TU Delft

Delft University of Technology

Quantifying Dark Web Shops' Illicit Revenue

Oosthoek, Kris; van Staalduinen, Mark; Smaragdakis, Georgios

DOI
10.1109/ACCESS.2023.3235409

Publication date
2023

Document Version
Final published version

Published in
IEEE Access

Citation (APA)
Oosthoek, K., van Staalduinen, M., & Smaragdakis, G. (2023). Quantifying Dark Web Shops' Illicit Revenue. *IEEE Access*, *11*, 4794-4808. https://doi.org/10.1109/ACCESS.2023.3235409

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 November 2022, accepted 29 December 2022, date of publication 9 January 2023, date of current version 18 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3235409

RESEARCH ARTICLE

Quantifying Dark Web Shops' Illicit Revenue

KRIS OOSTHOEK[1], (Member, IEEE), MARK VAN STAALDUINEN[2], AND GEORGIOS SMARAGDAKIS[1], (Senior Member, IEEE)

[1]Electrical Engineering, Mathematics and Computer Science (EEMCS), Delft University of Technology, 2628 CD Delft, The Netherlands
[2]CFLW Cyber Strategies, 2691 HB 's-Gravenzande, The Netherlands

Corresponding author: Kris Oosthoek (k.oosthoek@tudelft.nl)

This work was supported in part by the European Research Council (ERC) under Starting Grant ResolutioNet ERC-StG-679158.

ABSTRACT The Dark Web, primarily Tor, has evolved to protect user privacy and freedom of speech through anonymous routing. However, Tor also facilitates cybercriminal actors who utilize it for illicit activities. Quantifying the size and nature of such activity is challenging, as Tor complicates indexing by design. This paper proposes a methodology to estimate both size and nature of illicit commercial activity on the Dark Web. We demonstrate this based on crawling Tor for single-vendor Dark Web Shops, i.e., niche storefronts operated by single cybercriminal actors or small groups. Based on data collected from Tor, we show that just in 2021, Dark Web Shops generated at least 113 million USD in revenue. Sexual abuse is the top illicit revenue category, followed by financial crime at a great distance. We also compare Dark Web Shops' activity with a large Dark Web Marketplace, showing that these are parallel economies. Our methodology contributes towards automated analysis of illicit activity in Tor. Furthermore our analysis sheds light on the evolving Dark Web Shop ecosystem and provides insights into evidence-based policymaking regarding criminal Dark Web activity.

INDEX TERMS Computer crime, bitcoin, dark web.

I. INTRODUCTION

The World Wide Web (shortly Web) has been recognized as one of the greatest achievements of our times. It offers unprecedented opportunities for communication and commerce, and has truly revolutionized our lives. The original design of the Web did not have anonymity as a requirement. Any user browsing the Web leaves digital footprints that can be traced and unveil the user's identity [28], [55]. The public Web is also easy to crawl and index, hence collecting data to profile users. Users and administrators soon realized the privacy risks of the public Web and tried to protect content, user profiles, and communication with passwords and other authentication methods. Together with paywalls restricting access and thus indexing, this created the *Deep Web*, a part of the Web not indexed by search engines.

Over the years, many solutions were developed to offer anonymity to Web users ranging from end-to-end cryptography using public keys [18], [49], to Transport Layer Security (TLS) [1], [34], and anonymous communication [52]. While the first two communicate point-to-point, the latter is relayed, potentially better protecting user identity. The Onion Router (Tor) [48] is the most successful implementation for anonymous communication. Tor started as a US military project to protect the private communication of US military personnel deployed around the globe. Today, Tor is an independent overlay network of 7,000 nodes (relays) globally [56], [57].

Tor also is the infrastructure that supports the *Dark Web*, i.e., the Deep Web content that exists on overlay networks, called *darknets*, that operate on top of the public Internet. Darknets and Dark Web content can only be accessed with specific software, configurations, or authorization and often use a customized communication protocol. Moreover, Darknets can communicate and conduct business anonymously without revealing user information, e.g., the user's location or Internet Protocol (IP) address. The Dark Web became popular among activists as it protects the freedom of speech under duress and activists in different regions of the world, e.g., protesters in Arabic Spring [48], and whistle-blowers such as WikiLeaks [66].

Unfortunately, the anonymity by design facilitated by the Dark Web also was attractive to cybercriminals and terrorists.

The associate editor coordinating the review of this manuscript and approving it for publication was Yang Liu.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

By some estimates, the illicit activity on the Dark Web exceeds 2 billion USD [5], [11], [12], [58], [59]. However, such reports do not reveal information about their data sources. Usually, they focus on large *Dark Web Marketplaces* that provide a platform for the anonymous distribution of illegal goods, e.g., guns, drugs, sexual abuse material, and stolen financial data. Many Dark Web Marketplaces have been prosecuted and seized by law enforcement agencies, e.g., DarkMarket [24] and Hydra [62].

In recent years, small shops, called *Dark Web Shops*, single-vendor shops run by individuals or small-scale collectives, have been added to the Dark Web ecosystem. There are many reasons these small individually owned shops became popular: (i) Readily available webshop software has enabled Dark Web retailers to sell illicit goods directly, without paying a commission to Dark Web Marketplaces [44]; (ii) Retailers on the Dark Web increasingly avoid affiliation with notorious Dark Web Marketplaces, which are frequently involved in geo-political power games [61]; (iii) The take-down of Dark Web Marketplaces has affected business continuity and trust of some of the retailers, leading them to initiate self-hosted shops [22].

Previous research has analyzed the Dark Web and tried to quantify revenue from illicit trading on the Dark Web. Most of these studies focused on Dark Web Marketplaces as they have been popular during the last years [4], [7], [11], [12], [13], [20], [32], [39], [44], [54], [58], [59], [64]. Furthermore, focusing on a single Tor domain expedites data collection. In this paper, we focus on the evolving ecosystem of individually owned shops, as a specific subset of the whole Dark Web ecosystem. We attempt to understand its structure, operation, payment revenue, and laundering strategies. We also compare the structure and operation of Dark Web Shops with Dark Web Marketplaces and investigate differences and similarities.

The Dark Web Shops ecosystem is a less well-studied portion of the Dark Web that is also fueled with cryptocurrencies, especially Bitcoin [16], [38], [63], [64]. Our study sheds light on the evolving Dark Web ecosystem and is one of the first large-scale studies to estimate the illicit revenue generated by Dark Web Shops and understand the popularity of abuse types such shops facilitate.

We provide timely and valuable insights, as many Dark Web Shop transactions are suspicious. According to forthcoming market regulation legislation, suspicious cryptocurrency transactions must be reported to the authorities. For example, from 2024, the European Union will enforce the new Markets in Crypto-Assets (MiCA) rules [47]. MiCA requires cryptocurrency exchanges and other service providers to identify issuers of cryptocurrency transactions and owners of self-hosted hardware wallets for cryptocurrency transactions over 1,000 Euros. We hope the insights provided in this study contribute to informed policy-making in this area.

The contributions of this paper can be summarized as follows:

- To collect input data for our methodology, we develop a crawler for illicit Tor onions to collect Bitcoin addresses and characterize associated illicit activities.
- We develop a methodology to perform extensive data cleansing on a dataset of illicit Tor domains to filter out non-illicit and duplicate Tor domains, unrelated and incorrectly formatted Bitcoin addresses.
- Our analysis of the Tor crawler data based on our methodology shows that the revenue of Dark Web Shops was at least 113 million USD in 2021.
- Our analysis shows that the top category of illicit offerings by revenue is sexual abuse, totaling close to 94 million USD revenue; followed at large distance by financial crime, accounting for more than 10 million USD.
- Our investigation shows no overlap between Bitcoin addresses we discovered related to Dark Web Shops and those released after the take-down of the largest Dark Web Marketplace, Hydra (that by some measures had 80% of the Dark Market Revenue share). This suggests that shops and marketplaces are parallel Dark Web ecosystems.
- Our analysis shows that cryptocurrency exchange platforms are used by both owners of Dark Web Shop and Dark Web Marketplaces, which motivates the need for continuous monitoring and regulatory intervention.

II. BACKGROUND

A. TOR

Tor is an abbreviation of The Onion Router [17]. It is the most popular software for darknets and is widely used for implementing *onion routing*, i.e., relaying traffic through multiple servers (relays) and adding additional encryption at each hop. The Tor core software and Tor Browser are free and open source. As a network, Tor is maintained by many volunteers running Tor nodes, collectively providing an overlay network intended to facilitate increased user privacy over the regular Internet, effectively hiding user IP addresses. Next to many Tor domains (also called *onions*) serving hypertext similar to the regular Hypertext Transfer Protocol (HTTP), the Tor network is also used to facilitate other Transmission Control Protocol (TCP) based services such as email (OnionMail) and instant messaging (Ricochet Refresh), which uses Tor for its peer-to-peer transactions. Many popular browsers are also able to route traffic over Tor for anonymity. The Tor network further provides bridges to the regular Internet to defeat government censorship in several jurisdictions, e.g., during the Arab Spring in late 2010 [48]. Today, more than 7,000 Tor nodes are online [56], [57].

B. BITCOIN

Bitcoin is a digital currency based on peer-to-peer technology [40]. As opposed to government-issued (fiat) currencies such as the US dollar, the Euro, and the pound sterling, which central banks control, Bitcoin is not overseen by a central authority. Transactions between users and the issuing of new

Bitcoin are performed collectively by a global network of close to 15 thousand Bitcoin nodes [5], making it a decentralized currency. Bitcoin transactions, i.e., the transfer of value from one user to another, are effectively data structures broadcasted to the Bitcoin network, composed of at least one input and output. Inputs are quantities of Bitcoin controlled by the sender, with outputs specifying their destination. Every transaction represents a state transition in the blockchain, which is confirmed through mining, which leads to consensus. After confirmation, transactions are irreversible and are stored in the blockchain and propagated to all nodes in the network.

