Scraping and Analyzing Data of a Large Darknet Marketplace

York Yannikos*, Julian Heeger and Martin Steinebach

Fraunhofer Institute for Secure Information Technology SIT
National Research Center for Applied Cybersecurity ATHENE
Darmstadt, Germany
E-mail: york.yannikos@sit.fraunhofer.de; julian.heeger@sit.fraunhofer.de;
martin.steinebach@sit.fraunhofer.de
**Corresponding Author*

Received 01 December 2022; Accepted 05 December 2022;
Publication 28 April 2023

Abstract

Darknet marketplaces in the Tor network are popular places to anonymously buy and sell various kinds of illegal goods. Previous research on marketplaces ranged from analyses of type, availability and quality of goods to methods for identifying users. Although many darknet marketplaces exist, their lifespan is usually short, especially for very popular marketplaces that are in focus of law enforcement agencies.

We built a data acquisition architecture to collect data from White House Market, one of the largest darknet marketplaces in 2021. In this paper we describe our architecture and the problems we had to solve, and present findings from our analysis of the collected data.

Keywords: Darknet marketplaces, data acquisition, captcha solving, web crawling, convolutional neural networks.

Journal of Cyber Security and Mobility, Vol. 12_2, 161–186.
doi: 10.13052/jcsm2245-1439.1222
© 2023 River Publishers

1 Introduction

Darknets have become a widespread phenomenon on the Internet. Most reporting in the media focuses on the illegal aspects, citing the distribution of child pornography, drugs or weapons as typical use cases. However, advantages such as freedom from censorship in repressive political systems or anonymity when browsing or communicating in the clearnet are evident [16].

Due to their growing prevalence, the concept of darknet intelligence has become popular [13, 22]: The goal is to automatically extract data from the darknet, which is rather difficult to survey compared to the clearnet. The extracted data is then processed in such a way that relevant information can be derived from it.

This has become a business model for a number of international companies. They either add the service of darknet intelligence as part of their security or OSINT service or completely specialize on the subject. Subjects of interest for their services are:

- Data and leakage offerings, either commercial or free. This is relevant for anybody subject to data theft.
- Vulnerability updates, either due to disclosure on forums or to new tools available. This helps to plan defensive measures.
- Opinions about entities. Becoming the subject of aggressive discussions could hint towards future attacks.
- Market pricing data. The price asked for illegal goods may help to better understand market situations.
- Service offers. This helps to monitor the availability and quality of potential attackers hired to act against a potential target.

In this paper we do not address a specific darknet in general, but a selected marketplace. Our goal was to efficiently acquire and store data about the marketplace in order to perform extensive analyses of the data later.

Darknet marketplaces are usually found in Tor,[1] the most popular darknet today. Many marketplaces do not last long before they either suddenly shut down – usually performing an exit scam – or are taken offline by law enforcement agencies. In recent years, very popular marketplaces like Dream Market, AlphaBay, Hansa market, or Wall Street Market went offline, but others quickly took their spot as users looked for alternatives. Sometimes new iterations of closed marketplaces showed up again after a while using the same name and sometimes operated by the same individuals as before.

[1] https://www.torproject.org

For example, the second version of AlphaBay is now online since August 2021 after the predecessor was shut down in 2017.

White House Market

White House Market was one of the largest and most popular marketplaces recently with a relatively long lifespan of about two years (operating from August 2019 until October 2021). Notable about White House Market was that only Monero was accepted for payments, a cryptocurrency with strong focus on anonymity, and that the marketplace shut down in a very unusual way: The operators did not perform an exit scam. Instead, they announced their retirement and disabled any new orders, but also finalized any open orders and let the users withdraw their funds. Since its shutdown, no new version of White House Market has gone online yet.

When White House Market was still online, we built a microservice architecture to automatically acquire data from the marketplace. In this paper we describe the architecture we used to fully crawl the marketplace several times and present findings from our analysis of the collected data. We provide insight about the security mechanisms implemented by White House Market and give an overview about product categories, product offers, and vendor information.

This journal article is an extension of our previously published research paper *Data Acquisition on a Large Darknet Marketplace* [35]. We provide a broader overview about related work, describe an approach to automatically solve the second captcha on White House Market, and present additional findings from the data we collected.

The remainder of this paper is structured as follows: In Section 2 we describe the technical basics of Tor and in Section 3 specific properties of marketplaces in the Tor network. In Section 4 we provide an overview about other research regarding darknet marketplaces. In Section 5 we describe our data acquisition architecture in detail and present analysis results from the crawled data in Section 6. In Section 7 we summarize our work and give an outlook on further research.

2 Tor Basics

The Onion Router or *Tor* is an overlay network designed to provide technical anonymity for its participants. Tor implements a technique called *onion routing*, a smart way of routing packets between participants (*Tor nodes*):

Every packet in the Tor network is routed through a predefined path of at least three Tor nodes, building a *Tor circuit* such that source and destination of the packet are never directly connected. Before a packet is sent through a Tor circuit, it is wrapped in several layers of asymmetric encryption by using the public keys of all the nodes to be passed in the circuit. When the packet then arrives at the first node in the circuit, the node can decrypt and remove the outer layer of the packet to learn where to send it next. This continues for every node that is passed.

Tor nodes in a circuit usually have no knowledge of the purpose of the other nodes in the same circuit – exceptions are *entry nodes* that are directly connected to the source of a packet and *exit nodes* that are directly connected to the destination in the clearnet. However, Tor participants can not only access the clearnet anonymously where the destination itself is not anonymous to them: Tor also provides *onion services* which can only be accessed through Tor. In a circuit built to access an onion service, even the source of a packet does not know the destination node. Additionally, in the circuit even the node preceding the destination node does not know whether or not the next node is the actual destination. In this way, onion services can enable completely anonymous communication between Tor participants. Onion services are very often web sites but since Tor is an overlay network, every TCP-based protocol or service is supported, e.g. IRC, FTP, SMTP. All onion services are reachable using *.onion* addresses that can be resolved only via Tor.

3 Properties of Darknet Marketplaces

Darknet marketplaces in Tor are based on onion services, i.e. anonymously hosted sites only reachable within Tor. The marketplaces are typically built like large web shops with many different product categories, ranging from pharmaceuticals and drugs to stolen credit card data, botnet rentals, or fraud tutorials.

Most darknet marketplaces first show a specific type of captcha before allowing access to the site, usually in order to protect against DDoS attacks and automated web crawling. After solving the captcha, the user can continue to login or register an account. New users can register as customer or vendor by providing a username, a password, and typically a PIN to withdraw deposited funds. After the registration the user gets a short introduction into market rules and often security-related advice and is then able to browse the product offers. These are typically grouped in categories and accessible via a navigation menu.

Larger darknet marketplaces often look very similar with many different categories and subcategories, often ten-thousands of product offers, hundreds and sometimes thousands of vendors. Product offers typically contain information about the vendor, shipping, country of origin, payment, quantity and price. Vendor profile pages usually contain short descriptions, often PGP public keys and vendor ratings from customers with short review messages.

