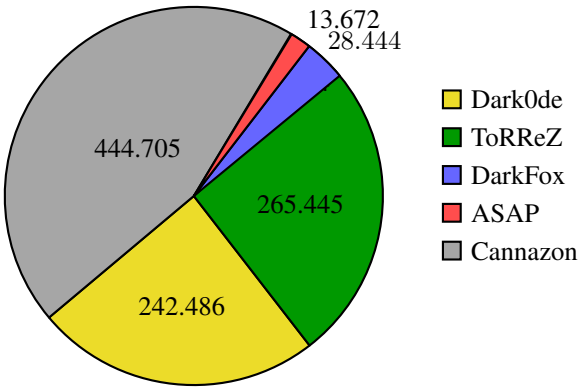


Figure 3: Sales of each market

Number of sales. Based on the large volume of sales in the *Cannazon* market shown in Table 4, the number of total sales by each market has been analyzed and, as can be seen in Fig. 3, although *Cannazon* has a catalogue approximately 20 times smaller than *Dark0de*, it has almost twice as many sales. It is similar with *ToRReZ*, but both *DarkFox* and *ASAP* lag far behind in terms of sales. In this case, the *Global Dreams* market has not been included because its sales number is 0.

Geography of the trading. The origin and destination of each of the products offered by the marketplaces is considered to understand the origin and destination of shipping. It might occur sellers spoof this information, but we believe this is unlikely since otherwise they would be scaming potential sellers, which would then be reflected in

Top sellers information						
#	Market	Seller	Category	N° of sales	Rating of seller	Number of ratings
1	Dark0de	Seller1*	Drugs	**29105**	98.00%	1020
2	Cannazon	Seller2	Drugs: Edibles, Hash, Weed, Other	**20575**	97.98%	18032
3	Cannazon	Seller3	Drugs: Hash, Weed	**18891**	98.45%	16364
4	Cannazon	Seller4	Drugs: Concentrates, Edibles, Other	**15589**	98.56%	12805
5	Cannazon	Seller5	Drugs: Edibles, Hash, Weed, Other	**13979**	98.28%	12562
6	Cannazon	Seller6	Drugs: Concentrates, Hash, Seeds, Weed, Other	**11981**	98.41%	10691
7	Cannazon	Seller7	Drugs: Concentrates, Hash, Weed, Other	**11567**	98.92%	10483
8	Cannazon	Seller8	Drugs: Hash, Weed	**11217**	99.49%	10009
9	Cannazon	Seller9	Drugs: Weed	**10795**	98.19%	9040
10	Cannazon	Seller10	Drugs: Concentrates, Edibles, Weed	**10660**	99.05%	9307
11	Cannazon	Seller11	Drugs: Concentrates, Edibles, Weed	**9593**	99.97%	8887
12	Cannazon	Seller12	Drugs: Hash, Weed	**9343**	98.20%	8478
...	...	...	...	...	...	...
22	Dark0de	Seller13	Digital Products	**5957**	97.80%	181
...	...	...	...	...	...	...
25	ToRReZ	Seller14	Digital Products, Drugs, Fraud	**5759**	27.78%	50
26	ToRReZ	Seller15	Fraud	**5338**	28.14%	47
...	...	...	...	...	...	...
90	DarkFox	Seller16	Digital Products, Fraud	**2201**	94.20%	-
...	...	...	...	...	...	...
262	ASAP	Seller17	Digital Products	**822**	94.69%	207

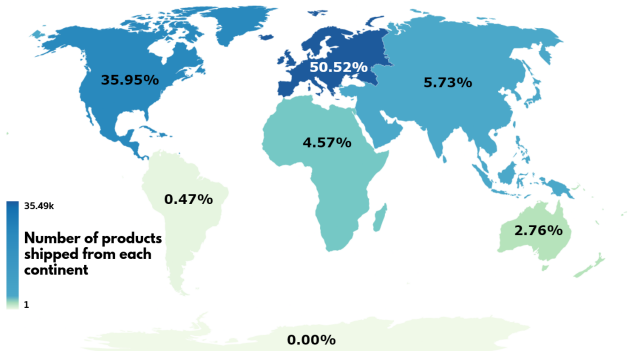
* We have anonymized the sellers' nickname for ethical reasons

TABLE 4: Information related to top sellers of all markets.

