Beyond the 'Silk Road': Assessing Illicit Drug Marketplaces on the Public Web

Richard Frank and Alexander Mikhaylov

Abstract Criminals take advantage of internet communications to amplify the impact of their actions and to form international criminal networks. At the same time, vast amounts of information generated by their online activities have become available for analysis. Open source web intelligence is a valuable methodology for understanding and responding to these new global criminal phenomena. Collecting data from websites, social media platforms and online discussion forums enables researchers, investigators and policy-makers to study and to develop appropriate responses to emerging threats. Automated web intelligence tools such as web crawlers can be used to extract relevant information from target websites and to map the threat landscape of criminogenic environments online. For the study presented in this chapter, we used our web-crawling software to download contents of 28 Russian online marketplaces for illicit drugs. Drug names, types, prices, quantities and geographical locations of sales were extracted and mapped to identify drug trafficking hotspots. Findings indicate such marketplaces can operate due to the ability of their clients to pay anonymously with virtual currencies (specifically Bitcoin and Qiwi) and to deliver the drugs through non-contact methods. This type of service is available in all large cities within Russia and provides to the seller with a safer and more anonymous alternative to "street-level" purchases. The method described in this study can be used to investigate and to prioritize online threats according to their location and severity.

Keywords Online drug trafficking · Anonymous payments · Novel psychoactive substances

R. Frank (✉) · A. Mikhaylov
School of Criminology, Simon Fraser University, Burnaby, BC, Canada
e-mail: rfrank@sfu.ca; amikhayl@sfu.ca

M. A. Tayebi et al. (eds.), *Open Source Intelligence and Cyber Crime*, Lecture Notes in Social Networks, https://doi.org/10.1007/978-3-030-41251-7_4

1 Introduction

The proliferation of internet connectivity throughout the world has facilitated the growth of legitimate and illegitimate economies alike. Globalization has made the international geographical borders, and the physical distance between buyers and suppliers, largely irrelevant. The digital economy has allowed consumers to pay bills, make money transfers and purchase goods with unprecedented convenience, whether locally or internationally. Online payments first emerged in the late 1990s in the form e-gold services, owned by Gold & Silver Reserve Inc. (G&SR), which allowed users to open online accounts and purchase gold (and other precious metals) in grams, then transfer them to other accounts as payments [1]. Whether used for legitimate business opportunities, or shady gambling websites and virtual Ponzi schemes, e-gold kickstarted the virtual currency industry [1]. Competitors quickly emerged, and the competition within the e-currency industry forced market actors to innovate and to offer increasingly varied types of exchange services for converting traditional fiat currency into virtual currency for online transfers/payments. The key feature which came to define the success or failure of a virtual currency is the ease of depositing money into an online account [1]. Increased regulatory and law enforcement attention to e-gold and novel digital currencies in the mid-2000s has resulted in traditional banking systems, especially in the United States, distancing themselves from e-currencies [1]. However, this merely meant e-currency companies would operate outside jurisdictions with strict regulations.

E-currencies and mobile payments often avoid anti-money laundering (AML) legislation as virtual currencies are not emitted or controlled by banks and as a result these organizations are not subject to the same regulations [2]. With the advent of e-commerce, the goods and services offered on underground markets have become increasingly diverse. While initially focused on trading in stolen credentials, stolen financial data and hacking, a new sector of the underground economy rose to prominence in 2011 with the creation of Silk Road—the first crypto-market for illicit drugs, with many successors to follow [3]. While a European survey showed that most illicit drug purchases among young people still relied on traditional retail distribution models, the United Nations Office on Drugs and Crime (UNODC) reported 90% of UNODC member states identified the internet as a source of drug trafficking [4]. It is especially true in relation to "novel psychoactive substances" (NPS), also known as "legal highs", which are chemically distinct analogs of traditionally used drugs not yet legally regulated.

Despite the rapid rise of crypto-markets, internet-based drug trafficking still represents only a small fraction of the world drug trade [5]. Traditional drugs remain the dominant part of international drug trafficking despite the rising popularity of NPS. Market diversification, however, has encouraged polydrug abuse, which poses risks to users as increasing numbers of different and previously unseen substances become available for purchase. Opioid users, for example, are at risk of consuming much stronger and deadlier fentanyl instead of traditional heroin, as synthetic opioids are more economically efficient to produce [4]. For instance, 260 substances classified as NPS were reported in 2012, and that number grew to 483 in 2015, with approximately 80 novel psychoactive substances securing a stable presence on the global drug market. Synthetic drug production is not tied to a particular locale, as these substances do not rely on extracting chemicals from specific plants and can be produced and distributed worldwide from anywhere [4].

Along with the emergence of NPS, “new payment methods” (NPM) have gained traction in legitimate and illegitimate economies alike. NPM include, but are not limited to, prepaid cards, mobile payment services and internet-based payment services which serve as an alternative to traditional payment methods, or offer an innovative way of transferring value between individuals and organizations [6]. For the study presented in this paper, we considered how NPM, specifically mobile payments and crypto-currencies, are used to enable the purchasing of illegal drugs online in a safe and secure fashion. We used our automated web data collection tool, called The Dark Crawler (TDC), to analyze advertisements from illicit drug marketplaces on the public web to assess business practices of this new emerging sector of the illicit drug market. TDC collects webpages from target websites, and then, using customized rules, extracts user-specified content from the pages [7–9]. However, these marketplaces frequently implement strategies to prevent the type of web-analysis we were trying to undertake. Specifically, they utilized Distributed Denial of Service (DDoS) attack protection and captchas. This required modifying our software to be able to satisfy these requirements and allow for (mostly-) automated data-capture. In this study, we used our modified TDC to analyze 28 online drug markets in order to identify types of substances sold, prices, payment methods used and geographic distribution of drug sales.

First, we provide a brief literature review on international drug trafficking and identify conditions which lead to the emergence of markets for new psychoactive substances, and review research on data capture from the internet. Then we outline our data collection methodology, considerations for data capture, followed by an analysis of the data we captured from online drug marketplaces. Finally, the results and limitations are discussed as the paper is concluded.

2 Literature Review

2.1 Drug Prohibition and Drug Market Evolution

The United Nations (UN) played a major role in the internationalization of drug prohibition. The 1961 Single Convention on Narcotic Drugs represents a major milestone of drug regulation which was widely adopted throughout the world [10]. The Single Convention was heavily influenced by the United States due to the American role in the creation of the UN. All 198 nations which signed the Single Convention implemented generally similar drug laws, criminalizing cultivating and refining drugs such as opioids and cannabis. The Single Convention was supplemented by the addition of psychotropic substances (e.g. LSD) to the list of prohibited drugs in the 1971 Convention on Psychotropic Substances [10]. Adherence to the Single Convention differs throughout the world, from lenient drug policies (e.g. Netherlands and Portugal), to stricter hardline approaches (e.g. the UK and Russia) [10–12]. Cannabis, opiates and amphetamine-type stimulants (ATS) are the three most traded and consumed types of drugs today [4].

Worldwide criminalization of traditional drugs has eventually produced a partial market shift to novel designer drugs, also called “legal highs”. Producers of such substances can get ahead of legislators and distribute psychoactive substances which are not yet criminalized. Some of the most known novel psychoactive substances are “spice” (synthetic cannabinoids) and “bath salts” (synthetic cathinones) [13–15]. “Spice” refers to an herbal mixture with added synthetic cannabinoids typically marketed as incense, which is sold online and in some countries in specialty stores such as tobacco shops. It first appeared in Europe and the United States in the early 2000s, signaling the rise of the market for “new psychoactive substances” (NPS) [16, 17]. “Spice” is used as an alternative to marijuana, as users reported effects similar to those achieved by smoking cannabis [17]. Along with different drug consumption patterns, emerging NPS markets have resulted in new distribution models. “Spice” and “bath salts” are available for purchase online on specialized websites that can be located through any search engine [15]. The internet has enabled easy access to outlets which sell prescription and non-designer illegal drugs due to regulatory difficulties inherent in controlling businesses which are dispersed among different jurisdictions [17]. Bath salts are typically advertised online as a legal alternative to amphetamine-type stimulant drugs. In the early 2000s, these substances were sold in retail stores owing to their legality at the time, but following their prohibition in 2010s in the US and Europe, the internet has become the primary source of distribution for bath salts and similar new psychoactive substances (NPS) [15, 18, 19]. In the US, synthetic cannabinoids were not widely known before 2009, but by 2010 the interest in “herbal incense” and other names for synthetic cannabinoids skyrocketed, based on internet searches [16].