Darknet marketplaces hosted as onion services provide technical anonymity for the participants. The operators usually employ strict security setups, e.g. enforcing all network traffic from and to the site to go through Tor, avoiding content service providers or hosting services in the clearnet, and requiring users to disable Javascript in their browser. Therefore, the captcha services used on darknet marketplaces are self-hosted and mostly self-developed, often based on publicly available open source libraries. This makes them usually much easier to solve automatically than professionally developed captchas like Google's ReCAPTCHA. Also, images on darknet marketplaces are often directly embedded in HTML as base64-encoded data URLs[2] which simplifies scraping such web pages.

4 Related Work

Since Silk Road gained global publicity as the largest darknet marketplace between 2011 and 2013, a wide range of research regarding such marketplaces has been done. In [21] Me and Pesticcio analyze 14 Tor marketplaces and derive connections between those by identifying PGP keys used by vendors in multiple marketplaces. The authors later provide a critical analysis of this approach [20]. Georgoulias et al. investigate 41 marketplaces and 35 vendor shops and also analyze discussions within forums of those marketplaces in [14].

Several papers proposed generic tools and methodologies to acquire data from darknet marketplaces: In [34] we propose a framework for collecting product data in Tor marketplaces and methods for enriching these findings by matching them with external sources from the clearnet. Ball et al. suggest another tool set for data collection and analysis in darknet marketplaces in [2].

There has also been research on analyzing specific information found on darknet marketplaces: In [6] Broséus et al. collect data from the Evolution marketplace and show which geographical information can be derived from

[2] RFC 2397 – https://datatracker.ietf.org/doc/html/rfc2397

it. Wang et al. focus on the detection of multiple identities of vendors in darknet marketplaces by utilizing deep learning based image matching in [30]. Goonetilleke et al. examines the Hydra [15] with respect to its mechanisms and execute a quantitative analysis of users. Today darknets are also created by messenger services, as e.g. Kempen shows in [18].

Many previous studies focused on drug markets or drug offerings in the darknet. In [7] Celestini et al. present a crawling method for AlphaBay, Nucleus and East India Company as well as a concept for data collection and cleansing. They also discuss which data they use for analysis and provide a number of results related to drug offers. In a study about Silk Road 2, Dolliver provides insight about the number of items for sale, the amount of drug offerings among them, and the number of vendors and countries involved [9]. Dolliver and Kenney provide a comparison of drug vendor characteristics on the darknet markets Evolution and Agora in [10]. In [12] Dolliver and Love compare Silk Road 2 and Evolution with respect to drug offerings, and in [11] Dolliver and Kuhns analyze drug offerings and vendors in Agora over a period of four months. Another comparison on drug markets in the darknet was published by Tzanetakis in [29], also addressing the impact on overall drug distribution. Bancroft discusses [3] the role of the darknet on drug availability. Karden and Strizek [17] look at web surveys and how they could help to learn about people who purchase drugs on cryptomarkets. Zambiasi [37] adresses the impact of the shutdown of darknet marketplaces on the amount of drugs traded in the streets. Bahamazava and Nanda [1] analyze the role of characteristics of cryptocurrencies in cryptomarkets for drug buyers.

Besides drugs, there exist many other offers on such marketplaces. Broadhrust et al. [5] show that during the COVID pandemic many products addressing COVID-19 were available in the 12 marketplaces investigated. Previous research also did not exclusively address large darknet marketplaces. In [36] we analyze transactions associated with bitcoin addresses of single-vendor shops and match them with products listed in the shop. In [19] Laferrière and Décary-Hétu investigate establishing trust in 108 single-vendor shops.

Own Previous Work

Beyond the works referenced in the section before, we have also discussed a number of additional aspects of the (Tor) darknet in previous papers. In [27] we show how to detect and analyze tor onion services. In [33] we compare

cyber attacks in clearnet and darknet, while in [28] we focus on phishing. The role of the darknet in cyber warefare is analyzed in [8]. Deanonymization of darknet users and its legal aspects are the subject of [31]. In [4] and [23] we address single-Vendor marketplaces and their discovery. The darknet is also used for file sharing, a topic we discuss in [26]. Our work [24] takes a more abstract point of view and identifies challenges in darknet research. Distribution of child pornography is one dire aspect of the darknet; in [32] we discuss the strategy of using synthetic material to infiltrate groups dealing with such material.

5 Data Acquisition on White House Market

Our data acquisition architecture comprises a database, multiple workers for crawling, scraping, and parsing data from the marketplace, a task queue to distribute worker tasks, and a captcha solver. Figure 1 shows the architecture with the different components.

We implemented each component as microservice in Python and used docker for deployment. In the following sections we describe the different microservices in detail.

5.1 Crawling

Since only logged in users could browse White House Market, we first created 20 different user accounts and stored the credentials in our database to be used

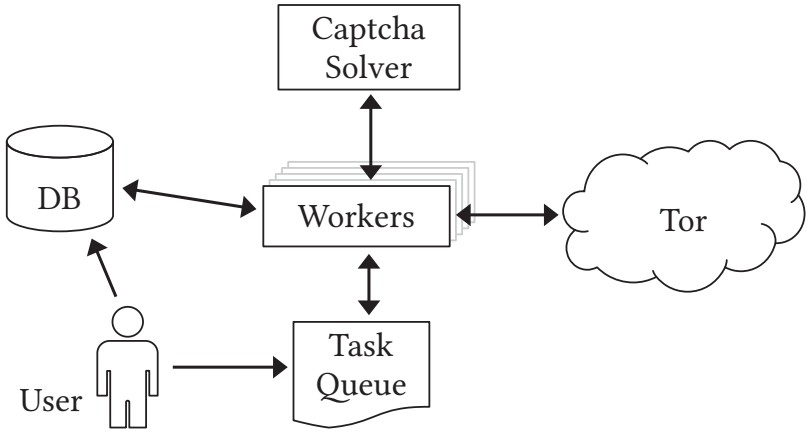


Figure 1 Overview of the architecture with microservices.

by our crawlers. The whole data acquisition process on White House Market involved the following tasks:

1. Login with provided user credentials
2. Collecting product categories and links from the navigation menu
3. Collecting all links to product offers per (sub)category
4. Collecting all content per product offers

For each of these tasks we implemented a worker class responsible for handling the task. We could then run several worker instances in parallel to crawl the marketplace. Workers handling product categories collected links to individual product offers and put them on the task queue. Other workers then scraped the product offers, stored the HTML content in the database, and parsed the content to extract relevant information which was also stored separately in the database. After all tasks were completed (i.e. the task queue was empty), the workers were shut down, ending the acquisition process. Figures 2 and 3 show examples of a product offer and a vendor page in the marketplace.

