Using Venom to Flip the Coin and Peel the Onion: Measurement Tool and Dataset for Studying the Bitcoin - Dark Web Synergy

Data/Toolset Paper

Lukas Ingemarsson
Linköping University
Linköping, Sweden

Karl Duckert Karlsson
Linköping University
Linköping, Sweden

Niklas Carlsson
Linköping University
Linköping, Sweden

Abstract

Bitcoin and the Dark Web present an interesting synergy that enables both legitimate anonymity and illicit activities, making it an important landscape to understand, especially as the Dark Web, with its hidden services, relies heavily on Bitcoin as a pseudonymous currency for transactions. However, a lack of scalable tools and timely datasets has limited systematic analysis of this ecosystem. To address this gap, we introduce Venom, a scalable framework for mapping Bitcoin activity on the Dark Web. Venom integrates multithreaded crawling, data extraction, and dataset generation, resulting in a comprehensive resource that allows us to easily collect snapshots of over 177,000 onion sites in roughly 24 hours. With the paper, we share both the tool and an example snapshot containing both per-site metadata and Bitcoin transaction data. Preliminary analysis reveals concentrated activity among key players and widespread content mirroring, offering new insights into the Dark Web's economic structure. Venom provides a critical resource for advancing research and monitoring in this domain.

CCS Concepts

• **Security and privacy** → *Distributed systems security*; • **Information systems** → **Web applications**;

Keywords

Dark Web, Tor, Bitcoin, Measurements

ACM Reference Format:
Lukas Ingemarsson, Karl Duckert Karlsson, and Niklas Carlsson. 2025. Using Venom to Flip the Coin and Peel the Onion: Measurement Tool and Dataset for Studying the Bitcoin - Dark Web Synergy: Data/Toolset Paper. In *Proceedings of the Fifteenth ACM Conference on Data and Application Security and Privacy (CODASPY '25), June 4–6, 2025, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3714393.3726488

1 Introduction

The rise of Bitcoin and the Dark Web offer both enhanced privacy and increased risks. While these technologies empower users to bypass censorship and surveillance through encrypted transactions and pseudonymous identities, they also provide a safe haven for a wide range of illicit activities.

CODASPY '25, June 4–6, 2025, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-1476-4/2025/06
https://doi.org/10.1145/3714393.3726488

The Dark Web is a concealed corner of the internet, hosting a variety of services ranging from anonymous communication platforms to illicit marketplaces. Bitcoin, on the other hand, is the dominant cryptocurrency in these transactions, drawing scrutiny from researchers and law enforcement. Despite its significance, there are limited free tools and datasets available to help capture the interplay between Dark Web services and Bitcoin usage. This limits researchers' ability to understand, monitor, and analyze this ecosystem effectively.

Existing studies on Bitcoin activity within the Dark Web are often constrained by the lack of systematic and scalable data collection methods and/or do not share the tools and datasets created. To address this gap, we present Venom, a novel data collection framework designed to systematically map Bitcoin activity across the Dark Web. By combining a multithreaded onion crawler with advanced data extraction and processing techniques, Venom enables the creation of a comprehensive dataset that captures detailed information about Bitcoin usage on onion sites. This framework not only addresses technical challenges in Dark Web crawling but also provides a valuable resource for understanding the ecosystem's structure and behavior. Our contributions are threefold:

- **Venom:** We present Venom, a scalable and efficient framework for crawling onion domains, extracting Bitcoin-related data, and consolidating the data into a structured dataset.
- **Dataset:** We provide a dataset comprising over 177,000 onion sites, including metadata, Bitcoin addresses, and associated transactional information, making it one of the most comprehensive open resources of its kind.
- **Insights:** We perform an initial analysis of the dataset, uncovering Bitcoin usage patterns such as the dominance of a few key players and the prevalence of mirrored content, shedding light on the Dark Web's economic dynamics.