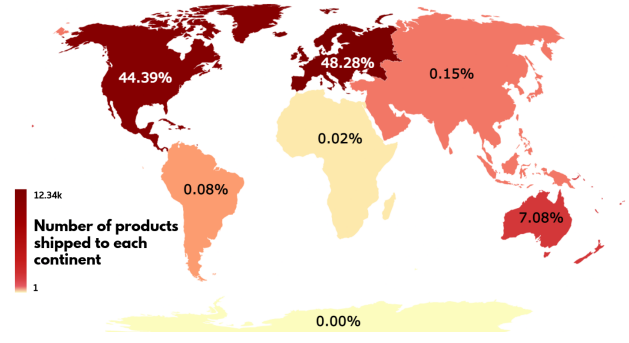
their reputation. Also, due to the different granularity offered by the marketplaces (countries, regions, continents, etc.), we have decided to process locations at the continent level, as shown in the map in Figure 4. In the case of the map 4a, it is shown that the majority of products (35,494) come from Europe, accounting for 50.52% of the total number of products whose origin is known. The second continent with the largest supply of products is North America with a percentage of 35.95%, followed by Asia with 5.73%.

As far as delivery is concerned, not all sellers choose to ship worldwide, limiting themselves to certain countries or continents. Of the 123,376 total products analysed, 59.79% allow shipping worldwide. Among the remaining products which have shipping limitations, 12341 allow shipping only to Europe as shown in Map 4b, followed by the continent of NA with a total of 11,346. For the remaining continents there are not a large number of products with limited shipping.

Therefore, the greatest traffic of these products, both in shipping and in receiving, takes place mainly in Europe and North America. The percentage of products whose possible destination is NA is 44.39%, compared to 35.95% of products that originate in this continent. However, there are 48.28% of products that allow shipment to Europe, compared to 50.52% of products that originate in this continent. The difference in percentages is greater on the American continent, which shows a greater reluctance to ship products from there, perhaps due to greater vigilance on the part of security forces.



(a) Number of products originating in each continent.



(b) Number of products with possible destination on each continent.

Figure 4: World map related to origin and possible destinations of products.

5. Case-study: the final days of Cannazon market

During our initial investigation, we decided to conduct temporal analysis in order to better understand the evolution of the markets. Thus, we extended the data collection by the crawler for one particular case, i.e., the marketplace *Cannazon*. We chose this market since it is the market where we observed most activity and variation, despite its reduced product offer (since it is focused on drugs). Thus, can be crawled entirely in a short time, while still giving insights on the market evolution. Concretely, the *Cannazon* market was crawled on a weekly basis. Our initial idea was to obtain these different snapshots for a larger period. However, we observed two main drawbacks. First, an alleged denial of service attack on *Cannazon* shot the market down for some weeks. Then, it was subsequently closed after two weeks since the attack. We show these events in Figure 5.

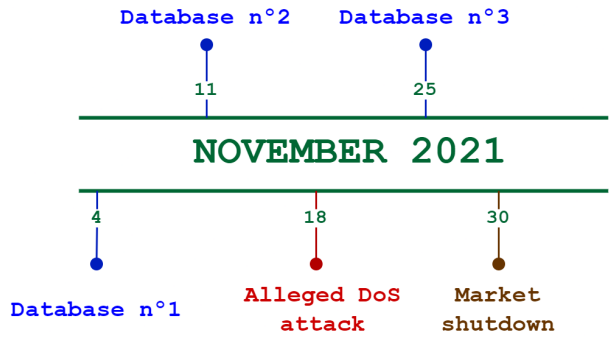


Figure 5: Timeline with databases obtained from Cannazon.