“Hot spots” of illegal drug trade have emerged on the internet due to the anonymity, global reach, and availability of impersonal payment methods which do not require client identification [20]. Silk Road, the archetypal crypto-market for illicit drugs, was structured similarly to eBay, providing a platform where any vendor can sell their product and any user can make purchases. This business model was made possible due to the ability of clients to make anonymous payments with the crypto-currency Bitcoin, while Silk Road itself was hosted in the encrypted Tor network to maximize anonymity and to minimize risks to participants. The business model pioneered by the Silk Road was revolutionary because it allowed users to access vendor reviews and detailed information about the product, which served to mitigate distrust in an environment of uncertainty [21–25]. Customers of Silk Road have listed a number of reasons for using the marketplace, such as access to a wider range of drugs, convenience, confidence in highly rated sellers and lower prices [21]. Additionally, users found desirable the lack of necessity to turn to street dealers to acquire drugs [26]. User feedback for products they received typically described quality (purity) of the product, delivery speed and packaging stealth [26, 27]. Convenience, high variety and low risk were the key reasons why internet users may purchase recreational drugs online.

While crypto-markets have received extensive attention from journalists, academics and law enforcement alike, there is another form of internet-mediated drug trafficking that has emerged in post-Soviet states, termed "noncontact drug dealing". Like crypto-markets, this method relies on an online storefront where advertisements for drugs are displayed, but the delivery method is different, as well as the fact these marketplaces are hosted openly in the public web and require no registration to access. Instead of shipping, noncontact drug dealing marketplaces utilize pre-arranged drug stashes located throughout the market's area of operation. Once the payment for drugs has been received, marketplace operators transfer instructions on how to locate the stash to clients as an instant message. The seller and the buyer never interact except online, and the buyer makes a money transfer using one of the new digital payment methods with minimal identification measures. This method is discussed in Russian-language literature on online drug trafficking, as well as a Financial Action Task Force (FATF) report, which emphasize the abuse of virtual currencies such as Qiwi, WebMoney and YandexMoney for making these illegal transactions [28–30]. The ability of these marketplaces to operate openly is possibly explained by bureaucratic and legislative obstacles encountered by Russian law enforcement in policing online drug dealing [31–33].

2.2 Noncontact Drug Dealing

The "noncontact" method of drug dealing has gained prominence due to the ability of individuals to use "new payment methods" with minimal identification measures. By communicating with drug vendors over the internet through discussion forums, online storefronts and marketplaces, drug users are able to procure prohibited substances without ever meeting the seller. The buyer and the seller agree on a place, which is used as a hiding spot for drugs, and a time for pickup. Once the buyer has made the payment, they receive instructions on where to find the stashed drugs [34]. Noncontact drug dealing appears to be predominantly used in Russia and neighboring Commonwealth of Independent States (CIS) countries. The Financial Action Task Force (FATF) report which included a description of the noncontact drug dealing method used an example that referred to Russia and Tajikistan [34]. Noncontact drug dealing has been described extensively in Russian academic journals. Drug market actors have started using mobile payment providers, such as Qiwi, WebMoney or YandexMoney, which enable money transfers between individuals [29]. A source in Russian law enforcement claimed that hand-to-hand drug dealing has become virtually non-existent and drug traffickers have largely moved on to noncontact drug dealing [35].

The essence of the noncontact drug dealing method can be described as follows: (1) the buyer makes an order online (through a website or an instant messenger);

(2) the buyer then makes a payment with a virtual currency or as a mobile money transfer; and (3) the buyer receives instructions as an SMS or an instant message on where to find the stash with drugs [36]. This method of drug distribution became commonplace because it significantly reduces the risk of becoming the target of buy-and-bust operations [30]. The current study analyzed 28 online drug marketplaces all of which used the noncontact drug dealing method. The data pertaining to geographic distribution of online marketplaces, drug types, prices, amounts and payment methods were extracted to evaluate criminal risks posed by the noncontact method of drug dealing coupled with the ability to make anonymous payments.

2.3 Data Capture from Web

The information age brought about persistent connectivity, and with it came the ever-present digital cataloging of our activities—whether legitimate or illegal. Taking advantage of these "digital traces" allows researchers to study online activities of criminals in their natural environment—so-called "convergence settings", rather than relying on potentially unreliable self-report data in surveys or interviews from a subset of "unsuccessful" criminals who were apprehended [5, 37]. Manual data collection is a labor-intensive process as it may require researchers to analyze hundreds and thousands of webpages which need to be saved locally, coded, and finally subjected to analysis. Despite this, manual data collection can provide large and rich datasets, which proved valuable for an in-depth analysis of online criminal activities [5, 37]. For example, Buskirk and colleagues sampled a crypto-market called Agora, at the time the largest "dark net" market, by manually saving webpages and extracting the data with a Microsoft Excel VBA macro [38]. This method allowed authors to automatically categorize nearly 80% of listings, which shows that manual data collection is not necessarily an obstacle to efficient internet-based research [38].

Automated data collection relies on software that creates local copies of the target content hosted online, a process called "mirroring", and extracting the relevant data from saved webpage content in a process called "scraping" [39]. Also known as web crawling, this process has typically been focused on online social networks such as blogs and forums, and also websites, often for consumer research and marketing purposes. In a review of research on web crawling most of studies were published in early 2000s and were concerned with conventional content such as cooking recipes and movies in the surface web [40]. The context of these studies is significantly different from criminal content hosted on the public web or the "dark web", where site administrators may take active steps to protect the website contents from access and indexing by automated means. The original Silk Road used cookies that allowed users to stay logged in for a week before expiring, which enabled researchers to bypass captchas while crawling the website [41]. Unlike the first iteration of Silk Road, Silk Road 2 did not use a captcha or require manipulating website cookies to perform data capture, enabling researchers to produce what the authors claims is a complete crawl of the website [23]. This approach however was criticized due to the instability of websites hosted in the Tor network undermining automated data collection methods which possibly result in incomplete datasets [38].

Web crawling, or “mirroring”, is typically performed in the following fashion: starting with a single webpage, the program indexes all hyperlinks within it then follows those links, repeating the process for each subsequent webpage within limits set by the user [39]. Ready-made solutions capable of mirroring webpages exist, such as HTTrack [23]. However, HTTrack and similar software will only mirror webpages, without analysis, thereby requiring subsequent steps for data cleaning and conversion of the unstructured data into structured data. Scraping can also be performed by separate programs, called scrapers, which can parse webpages to identify relevant content, such as usernames attached to forum posts or drug names [39]. Custom-written web crawlers tend to have more functionality than freely available solutions, such as the crawler DATACRYPTO which was specifically created for studying drug listings on the original Silk Road. This crawler both mirrored and scraped the webpages, building a database of drug listings, vendor information and buyer feedback. Creating a custom web crawler from scratch requires significant investments and may not be feasible for all researchers [5, 39].

For this study, the contents of target websites were downloaded using a custom-written web crawler and scraper called The Dark Crawler (TDC). TDC has previously been used to study hacker forums and extremist websites and to identify terrorist and child exploitation content on the dark web [9, 42–44]. In addition to automated data collection, TDC allows integrating other analytical tools for studying the extracted content. For example, natural language processing (NLP) supplemented with parts of speech (POS) tagging were used to calculate sentiment scores for posts on hacking forums that reference attacks against critical infrastructure such as government facilities, banks and hospitals [42]. Sentiment analysis showed discussions of prominent topics, such as financially motivated attacks against banks, or exploitation of vulnerabilities in government systems, corresponded to actual data on these incidents—what forum members discuss are types of attacks that are typically carried out [42]. Understanding the threat landscape of such attacks can help potential targets prepare for them and mitigate the damage. Another advantage of using an automated web crawler is the lack of necessity to manually review potentially disturbing media to establish their criminal content. For instance, identifying child exploitation and terrorism-related images is possible through image hashing, where each image is assigned a hash value that can be compared against law enforcement databases of known illegal images [9].