Outline: Sec. 2 presents background, and Sec. 3 details Venom, including crawling, data extraction, and consolidation. Sec. 4 summarizes the dataset, while Sec. 5 explores Bitcoin activity and marketplace trends. Sec. 6 concludes and discusses future directions.

2 Background

Bitcoin: Bitcoin (BTC) is a decentralized cryptocurrency enabling pseudonymous transactions without central authorities [16]. Balancing privacy and transparency, it attracts both legitimate users and malicious actors, some using it for illicit activities such as money laundering, ransomware payments, and darknet trading, often alongside tools like the Dark Web [11, 14].

The Dark Web: The Dark Web is a subset of the Deep Web intentionally hidden from standard search engines. While it enables

users to bypass censorship and protect their privacy, it also facilitates criminal activities such as drug sales, malware distribution, and forged document trades [10]. Access to the Dark Web requires special software like The Onion Router (Tor), which anonymizes users by encrypting and routing traffic through multiple servers.

Tor and Onion Addresses: Tor facilitates anonymous online communication by routing encrypted data through relay nodes, decrypting it layer by layer until reaching its destination [8]. Websites hosted via Tor use ".onion" addresses, bypassing traditional DNS. These onion sites, favored for their enhanced privacy, often host illicit Dark Web marketplaces. Tor thus serves as a key gateway for investigating these markets and Bitcoin-related activity.

Dark Web Marketplaces: Dark Web marketplaces like the infamous Silk Road, launched in 2011, revolutionized anonymous trading by allowing the exchange of illicit goods, primarily using Bitcoin [7]. After its 2013 FBI shutdown, successors like Silk Road 2.0, AlphaBay, and Agora emerged but were later dismantled or collapsed in exit scams, highlighting their volatility [9, 19].

3 Venom: Data Collection Framework

This section introduces Venom, our data collection framework named after the dark counterpart of Spider-Man. Venom (i) crawls the Dark Web to collect onion addresses, metadata, and associated Bitcoin addresses, (ii) retrieves detailed information about these Bitcoin addresses via a suitable API, and (iii) processes the data into a consolidated dataset. Fig. 1 provides a workflow overview.

3.1 Multithreaded Onion Crawler

To collect Bitcoin-related data from the Dark Web, we developed a Python-based onion crawler using Tor as a proxy. We started with 178 seed addresses obtained via the Ahmia search engine [1], by appending the Ahmia search URL [2] with 178 unique keywords relevant to Bitcoin (e.g., *market*, *mixer*, and Dark Web marketplace names). To manage the slow connection times typical of onion sites, we set a 120-second timeout threshold per request.

The crawler follows a traversal strategy where, for each onion address, it (1) saves and schedules all observed linked onion addresses for future crawling, and (2) extracts and stores all unique Bitcoin addresses encountered.

We used the Python library *BeautifulSoup4* to parse HTML content and to identify onion domains and Bitcoin addresses using the following basic regular expressions `"https?://\w+.onion"` and `"(bc1|[13])[a-zA-Z0-9]{25,61}"`.

Given the resource-intensive nature of crawling the Dark Web, we optimized the collection process in several ways:

- **Hosting:** The crawler ran on a Google Cloud VM [12] for uninterrupted operation.
- **Multithreading:** Using Python's *concurrent.futures*, we implemented a 60-thread configuration after testing the VM's limits, enabling concurrent crawling of multiple addresses.

With these optimizations, the crawler can quickly complete its traversal of all discovered onion sites (e.g., in approximately 24 hours during our tests). The output is saved as a map/dictionary in a text file, where each onion address served as a key, mapped to its title and associated Bitcoin addresses.

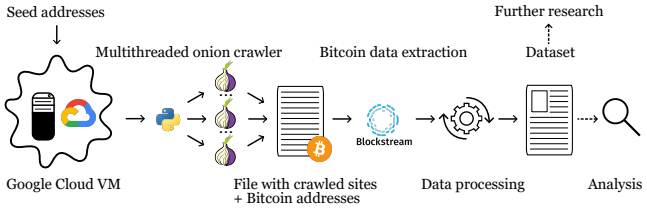


Figure 1: Overview of Venom, our data collection framework.