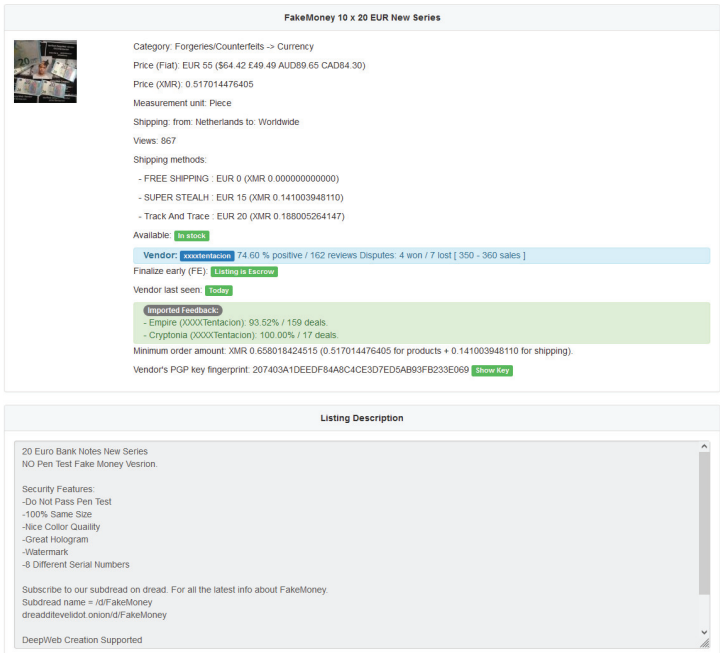


Figure 2 Example of a product page on WHM.

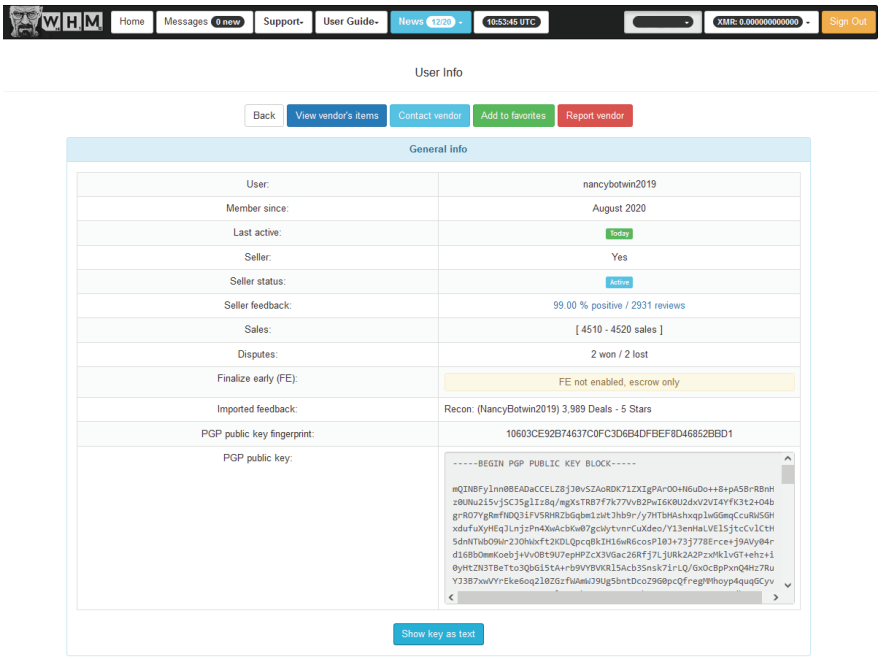
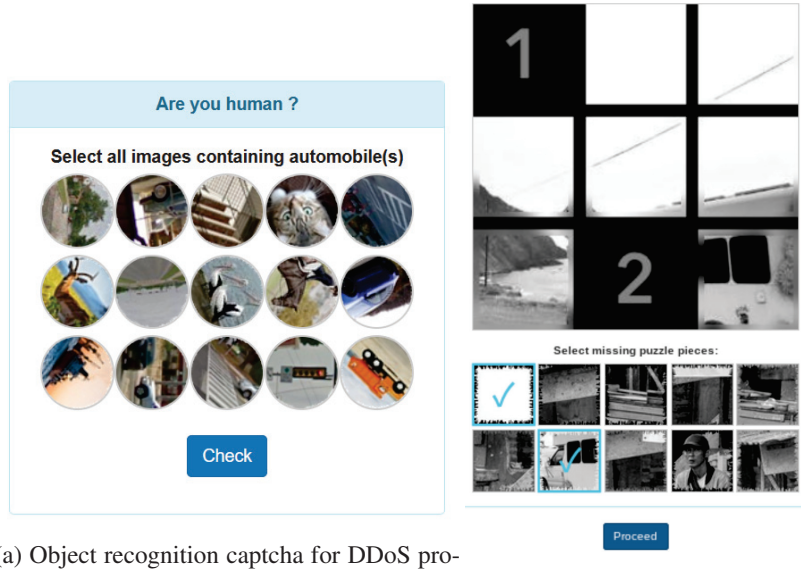


Figure 3 Examplle of a vendor page on WHM.

5.2 Captcha Solving

White House Market used two different sets of captchas a user had to solve: an object recognition captcha (Figure 4a) and a puzzle captcha (Figure 4b). The first captcha was used for DDoS prevention as seen on many other darknet marketplaces: Each user had to pass this captcha once before login or registration was possible. After submitting the correct solution, the captcha would not show up again – even logging out from the marketplace would not trigger the captcha again. This state remained until the user switched Tor circuits, effectively gaining a new Tor identity. In order to efficiently crawl data from the marketplace, we had to build a solver for this captcha.

The puzzle captcha was used as throttling mechanism to keep users from requesting too many pages (i.e. more than ten) at a time. For example, if a user went through all 23 pages of listings of a specific product category, the puzzle captcha had to be solved twice: after the 10th and also after the 20th page view. Since we wanted to crawl the marketplace completely with all product offers and vendor pages, we had to deal with the puzzle captcha, too.



(a) Object recognition captcha for DDoS protection

(b) Puzzle captcha for throttling

Figure 4 Different captcha types on White House Market.

5.2.1 Object recognition captcha

To build a solver for the object recognition captcha, we used Tensorflow's Keras API[3] to build a convolutional neural network (CNN) based on the VGG architecture proposed by Simonyan and Zisserman [25]. To train the CNN, We collected 1,000 captchas from White House Market. Each captcha showed 15 small images where objects of a specific category should be selected, e.g. "Select all images containing automobiles". While working with the captchas, we noticed that there existed 23 different object categories: 'airplane', 'automobile', 'banknote', 'bird', 'bridge', 'building', 'bus', 'cat', 'cloud', 'coin', 'crosswalk', 'deer', 'dog', 'fire hydrant', 'frog', 'gold bar', 'horse', 'road', 'ship', 'stair', 'traffic light', 'tree', 'truck'.

Before training the CNN we processed the collected captchas as follows:

1. Identification of the correct object images for each captcha, verification via White House Market
2. Labeling of correct object images with their corresponding class – result: 2,942 labelled object images
3. Conversion of all object images to RGBA color space

[3]https://www.tensorflow.org/tutorials/images/cnn

Figure 5 Examples of object images that were mostly incorrectly classified (asked for: 'road' / classified as: 'bridge').