2.4 Studying Online Drug Trafficking

Being the most notorious drug marketplace, the Silk Road attracted much attention from academics. This drug marketplace was one of the first to be studied in-depth before its shutdown by the authorities. The number of methods and approaches used was quite varied, from qualitative analyses of attitudes towards online drug purchases, to analyzing trends among drug listings by saving snapshots of the website for a given period [17, 38]. By collecting vendor names and PGP keys (cryptographic sequences used to authenticate a user) through an automated web crawler, researchers were able to describe organizational structure of online drug distribution networks [22]. In order to evaluate the impact of Silk Road's closure on the "darknet" drug trafficking economy, the number of vendors on competitor websites was analyzed through approximately one month period. Despite the shutdown of a major distribution hub, the drug market actors were able to quickly adapt and to relocate to competing markets, as demonstrated by an explosive growth of the number of vendors on them [38]. Multiple studies categorized the illicit goods offered on these marketplaces, ranging from drugs to pornography, stolen data and weapons [23, 24]. We chose to focus on extracting and analyzing sales data – specifically drug names, prices, quantities and regions where they were available. This methodology allows to present an overview of the noncontact drug dealing problem and its extent, as well as to consider implications of this method for combatting drug trafficking internationally.

Furthermore, as stated previously, noncontact drug dealing is a topic of extensive discussion among Russian-speaking academics and law enforcement professionals [28–30, 35, 36, 45–48]. However, the studies are primarily concerned with organizational aspects of noncontact drug dealing [28, 29, 36, 46–48], as well as investigative methods and practices [29, 46, 48]. Few studies attempt to characterize the drug market facilitated through these online marketplaces in terms of drug types, prices and volumes beyond general trends (e.g. a 9.5% increase in consumption of synthetic drugs between 2013 and 2015) [35, 36, 45]. By leveraging data collection capabilities of the web crawler, we were able to provide an overview of the market based on the publicly available drug listings data [23].

3 Methods

As this is an exploratory study, the intention was to identify market prices, the scale of the problem (i.e. how widespread non-contact drug dealing is), and which payment methods were used to pay for the drugs. Specifically, the goal is to systematically go through entire online websites and extract the required content so the drug prices can be analyzed. To do this, 28 websites were selected (Sect. 3.1). However, before data could be captured via our existing The Dark Crawler (TDC) infrastructure, several challenges had to be overcome as the target websites had implemented techniques to specifically prevent automated bots from entering the site. After these changes were implemented (Sect. 3.2) we were able to capture the required pages (Sect. 3.3), and defined rules (Sect. 3.4) which allowed us to extract the drug prices, locations and volumes into an analyzable dataset.

3.1 Data Selection

Exploratory research led us to public Russian drug forums as a starting point, where advertisements of drug markets in our sample were discovered. A clear pattern emerged among the websites found—their designers relied on largely the same webpage template with minor variations such as background images. Additionally, these websites were named similarly and hosted on the same top-level domain *.biz* (for example, *narco24.biz, kumar24.biz*, etc.). Although multiple domains were used, *.biz* appeared to be the most prevalent—searches for “24 biz” proved to be the most fruitful, while searches for “*24 pw*” or “*24 cc*” yielded less relevant results. Overall, over 47 open drug market websites were identified by searching “24 biz” from 100 search hits spread out on 10 pages of the results from the Russian search engine Yandex. Inclusion criteria were established, where an open drug market website had to (1) be public (accessible without registration) and (2) display drug advertisements. Between June 2016 and August 2016, only 28 out of 47 (59.5%) drug market websites remained available long enough to be captured, whereas 19 (40.5%) had to be omitted as no advertisements were displayed.

3.2 Data Collection Challenges

TDC has been used to crawl webpages and extract structured data from them (for examples, see [9, 42–44]), however the websites that were picked for this study differed from previously studied websites in that these were protected by DDoS browser checks.

There are multiple commercial solutions available to defend against DDoS attacks, for example, by shielding the target website with proxy servers that distribute the incoming traffic between themselves and balance the load, or requiring the browser to solve JavaScript calculations before the page is displayed. If this ability to solve JavaScript calculations is not present, then it is highly likely that the requestor is an automated web crawler. The online marketplaces for illegal drugs sampled for the current study utilized DDoS protection offered by a major commercial provider also relied on by legitimate businesses. By integrating the ability to bypass DDoS browser checks into the TDC, we have significantly expanded the pool of websites the data can be collected from. TDC initially was relying on simple GET requests executed in parallel via multiple threads to retrieve multiple HTML pages simultaneously (Fig. 1). However, while this method does retrieve the HTML served by the server, it does not interpret (i.e. execute) any of the content on it. Any JavaScript code that is required does not run (because it is not even retrieved, unless it is embedded into the HTML), thus the checks that run against the browser would fail. To get around this problem, the GET method was replaced by the customizable web-browser engine Chromium. See Fig. 2 for a high-level overview of the structure of TDC.

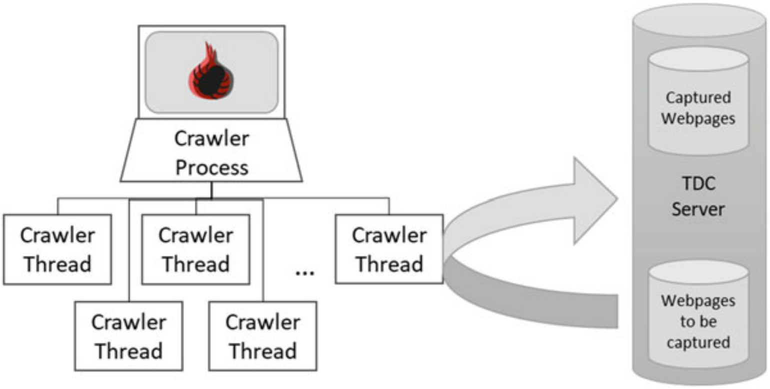


Fig. 1 The Dark Crawler high-level overview (original)

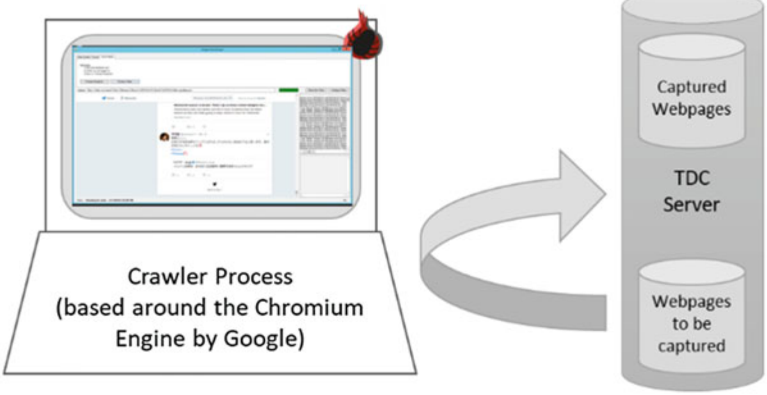
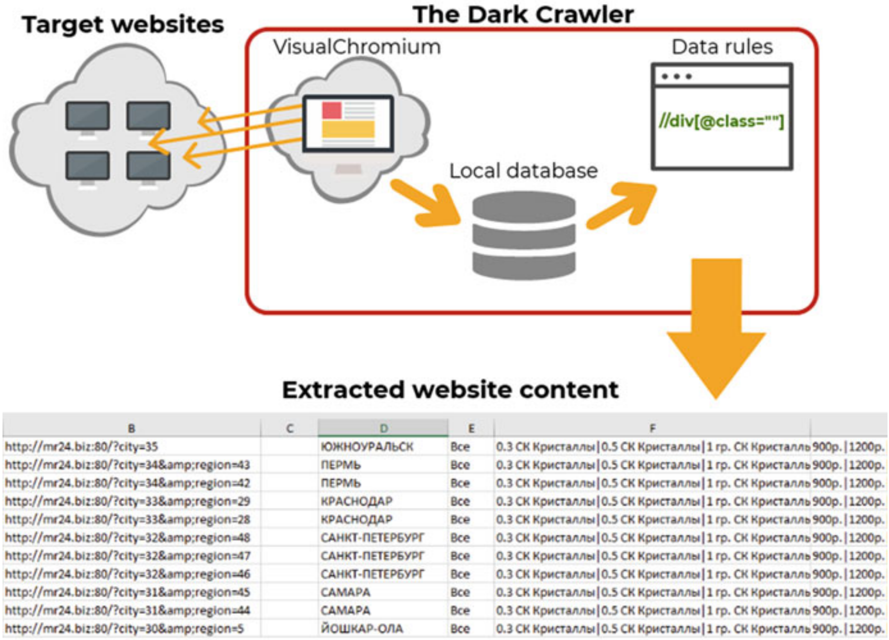