3.2 Bitcoin Data Extraction

Next, we retrieve data for collected Bitcoin addresses using a Python script that queries the Blockstream Bitcoin API [4, 5]. The API returns JSON responses, from which we extract "chain_stats", including the total Bitcoin sent, received, and confirmed transactions. (There are also "mempool_stats"; however, these pertain to unconfirmed transactions and are not relevant to our analysis.) Invalid addresses are filtered out based on unsuccessful API requests.

3.3 Data Processing and Consolidation

After data collection, the information is processed to create the final dataset. Valid Bitcoin addresses, initially stored as individual JSON objects, are reformatted into a unified text file per address. Problematic characters, such as line breaks (`\n` and `\r`) are safely removed from onion site titles without impacting their semantic meaning. The processed data is then consolidated into a structured dataset using a Python script and the *pandas* library.

Each website entry in the dataset was categorized by analyzing its title for specific keywords. Keywords were chosen based on their relevance to common topics, and those with flexible stems (e.g., "generat" for "generate" and "generator") were used to maximize coverage. For frequently recurring titles that could not be categorized programmatically, manual review was performed, excluding titles indicative of abusive content.

3.4 Dataset Format

After processing, structuring and consolidating the collected data, we form a dataset with parameters as follows:

- *onion_addr*: the onion website's address.
- *title*: the onion website's title.
- *topic*: the topic(s) of the onion website's content.
- *btc_addrs*: a list containing the Bitcoin addresses that appeared on the onion website, sorted in descending order, based on the sum of the Bitcoin amount spent and received.
- *btc_addrs_count*: the number of Bitcoin addresses that appeared on the onion website.
- *total_sent*: the total amount of Bitcoin sent, by all the Bitcoin addresses that appeared on the onion website.
- *total_received*: the total amount of Bitcoin received, by all the Bitcoin addresses that appeared on the onion website.
- *n_tx*: the total number of transactions made by all the Bitcoin addresses that appeared on the onion website.
- *comment*: extra information regarding the entry, e.g., the exception that occurred if the crawler's request to the onion website was unsuccessful.

Finally, the dataset is exported as a CSV file using a semicolon separator (avoiding conflicts with commas in fields or address lists).

3.5 Limitations

Our reliance on Ahmia as the primary source for crawler seeds introduced potential biases in the dataset due to Ahmia's indexing policies. Specifically, Ahmia excludes onion addresses that implement the Robots Exclusion Protocol, which restricts web crawler access. In addition, Ahmia filters out services associated with abusive material. Although such content was not represented in our seeds, it may have been encountered deeper in the crawl as the crawler traversed linked addresses.

To reduce data collection times, the crawler focused on main onion pages, excluding subpages, and multithreading was implemented. Crawling subpages or extending runtime could enhance dataset depth, accuracy, and coverage.

We did not filter Bitcoin addresses or try to assess whether they are more or less likely to be involved in illicit activity. We also note that the received amounts cannot be translated directly to revenues (e.g., sites often use mixers and/or otherwise transfer money across several accounts) and some addresses may be related to other services (e.g., cryptocurrency exchanges). Despite these limitations, the dataset provides valuable insights into patterns of Bitcoin use across onion sites.

4 Dataset Summary

Using Venom, we are able to collect a complete crawl in roughly 24 hours. Here, we present a snapshot collected on Apr. 24, 2024.

Seed Selection: For the dataset presented here, we used 178 seeds (as outlined above). In general, the web crawler's seed selection strategy proved highly effective, allowing us to use Ahima to expand the 178 seeds into 196,591 onion addresses. To identify this set, we first started with fewer than 50 seeds, and then refined the list using trial and error with relevant keywords. Over time, adding new seeds had minimal impact, suggesting an upper limit of retrievable onion domains through Ahmia. While our seeds excluded services outside Ahmia's indexing policies and did not traverse subpages, the inclusion of terms like "dark web directory" and "hidden service link list" helped capture a diverse and comprehensive dataset of onion services.