We then split the object images into 80% training and 20% validation data (ensuring a uniform distribution of classes in both data sets). Since White House Market seemed to randomly rotate object images in their captchas, we augmented our training data set with rotated versions of each object image (11 rotations by 30 degrees per image) and trained our CNN.

While working with the captcha, we noticed that we got more incorrect classifications for specific object categories than for others. For example, the correct category for all 8 object images shown in Figure 5 was 'road'. However, our CNN classified only one object image correctly, while the other 7 were incorrectly classified as 'bridge'. In most cases, we would argue about whether this was really a classification error: In our example, 6 of the incorrectly classified object images actually show a bridge besides a road and were very similar to object images from the 'bridge' category.

After training we collected another 1,000 new captchas from White House Market as test data and were able to achieve 78% accuracy in solving these captchas.

5.2.2 Puzzle captcha

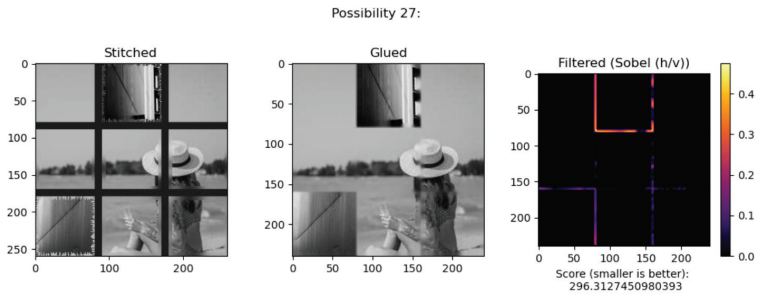
The next step was to focus on the puzzle captcha. While working with the captcha on White House Market, we quickly noticed that after the captcha was shown, a simple logout and login was enough to bypass the captcha without having to solve it. We then continued logging out and back in after every 10th page request. We used this (quite reliable) workaround while crawling until we finished developing the puzzle captcha solver.

To build an automatic solver for the puzzle captcha we first investigated how the captcha worked: Each shown captcha image was always divided into exactly 9 tiles with 2 tiles missing. The missing ones had to be chosen from a group of 10 "solution tiles" in any order. If the missing 2 tiles were correctly chosen, the captcha was solved. Therefore, 90 possible permutations of the image existed for each captcha and $\binom{10}{2} = 45$ different possibilities to choose 2 tiles from 10, so the probability of solving the captcha by random guess was $\frac{1}{45} \approx 2\%$.

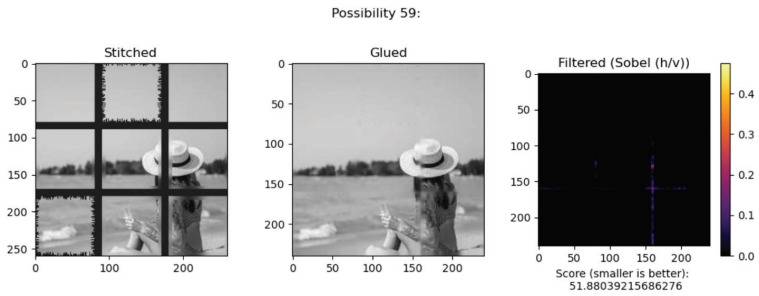
We developed the following algorithm to solve the puzzle captcha automatically:

1. Find the location of the missing tiles in the image
2. Create all 90 possible captcha solutions as follows:
    1. *Stitch:* place two solution tiles into the missing area of the image
    2. *Glue:* remove borders and distortions of the tiles with inpainting
    3. Calculate a score using the Sobel filter to detect edges within the stitched and glued image
3. Choose the solution with the lowest score

Figures 6a and 6b show the stitched and glued images as well as the corresponding scores for two possible solutions of a puzzle captcha. We tested our approach with 500 captchas taken from White House Market and achieved an accuracy of about 83% in solving the puzzle captcha.



(a) Higher score after stitching and gluing two incorrect tiles



(b) Lower score after stitching and gluing the correct tiles

Figure 6 Score calculation based on Sobel filter for different possible solutions of a puzzle captcha.

5.2.3 Captcha solving API

We deployed both captcha solvers as microservices with an API. If workers crawling the marketplace encountered the object recognition or the puzzle captcha, they could then use the API to send it to our solver and get the correct solution back in about 4 out of 5 cases. This was sufficient for us to crawl the marketplace without any other significant obstacles.

5.3 Database

In our architecture we used PostgreSQL as database backend to store the acquired data from White House Market. The data was stored in two steps:

1. Immediately after scraping a web page, the raw HTML content of that page was stored in a dedicated table in the database. This was done to ensure that we still have the raw content available in cases where the marketplace suddenly went offline or errors occurred during the following step.
2. The HTML content was parsed to extract relevant data like product name and description, country of origin, vendor name, price, vendor feedback. The extracted content was also stored in the database, structured in different tables for later analysis.

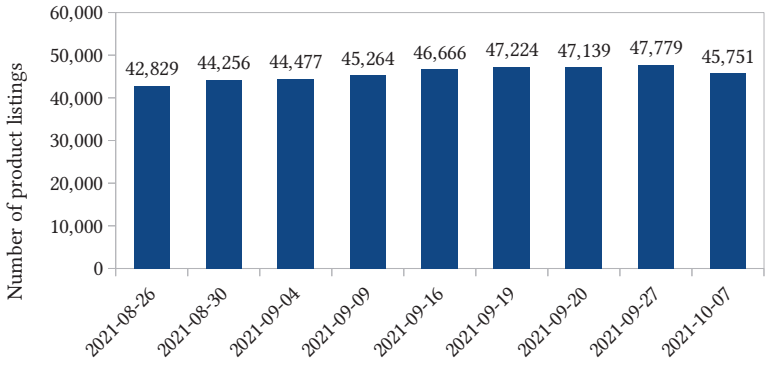
We implemented an API for the workers to access the database, i.e. to store acquired data and to retrieve and update login session information for the user accounts we used to crawl the marketplace. To avoid duplicate entries in the database, we designed the database schema for quick lookups regarding whether or not a web page had been scraped already by another worker during the same crawling process.

6 Results

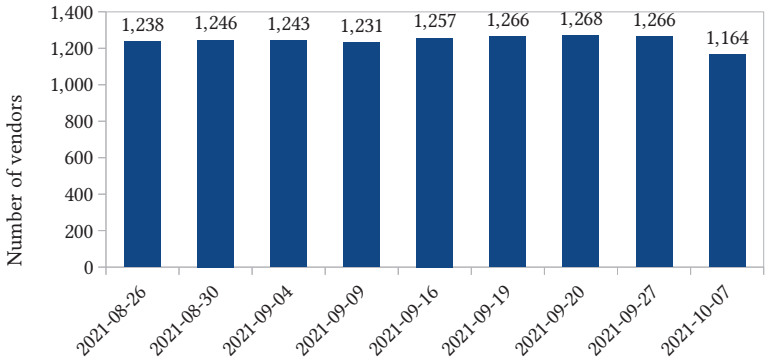
We crawled White House Market 9 times in 2021, from August 26 to October 7. In the following we present the results of our analysis of the acquired data.