Crawler Process (based around the Chromium Engine by Google)
Captured Webpages
TDC Server
Webpages to be captured

Fig. 2 The Dark Crawler high-level overview (modified)

To capture the data from these websites, the webpage-retrieval engine used by TDC was replaced with an actual browser capable of being automated. In this fashion, the automated browser, called VisualChromium, could access and render the webpages just like any other browser would, in the process solving any JavaScript required. When the browser-checks were done, and the webpage loaded, the automation kicked in and the HTML retrieved was analyzed. Some pages required the solving of captchas to confirm the user is human and the retrieval is not an automated request. As pages are being retrieved, TDC checks if each page meets certain conditions, such as specific text (not) being present. If these conditions are met, TDC moves onto the next page, and if not, the data capture stops and awaits user action. A condition which was present on "normal" pages but not captcha pages allowed the Crawler to recognize captcha pages and to wait for user action before



B	C	D	E	F	
http://mr24.biz:80/?city=35		ЮЖНОУРАЛЬСК	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=34&amp;region=43		ПЕРМЬ	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=34&amp;region=42		ПЕРМЬ	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=33&amp;region=29		КРАСНОДАР	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=33&amp;region=28		КРАСНОДАР	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=32&amp;region=48		САНКТ-ПЕТЕРБУРГ	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=32&amp;region=47		САНКТ-ПЕТЕРБУРГ	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=32&amp;region=46		САНКТ-ПЕТЕРБУРГ	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=31&amp;region=45		САМАРА	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=31&amp;region=44		САМАРА	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.
http://mr24.biz:80/?city=30&amp;region=5		ЙОШКАР-ОЛА	Все	0.3 СК Кристаллы|0.5 СК Кристаллы|1 гр. СК Кристалль	900р.|1200р.

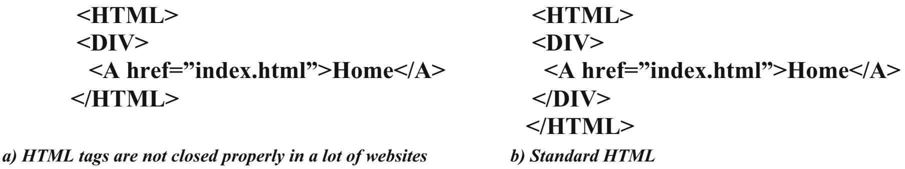
Fig. 3 A summary of the data capture process

continuing. In the future, automated captcha solving will be investigated, but for the sites downloaded in this study, the websites did not use captchas, and thus this was not an issue.

After the “mirroring” process was complete, the captured data was extracted based on user-specified data rules (see Fig. 3 for an overview, and Sect. 3.4 for more details). All rules were applied to all data, producing a .csv format file compatible with Microsoft Excel. Each column represents a data element (e.g. drug names) and rows represent single web-pages. Particularities of data collection and using the Dark Crawler are discussed below.

3.3 Final Data Collection

The Dark Crawler is an automated data collection tool which is able to download entire contents of webpages utilizing user-specified rules. Given a list of webpages, the Crawler will traverse them and parse them apart. The process is as follows:



```
<HTML>
<DIV>
  <A href="index.html">Home</A>
</HTML>
```

a) HTML tags are not closed properly in a lot of websites

```
<HTML>
<DIV>
  <A href="index.html">Home</A>
</DIV>
</HTML>
```

b) Standard HTML

Fig. 4 Real-life HTML vs the standard. (**a**) HTML tags are not closed properly in a lot of websites. (**b**) Standard HTML

- For each webpage within the Queue, repeat until the Queue is empty
  - the Crawler gets a webpage from the queue of webpages to retrieve;
  - it retrieves the webpage, resulting in an HTML string;
  - the webpage HTML is cleaned up, with invalid (or unclosed) tags fixed (see Fig. 4 for an example);
  - the HTML is parsed according to data capture rules and relevant information is identified, such as links or images
  - each link within the HTML is added to the queue of future pages to be retrieved
- Rules are applied to the cleaned HTML to extract user-defined pieces of data (in this case, drug names, prices, amounts, etc.) into a spreadsheet (see Sect. 3.4 for details).

As the captured data is stored in a structured database, the data extraction process can be repeated after the data capture process, and the data extracted in multiple ways, e.g. by searching for specific keywords or extracting certain types of data elements (e.g. drug names). The resultant information is then saved in the form of a matrix, with each webpage in a row, and each extracted data-element in a column. This data can then be subjected for further analysis by, for example, geo-mapping the distribution of drug sale locations.

3.4 Data Extraction

Landing pages of drug marketplaces in our sample were structured in a similar way. Each drug advertisement was placed inside a container that showed drug name, price, quantity, whether the goods were in stock, and a "Buy" button which would take the user to a page with transaction details. An example of such a webpage is shown in Fig. 5. Each webpage element can be targeted for capture by defining data rules. Some of the websites in the sample modified the default webpage template (e.g. by restructuring and renaming elements of the webpage), however the overall structure was consistent across all 28 sites studied. TDC downloaded each page, cleaning up unnecessary data, resulting in a structured table containing data from these webpage elements. This enables the user to quickly extract all the required information from a webpage, significantly reducing the amount of manual work required to preserve these data.

The Dark Crawler identifies relevant information within the webpage by relying on data rules specified by the user. A data rule is a {*path*, *pattern*} combination, which uses the XPaths standard for *path* querying trees and the XQuery standard for the *pattern*s, see Sects. 3.4.1 and 3.4.2 respectively for more details. This process allows selecting elements from a webpage to be downloaded and stored in a database.

Using Paths to Select Webpage Elements

A hierarchy of branches from the root makes up a path, which follows the *XPaths* standard, where a path results in zero or more nodes in a tree. To identify the drug name which is priced at “900” in Fig. 5, the Crawler identifies distinct containers on the webpage, traversing series of branches (*div, ul, li,* and *div)*, where the path would be “*/html/body/div/section/ul/li/div*”. Figure 6 shows how this process identifies 4 drug names located on this webpage, one per container. To select a specific drug name from the list, the path is modified to target one of the <*div*> tags under <*li*>. The resultant rule then would be “*/html/body/div/section/ul/li/div[2]*” which directs the Crawler to select the second *div* tag from the branch. By iterating through all the *div* tags in this fashion, all prices can be extracted for all the drugs on the webpage. A similar strategy is used to extract all other required content.