High-Level Statistics: In total, the crawler attempted 196,591 onion addresses, successfully reaching 177,127 (90.1%). It collected 42,696 Bitcoin address strings, of which 10,210 (23.9%) were unique. The high duplication rate resulted from mirrored content across multiple onion sites, such as "LordPay Market," mirrored on 1,648 domains. This practice likely arises from the ease and zero cost of creating onion domains. Mirroring may improve resilience against takedowns and boost market exposure by increasing visibility in directories of onion services, thus driving traffic and profit.

Among the unique Bitcoin-like strings, 4,382 (42.9%) were valid addresses. Invalid matches occurred primarily when strings resembling Bitcoin addresses were erroneously detected, often due to onion addresses missing the .onion suffix. Table 1 presents summary statistics on the crawler's output.

Of the 19,464 failed requests, the majority (17,561 or 90.2%) resulted from connection errors, likely indicating that onion sites were temporarily or permanently offline. Other failures included connection timeouts (1,438 or 7.4%), HTTP errors (365 or 1.9%), and rare miscellaneous errors (100 or 0.5%). Timeouts, caused by the preset request limit, helped terminate prolonged attempts unlikely to succeed. Although HTTP errors varied, their infrequent occurrence had minimal impact on the results.

Table 1: Statistics for the web crawler's output.

Category	*Frequency*
Extracted onion addresses	196,591
Successfully visited onion addresses	177,127
Extracted Bitcoin addresses	42,696
Unique Bitcoin addresses	10,210
Valid Bitcoin addresses	4,382

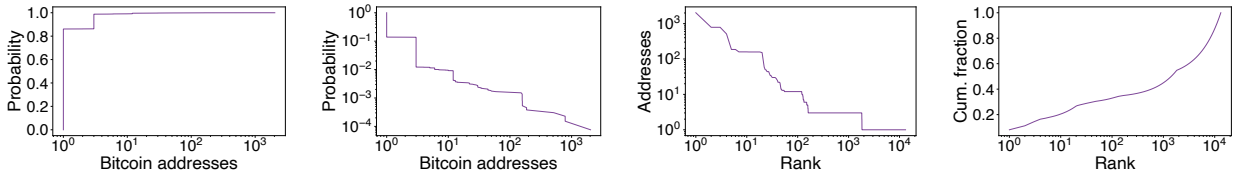
5 Example Results

5.1 Bitcoin Activity Statistics

Onion Site Usage of Bitcoin Addresses: Out of the 177,127 crawled onion sites, most of the sites (163,973 or 92.6%) did not contain any Bitcoin addresses on their landing page. While this can be seen as a lower bound, since the crawler's design limits access to main pages, excluding Bitcoin addresses found on subpages of onion domains, we note that the low fraction of sites with at least on Bitcoin address (13,154 or 7.4%) may in part be due to there simply being diverse content on the Dark Web, and many of them simply not relying on Bitcoins.