6.1 Number of Product Offers and Vendors

During our data acquisition we collected 57,007 different product offers and 1,448 vendors in total. Figure 7 shows the absolute numbers per crawling date. Because no directory existed on the site with a complete listing of all users, we could not determine the total number of all users. The marketplace



(a) Number of product offers



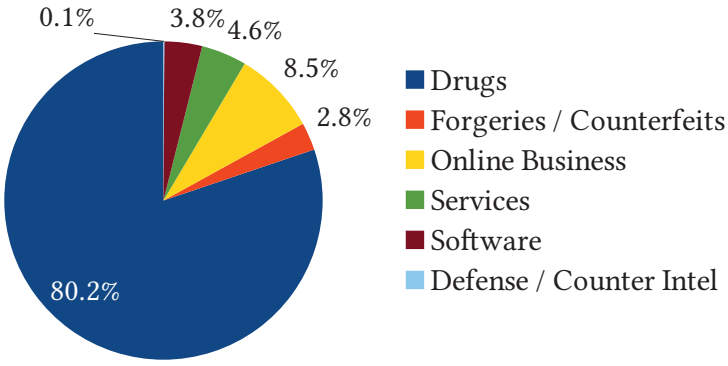
(b) Number of vendors

Figure 7 Absolute number of product offers and vendors per crawling date.

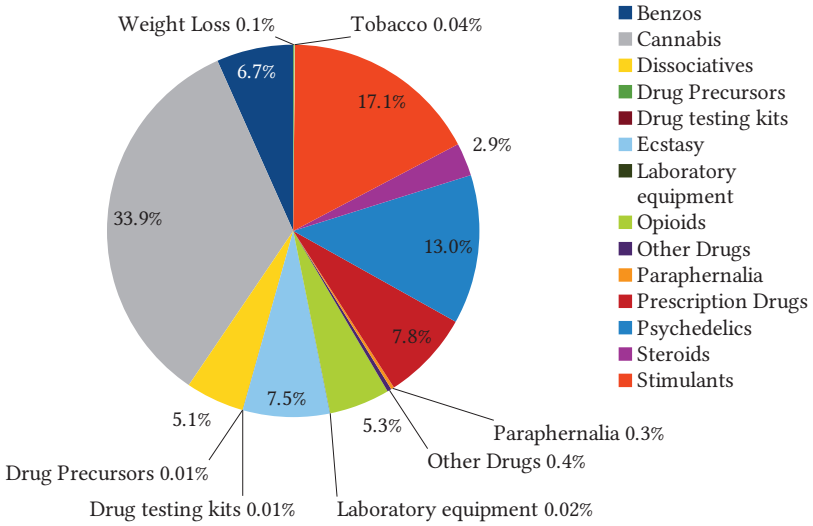
claimed that they had a user base of over 310,000 customers with more than 110,000 customers actively using the site.

6.2 Product Categories

White House Market listed many different product categories and subcategories. The largest category by far was "Drugs" comprising 80% of all products offered on the marketplace, followed by Online Businesses with 8.5% of all offers. Figure 8a shows the share of product offers for each of the main product categories.



(a) Share of product offers per main category



(b) Subcategories of drugs

Figure 8 Different product categories: main categories and subcategories for drugs.

We took a closer look at the "Drugs" category with several levels of subcategories. The share of product offers of the first subcategory level is shown in Figure 8b. While cannabis was the most popular drug with 34% of all offers, we observed significantly less offers of prescription drugs with only about 8%, as well as for stimulants (cocaine, meth) or psychedelics (LSD, DMT) with 17% and 13%, respectively.

6.3 Origin and Shipping Locations

For all scraped offers we analyzed the country of origin and the shipping location. Figures 9a and 9b show the distribution of the countries over all 57,007 product offers. Almost 55% of all products offered where shipped from either the US, UK, or Germany. More than 16% of all offers were digital items or services with no specific country of origin, but not all of those offered digital shipping.[4] For almost 28% of all products worldwide shipping was offered and 50% of all products could be shipped to the US.

6.4 Product Sales

We analyzed the number of product sales for all vendors and grouped them into different sale intervals, shown in Figure 10. Here we could see that more than half (52.6%) of all vendors on White House Market sold between 101 and 1000 products. A little more than a quarter (26.6%) sold 1–100 products about 7% of the vendors sold more than 2000 products. The top vendor had 19600–19610 product sales mentioned on his profile.

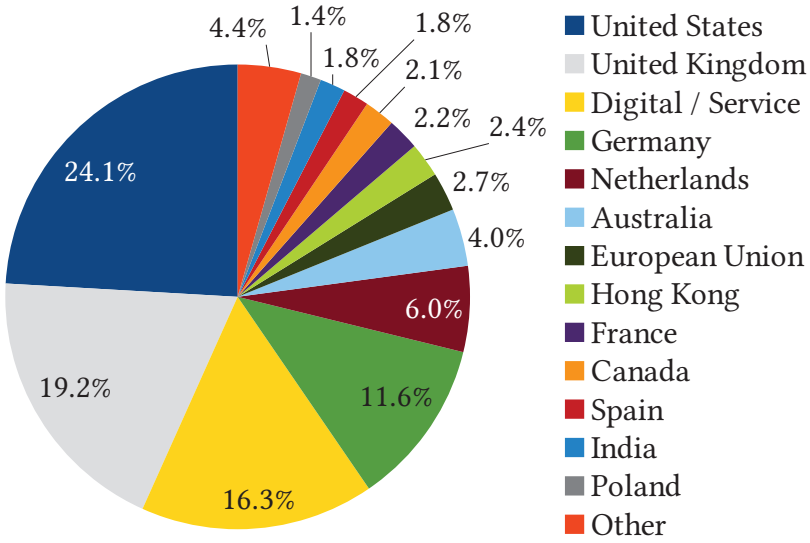
6.5 Activity on Other Markets

Based on the "imported feedback" field of the vendor profiles (where external feedback was shown), we could gather information about vendor activity on other marketplaces.[5] Figure 11a shows the amount of vendors that were active on other markets besides White House Market. Almost half (48.8%) of all vendors did not provide information about other markets while nearly a third (32.4%) mentioned activity on only one, and roughly 9% on two other marketplaces.