Due to the aforementioned issues, we were able only to get three snapshots of the market, i.e., in November 4th, in November 11th, and in November 25th (just 5 day before the closure of the markets). While these data do not allow us to conduct a longitudinal analysis of the market evolution, we still observe an interesting **variation in the market** between the different weeks. As it can be observed in Table 5 the total number of products vary in one week. For example, a total of 173 products were removed in the first week of study, which account for around 7% of the stock. This might be due to the products being sold, or the vendors removing the items for any other reason, since we don't have evidence of actual trading. We also observe a total of 331 new products in one week (increasing the stock by 13% respect to the previous snapshot). Again, this suggest the market was in expansion.

We observe, however, a different pattern during the period of attack, i.e., between our second an third snapshots (a span of two weeks). Regarding the selling of products, a total of 349 products were removed (13.5%), which is almost the double of the trading from the first week, which seems reasonable. However, different from the first week, in the two weeks where the market claimed being victim of an attack, only 248 products were added (9%). Thus, the amount of products added was severely affected by the attack.

Our analyses confirm that *Cannazon* was a very active market, where in just three weeks 579 new products have been added, and 522 were removed. This also shows that, compared to other markets with tens of thousands of products, this small market manages to renew its catalog to the point where there is a variation of its products in three weeks that reaches 22.47%. Despite this factor of success, the operators of the market decided to close it, justifying their choice due to the attacks being received.

Date	No. Products	Removed products*	Added products*
4th November	2,519	-	-
11th November	2,677	173	331
25th November	2,576	349	248

* Compared with previous week

TABLE 5: Evolution in number of products in Cannazon.

In order to understand not only the stock of the *Cannazon* market, but also the economic variance, we look at the possible **price variation** over time. For this analysis, it was decided to take into account those products that existed both in the snapshot of the first week (4th November) and in the snapshot of the third week (25th November), being a total of 2060 products.

During price analysis, we realized that some products experienced high variations, e.g., one product modifying its price from 933,33€ to 113,33€. We also observe that the quantity offered was changed, from 100gr to 10gr. Thus, the price reduction is justified. Accordingly, those products whose quantity has changed from the first to the third week have not been taken into account, being a total of 25 products.

Of the remaining 2035 products whose quantities are the same, 996 products (49% aprox.) have not changed its price. From the 1039 products whose prices have changed (51.05% of the total), we find the following values:

- The total average price variation is -7.73€, representing an average price reduction of 1.21%.
- Considering only price reduction, the average of products decreased in -107.72€, representing an average reduction of 26.24%.
- The average of those products whose price has increased was +1.84€, which means an increase of 2.00% on average.

Our analysis suggest that the products that price reduction is done at a larger percentage, and those products becoming more expensive were increased in a less significant way. It can also be seen that in this market, from the first week (at full capacity), to the week after suffering a DDoS attack, the price trend has been downward on average. This shows how the attack also affected the market economy. A possible reason is due to the loss of trust from buyers on the market, which results in a lower number of purchases. These, in turn, might have been compensated by a price reduction by the sellers, leading to a final decision of the administrators to close the market.

6. Conclusions

The study of Dark Web markets is currently a challenging subject due to challenges for scraping data, which often makes that information used for research is obsolete or not available for large-scale analyses. In this paper, we describe a tool which allowed us to collect a database with recent data. The final database contains data from 6 different marketplaces, with a total of 123,376 listings, 2,760 sellers and almost one million sales made according to the information collected from these pages.

We confirm that the most traded products are drugs, especially cannabis and its derivatives: cannabis, seeds, etc. This is represented in the market with the highest number of sales, which is *Cannazon*, dedicated exclusively to this type of drug. Indeed, 9 of the top 10 sellers in this study belong to this market as well as all of the top sellers are focused on the sale of drugs.

Also, we observe that these markets rapidly evolve with societal needs. Even where they were few COVID vaccines in the market, several types of vaccines were already being offered, most probably fake ones. However, this gives an idea of the rapidity with which product catalogues are updated, offering practically any type of product. Finally, according to the research performed on product locations, there are continents from which exports are much higher than others, such as Europe or North America. While more than half of the products allow worldwide shipping, the other half are limited to certain countries or continents. Considering these limitations, the continent with highest flexibility (i.e., allowing for product reception) is Europe.