C. BITCOIN: REGULATION AND MARKET CAPITALIZATION

While Bitcoin was designed to function anonymously, its current mainstream usage has effectively made it pseudonymous. Based on Know Your Customer (KYC) legislation [46] rolled out in many jurisdictions, people are required to legally identify themselves when signing up with an exchange platform to be able to buy Bitcoin. The disclosure of their names makes it difficult to achieve complete anonymity when a transaction shows up in an investigation. Law enforcement investigators can link several steps back to their origin. Suppose this is an exchange platform that is registered as a benign financial service provider in a jurisdiction. In that case, they can order the exchange to disclose the user's identity behind the specific transaction. This opens up possibilities for forensic investigation through blockchain analysis.

In the current bear market (Fall 2022), Bitcoin's market capitalization is 400 billion USD on average [67], which is significantly less than its record market cap of 1,156 billion USD in November 2021. The illicit activity in Bitcoin is estimated at 2 billion USD, i.e., less than 1 percent as reported (lower bound estimations) by blockchain analytics firms, e.g., Chainanalysis [11], the nominal value is still considerable. Especially when taking into account that criminal activity like money laundering usually increases in times of economic downturn [25] and geopolitical tension [60]. Fortunately, Bitcoin's open ledger is a robust forensic tool, enabling unprecedented opportunities to track funds, especially when compared to tracing cross-border bank transactions.

III. RELATED WORK

Previous research studied the Dark Web and tried to quantify revenue from illicit trading on the Dark Web. Most authors have focused on Dark Web Marketplaces as they have been popular during the last years [4], [7], [11], [12], [13], [20], [32], [39], [44], [54], [58], [59], [64]. Relatively few studies focused on other parts of the Dark Web [7], [33], [38].

Christin [13] crawled the Silk Road Marketplace and found it was primarily drug-oriented. Meiklejohn et al. [39] purchased items from various Dark Web Marketplaces to obtain seller Bitcoin addresses as input to clustering heuristics. Hiramoto and Tsuchiya [32] have analyzed Bitcoin transactions of addresses associated with seven Dark Web Marketplaces based on Bitcoin addresses gathered via walletexplorer.com [65]. Their analysis, however, didn't check if the addresses appeared on the actual Dark Web Marketplace. Hence they work with an indirect data source, solely relying on a clustering algorithm. Elbahrawy et al. [20] have focused on customer migration between different Dark Web Marketplaces based on pre-processed vendor data.

Bracci et al. [7] studied the selling of COVID-19 products in 194 different Tor outlets, specifically on selling vaccines. In earlier work, authors performed similar focused research into cybercriminal capabilities [64], stolen identity documents [54], firearms [44], and drugs [4].

Lee et al. [38] analyzed Bitcoin transactions to addresses scraped from Tor. The set of addresses was relatively small, but important insights about the Dark Web between 2013 and 2018 could be extracted. The scraped domains were categorized into several categories. Their analysis showed that over 80% of the Bitcoin addresses in the Dark Web were indeed used with malicious intent. Their study estimates the Dark Web revenue in their dataset to be around 180 million USD for the period between 2013 and 2018. Their seed dataset contained 85 Bitcoin addresses.

Paquet-Clouston et al. [45] used the co-spending heuristic [37] to estimate ransomware payments in Bitcoin. Based on an analysis of Bitcoin addresses from 35 ransomware families, they quantify the minimum worth of the ecosystem at over 12 million USD. However, they included addresses that represented 2 million USD in revenue afterward attributed to the Silk Road black market and thus cannot be fully accounted for as ransomware payments. A recent work by Oosthoek et al. [42] analyzed ransomware payments worth around 101 million USD in recent years, and they showed that there is no overlap between the Bitcoin addresses used for ransomware and those used in reported Bitcoin addresses from studies in the Dark Web.

Chainalysis [11], [12] publishes annual reports with estimations about the total revenue of illicit activity on the Dark Web and per category. The estimate for 2021 was 2.1 billion USD. Although the analysis provides valuable policy-making insights, their methodology is proprietary. Moreover, they focus on Dark Web Marketplaces exclusively. United Nations [35], [58], [59] and Interpol [35] also publish reports for the revenue in the Dark Market, again focusing on notorious Dark Web Marketplaces and illicit activity such as drugs, trafficking, and guns. The sources of the data are also proprietary.

To our knowledge, our analysis is the first to provide a thorough methodology for the analysis of crawled Tor data. In our demonstration of its application, we shed a unique light on the evolving ecosystem of Dark Web Shops, based on a dataset with much higher coverage than previous studies.

IV. METHODOLOGY

This section describes the Tor crawler we developed and implemented to collect content from onions. It also describes the Bitcoin address clustering methodology we used in our

analysis. The cleansing methodology explicitly developed for this analysis is discussed separately in Section V.

A. TOR CRAWLER

While search engines index the content of the regular Internet, such indexing is not possible on the Dark Web. Access to Dark Web data requires using specialized software, such as the Tor browser and the Tor relay client. Indexing of Dark Web content is further complicated by the fact that Dark Web domains are usually short-lived [51].

The Tor crawler that we utilize for our collection and analysis of Dark Web data was launched in 2013 as part of a research project [53] to increase the coverage of Dark Web that can be indexed beyond a small number of seed Tor domains that can be found on the publicly accessible Web (clearnet). Today, the data collected by the crawler is available as a commercial product, called Dark Web Monitor (DWM), mainly to law enforcement agencies worldwide by CFLW Cyber Strategies. The crawler has provided insights that law enforcement agencies and prosecutors have utilized in recent years.

The crawler maintains a list of onions and adds new domains when they are discovered in the crawling process. Every onion is crawled at least every 18 hours. This ensures that even short-lived domains are crawled and indexed. For each onion indexed, the crawler follows all address paths from pages available within the domain (page tree). If a previously unseen domain is discovered, the crawler will automatically crawl that URL to add it to the archive and schedule for automatic crawling of the new URL. One of the main challenges is to have a complete overview of onions, as this is not facilitated and, on a technical level, not supported by the Tor network itself. This 'snowballing' approach of scanning all pages for new URL entries recursively leads to new entries which each crawl.

When a Tor domain is offline, either because it is not active anymore or due to temporal unavailability, e.g., outage or routing issues, the Tor domain is revisited with an inter-visit interval of 1 hour. In the case that the Tor domain continues to be unavailable after three attempts, the crawling schedule for this domain follows an exponential back-off, i.e., the Tor domain is visited after 18, 36, 72 hours up to a maximum revisit regularity of 10 days.

For the content analysis, the crawler uses regular expressions. It automatically extracts cryptocurrency addresses, PGP keys, and email addresses that can be used for attribution. The raw data is archived in cloud storage buckets. Since its launch, the data accounts for 25 Terabytes (until end of first semester 2022), 15 Terabytes collected during 2021 alone.

Our analysis shows that multiple cryptocurrencies are used for illicit activity, namely, Bitcoin, Bitcoin Cash, Litecoin, Monero, Ethereum, and Binance Coin. However, our analysis confirms previous results [38] that, by far, the most popular cryptocurrency is Bitcoin. Indeed, around 99% of all addresses discovered in our dataset is Bitcoin.

TABLE 1. Categorization of abuse types used to classify Tor domains based on content for the Tor domains indexed by our crawler.

Abuse Type	Description
Cybercrime	DDoS, bulletproof hosting, exploit development
Drugs / Narcotics	Cannabis, synthetic drugs, pharma
Extremism	Extreme right, radical islam, anarchists, neo-nazi
Financial Crime	Carding, hacked accounts, stolen giftcards
Goods and Services	Marketplaces, counterfeit, gambling, firearms
No Abuse	Whistleblower, communities, how-to's, non-profit orgs.
Sexual Abuse	Child sexual abuse, animal sexual abuse, sextortion
Violent Crime	Assasination, hate crime

For page content classification, we use human crowdsourcing to categorize the content. Each newly crawled domain is inspected by a team of analysts that, based on the page content, assigns a label indicating the primary type of abuse observed on that particular domain. A domain may be assigned to more than one human analyst to improve the accuracy of the labeling. An overview of the so-called ''abuse types'' used in our study is available in Table 1. We notice that our categorization and description do not follow other proposed, but not yet standardized categorizations [36].

For our study, we utilize the latest version of our Tor crawler [53], introduced in 2020. The latest version of the Tor crawler establishes over 100 parallel connections and makes it possible to scan all known Tor domains within 24 hours. Since its launch, the crawler is estimated to have indexed about a fifth of Tor domains based on statistics published by the Tor project [56], i.e., approximately 1.5 million unique Tor domains. The un-indexed domains are primarily onions serving non-HTTP protocols. Approximately 100 thousand new unique online domains were crawled and indexed in the first semester of 2022. This figure accounts for the many mirrors used by actors to increase the resiliency of their operations. Such duplicates are aggregated within a single domain ID if the HTML source code is identical to another domain.

Our crawler has certain limitations. Onions may be protected by CAPTCHA [7], [15], making crawling and indexing challenging. This is typically true for Dark Web Marketplaces but not for Dark Web Shops. Indeed, popular Dark Web Marketplaces are usually protected with CAPTCHA or user passwords, e.g., Hydra Market, and are not indexed partially or not indexed at all. A recent study [16] shows that coverage of scrapers of Dark Web Marketplaces is usually low, missing on average 46% of the listings. Due to this, the actual revenue of Dark Web Marketplaces is systematically underestimated. On the contrary, due to the implementation of standard off-the-shelf software suits that do not support CAPTCHAs by default, the majority, more than 80%, of single vendor shops has not (yet) implemented CAPTCHAs. A few examples of such stores at the time of writing (December 2022) are DrugzFromNL, Firearms72, Deep Web Guns Store, Patron Cocaine, WeAreAMSTERDAM, Tom and Jerry Shop and

GammaGoblin. We are aware that particular single vendor shops, primarily those facilitating *serious crime* such as homicide, might be additionally protected by CAPTCHA and are thus not included in our analysis. The current version of our crawler does not crawl nor index domains protected by a user login. However, it crawls and indexes the front page of the Tor domain. This leads to a partial view, meaning that only non-protected onions are fully indexed. Our analysis provides a lower bound of the estimate of illicit revenue by Dark Web Shops. Moreover, not all the Tor domains are scanned with the same frequency. Thus, it is possible to have a less accurate index for high dynamic content domains when compared with static content domains. This is a limitation of any crawling process and also applies to many crawlers that index the publicly accessible Web.