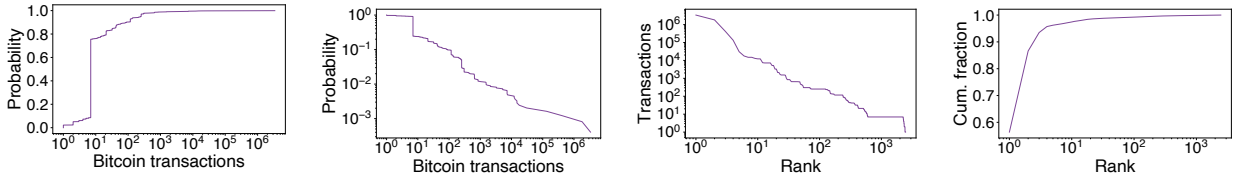
Among the 13,154 (7.4%) onion sites that did contain at least one Bitcoin address (on their landing page), we observed a high skew in the number of Bitcoin addresses per onion site. Using the distribution plots shown in Fig. 2 we make several observations. First, the Cumulative Distribution Function (CDF) of the number of Bitcoin addresses per site with at least one Bitcoin address captures that most of the onion sites with at least one Bitcoin address (11,328 out of 13,154 or 86.1%) only have one single Bitcoin address. Second, the Complementary CDF (CCDF) of the number of Bitcoin addresses per site highlight that there is a heavy tail of sites with many Bitcoin address, that this tail is power-law like (as indicated by the tail being approximate linear on log-log scale) and the onion site with most Bitcoin addresses having 2,028 addresses. Third, and in contrast, the flattening of the tail of the rank plot (rather than straight-line behavior all the way into the tail) highlights the longer tail of sites that only have one or two addresses (also somewhat captured in the CDF) than what would be expected in the case of a Zipf-like tail. As shown later, this in part is due to replicated sites. We further note that this is captured by the concentration plot, which shows the fraction of all observed Bitcoin addresses that are covered by the R top-ranked sites as a function of the rank threshold R, where we observe a convex behavior on log-lin scale.

Number of Transactions and BTCs Received per Onion Site: Of the 13,154 onion sites with at least one observed Bitcoin address, only 2,463 of the sites (18.7%) saw at least one transaction; most recorded zero transactions and received zero BTC. This heavy skew toward inactivity suggests that many onion sites do not generate much Bitcoin activity. However, when looking closer at those with activity, we observe significant variability in the number of Bitcoin transactions and total amount received. To better understand these distributions, we analyzed their respective CDF, CCDF, rank plot, and concentration plot; shown in Figs. 3 and 4.



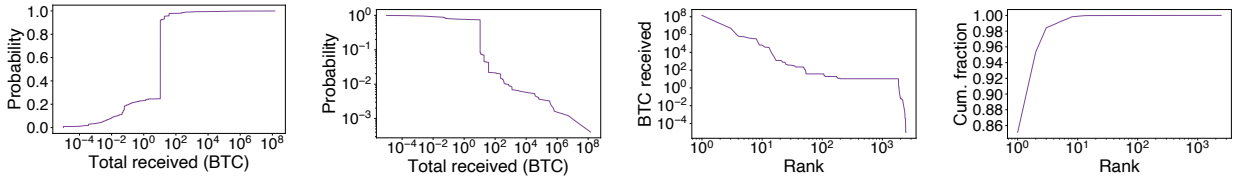
(a) CDF (b) CCDF (c) Rank plot (d) Concentration plot

Figure 2: Distribution statistics of the number of Bitcoin addresses per onion website (with at least one Bitcoin address).



(a) CDF (b) CCDF (c) Rank plot (d) Concentration plot

Figure 3: Distribution statistics of the number of associated Bitcoin transactions per onion website (with transactions).



(a) CDF (b) CCDF (c) Rank plot (d) Concentration plot

Figure 4: Distribution statistics of the total received BTC per onion website (that received BTCs).

First, looking at the CDFs, we observe an interesting anomaly: among the active sites, the most common values were seven transactions (1,652 sites), out of most received 10.77 BTC (1,648 sites). These two peaks (one visible in each CDF) are attributable to content mirroring, where multiple onion domains replicate the same Bitcoin addresses across many mirrors.

Second, the CCDF highlights the heavy-tail behavior of the distribution, with a small number of highly active sites accounting for a disproportionate share of transactions and Bitcoin volume. The rank-plot further illustrates this disparity, showing that the top-ranked site alone handled 3,396,913 transactions and received 141,915,522.844 BTC, dominating the dataset. (We look closer at this site in Sec. 5.3.) The general dominance of a few dominating onion sites is further captured by the concentration plot, which shows that the top-four sites are responsible for more than 95% of the transactions and 99% of the received BTCs.

This example analysis reveals a highly imbalanced distribution of Bitcoin activity on the Dark Web, with most onion sites showing no transactions or Bitcoin volume, while a few dominate the ecosystem. This concentration underscores the significance of a small subset of services in driving Bitcoin-related activity on the Dark Web.