An interesting insight from our analysis regards the analysis on the *Cannazon* market. We have observed that this was one of the most active market during our initial phase of research, and thus decided to conduct a longitudinal analysis. However, the attack suffered by the market might have lead to a lack of trusts from sellers and buyers, ultimately leading to the decision of the operators to close the market. While we do not have factual evidence of this connection, this case study shows that activities that damage the reputation an trust of the users on a market might serve as a potential intervention approach for law enforcement.

This approach can be applied to future market research, both to obtain data in new markets and to update data in existing markets. In the future, this tool could be integrated with artificial intelligence and natural language processing techniques in order to further automate the scraping of the data, obtaining much larger databases and therefore being able to analyze with greater precision all the data in the Dark Web.

Acknowledgments

As part of the open-report model followed by the Workshop on Attackers & CyberCrime Operations (WACCO), all the reviews for this paper are publicly available at https://github.com/wacco-workshop/WACCO/tree/main/WACCO-2022. This work is supported by the Spanish grants ODIO (PID2019-111429RB-C21, PID2019-111429RB) and the Region of Madrid grant CYNAMON-CM (P2018/TCS-4566), co-financed by European Structural Funds ESF and FEDER, and Excellence Program EPUC3M17.

References

[1] S. Nazah, S. Huda, J. Abawajy, and M. M. Hassan, "Evolution of dark web threat analysis and detection: A systematic approach," *IEEE Access*, vol. 8, pp. 171 796–171 819, 2020.

[2] "The hidden wiki," Jul 2021. [Online]. Available: https://en.wikipedia.org/wiki/The_Hidden_Wiki

[3] "Ahmia," Sep 2021. [Online]. Available: https://ahmia.fi/

[4] P. Ranakoti, S. Yadav, A. Apurva, S. Tomer, and N. R. Roy, "Deep web amp; online anonymity," in *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, 2017, pp. 215–219.

[5] S. He, Y. He, and M. Li, "Classification of illegal activities on the dark web," in *Proceedings of the 2019 2nd International Conference on Information Science and Systems*, ser. ICISS 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 73–78.

[6] O. Cherqi, G. Mezzour, M. Ghogho, and M. El Koutbi, "Analysis of hacking related trade in the darkweb," in *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2018, pp. 79–84.

[7] A. Baravalle, M. S. Lopez, and S. W. Lee, "Mining the dark web: Drugs and fake ids," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, 2016, pp. 350–356.

[8] "Covid-19 and drugs: Drug supply via darknet markets," European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Tech. Rep., 2020.

[9] M. W. Al Nabki, E. Fidalgo, E. Alegre, and I. de Paz, "Classifying illegal activities on tor network based on web textual contents," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 35–43. [Online]. Available: https://aclanthology.org/E17-1004

[10] T. Fu, A. Abbasi, and H. Chen, "A focused crawler for dark web forums," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 6, pp. 1213–1231, 2010.

[11] S. Pastrana, D. R. Thomas, A. Hutchings, and R. Clayton, "Crimebb: Enabling cybercrime research on underground forums at scale," *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 2018.

[12] A. Biryukov, I. Pustogarov, F. Thill, and R.-P. Weinmann, "Content and popularity analysis of tor hidden services," in *2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, 2014, pp. 188–193.

[13] M. Singh, "Deep web structure," *IEEE Internet Computing*, vol. 6, no. 5, pp. 4–5, 2002.

[14] R. Ehney and J. D. Shorter, "Deep web, dark web, invisible web and the post isis world," *Issues In Information Systems*, 2016.

[15] I. Hernández, C. R. Rivero, and D. Ruiz, "Deep web crawling: a survey," *World Wide Web*, vol. 22, no. 4, pp. 1577–1610, Jul 2019. [Online]. Available: https://doi.org/10.1007/s11280-018-0602-1

[16] L. Jiang, Z. Wu, Q. Feng, J. Liu, and Q. Zheng, "Efficient deep web crawling using reinforcement learning," *Advances in Knowledge Discovery and Data Mining*, p. 428–439, 2010.

[17] J. Lu, Y. Wang, J. Liang, J. Chen, and J. Liu, "An approach to deep web crawling by sampling," in *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, 2008, pp. 718–724.