B. BITCOIN ADDRESS CLUSTERING

Bitcoin address clustering aims to break pseudoanonymity in blockchain by linking Bitcoin addresses that are controlled by the same entity based on the information available from blockchain transaction analysis. Several heuristics have been proposed to achieve Bitcoin address clustering based on different assumptions of how users transact in a blockchain [30], [39], [41]. To discover whether a Bitcoin address belongs to a cluster of multiple addresses, we use GraphSense [31], which builds on BlockSci [37]. To discover Bitcoin address clusters, called entities, GraphSense exclusively uses the co-spending heuristic, also known as multi-input, which high effectiveness has been shown empirically [2], [30], [39]. The co-spending heuristic recursively queries addresses that were used to combine funds in a transaction. If a transaction has input from multiple addresses, these are all likely controlled by the same actor (individual or group). Figure 1 provides a graphical representation of this hypothesis.

While the co-spending heuristic is generally reliable, it might lead to false positives caused by CoinJoin and PayJoin transactions [31]. CoinJoin and PayJoin are privacy-preserving transaction methods that combine payments of multiple parties into one transaction to obfuscate ownership. GraphSense uses the algorithm proposed by Goldfeder et al. [27] to identify the most common types of CoinJoin transactions and exclude these from input into the clustering heuristic. Another common heuristic, the change address heuristic, isn't implemented in GraphSense as its reliability has been proven inconsistent due to its dependence on critical characteristics in end-user wallet software [37].

V. CLEANSING METHODOLOGY

Data crawled from Tor is inherently noisy. Proper filtering will provide a more accurate portrayal of the relevant ecosystem. In this section, we present our methodology to remove corrupted, incorrectly formatted, duplicate, or incomplete data, i.e., to perform extensive data cleansing, resulting in a dataset that can serve as a basis for dependable lower-bound estimates. Our methodology, described in detail below, focuses on cleansing three core aspects of our data

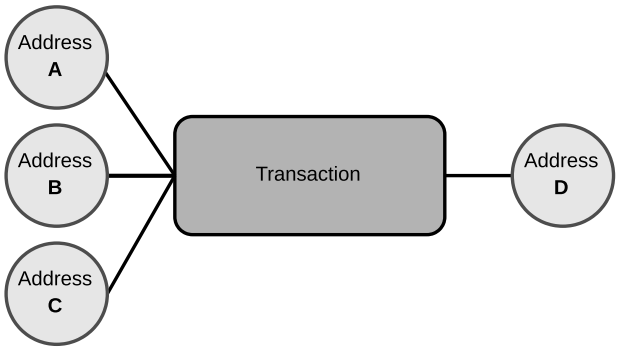

FIGURE 1. Bitcoin address co-spending heuristic: Example of three Bitcoin addresses with common spending in one transaction.

set: Tor domains (onions), Bitcoin addresses, and Bitcoin address clusters detected with the co-spending heuristic. For an illustration of the pipeline of our methodology, we refer to Figure 2. We hope to contribute to the standardization and replication of analyses like ours by providing a detailed design and evaluation of our methodology.

A. TOR DOMAINS

A portion of Tor domains is legal, with the facilitation of anonymous, licit services as the sole intention. Our analysis exclusively focuses on *illicit*, i.e., unlawful criminal activity. This means we solely regard pairs of Tor domains and Bitcoin addresses linked to suspicious, or likely illegal, activity, which we confirm through inspection of each pair. This inherently leads to lower-bound results, as the relationship between many domains and addresses needs to be clarified, leading to exclusion from analysis. We only include domain-address pairs which are manually validated as illicit.

The initial stage of our cleansing methodology focuses on filtering out non-illicit or otherwise unwanted domains. Each domain represents a unique address in the `.onion` special-use top-level domain. The key objective of the first cleansing phase is to establish relationships between a Tor page with an illicit offering and a Bitcoin address. These relationships can be one-to-one, meaning an individual domain contains a single valid Bitcoin address or one-to-many, i.e., it contains more than one address.

We focus exclusively on the entire year of 2021, as the latest crawler version was introduced in 2020. From our crawler, we obtained Tor domains, also referred to as *onions*, which appeared online between January 1 and December 31, 2021. The content was collected and indexed for each Tor page crawled. Domain names, Bitcoin addresses, and page titles were parsed from the crawler collection. Other metadata and page sources were stored separately for reference. To each Tor domain, a label was added indicating the abuse type as listed in Table 1. These labels are assigned by a team of analysts that manually inspect newly crawled pages. Domains clearly non-illicit, i.e., of civil rights organizations, political parties, or whistle-blower sites, are classified as *No Abuse* in our study. Note that this provides a two-step approach to

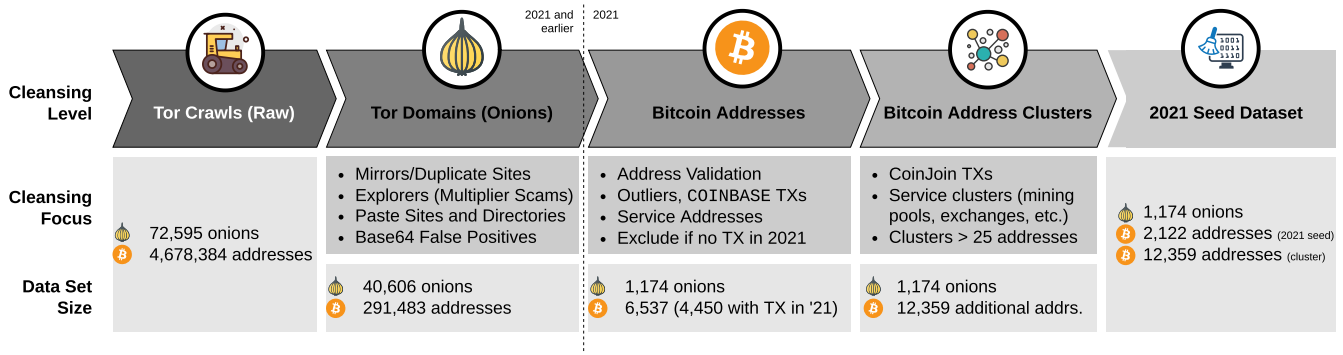

FIGURE 2. Pipeline of our dataset collection and cleansing methodology.

establish the illicit nature of domains: (i) during the labeling of newly scraped domains and (ii) in our manual analysis of remaining domain-address pairs after completion of all steps of the methodology.

The corpus of collected raw data analyzed for this paper includes 72,595 unique domains which appeared online in the Tor network at some point in 2021. The crawler collected and indexed content from these domains for 710,484 pages (URLs). After analyzing the content, 138,967,218 non-unique cryptocurrency addresses were extracted. A single cryptocurrency address can be detected within multiple Tor domains. This primarily occurs due to mirrored domains and the presence of blockchain explorers, which display recently mined blocks, addresses, and transactions. After our analysis, we identified 4,730,419 unique cryptocurrency addresses, of which the vast majority, i.e., 4,678,384 were Bitcoin addresses. These addresses are unverified, meaning that they are formatted as a Bitcoin address but not yet sanity-checked and confirmed by a Bitcoin node as valid. This happens in a consecutive cleansing phase. With Bitcoin dominating our dataset, with 98.9% addresses being Bitcoin, dominant over other detected cryptocurrencies, we focused on Bitcoin exclusively as the dominant currency in the Dark Web.

1) MIRRORS

Our raw dataset contains over 70 thousand unique Tor domains. However, owners of Tor sites use multiple redundant domains, and often infrastructure is taken offline and made available again on a new domain. As our crawler saves the page tree with each visit, we were able to filter out full-match duplicates based on both the title of the front page and contents of the page source, based on a hash generated at crawler runtime. Based on this, we identified 51,324 unique onions in our dataset for the year 2021.

2) NON-ILLICIT/UNWANTED DOMAINS

We excluded Tor domains that did not fit our classification of illicit activity in Table 1, focusing on outliers by rank-ordering the domains in our dataset based on the number of Bitcoin addresses per individual domain. This reduces the initial 51,234 domains to 40,606 unique domains, primarily due to the exclusion of three categories:

(i) Explorers: Our crawler output contains domains that automatically post block mining output, similar to blockchain explorers such as *blockchain.com*. These sites are advertised as Bitcoin multipliers, displaying recent transaction data as proof of their supposed capabilities, a tactic also observed by previous studies [21]. While these apparent scams extort money from unaware victims and, thus, are illicit, the Bitcoin addresses advertised are unrelated. Hence we excluded such domains from our analysis. We did this based on a rank order of address quantity per domain and manual inspection.

(ii) Indexes and Directories: We also exclude index sites and Tor directories. These sites, which also exist on the public Web, serve as springboards linking to various Tor hidden services. Some of these host copies of specific pages they are linking to, causing duplicate pages found on different sites. We also removed non-illicit pages that appeared on illicit domains, as these also cause duplicates.

(iii) Paste sites and Forums: The set of Tor domains was manually inspected to remove further sites that weren't clearly illicit. Notable examples of excluded domains are paste sites listing Bitcoin addresses without clear context and forum posts referring to Bitcoin addresses without clear intent. Messages in foreign languages were automatically translated and manually inspected to understand the context, and the corresponding Bitcoin addresses were only preserved when in scope.

3) FALSE POSITIVES

In this step, we removed false positives caused by domains using inline Base64-encoded images often used to slow down crawlers [13], of which a portion was detected as a Bitcoin address by our crawler. We also checked whether each domain had a label indicating the abuse type attached and additionally checked each abuse type for correctness using a random sample of 100 domains. This first phase of cleansing results in an intermediate dataset of 40,606 unique Tor domains with 291,483 unique Bitcoin addresses.