5.2 Topic-Based Website Statistics

To glean some initial insights into what types of onion websites were responsible for most Bitcoin activity, we categorized each website in the dataset into one of nine categories. For this analysis, we first searched for common keywords, and then assigned topic labels based on the combination of keywords that appeared in the title. Then, for websites that could not be programmatically categorized, but whose title reappeared frequently across different sites, we manually reviewed and assigned topics (except for those with titles indicative of abusive content). Finally, the websites that we could not label ourselves were placed in an *Other* category. The frequency of each category is shown in Fig. 5, with an explanation for each category listed next in order of frequency (with the *Other* category placing sixth): (1) *Stolen funds*: Websites offering stolen credit cards, PayPal accounts, and Western Union accounts. (2) *Abusive content*: Websites containing abusive materials, such as child pornography, rape, or other atrocities. (3) *Market*: Websites functioning as marketplaces, selling a variety of products or services. (4) *Pornography*: Websites featuring pornographic material. (5) *Website list*: Websites providing directories of hidden services. (7) *Hacking*: Websites offering hacking services. (8) *Stolen bitcoins*: Websites offering stolen Bitcoin wallets or private keys. (9) *Bitcoin generator*: Websites advertising Bitcoin miners, generators, or multipliers. (10) *Bitcoin mixer*: Websites offering Bitcoin mixing services.

Most Frequent Topics: The largest category was *stolen funds*, accounting for 109,169 (61.6%) of websites. The broad scope of this category, which groups various but fundamentally similar activities, reflect the prevalence of financial crime in illicit Bitcoin transactions [11]. Perpetrators leverage the Dark Web for anonymous transactions, further facilitating these activities.

The next three most common categories all are observed with a similar frequency: *Abusive content* appeared on 30,955 (17.5%), highlighting the grim reality of illicit and harmful activities online. *Market* appeared on 27,921 (15.8%) sites, capturing that there is a broad range of marketplaces, products, and services available via the Dark Web. Finally, the large number of *Pornography* sites (27,254 or 15.4%) captures that the high prevalence of such sites also on the Dark Web (in addition to high usage on the regular web).

Topics Associated with Most Bitcoin Activity: There are significant differences in the total amount of BTCs received per

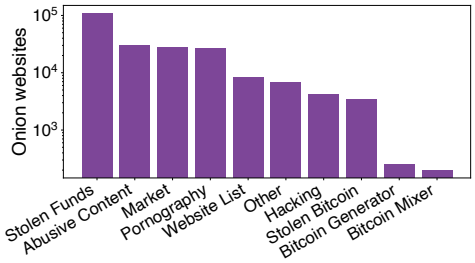


Figure 5: The number of onion websites per topic category.

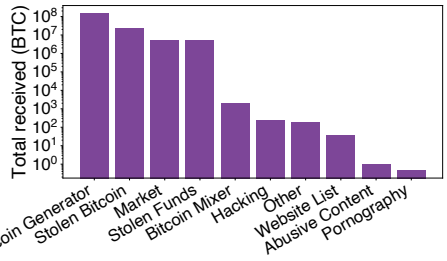


Figure 6: Total BTC received across different website topics.

topic, as visualized in in Fig. 6. The top three topics in this metric were *Bitcoin generator* (148,339,475.1548 BTC), *Stolen bitcoin* (23,989,323.4991 BTC), and *Market* (5,386,913.2111 BTC). While we see significant amounts received across all categories, it is interesting that the top two categories are two of the three least frequent topics (Fig. 5). However, these large funds can be explained by these services directly relating to Bitcoins. In some cases, these sites also contain lists including some "active" Bitcoin addresses not in their own possession, in an attempt to instill trust in their service.