[18] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem," in *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, Aug. 2015, pp. 33–48.

[19] Gwern, "Darknet market archives (2013–2015)," Jul 2015. [Online]. Available: https://www.gwern.net/DNM-archives

[20] R. Munksgaard, J. Demant, and G. Branwen, "A replication and methodological critique of the study "evaluating drug trafficking on the tor network"," *International Journal of Drug Policy*, vol. 35, p. 92–96, 2016.

[21] J. Demant, R. Munksgaard, and E. Houborg, "Personal use, social supply or redistribution? cryptomarket demand on silk road 2 and agora," *Trends in Organized Crime*, vol. 21, no. 1, pp. 42–61, Mar 2018. [Online]. Available: https://doi.org/10.1007/s12117-016-9281-4

[22] K. Turk, S. Pastrana, and B. Collier, "A tight scrape: Methodological approaches to cybercrime research data collection in adversarial environments," in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2020, pp. 428–437.

[23] Fleigst, "Fleigst/darkweb: Darkweb scraping," Sep 2021. [Online]. Available: https://github.com/fleigst/DarkWeb

[24] WebScrape, "Dark and deep web data scraping," Sep 2021. [Online]. Available: https://webscraping.us/dark-web-data-scraping

[25] P. S. Narayanan, R. Ani, and A. T. L. King, "Torbot: Open source intelligence tool for dark web," in *Inventive Communication and Computational Technologies*, G. Ranganathan, J. Chen, and Á. Rocha, Eds. Singapore: Springer Singapore, 2020, pp. 187–195.

[26] R. Rawat, A. S. Rajawat, V. Mahor, R. N. Shaw, and A. Ghosh, "Dark web—onion hidden service discovery and crawling for profiling morphing, unstructured crime and vulnerabilities prediction," in *Innovations in Electrical and Electronic Engineering*, S. Mekhilef, M. Favorskaya, R. K. Pandey, and R. N. Shaw, Eds. Singapore: Springer Singapore, 2021, pp. 717–734.

[27] M. Ball, R. Broadhurst, A. Niven, and H. Trivedi, "Data capture and analysis of darknet markets," *SSRN Electronic Journal*, Mar 2019.

[28] B. AlKhatib and R. Basheer, "Crawling the dark web: A conceptual perspective, challenges and implementation," *Journal of Digital Information Management*, vol. 17, no. 2, p. 51, 2019.

[29] B. Alkhatib and R. S. Basheer, "Mining the dark web: A novel approach for placing a dark website under investigation," *International Journal of Modern Education and Computer Science*, vol. 11, no. 10, p. 1–13, Oct 2019.

[30] A. V. Vu, J. Hughes, I. Pete, B. Collier, Y. T. Chua, I. Shumailov, and A. Hutchings, "Turning up the dial: The evolution of a cybercrime market through set-up, stable, and covid-19 eras," in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 551–566. [Online]. Available: https://doi.org/10.1145/3419394.3423636

[31] A. Bracci, M. Nadini, M. Aliapoulios, D. McCoy, I. Gray, A. Teytelboym, A. Gallo, and A. Baronchelli, "Dark web marketplaces and covid-19: Before the vaccine," *EPJ Data Science*, vol. 10, no. 1, 2021.

[32] A. Bracci, M. Nadini, M. Aliapoulios, I. Gray, D. McCoy, A. Teytelboym, A. Gallo, and A. Baronchelli, "Dark web marketplaces and covid-19: The vaccines," *SSRN Electronic Journal*, 2021.

[33] D. R. Thomas, S. Pastrana, A. Hutchings, R. Clayton, and A. R. Beresford, "Ethical issues in research using datasets of illicit origin," in *Proceedings of the Internet Measurement Conference (IMC)*. ACM, November 2017.

[34] D. Dittrich, M. Bailey, and E. Kenneally, "Applying ethical principles to information and communication technology research: A companion to the Menlo Report," U.S. Department of Homeland Security, Tech. Rep., Oct 2013.

[35] J. Martin and N. Christin, "Ethics in cryptomarket research," *International Journal of Drug Policy*, vol. 35, pp. 84–91, 2016.