B. BITCOIN ADDRESSES

After filtering out non-illicit and unwanted Tor domains, we also need to filter out Bitcoin addresses unrelated to illicit activity. We assume the exact requirement that an address and the majority of its holdings in Bitcoin should be confidently classified as illegal. This isn't straightforward due to Bitcoin's privacy and pseudonymous characteristics. Hence we opted for a lower-bound estimate, excluding all addresses which can be attributed as belonging to a Bitcoin exchange platform. For such addresses, a portion of holdings is likely illicit, but the proportion cannot be reliably established. The domain itself was excluded from further analysis when all Bitcoin addresses detected in a single Tor domain were excluded.

1) ADDRESS VALIDATION

We checked the remaining 291,483 Bitcoin addresses against a Bitcoin node for validity. 38,212 addresses were reported as invalid, i.e., sanity checks such as for address formatting did not pass the test, or the existence of the address hasn't yet been confirmed in block mining. Out of the 253,271 valid addresses, a remarkable quantity of 246,187 (97.2%) had no transactions, meaning they were never used according to the blockchain data. These addresses cannot represent any illicit activity, so they were also disregarded, resulting in 7,084 valid addresses with one or more transactions.

2) OUTLIERS

This step excluded outliers based on the number of observations of individual Bitcoin addresses in different Tor domains and the total holdings of these addresses. Based on this, we excluded Bitcoin addresses found in Tor domains with Bitcoin ''Rich'' Lists, i.e., displaying Bitcoin addresses with the biggest holdings. We also excluded several Bitcoin addresses if they historically only received COINBASE transactions, which indicates they belong to mining pools. COINBASE transactions (not to be confused with the exchange platform of the same name) are newly mined coins issued as a block reward, which cannot be related to illicit activity. This was furthermore validated by excluding mining-related addresses shared by Romiti et al. [50] and GraphSense [29].

3) SERVICE ADDRESSES

We refer to addresses controlled by centralized exchange platforms such as Coinbase and Kraken as service addresses, as the exchange service owns the private key of the addresses used for deposit and withdrawl. This also includes addresses associated with Bitcoin-accepting payment providers and gambling sites, which store user-owned Bitcoin in custody [43]. Exchange platforms are of great importance to blockchain analysts because they provide an opportunity to identify real-world actors behind Bitcoin transactions if the exchange adheres to Know Your Customer (KYC) legislation. However, addresses operated by exchanges likely represent the holdings of more than one user. Furthermore, ownership of funds can be transferred without on-chain evidence through paper wallets or shared credentials.

As we cannot reliably classify funds terminating at exchanges as illicit, we have excluded these from our analysis based on two metrics. First identified exchanges using labels from GraphSense [31], walletexplorer.com [65], and BitRank [6] (a commercial service with a free daily allowance). If one or more of these services identified an address controlled by an exchange, it was excluded. Addresses with more than 1,000 incoming transactions were also excluded. In total, 547 addresses were removed, further decreasing our set of addresses to 6,537.

By filtering out exchanges and mining-related addresses, we likely also exclude from our dataset the portion of revenue sent to that address. Filtering out addresses with over 1,000 transactions may also exclude non-exchange addresses. This is a well-considered step in our approach to a conservative but clean estimate. We strive to exclude any funds that cannot reliably be attributed to an illicit offering on Tor.

4) BITCOIN TRANSACTIONS IN 2021

For our analysis, we focus on the year 2021, which is an entire year with the latest version of the crawler. To get an impression of what 2021 looked like in terms of illicit revenue by Dark Web Shops, we only regarded transactions between January 1 and December 31, 2021. We filtered for addresses 'active' in 2021, i.e., with one or more transactions during the above period. This filter reduced the corpus of Bitcoin addresses from 6,537 to 4,450. Tor domains with exclusively Bitcoin addresses that didn't have any transactions in 2021 were also excluded. As a result of the last filter, the amount of Tor domains included dropped to 1,174.

C. BITCOIN ADDRESS CLUSTERS

For Bitcoin address clustering, we used GraphSense [31], which builds on BlockSci [37]. GraphSense uses BlockSci's ability to detect the most common types of CoinJoin and does not detect any when we apply it to our dataset. According to labels from various sources described earlier, using privacy wallets such as Wasabi was also non-existent. Previous reports also mentioned that off-the-shelf Dark Web store front-end software such as Eckmar [19] and TradeMed [3] have become more sophisticated and generate new Bitcoin addresses for each purchase by default. This makes address clustering more challenging.

We excluded probable service clusters if one or more of the following two criteria were met: (i) the cluster contains more than 1,000 addresses and (ii) if one or more of three unique sources (Graphsense [31], walletexplorer.com [65], BitRank [6]) attributes the cluster itself or one or more addresses in a cluster to an exchange platform.

The most significant effect due to this exclusion of service clusters occurred in the Financial Crime category. The identification and subsequent exclusion of clusters of exchanges,

Service Clusters, also leads to the exclusion of service addresses in the seed dataset. Because of this, our final number of seed addresses used for analysis is 2,122. This is a significant reduction, the process of which is represented in Section V. The illicit revenue represented by this set of addresses is a lower-bound estimate of overall illicit revenue in Tor related to Dark Web Shops. However, due to the various steps taken, we are confident that as opposed to the initial 291 thousand addresses, the 2,122 seed addresses provide a robust representation of payment size, buyer activity, and distribution between different types of illicit activity related to the Dark Web Shops.

VI. QUANTIFYING ILLICIT REVENUE

Based on the methodology outlined in Section V, in this section, we provide an overview of illicit revenue made by Dark Web Shops in the entire year of 2021. We discuss results from the analysis of incoming and outgoing Bitcoin transactions to the set of Bitcoin seed addresses, as well as based on an expanded set of addresses, using the heuristics discussed in Section IV.

A. SEED ADDRESS REVENUE PER ABUSE TYPE

Table 2 provides an overview of the results of our analysis by type of abuse, being the type of illicit activity (in the first column). We refer to Table 1 for a description of each abuse type. In the second column, we provide the number of onions and affiliated pages per abuse type. Although the number of domains is in the order of tens, the number of affiliated pages is typically in the order of thousands. Sexual abuse and financial crime are the two categories with the highest number of Tor domains or onions and pages, with around thirty thousand affiliated pages each, i.e., around 82% of the domains are associated with these two categories. Notice also that the No Abuse category is very small. As discussed in Section V it only contains a small number of civil rights organizations and whistle-blower sites; onions not evidently non-abusive were not considered in our analysis. As shown in Table 2 total, for our analysis, we consider 1,197 Tor domains and 73,209 pages.

The third column presents the number of seed Bitcoin addresses included per abuse type. Unfiltered is the raw crawler result, and the filtered number is after the application of our methodology. For our analysis, we utilize the set of filtered seed addresses after the cleansing data process described in the previous section. In parentheses, we provide the results of Bitcoin address clustering. For completion, we report both the output of the clustering (unfiltered) and the results after cleansing (filtered). In our analysis, we take a conservative approach by only considering the filtered set of Bitcoin addresses and filtered clusters. Again, the popular categories are sexual abuse and financial crime, with more than a million and half a million associated Bitcoin addresses. More than 270 thousand Bitcoin addresses are also associated with the drugs/narcotics category. Overall, in our study, we consider 2,122 seed Bitcoin addresses and, in total, 2,079,173 Bitcoin addresses after address clustering and cleansing.

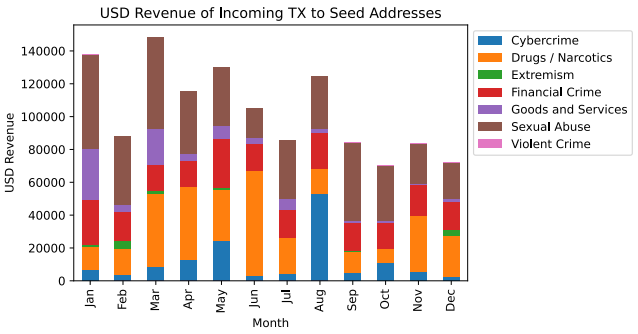

FIGURE 3. USD Revenue of incoming transactions to seed addresses found in the Dark Web Shops in 2021 using our crawler.

For the analysis of transactions to and from seed addresses, we focus on the set of transactions without parentheses in columns four and five. Transactions for sexual abuse and financial crime dominate, with about half of the total incoming and outgoing transactions being attributed to these two types of abuse. We also notice that there is a significant imbalance between the number of incoming (14,119) and outgoing transactions (6,008). This is also the case for incoming/outgoing transactions for each and every individual category. This is to be expected as the payments are at a given price of the product, and the outgoing transactions (laundering) are typically aggregated into bulk transactions.

The last two columns of Table 2 show the revenue per category for the incoming and the outgoing transactions, respectively. Our estimation of the revenue in USD is based on the daily average Bitcoin-USD exchange rate extracted from CoinGecko's API [14]. All USD values are rounded to the closest USD. We focus again on the values in the parentheses that correspond to the revenues of the transactions of Bitcoin addresses after clustering and cleansing (filtered dataset). For a complete reference, we provide in Table 8 (in Appendix I) the results when we consider Bitcoin address clustering without filtering (unfiltered dataset). The total revenue of both the incoming and outgoing transactions exceeds trillions which are totally unrealistic. Even for individual categories, e.g., sexual abuse and financial crime is in the order of hundreds of billions, again not realistic. This further justifies our decision to take a conservative approach and use the filtered data following the cleansing process introduced in V.

B. LONGITUDINAL ANALYSIS OF SEED ADDRESS TRANSACTIONS

We also have examined the longitudinal revenue of the shops in our dataset per individual abuse category. In Figure 3 and 4 we plot the revenue per abuse type per month for all the abuse types provided by the Dark Web Shops in our study. Sexual abuse and financial crime again appear as the most

TABLE 2. Overview of our analysis for Dark Web Shops in 2021. Revenue is in USD, rounded to the nearest whole USD. The initial values correspond to the seed Bitcoin addresses. The values in parentheses correspond to the values after Bitcoin address clustering and data cleansing ("Filtered Dataset").