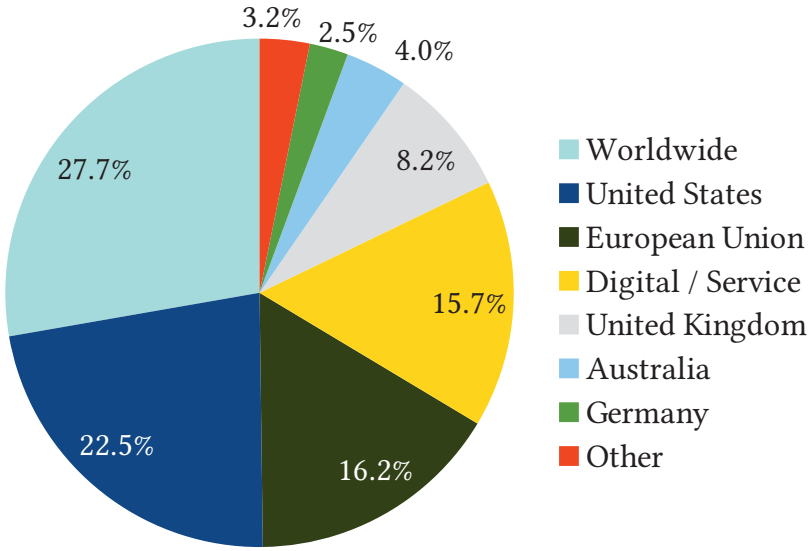
Figure 11b shows the individual marketplaces where the 742 (51.2%) of the vendors were active. A third (32.9%) of these 742 were active on Empire Market, 19% had a vendor profile on the Recon search engine and 12.6% previously used Dream Market. Dream Market was taken offline by law enforcement agencies in 2019, Empire conducted an exit scam in 2020 and Recon is currently also offline (since October 2022).

[4]For 325 items worldwide shipping was offered (probably including digital shipping), for 8 items only shipping to France was offered, and for another 8 shipping was available to the US only.

[5]Imported feedback included ratings from the Recon search engine in Tor that was mentioned on several vendor profiles. We counted this as *market activity* in our evaluation.



(a) Countries of origin



(b) Shipping countries

Figure 9 Country of origin and shipping locations for product offers.

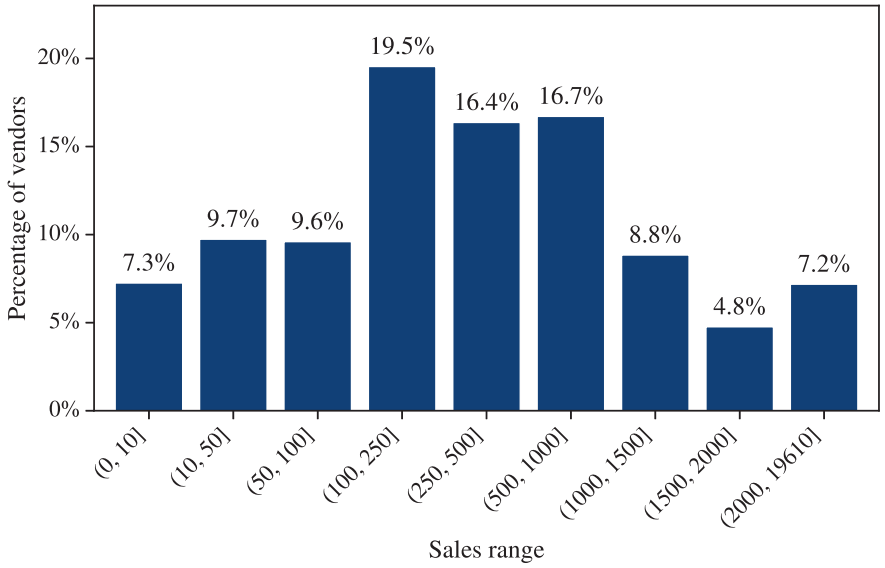


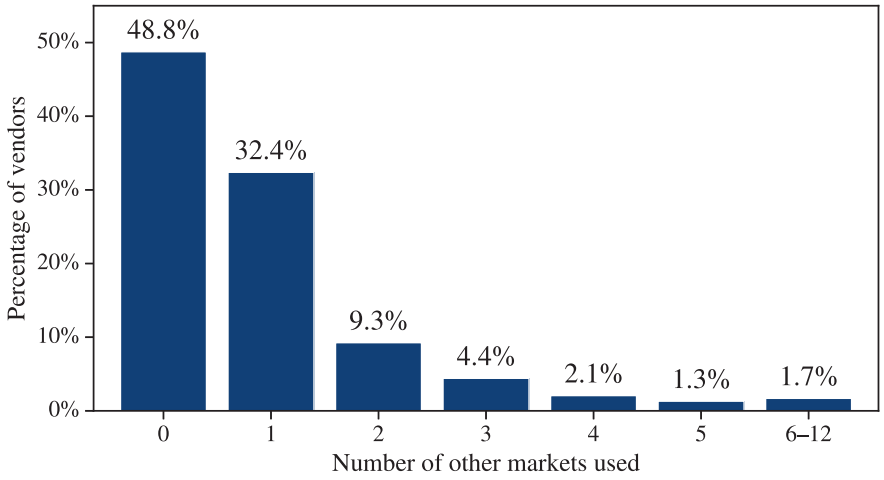
Figure 10 Percentage of vendors per sales range.

6.6 Information From PGP Public Keys

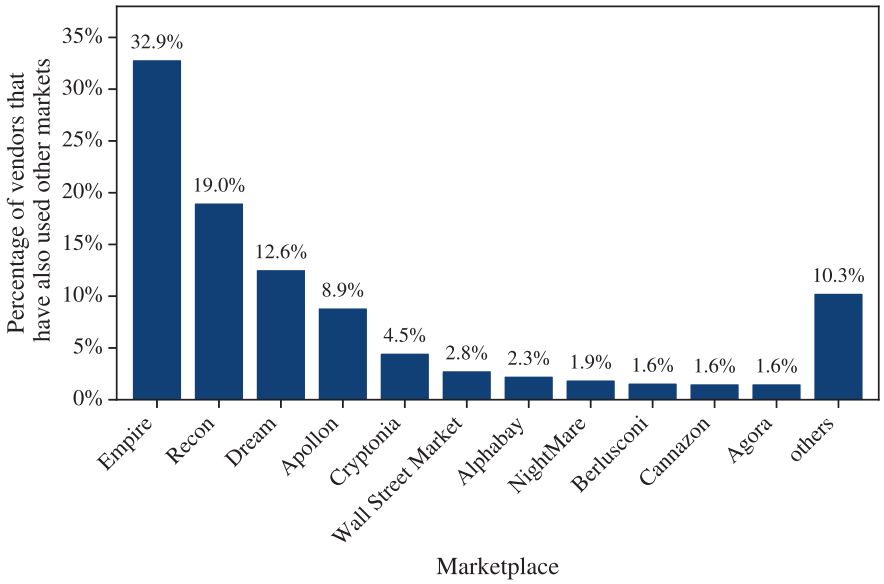
Every vendor offering products on White House Market had to provide a PGP public key for communication. We scraped all public keys from the vendor pages and extracted information about email addresses and creation date from the keys. We noted that many keys were created long before White House Market went online, some almost 7 years earlier. Although we cannot trust the creation dates of the public keys, this could be an indication that those keys were also used elsewhere. Figure 12 shows the graphs of the cumulative numbers of created user accounts and PGP keys. We can also clearly see an increase of user account creations on the marketplace in August and September 2020.

Regarding the individual PGP key lengths we found that more than half of the vendors used 4096-bit RSA keys and nearly another third used 2048-bit RSA keys. Figure 13 shows the distribution of different key lengths among the vendors. More than 98% of the vendors used RSA keys with standard lengths while roughly 2% used RSA with untypical key lengths, DSA or EdDSA keys.