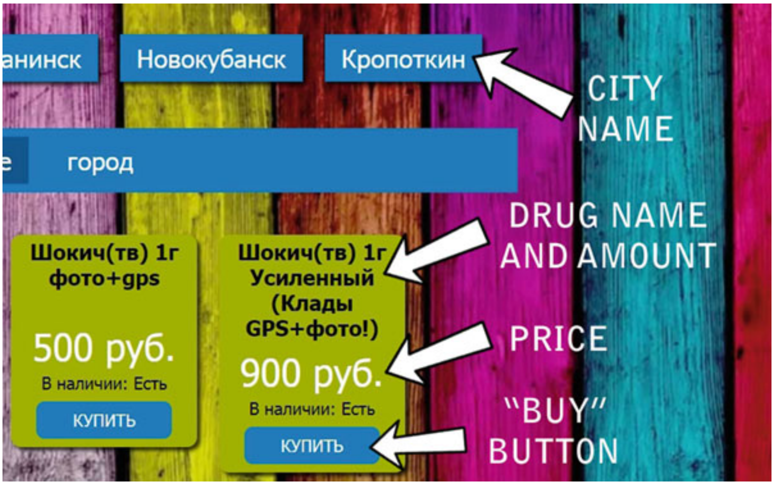
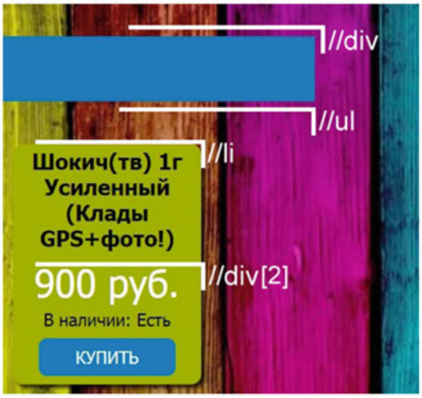
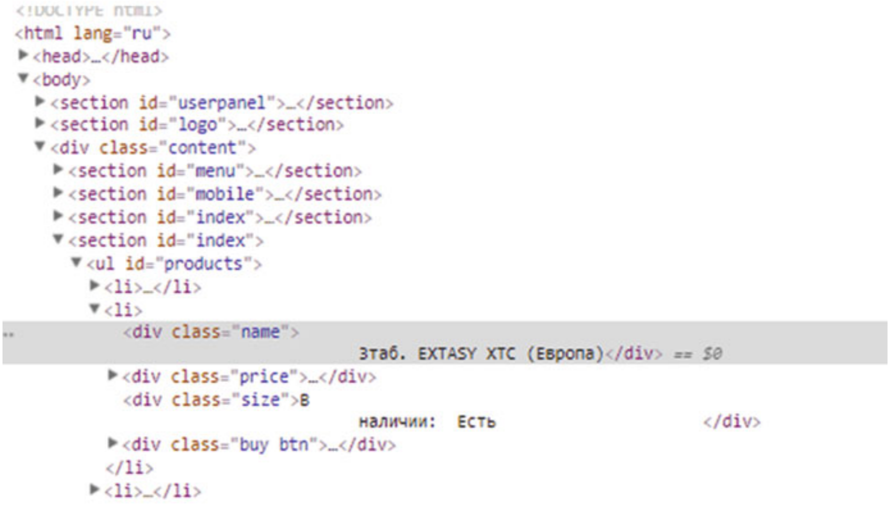


Fig. 5 A typical drug marketplace landing page with advertisements displayed



a) An illustration of how TDC navigates through webpage elements



```
<!DOCTYPE html>
<html lang="ru">
▶<head>…</head>
▼<body>
  ▶<section id="userpanel">…</section>
  ▶<section id="logo">…</section>
  ▼<div class="content">
    ▶<section id="menu">…</section>
    ▶<section id="mobile">…</section>
    ▶<section id="index">…</section>
    ▼<section id="index">
      ▼<ul id="products">
        ▶<li>…</li>
        ▼<li>
..          <div class="name">
                                     3таб. EXTASY XTC (Европа)</div> == $0
          ▶<div class="price">…</div>
           <div class="size">В
                                     наличии:  Есть                      </div>
          ▶<div class="buy btn">…</div>
          </li>
        ▶<li>…</li>
```

b) The HTML code of a webpage, shown as a tree-like structure

Fig. 6 HTML content as a tree on a similar drug market website. (**a**) An illustration of how TDC navigates through webpage elements. (**b**) The HTML code of a webpage, shown as a tree-like structure

Applying Patterns to Filter Results of Paths

Targeting the webpage element allows the Dark Crawler to download the desirable content, but not necessarily in the proper format. The *rule* selects the necessary element, but cannot select any specific parts of it. This is problematic, for example, where the price would be extracted as “900 руб.”, and not just “900”. To remove unnecessary text or to perform calculations with the result of operations with the *path* rule, *patterns* are used, which in turn follow the standard language of XQuery. For an example, working with Fig. 5, the currency is specified after the numeric value, which is in rubles for all the advertisements in the sample. Moving down the HTML branches to */li/div class*=“*price*” would produce the result “900 руб.”, where only “900” is necessary. Specifying a pattern in addition to the path will prune any unnecessary data. The pattern “(?<RESULT1>.*) руб.” is created to remove the text following the numeric value.

4 Results

4.1 *eDrugs*

Once the 28 websites were captured, and rules set up, 935 drug advertisements were extracted. Most of the sample (N = 839, 89.7%) was represented by small quantities intended for personal use ranging between 0.3 and 10 g per order. However, a fraction of the sample (N = 97, 10.3%) contained larger quantities of drugs, advertisements for which were marked as “wholesale” or “pre-order”. Drug amounts in “wholesale” orders ranged between 3 and 1000 g, where smaller marketplaces typically sold a few grams of substances per order, and larger marketplaces offered amounts upwards of tens and hundreds of grams per posting. Each website displayed between 1 and 250 drug postings, with 29 on average. Among types of substances being sold, wholesale orders and consumer-sized orders taken together, amphetamine-type stimulants made up 58.5% (n = 547), synthetic cannabinoids—22.4% (n = 210) and natural cannabinoids, i.e. marijuana products—16.3% (n = 153). The rest of the sample, 2.9% (n = 25) was represented by hallucinogens, advertised by a single marketplace as “acid”. This is shown in Figs. 7 and 8. Regional differences between drugs offered for sale are visible in Fig. 8. More variation in substance types can be observed in the Urals region, and natural cannabinoids appear more frequently.

4.2 *Markets*

Three categories of drug marketplaces become immediately evident – large (sales revenue over $4000), medium ($800–$1600) and small (under $800). The distribution of total drug prices across marketplaces and average drug prices across the Central region respectively are shown in Figs. 9 and 10. While large marketplaces conducted business in multiple cities and different geographical areas, medium marketplaces were limited to a more meager area of operation, and small marketplaces focused exclusively on a single city and/or small suburban towns located closest to it. Drug prices across the country remained surprisingly consistent, which could be explained by established equilibrium prices or price fixing.