	Dataset Statistics		Filtered Dataset - Transactions and Revenue			
			Transactions		USD Revenue	
Category	# Tor domains (# pages)	# BTC addresses seed unfiltered / filtered (co-spent unfiltered / filtered)	incoming 2021 seed (co-spent)	outgoing 2021 seed (co-spent)	USD received 2021 seed (co-spent)	USD sent 2021 seed (co-spent)
Cybercrime	46 (1,287)	236 / 110 (1,186,952 / 132,092)	577 (6,338)	264 (621)	$141,329 ($1,389,204)	$141,177 ($1,390,486)
Drugs / Narcotics	17 (392)	104 / 45 (493,877 / 271,362)	1,161 (4,553)	290 (597)	$330,983 ($1,594,520)	$328,764 ($1,414,285)
Extremism	12 (7,683)	19 / 17 (5880 / 231)	150 (4,001)	33 (246)	$13,053 ($577,574)	$8,461 ($509,745)
Financial Crime	227 (31,785)	3968 / 397 (35,272,512 / 548,051)	1,948 (45,768)	3,305 (4,337)	$231,308 ($10,164,827)	$213,702 ($8,062,361)
Goods and Services	41 (2,107)	112 / 67 (41,632 / 18,645)	331 (11,172)	231 (1,072)	$86,766 ($1,082,019)	$85,970 ($1,028,882)
No Abuse	15 (44)	160 / 79 (501 / 418)	3,303 (152,958)	319 (896)	$1,795,163 ($3,845,552)	$1,791,164 ($3,845,551)
Sexual Abuse	836 (29,870)	1945 / 1403 (5,532,538 / 1,108,367)	6,636 (61,400)	1,563 (5,151)	$441,363 ($94,257,825)	$390,682 ($94,241,807)
Violent Crime	3 (41)	29 / 4 (1124 / 7)	13 (15)	3 (4)	$1,401 ($10,601)	$1,109 ($4,532)
Total	**1,197 (73,209)**	**6537 / 2,122 (42,535,016 / 2,079,173)**	**14,119 (286,205)**	**6,008 (12,924)**	**$3,041,376 ($112,922,122)**	**$2,961,029 ($110,497,649)**

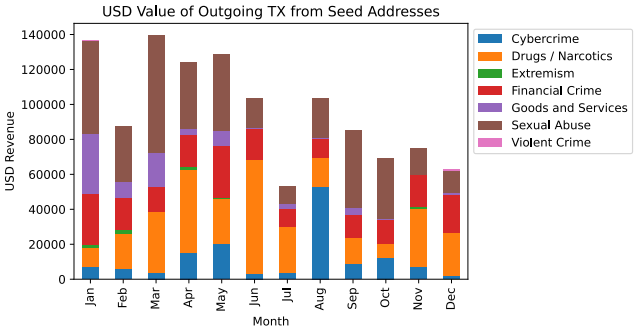

FIGURE 4. USD Value of outgoing transactions from seed addresses found in the Dark Web Shops in 2021 using our crawler.

high-ranking categories over the entire year, but without significant variation. The contribution of the other categories is relatively stable over time. Regarding overall revenue, although there is more activity during the first part of the year, an evident seasonal trend is absent. We note that some of the fluctuations may be related to the take-down of shops or the launch of new in some categories that are beyond the scope of this study. One example of such fluctuation is the outlier for Cybercrime in August, which is related to the purchasing of a stolen Bitcoin wallet. We analyzed this and left it in because, based on blockchain transaction data, it seemed authentic.

C. REVENUE PER ABUSE TYPE AFTER ADDRESS CLUSTERING

The last two columns of Table 2 also provide the revenue per abuse type after retrieving additional addresses based on our clustering algorithm, as discussed in Section IV. The aggregate estimated incoming revenue is around 113 million USD. The estimated total outgoing revenue is around 110.5 million USD. This shows that although there is an asymmetry in the number of transactions, the incoming/outgoing revenue is rather balanced. Thus, the outgoing transactions are made in bulk, but almost the total incoming revenue is laundered within a year. Notice that some incoming or outgoing transactions may occur in the previous or following year, respectively. Then we focus on the individual categories. Sexual abuse contributes by far the most to the incoming illicit activity revenue of Dark Web Shops. Around 94.2 of 112.9 million incoming revenue is associated with sexual abuse, i.e., more than 83% of the illicit revenue of Dark Web Shops. The second contributor is financial crime, with 10.1 million USD, i.e., around 9% of the illicit revenue. The rest of the contributors in the top 5 list are drugs/narcotics, cybercrime, and goods and services, with approximately 1.6, 1.4, and 1.1 million USD in revenue, respectively.

In Table 3, we show the distribution of payments (incoming transactions) to the Bitcoin seed addresses in 2021 per abuse type. We observe that there is a significant difference between

the minimum and maximum transaction values. Indeed, the minimum value is typically cents, while the maximum value is multiple thousands of USD. The median values, however, are more representative of the type of business for Dark Web Shops, in the orders of tens of USD. The 75-percentile values are similar to the median values, which is another indicator that the product's price is in the range of 50 to 500 USD. Our observations concur with independent studies for the individual use of drug unit prices and unit prices for other illicit activities [58].

VII. QUANTIFYING SHOP VS MARKETPLACE REVENUE

In this section, we compare the revenue characteristics, operation, and laundering practices of Dark Web Shops with those observed for Dark Web Marketplaces. Recall that Dark Web Shops are run by individual actors and small groups, selling illicit merchandise to customers directly. On the contrary, Dark Web Marketplaces are run by criminal conglomerates, offering themselves and, against a commission, other criminal actors a marketplace to sell, typically, illicit goods.

A. THE HYDRA MARKETPLACE AND ITS TAKE-DOWN

Hydra was launched in 2015 and has been recognized as one of the largest Dark Web Marketplaces primarily selling drugs in former Soviet bloc countries such as Russia, Ukraine, Belarus, and Kazakhstan. According to an industry report by Chainanalysis [12] Hydra was the dominant Dark Web Marketplaces in 2021. This report estimated that the total revenue of Dark Web Marketplace was around 2.1 billion USD, and Hydra's market share was around 80%.

After being the target of law enforcement scrutiny for many years, at least a large part of Hydra infrastructure was taken down in April 2022 by German authorities [9]. The seized server infrastructure reportedly contained more than 17 million user accounts and 19 thousand seller accounts [8]. While many accounts might be superfluous, as Dark Web marketplaces usually do not provide account password reset functionality, these numbers provide an idea of the scale of its customer base. The US Treasury Department publicly released 117 associated Bitcoin addresses associated with Hydra after its take-down by German authorities [61], [62]. The press release by German authorities also claimed Hydra's role as the biggest marketplace [9]. According to data from our crawler, Hydra still partially remains online.

The release of Bitcoin addresses seized by law enforcement allowed us to extract Hydra's transactions in 2021 and use these as input to our clustering algorithm. Based on that, we are thus able to establish a reliable sample of Hydra's revenue in 2021. In filtering, we excluded transactions to the address of Garantex Exchange, also included in the press release [62]. Garantex was an affiliated money laundering service seized simultaneously with Hydra. Inclusion of its Bitcoin address would wrongly multiply reported revenue.

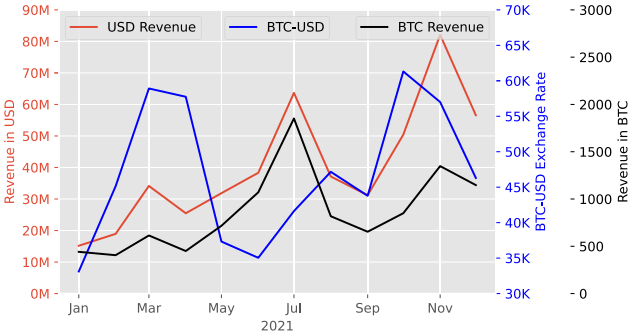

FIGURE 5. Incoming Transaction Revenue to Hydra Address Clusters in 2021.

The revenue in USD of incoming transactions to the seed addresses reported by the US Office of Foreign Assets Control (OFAC) was 792.6 million USD, with the earliest incoming transaction on April 25, 2015. Based on the set of seed addresses, 64 Bitcoin address clusters were discovered, of which the largest had over 6,028,684 Bitcoin addresses. 40 Bitcoin addresses reported by OFAC did not belong to a cluster, which means co-spending did not take place.

In Table 7 (see Appendix I), we provide the revenue with Bitcoin address clustering without filtering. Again, the number is in the order of multiple billions, and although this is mentioned in some reports [23] as correct, we deem this is caused by address clusters of the Garantex Exchange previously mentioned. The revenue flowing into Garantex can not be fully attributed to Hydra. Without the removal of this cluster, the total incoming payments would have been around 7.6 billion USD.

Some industry reports claim that Hydra was involved in ransomware operations [23]. However, when we compared the Hydra-associated addresses with the publicly available Bitcoin addresses used in ransomware campaigns [42], we did not find any match.

B. DARK WEB SHOPS VS. HYDRA TRANSACTIONS REVENUE

In Table 4, we report Hydra's revenue (in USD) of incoming and outgoing transactions. A first observation is that the median transaction value for Hydra is in the orders of thousands of USD compared to the tens of USD in the Dark Web Shops. The maximum value of Hydra transactions is also multiple orders higher than these of Dark Web Shops, reaching 6 million USD. From these values, we can be confident that the structure and customers of the two markets, namely, the Dark Web Shops and the Dark Web Marketplaces, are quite different. The overall incoming revenue for Hydra during 2021 is around 485 million USD, much higher than our lower-bound revenue estimate of Dark Web Shops of 113 million USD. However the reported Hydra revenue is probably partial, as this is the part of the revenue affected by the take-down. We also notice that there is a

TABLE 3. Dark Web Shop payment statistics in 2021 based on seed addresses from our crawler output.