At the other end of the spectrum are the topics *Website list*, *Abusive content*, and *Pornography*, which received the lowest totals of Bitcoins: 37.46 BTC, 0.96 BTC, and 0.47 BTC, respectively. Despite *Abusive content* and *Pornography* being among the most frequent topics, their minimal Bitcoin presence likely reflects limitations in our methodology. Since the crawler only accessed main pages, it may have missed payment-related content on subpages, especially for services hidden behind paywalls. The actual profitability of these services could be significantly higher than our data suggests.

5.3 Example Marketplaces of Interest

Here, we use some example marketplaces to highlight some interesting observations from our preliminary analysis of the dataset. For this analysis, we compiled profiles for selected onion addresses by manually accessing and reviewing websites from the dataset. Given the uncertainty surrounding the legitimacy of these services and the true ownership of the Bitcoin addresses, we censored all addresses appearing in the overview images.

Top Bitcoin Related Sites: We highlight two notable marketplaces with significant Bitcoin activity. "Bitcoin Doubler 2020" (Fig.7a), which ranked highest in Bitcoin volume, claimed to offer a multiplier service. However, its static "proof" transactions with fixed dates, despite some matching Blockstream API data, suggest it is a scam due to inconsistencies. "BitSale" (Fig.7b) listed 787 Bitcoin wallets for sale. While some wallet balances initially aligned with API data, discrepancies in transfers point to abandonment or fraudulent intent.

Mirrored Sites: Several marketplaces operated across multiple mirrored onion sites with identical content. "Bitcoin BTC Mixer" (Fig.8a) appeared on over 50 mirrors, using a single donation address that received small, irregular Bitcoin transactions, suggesting potential legitimacy. "Bitcoin Private Key Shop" (Fig.8b) was mirrored on 18 sites, listing private keys for sale with static updates marking some as sold. Unlike "Bitcoin Doubler 2020", this static information may reflect a limited inventory rather than fraud. However, without access to paywalled content, the legitimacy of these services remains uncertain.

Same Bitcoin Address for Different Services: Some marketplaces share the same Bitcoin address while offering different services, suggesting a single operator behind them. For example, "Choose Better" (Fig.9a) and "Hack Twitter and Instagram Accounts" (Fig.9b) both used the same address across nine onion sites. "Choose Better" claims to detect scams and sells access to "legitimate" services, while the hacking site offers Instagram and Twitter credentials for payment. This shared address and cross-promotion likely funnel users toward other services controlled by the same entity, leveraging a coordinated business model to attract customers.

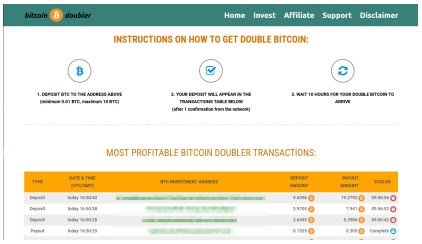
6 Related Work

The Dark Web and Bitcoin usage associated with illicit activities have been studied both individually, and in combination.

The Tor Ecosystem: Biryukov et al. [3] analyzed hidden services, finding a roughly equal distribution between those with and without illegal content, though the most popular sites tended to involve criminal activity. Chertoff [6] examined government regulation of the Dark Web, emphasizing the need to preserve user privacy while targeting illegal sites rather than individual users.

Bitcoin Abuse: The misuse of Bitcoin for criminal purposes is well-documented. Möser et al. [15] conducted the first systematic study on Bitcoin mixers, showing that they effectively hinder identification. Huang et al. [13] traced Bitcoin payment flows in ransomware attacks, from victim acquisition to operator collection. Paquet-Clouston et al. [17] analyzed sextortion campaigns, using Bitcoin addresses from spam to map monetary flows and uncover the structure of these schemes. Rosenquist et al. [18] examined Bitcoin flows linked to abusive activity, identifying patterns and a significant rise in transfers to reported addresses over time.