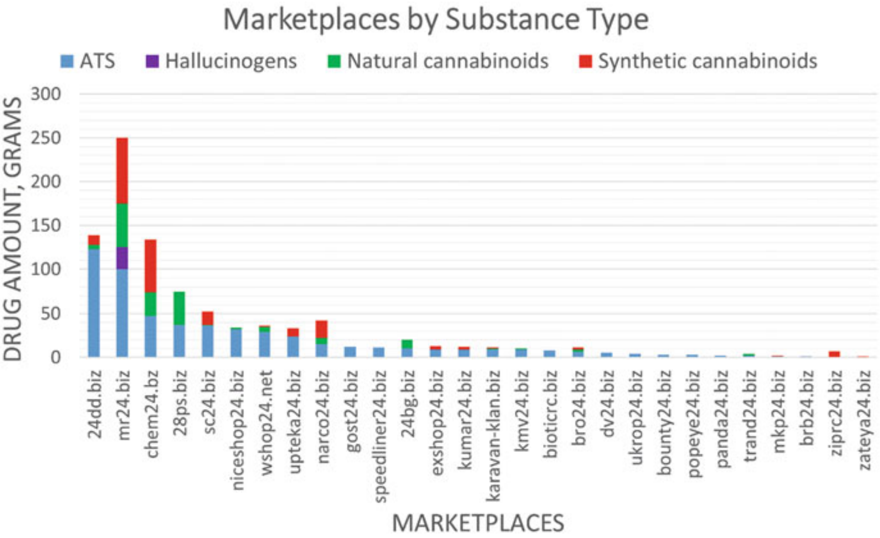


Fig. 7 Marketplaces by substance type

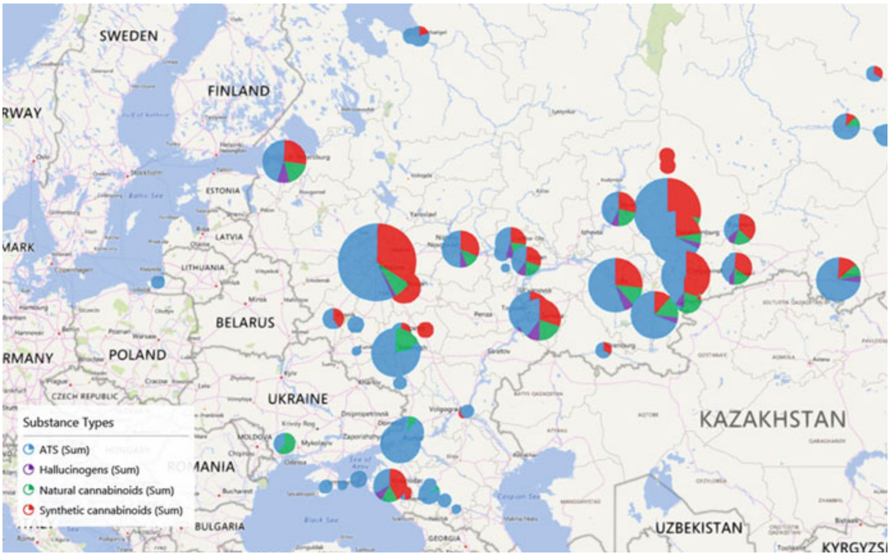


Fig. 8 Substance types spread across the region

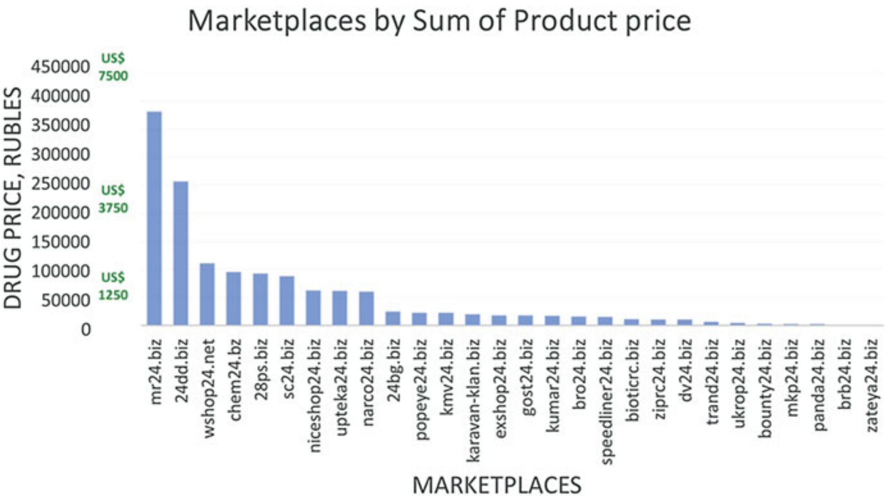


Fig. 9 Marketplaces by total price

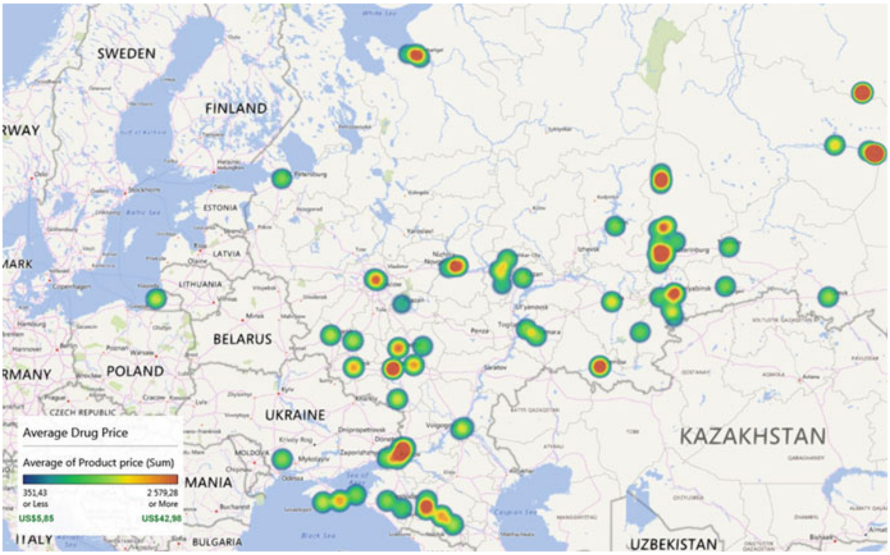


Fig. 10 Average prices across the region

4.3 Payment Methods

Noncontact drug dealing literature implicates virtual currencies such as Qiwi, WebMoney, YandexMoney and Bitcoin in internet-mediated drug trafficking, as these virtual currencies offer convenient ways to pay without being subjected to the same standards for identification as bank payments would be. However, in our sample of open web drug markets only Qiwi and Bitcoin were used. Furthermore, Qiwi, a mainstream virtual currency marketed as a vehicle for personal money transfers and utility bill payments, was used more often—on 28 marketplaces, as opposed to 20 for Bitcoin (see Figs. 11 and 12). This is perhaps due to the fact Qiwi is much more "user-friendly", where anyone can create an online wallet without any identification. Anonymous Qiwi wallets are limited to $250 per transaction up to $665 per month, and personal money transfers (person-to-person transactions) are prohibited by federal legislation. Nevertheless, these personal money transfers play a key role in facilitating online drug trafficking, as open web drug markets' business model relies on simple and straightforward payments available to any internet user without a level of technical and financial competency required to use crypto-currencies. Even without taking into account the illegal nature of goods being sold on these marketplaces, payments for these goods themselves are against the law. A number of procedural and legislative obstacles exists which prevent efficient policing of the internet environment by the Russian law enforcement, as mentioned above.