To see if we could find information about PGP keys in the clearnet, we conducted a Google search for the PGP keys of the top 20 vendors



(a) Amount of vendors with activity on other marketplaces



(b) Amount of vendors per marketplace with activity on at least one other marketplace

Figure 11 Vendor activity on other marketplaces.

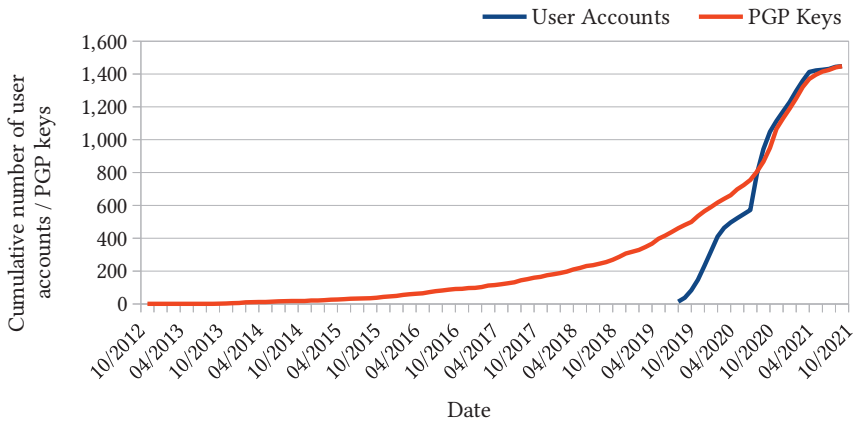


Figure 12 Cumulative number of created vendor accounts and associated PGP public keys over time (creation date).

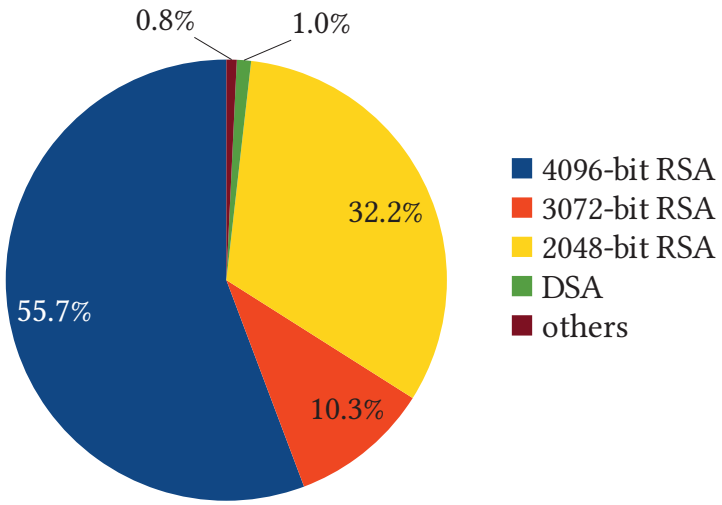


Figure 13 Distribution of PGP key lengths among vendors.

(regarding products sales) as well as for another 50 randomly chosen vendors. However, we could only find a very small number of keys mentioned on clearnet websites that just advertised other (mostly already offline) darknet marketplaces, so our findings did not provide much additional insight about the vendors.

We also analyzed any email addresses extracted from the user id fields of the PGP public keys and observed that many popular email providers

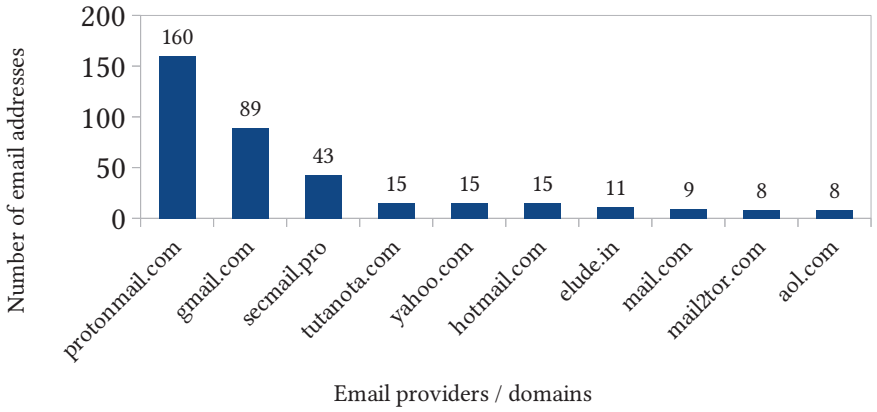


Figure 14 Most used email providers associated with PGP public keys.

like protonmail.com or gmail.com were used. Figure 14 shows the top 10 email providers extracted from the PGP public keys. Information about email addresses and providers could be helpful e.g. in law enforcement investigations regarding vendor identities.

7 Conclusion

In this paper we described a data acquisition architecture to collect data from White House Market, a large darknet marketplace in Tor that started in August 2019 and ceased operations in October 2021. In order to build our architecture, we first needed to discover and analyze the security mechanisms White House Market used to prevent web crawling and scraping. We encountered two different types of captchas and developed an automatic captcha solver for both. For the object recognition captcha we used a convolutional neural network, built a data corpus for training and subsequently achieved almost 78% accuracy. For the puzzle captcha we used image processing methods to find the captcha solution and achieved 83% accuracy.

Using the captcha solvers as microservice in our architecture, we could fully crawl the marketplace 9 times until it shut down. Based on the data we collected we presented findings regarding type and amount of products offered, about vendors, e.g. number of sales, shipping information, and activity on other marketplaces. We extracted potentially interesting data from the PGP public keys of the vendors giving insight about key reuse, key lengths, and popular email providers used.

In the future we will extend our architecture by developing additional microservices to crawl other popular marketplaces. We also plan to further look into the different types of captchas used on those marketplaces and develop strategies to quickly build automatic captcha solvers.

Acknowledgments

The research for this publication was conducted as part of the PANDA[6] project funded by the German Federal Ministry of Education and Research (BMBF), code 13N14355.

References

[1] Katsiaryna Bahamazava and Rohan Nanda. The shift of darknet illegal drug trade preferences in cryptocurrency: The question of traceability and deterrence. *Forensic Science International: Digital Investigation*, 40:301377, 2022.

[2] Matthew Ball and Roderic Broadhurst. Data capture and analysis of darknet markets. *Available at SSRN 3344936*, 2021.

[3] Angus Bancroft. Potential influences of the darknet on illicit drug diffusion. *Current Addiction Reports*, pages 1–6, 2022.

[4] Fabian Brenner, Florian Platzer, and Martin Steinebach. Discovery of single-vendor marketplace operators in the tor-network. In *The 16th International Conference on Availability, Reliability and Security*, pages 1–10, 2021.

[5] Roderic Broadhurst, Matthew Ball, and Chuxuan Jessie Jiang. Availability of covid-19 related products on tor darknet markets. *Australasian Policing*, 12(3):8–13, 2020.