						Percentile		
Abuse Type	Payments in 2021	Min	Max	Median	Std Dev	75%	90%	99%
Cybercrime	552	$0.18	$50,013.85	$55.90	2,175.98	$138.14	$299.69	$1,583.66
Drugs / Narcotics	1141	$0.19	$6,817.84	$121.94	534.32	$285.13	$617.34	$2,780.16
Extremism	146	$0.31	$2,554.18	$17.56	260.00	$62.38	$105.35	$1,022.45
Financial Crime	1744	$0.18	$9,480.63	$58.64	418.71	$102.47	$178.81	$1,288.35
Goods and Services	323	$0.18	$8,215.41	$51.28	696.16	$275.26	$699.69	$2,535.59
No Abuse	3303	$0.16	$157,043.02	$3.03	1,637.95	$4.28	$11.66	$640.63
Sexual Abuse	3,946	$0.17	$17,957.43	$31.67	327.51	$49.23	$85.56	$599.64
Violent Crime	13	$4.92	$413.59	$66.67	121.23	$149.03	$247.12	$395.32

TABLE 4. Hydra's Revenue in 2021, based on 2021 transactions to co-spent clusters of Hydra addresses.

	Lower Bound, Filtered Clustering	
	USD Received	USD Sent
Min	$1	$1
Max	$6,095,231	$6,095,013
Median	$9,890	$3,514
Std Dev	$164,802	$79,969
Total	**$485,005,353**	**$380,433,794**

TABLE 5. Outgoing Transactions and USD Value to WalletExplorer Entities (Dark Web Stores).

Entity Label	USD Received	#Transactions
binance.com	15,694,336	29,724
huobi.com	7,324,820	13,548
coincheck.com	966,937	2,664
bitzlato.com	349,483	554

TABLE 6. Outgoing Transactions and USD Value to WalletExplorer Entities (Hydra).

Entity Label	USD Received	#Transactions
huobi.com	4,248,876	677
bittrex.com	286,292	22
poloniex.com	84,382	8
btc-e.com	57,222	12
localbitcoins.com	15,690	15
bitzlato.com	5,662	7
cryptonator.com	2,897	3
matbea.com	2,724	1

substantial imbalance between the incoming and outgoing transaction revenue, most likely due to commissions and other complex transactions that occur in large Dark Web Marketplaces.

In Figure 3, we plot the revenue of Hydra per month. Although there is no clear trend, the revenue of Hydra has been increasing over time. This was not the case with the monthly revenue evolution for the Dark Web Shops, see Figures 3 and 4. The Bitcoin-USD rate seems to have some influence on Hydra's revenue, but there is not always a strong correlation between revenue and the Bitcoin-USD rate. Recall that the crawler did not scrape Hydra as it was protected by CAPTCHA [68]. Thus, we can not analyze the revenue per type of abuse.

C. DARK WEB SHOPS VS. HYDRA BITCOIN ADDRESS AND LAUNDERING OVERLAP

We also investigate if there is any overlap between the Bitcoin addresses associated with Dark Web Shops that we identified after cluster and cleansing with these identified with the same technique for Hydra. Our analysis shows that there is no overlap, which is another indication that Dark Web Shops and Marketplaces are parallel underground markets. We acknowledge that for our comparison, we take a very conservative approach.

However, when we turn our attention to laundering by Dark Web Shops and Hydra, we notice that they both utilize exchange points. Previous works also confirm that Dark Web Shops utilize sophisticated techniques to laundry money using exchanges and wallets [26]. In Table 5, we present the total revenue and number of transactions for one-hop outgoing transactions (laundering) of Dark Web Shops per exchange point in our study. For the analysis of transactions, we used GraphSense [31]. In Table 6, we repeat the same for Hydra. We notice that Dark Web Shops and Marketplaces not only utilize exchanges but also share two common ones, namely Huobi and Bitzlato. The two common exchanges have repeatedly reported that they participate in the laundering of illicit activity [10]. We recognize potentially more transactions with exchanges can be uncovered with commercial tools. Our labels were sourced from open sources with outdated, limited datasets.

D. DISCUSSION

Our analysis shows the necessity of continuously monitoring payments to Dark Web Shops. Our results indicate that based on such monitoring, potentially at least 113 million USD worth of illicit activity, primarily in sexual abuse and financial crime, can be tackled, which is a significant fraction of the overall estimated Dark Web market, by some measures, 5% to 10% [11] in 2021. Our analysis also shows that Dark Web Shops utilize cryptocurrency exchanges to launder money.

Most likely, more advanced laundering mechanisms not (yet) recognized in open-source address labels, such as Bitcoin tumblers, are employed. Our methodology offers a scalable way for cryptocurrency services to monitor illicit activity and exclude them from their operation. It also provides insights about the evolving Dark Web Shops ecosystem to authorities towards evidence-based policymaking.

Identifying legal entities behind a Bitcoin address makes it possible to attribute transactions to human beings. This is accelerated by address clustering technology as well as existing and forthcoming European Union KYC legislation [47]. Based on co-spending, joint ownership of addresses can be established [31], [37], [39]. If the individual or legal entity behind at least one of the addresses in a cluster is known, the ownership of the whole cluster is known. As exchange platforms are bound to the legislation of their particular jurisdiction, most of them nowadays adhere to KYC legislation. Based on this, they require customers signing up for an account to present proof of identity and, in some cases, even share their home addresses. Through this legislation, law enforcement investigators can now request the personal details of someone behind a deposit or withdrawal from an exchange account.

Our ongoing research shows that while the coverage of public labels attributing Bitcoin addresses to their controlling entity is scarce, some coverage in publicly accessible sources does exist. We confirm that we are able to run a similar analysis for some of the large Dark Web Marketplaces. This capability is important for several reasons. These labels not only reveal the exchange platforms that were potentially involved in leading law enforcement to take down Hydra's infrastructure [62] but also show that it is possible to bootstrap our Dark Web crawler to crawl different parts of the Dark Web.

VIII. CONCLUSION

Difficulties with scraping and indexing onions complicate Tor's analysis of illicit offerings. One way to sidestep such challenges in research efforts is to focus on a single onion representing big clusters of illicit activity, namely Dark Web Marketplaces. Many researchers have focused on such marketplaces in the past. Much still needs to be discovered regarding the expanding ecosystem of Dark Web Shops, i.e., single-vendor shops operated by individuals or small groups. For the analysis of this, the difficulties above need to be tackled.

In this paper, we develop and apply a methodology to collect and analyze the content and involved Bitcoin addresses in Dark Web Shop websites. In the process, we rely on experts to annotate the illicit activity associated with each Dark Web Shop page. Part of our methodology is a detailed data cleansing process to reliably estimate a lower bound of the revenue of Dark Web Shops by analyzing their incoming transactions. Our analysis shows that the Dark Web Shop revenue was at least 113 million USD in 2021. The top illicit category facilitated by Dark Web Shops is sexual abuse (with revenue close to 94 million USD, or 83% of the total revenue) and financial crime (with around 9% of the total revenue). Furthermore, our analysis does not show an overlap between Bitcoin addresses associated with Dark Web Shops and those large ones exposed in the (partial) takedown of one of the largest Dark Web Marketplaces, namely, Hydra. This indicates that Shops and Marketplaces are parallel Dark Web economies. However, when we examine the laundering (outgoing) transactions, our analysis shows that both Dark Web Shops and Marketplaces utilize exchanges, in some cases, the same ones (Huobi, Bitzlato). The insights, tools, and analysis we develop in our work will seed future work in the area and will help computer scientists, economists, and policymakers alike to understand the evolving Dark Web ecosystem.

TABLE 7. Hydra revenue based on 117 seed addresses from Office of Foreign Assets Control [62].

	Seed Addresses		Upper Bound (all clusters)		Lower Bound (excl. exchanges)	
	USD Received	*USD Sent*	*USD Received*	*USD Sent*	*USD Received*	*USD Sent*
Min	$29	$0	1	1	1	1
Max	$153,547,344	$154,057,456	$4,671,650,304	$4,722,737,152	$155,586,032	$155,635,424
Median	$2,573,356	$2,495,489	$3,215,183	$3,301,529	$3,197,913	$2,834,768
Std Dev	$192,324,151	$19,381,411	$484,036,683	$488,907,653	$22,859,744	$22,879,493
Total	**$792,325,710**	**$792,563,212**	**$7,690,539,907**	**$7,752,677,732**	**$930,687,683**	**$930,844,516**

APPENDIX I. SUPPLEMENTARY TABLES

For a complete reference, Table 7 provides Hydra's entire 2016-2022 transaction revenue to Bitcoin addresses shared by OFAC [62]. This information is presented to compare against Hydra's Revenue in 2021 which is available in Table 4 (Section VII of the paper).

TABLE 8. Overview of the dataset statistics with filtered and unfiltered address clustering.