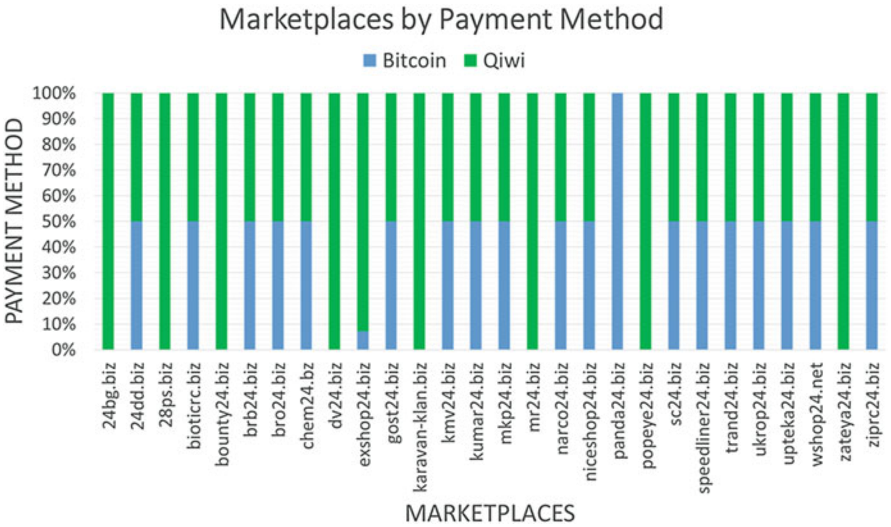


Fig. 11 Marketplaces by payment methods

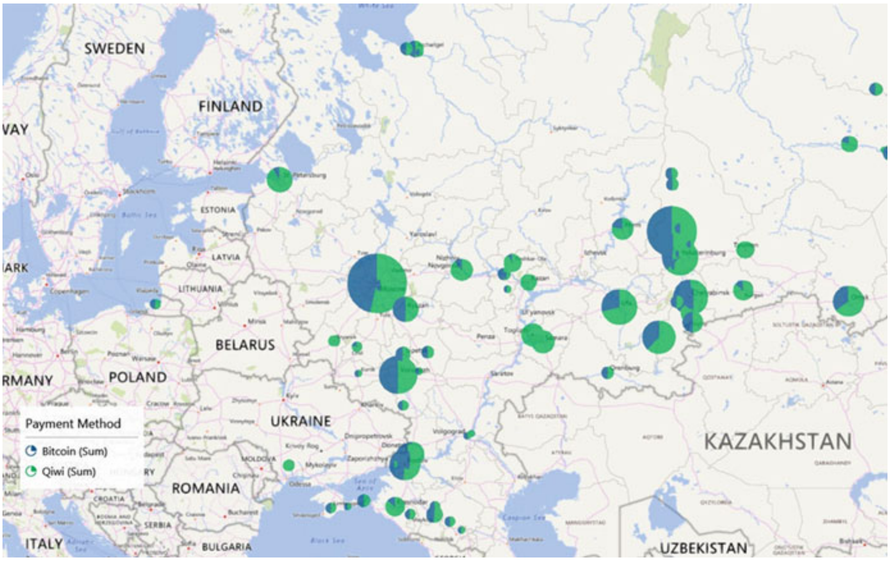


Fig. 12 Payment methods across the region

5 Discussions and Conclusion

The digital economy has reshaped trade in both conventional and illicit goods. International drug prohibition regimes criminalizing traditional drugs throughout most of the world forced drug markets to adapt and to innovate, resulting in the rise of novel synthetic drugs acting as analogs of traditional substances capable of being produced anywhere. Although the majority of world drug trafficking is still made up of traditional drugs, the rising popularity of novel psychoactive substances and innovative methods of their distribution emerging online represent a new worrisome trend. Taking advantage of opportunities afforded by the digital economy and e-commerce, drug traffickers have capitalized on the ability to sell novel, not yet regulated substances with the use of new payment technologies that enable their clients to pay with relative anonymity. We analyzed 28 online drug marketplaces which sold both traditional and novel psychoactive substances. While amphetamine-type stimulants and synthetic cannabinoids dominated the sample, traditional marijuana products and some hallucinogens were also advertised. These marketplaces primarily targeted the average consumer with drug amounts intended for personal use, however larger marketplaces also offered business-sized amounts of drugs for purchase.

Drug marketplaces which utilize noncontact drug dealing advertise and distribute drug products through internet sites, similarly to crypto-markets, but rely on a different delivery method, i.e. pre-arranged stashes. The ability of marketplace clients to carry out anonymous transactions via Qiwi and Bitcoin enables this business model to exist. Relative to crypto-markets, these websites are even easier to access, and the ability to pay with a common virtual currency enables anyone with a small amount of cash and a mobile phone to make a purchase. Operators of drug marketplaces in our sample took active steps to prevent automated access to their websites, which were circumvented by designing a separate application acting as web browser that can be used to solve the captcha and pass a DDoS protection browser check so that the crawl may continue. Automated data collection enabled us to download and analyze 935 advertisements for drugs, identifying contours and practices of the market. Although some of the websites identified as online drug marketplaces became unavailable during the data collection period, the information from the remainder of the sample was sufficient to chart and map the extent of the problem. Online drug marketplaces appear to operate throughout major cities and population clusters, where a higher variety of drugs is offered. Expanding the sample to include more drug marketplaces and/or locations would result in a more complete picture of the market. Such studies can be used to identify innovations in drug distribution, as well as to gauge consumption patterns among drug users.

The total number of noncontact drug-dealing marketplaces is unknown, therefore representativeness of our sample cannot be stated with certainty (935 postings across 28 marketplaces). Nearly half of the drug marketplaces initially identified for data capture became unavailable or had no postings displayed during the data collection period, pointing to instability of these websites perhaps due to market forces, business practices such as exit scams or law enforcement pressure. While these results may not paint a comprehensive picture, this study offers a glimpse into the noncontact drug dealing market which exists due to the ability of marketplaces operators to exploit virtual currencies with low (or none, in case of Bitcoin) customer identification standards. Since this study only considered Russian-language online drug marketplaces, external validity of these results may be low due to differences in legislation, availability of payment providers or drug trafficking routes. Nevertheless, the delivery method, i.e. pre-arranged stashes, represents an innovation in drug distribution that drug traffickers in other countries may also utilize assuming there are easy payment methods in place for internet users to transfer value (digital currency in most cases) from one person to another.

From a law enforcement perspective this type of non-contact drug purchase poses several challenges in identifying the dealer. While the delivery method would prevent the de-anonymization of the seller, the payment method, both Qiwi and bitcoin, could be traced. Although the money could be traced when the buyer transfers Qiwi to the dealer, as both the sender and recipient use mobile phones, the dealer can add a layer of security to the transaction by immediately converting Qiwi to physical currency through ATM withdrawals, and through the use of burner-phones. This would provide only a small window of opportunity for law enforcement to identify the identity and location of the dealer, before the money is withdrawn and the burner phone disposed. In the event the buyer pays with bitcoins, the identification of the dealer becomes much more challenging, as it takes significant effort to trace Bitcoin payments through tumblers, and eventually to a Bitcoin exchange, where the money might not be withdrawn for years. Finally, law enforcement could try to identify and to apprehend the author/owner of the website which peddles NPS, although, given the large number of cities where NPS is available through online non-contact methods, it is very likely that the owner of the website is acting as a middle-man between the buyer and the actual on-site dealer. Like crypto-markets, larger drug marketplaces in our study also offered business-to-business purchases with drugs intended for resale, emphasizing the scalability of this business model which can serve both individuals and subsequent drug traffickers in the chain of distribution.

References

1. P. Mullan, Who uses digital currency? in *The Digital Currency Challenge: Shaping Online Payment Systems Through US Financial Regulations* (Palgrave Pivot, New York, 2014), pp. 13–15
2. T. Tropina, Fighting money laundering in the age of online banking, virtual currencies and internet gambling. ERA Forum **15**(1), 69–84 (2014)
3. D. Décary-Hétu, L. Giommoni, Do police crackdowns disrupt drug cryptomarkets? A longitudinal analysis of the effects of Operation Onymous. Crime Law Soc. Chang. **67**(1), 55–75 (2017)
4. United Nations Office on Drugs and Crime, World Drug Report 2017 (2017), https://www.unodc.org/wdr2017/index.html
5. J. Aldridge, D. Décary-Hétu, Hidden wholesale: the drug diffusing capacity of online drug cryptomarkets. Int. J. Drug Policy **35**(C), 7–15 (2016)
6. Financial Action Task Force, *Guidance for a Risk-Based Approach to Prepaid Cards, Mobile Payments and Internet-Based Payment Services* (FATF/GAFI, Paris, 2013), http://www.fatf-gafi.org/publications/fatfrecommendations/documents/rba-npps-2013.html
7. B. Monk, J. Mitchell, R. Frank, G. Davies, Uncovering Tor: an examination of the network structure. Secur. Meas. Cyber Netw. **2018**, 4231326 (2018)
8. B. Westlake, M. Bouchard, R. Frank, Assessing the validity of automated webcrawlers as data collection tools to investigate online child sexual exploitation. Sex. Abus. **29**, 685 (2015)
9. A.T. Zulkarnine, R. Frank, B. Monk, J. Mitchell, G. Davies, Surfacing collaborated networks in dark web to find illicit and criminal content, in *Intelligence and Security Informatics (ISI) Conference*, Arizona (2016)
10. F. Mena, D. Hobbs, Narcophobia: drugs prohibition and the generation of human rights abuses. Trends Organised Crime **13**(1), 60–74 (2010). https://doi.org/10.1007/s12117-009-9087-8
11. D.R. Bewley-Taylor, The American crusade: the internationalization of drug prohibition. Addict. Res. Theory **11**(2), 71–81 (2003). https://doi.org/10.1080/1606635021000021377
12. E. Crick, Drugs as an existential threat: an analysis of the international securitization of drugs. Int. J. Drug Policy **23**(5), 407–414 (2012). https://doi.org/10.1016/j.drugpo.2012.03.004
13. J. Buchanan, Ending drug prohibition with a hangover? Br. J. Community Justice **13**(1), 55–74 (2015)
14. D. Perrone, R.D. Helgesen, R.G. Fischer, United States drug prohibition and legal highs: how drug testing may lead cannabis users to spice. Drugs: Educ., Prev. Policy **20**(3), 216–224 (2013). https://doi.org/10.3109/09687637.2012.749392
15. K. Meyers, Ö. Kaynak, E. Bresani, B. Curtis, A. McNamara, K. Brownfield, K.C. Kirby, The availability and depiction of synthetic cathinones (bath salts) on the Internet: do online suppliers employ features to maximize purchases? Int. J. Drug Policy **26**(7), 670–674 (2015). https://doi.org/10.1016/j.drugpo.2015.01.012.GLOBAL CRIME 25