[6] Julian Broséus, Damien Rhumorbarbe, Marie Morelato, Ludovic Staehli, and Quentin Rossy. A geographical analysis of trafficking on a popular darknet market. *Forensic science international*, 277:88–102, 2017.

[7] Alessandro Celestini, Gianluigi Me, and Mara Mignone. Tor marketplaces exploratory data analysis: The drugs case. In *International conference on global security, safety, and sustainability*, pages 218–229. Springer, 2017.

[6] https://panda-projekt.de

[8] Kai Denker, Marcel Schäfer, and Martin Steinebach. Darknets as tools for cyber warfare, 2019.

[9] Diana S Dolliver. Evaluating drug trafficking on the tor network: Silk road 2, the sequel. *International Journal of Drug Policy*, 26(11):1113–1123, 2015.

[10] Diana S Dolliver and Jennifer L Kenney. Characteristics of drug vendors on the tor network: a cryptomarket comparison. *Victims & Offenders*, 11(4):600–620, 2016.

[11] Diana S Dolliver and Joseph B Kuhns. The presence of new psychoactive substances in a tor network marketplace environment. *Journal of psychoactive drugs*, 48(5):321–329, 2016.

[12] Diana S Dolliver and Katherine L Love. Criminogenic asymmetries in cyberspace: a comparative analysis of two tor marketplaces. *Journal of Globalization Studies*, 6(2):75–96, 2015.

[13] Claude Fachkha and Mourad Debbabi. Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization. *IEEE Communications Surveys & Tutorials*, 18(2):1197–1227, 2015.

[14] Dimitrios Georgoulias, Jens Myrup Pedersen, Morten Falch, and Emmanouil Vasilomanolakis. A qualitative mapping of darkweb marketplaces. In *2021 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–15. IEEE, 2021.

[15] Priyanka Goonetilleke, Alex Knorre, and Artem Kuriksha. Hydra: A quantitative overview of the world's largest darknet market. *Available at SSRN 4161975*, 2022.

[16] Eric Jardine. Privacy, censorship, data breaches and internet freedom: The drivers of support and opposition to dark web technologies. *new media & society*, 20(8):2824–2843, 2018.

[17] Alexandra Karden and Julian Strizek. The potential for using web surveys to investigate drug sales through cryptomarkets on the darknet. *Monitoring drug use in the digital age: Studies in web surveys, EMCDDA Insights. Available at www.emcdda.europa.eu/publications/insights/web-surveys/potential-web-surveys-investigate-drug-sales-cryptomarkets-darknet_en*, 2022.

[18] Annalise Kempen. The drug dealer lurking inside your smartphone and computer. *Servamus Community-based Safety and Security Magazine*, 115(6):10–13, 2022.

[19] Dominique Laferrière and David Décary-Hétu. Examining the uncharted dark web: Trust signalling on single vendor shops. *Deviant Behavior*, pages 1–20, 2022.

[20] Gianluigi Me and Liberato Pesticcio. Tor black markets: economics, characterization and investigation technique. In *Cyber Criminology*, pages 119–140. Springer, 2018.

[21] Gianluigi Me, Liberato Pesticcio, and Paolo Spagnoletti. Discovering hidden relations between tor marketplaces users. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 494–501. IEEE, 2017.

[22] Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin, Vineet Mishra, Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, and Paulo Shakarian. Darknet and deepnet mining for proactive cybersecurity threat intelligence. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pages 7–12. IEEE, 2016.

[23] Florian Platzer, Fabian Brenner, and Martin Steinebach. Similarity analysis of single-vendor marketplaces in the tor-network. *Journal of Cyber Security and Mobility*, pages 205–238, 2022.

[24] Florian Platzer and Alexandra Lux. A synopsis of critical aspects for darknet research. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, pages 1–8, 2022.

[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[26] Martin Steinebach. File-sharing and the darknet, 2020.

[27] Martin Steinebach, Marcel Schäfer, Alexander Karakuz, and Katharina Brandl. Detection and analysis of tor onion services, 2020.

[28] Martin Steinebach, Sascha Zenglein, and Katharina Brandl. Phishing detection on tor hidden services. *Forensic Science International: Digital Investigation*, 36:301117, 2021.

[29] Meropi Tzanetakis. Comparing cryptomarkets for drugs. a characterisation of sellers and buyers over time. *International Journal of Drug Policy*, 56:176–186, 2018.

[30] Xiangwen Wang, Peng Peng, Chun Wang, and Gang Wang. You are your photographs: Detecting multiple identities of vendors in the darknet marketplaces. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ASIACCS '18, page 431–442, New York, NY, USA, 2018. Association for Computing Machinery.

[31] Sandra Wittmer, Florian Platzer, Martin Steinebach, and York Yannikos. Deanonymisierung im tor-netzwerk–technische möglichkeiten und rechtliche rahmenbedingungen. In *Selbstbestimmung, Privatheit und Datenschutz*, pages 151–169. Springer Vieweg, Wiesbaden, 2022.

[32] Sandra Wittmer and Martin Steinebach. Verwendung computergenerierter kinderpornografie zu ermittlungszwecken im darknet, 2019.

[33] York Yannikos, Quang Anh Dang, and Martin Steinebach. Comparison of cyber attacks on services in the clearnet and darknet. In *IFIP International Conference on Digital Forensics*, pages 39–61. Springer, 2021.

[34] York Yannikos, Julian Heeger, and Maria Brockmeyer. An analysis framework for product prices and supplies in darknet marketplaces. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, pages 1–7, 2019.

[35] York Yannikos, Julian Heeger, and Martin Steinebach. Data acquisition on a large darknet marketplace. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, pages 1–6, 2022.

[36] York Yannikos, Annika Schäfer, and Martin Steinebach. Monitoring product sales in darknet shops. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, page 59. ACM, 2018.

[37] Diego Zambiasi. Drugs on the web, crime in the streets. the impact of shutdowns of dark net marketplaces on street crime. *Journal of Economic Behavior & Organization*, 202:274–306, 2022.

Biographies

York Yannikos is a research associate in the Media Security and IT Forensics department at Fraunhofer SIT and a researcher at ATHENE. He holds a (equiv. Master's) degree in computer science from the University of Rostock, Germany. His research interests include digital forensic tool testing, darknet marketplaces, and open source intelligence.

Julian Heeger is a research associate in the Media Security and IT Forensics department at Fraunhofer SIT and a researcher at ATHENE. He holds a Master's degree in IT security from the Technical University of Darmstadt.

Martin Steinebach is the head of the Media Security and IT Forensics department at Fraunhofer SIT. He studied computer science and received his PhD from the Technical University of Darmstadt for this work on digital audio watermarking in 2003. From 2003 to 2007 he was head of the Media Security in IT department at Fraunhofer IPSI. In 2016 he became honorary professor at the TU Darmstadt and gives lectures on multimedia security as well as civil security. He is principle investigator at ATHENE and represents IT Forensics and AI Security. Previously, he was principle investigator at CASED with the topics multimedia security and IT forensics.