			Unfiltered Dataset (excl. explorers)				Filtered Dataset			
			Transactions		USD Revenue		Transactions		USD Revenue	
Category	# Tor domains (# pages)	# BTC addresses seed unfiltered / filtered (co-spent unfiltered / filtered)	incoming 2021 seed (co-spent)	outgoing 2021 seed (co-spent)	USD received 2021 seed (co-spent)	USD sent 2021 seed (co-spent)	incoming 2021 seed (co-spent)	outgoing 2021 seed (co-spent)	USD received 2021 seed (co-spent)	USD sent 2021 seed (co-spent)
Cybercrime	46 (1,287)	236 / 110 (1,186,952 / 132,092)	5,721 (1,956,150)	4,271 (26,631)	$26,670,066 ($78,176,474)	$25,854,339 ($72,241,561)	577 (6,338)	264 (621)	$141,329 ($1,389,204)	$141,177 ($1,390,486)
Drugs / Narcotics	17 (392)	104 / 45 (493,877 / 271,362)	4,055 (800,979)	1,232 (81,204)	$470,937 ($679,114,794)	$348,272 ($678,662,966)	1,161 (4,553)	290 (597)	$330,983 ($1,594,520)	$328,764 ($1,414,285)
Extremism	12 (7,683)	19 / 17 (5880 / 231)	197 (364,433)	153 (5,348)	$21,129 ($2,570,640,029)	$14,152 ($2,168,525,109)	150 (4,001)	33 (246)	$13,053 ($577,574)	$8,461 ($509,745)
Financial Crime	227 (31,785)	3968 / 397 (35,272,512 / 548,051)	115,315 (86,834,788)	3,391 (21,916,201)	$30,106,149,207 ($446,239,322,493)	$29,269,595,384 ($445,208,550,108)	1,948 (45,768)	3,305 (4,337)	$231,308 ($10,164,827)	$213,702 ($8,062,361)
Goods and Services	41 (2,107)	112 / 67 (41,632 / 18,645)	3,320 (95,602)	2,547 (26,396)	$421,932 ($628,863,605)	$127,401 ($520,894,173)	331 (11,172)	231 (1,072)	$86,766 ($1,082,019)	$85,970 ($1,028,882)
No Abuse	15 (44)	160 / 79 (501 / 418)	74,807 (158,413)	926 (1,917)	$912,433,683 ($15,290,447,315)	$832,527,513 ($14,164,154,157)	3,303 (152,958)	319 (896)	$1,795,163 ($3,845,552)	$1,791,164 ($3,845,551)
Sexual Abuse	836 (29,870)	1945 / 1403 (5,532,538 / 1,108,367)	88,635 (96,645,013)	51,382 (14,647,957)	$3,001,567,051 ($599,737,527,770)	$2,907,675,309 ($598,208,550,108)	6,636 (61,400)	1,563 (5,151)	$441,363 ($94,257,825)	$390,682 ($94,241,807)
Violent Crime	3 (41)	29 / 4 (1124 / 7)	103 (3,990)	43 (1687)	$2,501 ($9,532)	$1,109 ($4,607)	13 (15)	3 (4)	$1,401 ($10,601)	$1,109 ($4,532)
Total	**1,197** **(73,209)**	**6537 / 2122** **(42,535,016 / 2,079,173)**	**292,153** **(186,859,368)**	**63,945** **(36,707,341)**	**$34,047,745,506** **($1.065,224,102x10^12)**	**$33,036,143,479** **($1.061,021,583x10^12)**	**14,119** **(286,205)**	**6008** **(12,924)**	**$3,041,376** **($112,922,122)**	**$2,961,029** **($110,497,649)**

We include Table 8 with raw results for a complete reference. The table complements Table 2, which appears in paper Section VI and is also included here. This table includes seed and cluster revenues before cleansing in the inner segment. Even though problematic domains such as Bitcoin multiplier scams showing unaffiliated Bitcoin addresses are already filtered out, the reported revenues are still heavily influenced by unclean data. With this, we show the importance of thorough cleansing to arrive at a reliable estimation of illicit revenue due to filtering a lower bound.

REFERENCES

[1] J. Aas, R. Barnes, B. Case, Z. Durumeric, P. Eckersley, A. Flores-López, J. A. Halderman, J. Hoffman-Andrews, J. Kasten, E. Rescorla, S. Schoen, and B. Warren, "Let's encrypt: An automated certificate authority to encrypt the entire web," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2019, pp. 2473–2487.

[2] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in Bitcoin," in *Financial Cryptography and Data Security* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2013.

[3] B0bbyB0livia. (2018). *TradeMed*. [Online]. Available: https://github.com/B0bbyB0livia/trademed

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

[5] Bitnnodes, (2022). *Reachable Bitcoin Nodes*. [Online]. Available: https://bitnodes.io/

[6] BitRank Verified. (2022). *BitRank: Crypto Tracking to Meet Cryptocurrency Regulations*. [Online]. Available: http://www.bitrankverified.com/

[7] A. Bracci, M. Nadini, M. Aliapoulios, I. Gray, D. McCoy, A. Teytelboym, A. Gallo, and A. Baronchelli, "Dark web marketplaces and COVID-19: After the vaccines," *EPJ Data Sci.*, vol. 10, no. 1, pp. 1–27, 2021.

[8] British Broadcasting Corporation (BBC). (2022). *Hydra: How German Police Dismantled Russian Darknet Site*. [Online]. Available: https://www.bbc.com/news/technology-61002904

[9] Bundeskriminalamt. *Ilegaler Darknet-Marktplatz 'Hydra Market' Abgeschaltet*. Accessed: Apr. 10, 2022. [Online]. Available: https://www.bka.de/DE/Presse/Listenseite_Pressemitteilungen/2022/Presse2022/220405_PM_IllegalerDarknetMarktplatz.html

[10] Chainalysis. *Crypto Money Laundering: How Criminals Cash Out Billions in Bitcoin and Other Cryptocurrencies*. Accessed: Jul. 11, 2022. [Online]. Available: https://blog.chainalysis.com/reports/crypto-laundering/

[11] Chainalysis. (2021). *The 2021 Crypto Crime Report*. [Online]. Available: https://go.chainalysis.com/2021-Crypto-Crime-Report.html

[12] Chainalysis. (2022). *The 2022 Crypto Crime Report*. [Online]. Available: https://go.chainalysis.com/2022-Crypto-Crime-Report.html

[13] N. Christin, "Traveling the silk road: A measurement analysis of a large anonymous online marketplace," in *Proc. Web Conf. (WWW)*, 2013, pp. 213–224.

[14] CoinGecko. *The Most Comprehensive Cryptocurrency API*. Accessed: Apr. 2, 2022. [Online]. Available: https://www.coingecko.com/en/api

[15] B. Covrig, E. B. Mikelarena, C. Rosca, C. Goanta, G. Spanakis, and A. Zarras, "Upside down: Exploring the ecosystem of dark web data markets," in *Proc. IFIP SEC*, 2022, pp. 489–506.

[16] A. Cuevas, F. Miedema, K. Soska, N. Christin, and R. V. Wegberg, "Measurement by proxy: On the accuracy of online marketplace measurements," in *Proc. USENIX Secur.*, 2022, pp. 2153–2170.

[17] Naval Research Lab Washington DC. *TOR: The Second-Generation Onion Router*. Accessed: Jul. 25, 2022. [Online]. Available: https://apps.dtic.mil/sti/citations/ADA465464

[18] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 644–654, Nov. 1976.

[19] Eckmar Community. (2022). *Eckmar (Eckmar's Marketplace Script)*. [Online]. Available: https://github.com/eckmarcommunity/eckmar

[20] A. ElBahrawy, L. Alessandretti, L. Rusnac, D. Goldsmith, A. Teytelboym, and A. Baronchelli, "Collective dynamics of dark web marketplaces," *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, Nov. 2020.

[21] K. Eldefrawy, A. Gehani, and A. Matton, "Longitudinal analysis of misuse of Bitcoin," in *Proc. Int. Conf. Appl. Cryptogr. Netw. Secur.*, 2019, pp. 259–278.

[22] Elliptic. *Preventing Financial Crime in Cryptoassets: Typologies Report 2022*. Accessed: Jul. 25, 2022. [Online]. Available: https://www.elliptic.co/resources/typologies-report-2022

[23] Elliptic. *US Sanctions Garantex Exchange and Hydra Dark Web Marketplace Following Seizure of Hydra by German Authorities*. Accessed: Jul. 11, 2022. [Online]. Available: https://www.elliptic.co/blog/5-billion-darknet-market-hydra-seized-by-german-authorities

[24] Europol. (2022). *DarkMarket: World's Largest Illegal Dark Web Marketplace Taken Down*. [Online]. Available: https://www.europol.europa.eu/media-press/newsroom/news/darkmarket-worlds-largest-illegal-dark-web-marketplace-taken-down

[25] G. Fabre, *Criminal Prosperity: Drug Trafficking, Money Laundering and Financial Crisis After the Cold War*. London, U.K.: Routledge, 2013.

[26] Financial Action Task Force. (2018). *Professional Money Laundering*. [Online]. Available: https://www.fatf-gafi.org/media/fatf/documents/Professional-Money-Laundering.pdf

[27] S. Goldfeder, H. Kalodner, D. Reisman, and A. Narayanan, "When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies," in *Proc. PETS*, 2018, pp. 1–19.

[28] M. Gotze, S. Matic, C. Iordanou, G. Smaragdakis, and N. Laoutaris, "Measuring web cookies in governmental websites," in *Proc. 14th ACM Web Sci. Conf.*, Jun. 2022, pp. 44–54.

[29] GraphSense. (2022). *Miner Tagpack*. [Online]. Available: https://github.com/graphsense/graphsense-tagpacks/blob/master/packs/miners.yaml

[30] M. Harrigan and C. Fretter, "The unreasonable effectiveness of address clustering," in *Proc. IEEE Conf. Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People, Smart World Congr. (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, Jul. 2016, pp. 368–373.

[31] B. Haslhofer, R. Stutz, M. Romiti, and R. King, "GraphSense: A general-purpose cryptoasset analytics platform," 2021, *arXiv:2102.13613*.

[32] N. Hiramoto and Y. Tsuchiya, "Measuring dark web marketplaces via Bitcoin transactions: From birth to independence," *Forensic Sci. Int., Digit. Invest.*, vol. 35, Dec. 2020, Art. no. 301086.

[33] D. Y. Huang, M. M. Aliapoulios, V. G. Li, L. Invernizzi, E. Bursztein, K. McRoberts, J. Levin, K. Levchenko, A. C. Snoeren, and D. McCoy, "Tracking ransomware end-to-end," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2018, pp. 618–631.

[34] Internet Engineering Task Force (IETF). *The Transport Layer Security (TLS) Protocol Version 1.3*. Accessed: Jul. 25, 2022. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc8446

[35] Interpol. (2020). *Combatting Cyber-Enabled Financial Crimes in the Era of Virtual Asset and Darknet Service Providers*. [Online]. Available: https://cflw.com/download/20200701_Assessment_Report_Cyber_Enabled_Financial_Crime.pdf

[36] Interpol. (2020). *INTERPOL Darknet and Cryptocurrencies Working Group—Abuse Taxonomy*. [Online]. Available: https://interpol-innovation-centre.github.io/DW-VA-Taxonomy/taxonomies/abuses

[37] H. Kalodner, M. Moser, K. Lee, S. Goldfeder, M. Plattner, A. Chator, and A. Narayanan, "BlockSci: Design and applications of a blockchain analysis platform," in *Proc. USENIX Secur. Symp.*, 2020, pp. 2721–2738.