16. B. Curtis, K. Alanis-Hirsch, Ö. Kaynak, J. Cacciola, K. Meyers, A.T. McLellan, Using Web searches to track interest in synthetic cannabinoids (aka 'herbal incense'). Drug Alcohol Rev. **34**(1), 105–108 (2015). https://doi.org/10.1111/dar.12189
17. P. Griffiths, R. Sedefov, A. Gallegos, D. Lopez, How globalization and market innovation challenge how we think about and respond to drug use: 'Spice' a case study. Addiction **105**, 951–953 (2010). https://doi.org/10.1111/j.1360-0443.2009.02874.x
18. J. Gershman, A. Fass, Synthetic cathinones ('bath salts'): legal and health care challenges. P T **37**(10), 571–595 (2012)
19. L. Karila, B. Megarbane, O. Cottencin, M. Lejoyeux, Synthetic cathinones: a new public health problem. Curr. Neuropharmacol **13**(1), 12–20 (2015)
20. A. Lavorgna, Internet-mediated drug trafficking: towards a better understanding of new criminal dynamics. Trends Organised Crime **17**(4), 250–270 (2014). https://doi.org/10.1007/s12117-014-9226-8
21. M.J. Barratt, J.A. Ferris, A.R. Winstock, Use of Silk Road, the online drug marketplace, in the United Kingdom, Australia and the United States. Addiction **109**(5), 774–783 (2014). https://doi.org/10.1111/add.12470
22. J. Broséus, D. Rhumorbarbe, C. Mireault, V. Ouellette, F. Crispino, D. Décary-Hétu, Studying illicit drug trafficking on Darknet markets: structure and organisation from a Canadian perspective. Forensic Sci. Int. **264**, 7–14 (2016). https://doi.org/10.1016/j.forsciint.2016.02.045
23. D.S. Dolliver, Evaluating drug trafficking on the Tor Network: Silk Road 2, the sequel. Int. J. Drug Policy **26**(11), 1113–1123 (2015). https://doi.org/10.1016/j. drugpo.2015.01.008
24. A. Phelps, A. Watt, I shop online – recreationally! Internet anonymity and Silk Road enabling drug use in Australia. Digit. Investig. **11**(4), 261–272 (2014). https://doi.org/10.1016/j.diin.2014.08.001
25. M.V. Hout, T. Bingham, 'Surfing the Silk Road': a study of users' experiences. Int. J. Drug Policy **24**(6), 524–529 (2013). https://doi.org/10.1016/j.drugpo.2013.08.011
26. M.C. Van Hout, T. Bingham, Responsible vendors, intelligent consumers: Silk Road, the online revolution in drug trading. Int. J. Drug Policy **25**(2), 183–189 (2014). https://doi.org/10.1016/j.drugpo.2013.10.009
27. J. Van Buskirk, A. Roxburgh, M. Farrell, L. Burns, The closure of the Silk Road: what has this meant for online drug trading? Addiction **109**, 517–518 (2014). https://doi.org/10.1111/add.12422
28. A.L. Osipenko, P.V. Minenko, Investigative counteraction to illegal drug trafficking by telecommunication devices (in Russian). Bull. Voronezh Institute of MVD **1**, 151–155 (2014)
29. A.V. Puptseva, Problematic issues of detecting crimes in the domain of illegal drug trafficking with the use of wireless communication devices (in Russian), in *VIII International Scientific and Practice Conference*, Tyumen, 15 Feb 2016
30. A.V. Ryasov, The issues of illegal sale of drugs using information and telecommunication networks (in Russian). Vestnik SevKavGTI **16**, 197–199 (2014)
31. A.N. Kolycheva, Certain aspects of preserving evidentiary information stored on Internet resources (in Russian). Bull. Udmurt University. Economics Law **2**(27), 109–113 (2017)
32. L.M. Kryzhanovskaya, Interaction investigators and officers inquest in the production of selected investigative actions on cases of illegal trafficking in narcotic drugs (in Russian). Theory Pract. Soc. Dev. **1**, 1–4 (2008)
33. F.P. Vasilyev, Modern features of interpretation of law enforcement interaction in Russia and the need for improvement (in Russian). Innov. Sci **3**(2), 102–110 (2017)
34. Financial Action Task Force, *Financial Flows Linked to the Production and Trafficking of Afghan Opiates* (FATF/GAFI, Paris, 2014), http://www.fatf-gafi.org/documents/news/financial-flows-afghan-opiates.html
35. I.A. Sementsova, A. I. Fomenko, Internet environment as a method of trafficking in illegal drugs, psychoactive substances or their analogs (in Russian), in *VIII International Scientific and Practice Conference*, Tyumen, 10 Feb 2016
36. D.A. Donika, Noncontact method of drug distribution (in Russian). Int. J. Exp. Educ. **6**, 17–18 (2014), http://cyberleninka.ru/article/n/sbyt-narkoticheskih-sredstv-beskontaktnym-sposobom

37. M. Barratt, J. Aldridge, Everything you always wanted to know about drug cryptomarkets (but were afraid to ask). Int. J. Drug Policy **35**, 1–6 (2016)
38. J. Van Buskirk, S. Naicker, A. Roxburgh, R. Bruno, L. Burns, Who sells what? Country specific differences in substance availability on the Agora cryptomarket. Int. J. Drug Policy **35**, 16–23 (2016)
39. J. Aldridge, D. Decary-Hetu, Sifting through the net: monitoring of online offenders by researchers. Eur. Rev. Organised Crime **2**(2), 122–141 (2015)
40. H. Chen, *Dark Web*, Integrated Series in Information Systems, vol 30 (Springer, New York, 2012)
41. N. Christin, Traveling the Silk Road: a measurement analysis of a large anonymous online marketplace (2012). https://doi.org/10.21236/ada579383
42. M. Macdonald, R. Frank, J. Mei, B. Monk, Identifying digital threats in a hacker web forum, in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2015), pp. 926–933
43. M. Wong, R. Frank, R. Allsup, The supremacy of online white supremacists – an analysis of online discussions by white supremacists. Inf. Commun. Technol. Law **24**, 41–73 (2015)
44. K. Joffres, *Disruption Strategies for Online Child Pornography Networks* (Library and Archives Canada, Ottawa, 2012)
45. A.I. Anapolskaya, Characterizing typical methods and traces of crimes related to illegal drug trade (in Russian). Eur. Union Sci.: Jurisprud. **8**(17), 136–138 (2015), http://cyberleninka.ru/article/n/harakteristika-tipichnyh-sposobov-i-sledov-sovershenIya-prestupleniy-svyazannyh-s-nezakonnym-oborotom-narkotikov
46. M.V. Kondratiev, V.K. Znikin, Operational investigative characteristics of crimes related to the illegal sale of drugs (in Russian). Bull. Kemerovo State University **1**(61), 248–255 (2015), http://cyberleninka.ru/article/n/operativno-rozysknaya-harakteristika-prestupleniy-svyazannyh-s-nezakonnym-sbytom-narkoticheskih-sredstv
47. O.N. Korchagin, D.K. Chirkov, A.C. Litvinenko, Synthetic drugs in Russia as a threat to national security (in Russian). Actual Probl. Econ. Law **1**(33), 245–253 (2015), http://cyberleninka.ru/article/n/sinteticheskie-narkotiki-v-rossii-kak-realnaya-ugroza-natsionalnoy-bezopasnosti
48. A.V. Shebalin, Specifics of conducting a preliminary inquiry into illegal sales of drugs via noncontact methods (in Russian). Curr. Issues Fighting Crim. Other Offenses. **1**, https://xn%2D%2D90ao9d.xn%2D%2Db1aew.xn%2D%2Dp1ai/science/Izdanija_BJUI_MVD_Rossii/Materiali_nauchno_prakticheskih_konferen/AKTUALNIE_PROBLEMI_BORBI_S_PRESTUPLENIJA