[38] S. Lee, C. Yoon, H. Kang, Y. Kim, Y. Kim, D. Han, S. Son, and S. Shin, "Cybercriminal minds: An investigative study of cryptocurrency abuses in the dark web," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2019, pp. 1–15.

[39] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of Bitcoins: Characterizing payments among men with no names," in *Proc. Conf. Internet Meas. Conf.*, Oct. 2013, pp. 127–140.

[40] S. Nakamoto, "Bitcoin whitepaper," Bitcoin.org, Tech. Rep., 2008.

[41] J. D. Nick, "Data-driven de-anonymization in Bitcoin," M.S. thesis, Comput. Eng. Netw. Lab., ETH Zürich, Zürich, Switzerland, 2015.

[42] K. Oosthoek, J. Cable, and G. Smaragdakis, "A tale of two markets: Investigating the ransomware payments economy," 2022, *arXiv:2205.05028*.

[43] K. Oosthoek and C. Doerr, "Cyber security threats to Bitcoin exchanges: Adversary exploitation and laundering techniques," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 2, pp. 1616–1628, Jun. 2021.

[44] G. P. Paoli, J. Aldridge, R. Nathan, and R. Warnes. (2017). *Behind the Curtain: The Illicit Trade of Firearms, Explosives and Ammunition on the Dark Web*. [Online]. Available: https://www.rand.org/pubs/research_reports/RR2091.html

[45] M. Paquet-Clouston, B. Haslhofer, and B. Dupont, "Ransomware payments in the Bitcoin ecosystem," *J. Cybersecurity*, vol. 5, no. 1, Jan. 2019, Art. no. tyz003.

[46] F. Poskriakov, M. Chiriaeva, and C. Cavin, "Cryptocurrency compliance and risks: A European KYC/AML perspective," *Blockchain Cryptocurrency Regulation*, 2nd ed. Oct. 2020, pp. 1–16.

[47] European Parliament (Press Releases). (2022). *Crypto Assets: Deal on New Rules to Stop Illicit Flows in the EU*. [Online]. Available: https://www.europarl.europa.eu/news/en/press-room/20220627IPR33919/crypto-assets-deal-on-new-rules-to-stop-illicit-flows-in-the-eu

[48] The Tor Project. *Tor Project | Anonymity Online*. Accessed: Jul. 25, 2022. [Online]. Available: https://www.torproject.org/

[49] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digitalsignatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[50] M. Romiti, A. Judmayer, A. Zamyatin, and B. Haslhofer, "A deep dive into Bitcoin mining pools: An empirical analysis of mining shares," in *Proc. WEIS*, 2019, pp. 1–19.

[51] A. Sanatinia, J. Park, E.-O. Blass, A. Mohaisen, and G. Noubir, "A privacy-preserving longevity study of Tor's hidden services," 2019, *arXiv:1909.03576*.

[52] M. Shapiro, "Structure and encapsulation in distributed systems: The proxy principle," in *Proc. IEEE ICDCS*, Jun. 1986, pp. 198–204.

[53] M. Spitters, S. Verbruggen, and M. Van Staalduinen, "Towards a comprehensive insight into the thematic organization of the TOR hidden services," in *Proc. IEEE Joint Intell. Secur. Informat. Conf.*, Sep. 2014, pp. 220–223.

[54] M. S. C. Steel, "Stolen identity valuation and market evolution on the dark web," *Int. J. Cyber Criminology*, vol. 13, no. 1, pp. 70–83, 2019.

[55] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1388–1401.

[56] Tor Project. (2022). *Tor Metrics—Onion Services*. [Online]. Available: https://metrics.torproject.org/hidserv-dir-v3-onions-seen.html

[57] Tor Project. (2022). *Tor Metrics—Servers*. [Online]. Available: https://metrics.torproject.org/networksize.html

[58] United Nations Office on Drugs and Crime (UNODC). (2020). *2020 World Drug Report—In Focus: Trafficking Over the Darknet*. [Online]. Available: https://www.unodc.org/documents/Focus/WDR20_Booklet_4_Darknet_web.pdf

[59] United Nations Office on Drugs and Crime (UNODC). (2022). *2022 World Drug Report—Global Overview Drug Demand Drug Supply*. [Online]. Available: https://www.unodc.org/unodc/en/data-and-analysis/world-drug-report-2022.html

[60] Office of Foreign Assets Control US Department of the Treasury. (2022). *FAQ-1021, Do the prohibitions of Executive Order (E. O.) 14024 and Other Russia-Related Sanctions Extend to Virtual Currency?*. [Online]. Available: https://home.treasury.gov/policy-issues/financial-sanctions/faqs/1021?s=09

[61] US Department of the Treasury Office of Foreign Assets Control. *Russia-Related Designation; Cyber-Related Designation*. Accessed: Jun. 9, 2022. [Online]. Available: https://home.treasury.gov/policy-issues/financial-sanctions/recent-actions/20220405

[62] US Department of the Treasury Office of Foreign Assets Control. *Treasury Sanctions Russia-Based Hydra, World's Largest Darknet Market, and Ransomware-Enabling Virtual Currency Exchange Garantex*. Accessed: Jun. 9, 2022. [Online]. Available: https://home.treasury.gov/news/press-releases/jy0701

[63] J. V. D. Laarschot and R. V. Wegberg, "Risky business? Investigating the security practices of vendors on an online anonymous market using ground-truth data," in *Proc. USENIX Secur. Symp.*, 2021, pp. 4079–4095.

[64] R. V. Wegberg, S. Tajalizadehkhoob, K. Soska, U. Akyazi, C. H. Ganan, B. Klievink, N. Christin, and M. V. Eeten, "Plug and prey? Measuring the commoditization of cybercrime via online anonymous markets," in *Proc. USENIX Secur. Symp.*, 2018, pp. 1009–1026.

[65] WalletExplorer.com. (2022). *WalletExplorer.com: Smart Bitcoin Block Explorer*. [Online]. Available: http://www.walletexplorer.com/

[66] WikiLeaks. *WikiLeaks: Tor*. Accessed: Jul. 25, 2022. [Online]. Available: https://www.wikileaks.org/wiki/WikiLeaks: Tor?

[67] YCharts. *Bitcoin Market Cap*. Accessed: Apr. 2, 2022. [Online]. Available: https://ycharts.com/indicators/bitcoin_market_cap

[68] G. Ye, Z. Tang, D. Fang, Z. Zhu, Y. Feng, P. Xu, X. Chen, and Z. Wang, "Yet another text captcha solver: A generative adversarial network based approach," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2018, pp. 332–348.

KRIS OOSTHOEK (Member, IEEE) received the M.Sc. degree from Erasmus University. He is currently pursuing the part-time Ph.D. degree with the Delft University of Technology. He worked in various technical positions based on USA, U.K., and Afghanistan. He is also an Investigator with the Government Agency, The Netherlands. His research interest includes cybercriminal abuse of bitcoin. He received a several commercial cyber security certifications, such as CISSP, GICSP, GCTI, GXPN, and GRID.

MARK VAN STAALDUINEN received the M.Sc. and Ph.D. degrees in electrical engineering from the Department of Information and Communication Theory, Delft University of Technology (TU Delft). From 2007 to 2019, he worked at The Netherlands Organization for Applied Scientific Research (TNO), as a Consultant and an Innovation Manager focusing on cybercrime and national security. From January 2016 to December 2019, he was posted to Singapore to strengthen international cooperation and partnerships in Southeast Asia, and seconded as a Cybercrime Expert to the INTERPOL Global Complex for Innovation. In April 2022, he was also appointed to the Dutch Blockchain Coalition (DBC) as a Safety Theme Leader, responsible for fostering a safe and secure blockchain ecosystem. Since January 2020, he has been the Managing Director and the Founder of CFLW Cyber Strategies (CFLW).

GEORGIOS SMARAGDAKIS (Senior Member, IEEE) received the Diploma degree in electronic and computer engineering from the Technical University of Crete and the Ph.D. degree in computer science from Boston University, in 2009. In 2008, he was a Research Intern at Telefonica Research. From 2008 to 2014, he acted as a Senior Researcher at the Deutsche Telekom Laboratories. From 2014 to 2017, he was a Marie Curie Fellow at the Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology (MIT), and a Research Affiliate with the MIT Internet Policy Research Initiative (IPRI), from 2015 to 2018. He was a Professor at TU Berlin, from 2017 to 2021, and a Research Collaborator with Akamai Technologies, from 2014 to 2021. He is currently a Full Professor, the Chair, and the Section Head of Cybersecurity with the Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology (TU Delft). He is also a Researcher with the Max Planck Institute for Informatics (MPI-INF) and a Principal Investigator and a fellow of the Berlin Institute for the Foundations of Learning and Data (BIFOLD). Since 2021, he has been a Scientific Coordinator of AI for peace, justice, and security initiative with TU Delft. His research interests include data- and measurement-driven approach to the study of the internet security, resilience, state, and performance, and the enhancement of web privacy and security. He is an ACM Distinguished Member. His research was recognized with a European Research Council (ERC) Starting Grant Award, in 2015, a Marie Curie International Outgoing Fellowship, in 2013, Best Paper Awards at ACM SIGCOMM, in 2021, ACM IMC, in 2018, 2016, and 2011, ACM CoNEXT, in 2019 and 2015, IEEE INFOCOM, in 2017, three IETF/IRTF Applied Networking Research Prizes, in 2022, 2020, and 2019, "Best of ACM SIGCOMM Computer Communication Review," in 2019, and selected for Communications of the ACM (CACM) Research Highlights, in 2021. He served as the Technical Program Chair for TMA, in 2020, ACM CoNEXT, in 2019, and PAM, in 2018. He has also been involved in the Organization and Technical Program Committees of many conferences, including, ACM SIGCOMM, ACM IMC, ACM SIGMETRICS, ACM CoNEXT, ACM Web Conference, ACM Web Science, ACM HotNets, IEEE EuroS&P, IEEE Infocom, IEEE HotWeb, USENIX ATC, PETS, and ESORICS.