Bitcoin Usage on the Dark Web: Research into cryptocurrency abuse on the Dark Web is relatively more limited [11, 14]. Lee et al. [14] pioneered this area by developing MFScope, a framework that collected 27 million onion webpages and 10 million Bitcoin addresses, revealing that 80% of Bitcoin addresses on the Dark Web were linked to malicious activity. Compared to their crawler, our multithreaded web crawler, which only visits landing pages (not subpages), allows us to gather data for more onion domains (196,591 vs. 36,864) in much shorter time (24 hours vs. 15 month) and is made open source. Foley et al. [11] estimated that 25% of all Bitcoin users engage in illegal activity, highlighting cryptocurrency's transformative impact as an alternative payment method on Dark Web marketplaces. While these studies have advanced the field, our work introduces a highly replicable and efficient methodology, leveraging multithreading to expedite the typically slow process of crawling the Dark Web; a novel contribution in this domain.



(a) *Bitcoin Doubler 2020.* 141,915,440.2895 sent, 141,915,522.8440 received.

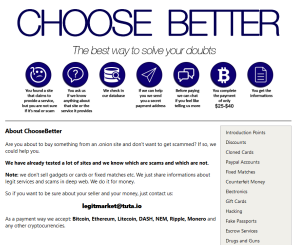
(b) *BitSale.* 34,308.7276 sent, and 36,398.6252 received.

Figure 7: Examples of marketplaces with (a) most received BTC and (2) second-most Bitcoin addresses.

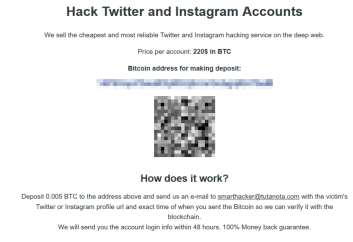
(a) *Bitcoin BTC Mixer* 35.7420 sent, and 37.6483 received.

(b) *Bitcoin Private Key Shop.* 10.8981 sent, and 11.9599 received.

Figure 8: Examples of mirrored marketplaces.



(a) *Choose Better* 0.1203 sent, and 0.1215 received.



(b) *Hack Twitter and Instagram Accounts.* 0.1203 sent, 0.1215 received.

Figure 9: Examples sites using the same address as payment option for different services.

7 Conclusion

The combination of the Dark Web and Bitcoin creates a complex ecosystem balancing anonymity and misuse. To facilitate data-driven analysis, we introduced Venom, a framework for systematically collecting and analyzing Bitcoin-related data from onion sites. Using Venom, we generate a dataset of over 177,000 onion sites, uncovering key patterns in Bitcoin usage and providing insights into the Dark Web's economic dynamics. Our framework advances research and offers practical tools for monitoring illicit activities.

Future work could focus on analyzing transactional relationships, enhancing real-time monitoring, and tracking long-term trends through daily snapshots. Other interesting future includes exploring dataset augmentation through follow-the-money-style Bitcoin analysis tools. While we prioritized Venom's core functionality and dataset integrity, such tools can be seamlessly incorporated to extend analyses and create complementary datasets.

Finally, we note that the crawler we created for this project had the sole focus of collecting the main page of onion domains; not their many subpages. Accordingly, there are definitely many more Bitcoin addresses to discover, analyze, and track on the Dark Web. However, for effective data collection, interesting future work must balance the depth that each domain is crawled against the data collection period of each snapshot. Code and dataset can be found here: www.ida.liu.se/~nikca89/papers/venom.html.

Acknowledgments

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

[1] Ahmia. 2024. Ahmia – Search Tor Hidden Services. https://ahmia.fi/.
[2] Ahmia. 2024. Search results for – Ahmia. https://ahmia.fi/search/?q=.
[3] Alex Biryukov, Ivan Pustogarov, Fabrice Thill, and Ralf-Philipp Weinmann. 2014. Content and Popularity Analysis of Tor Hidden Services. In *ICDCS Workshops*.
[4] Blockstream. 2024. Blockstream API. https://blockstream.info/api/address/.
[5] Blockstream. 2024. Blockstream: Bitcoin and digital asset infrastructure. https://blockstream.com/
[6] Michael Chertoff. 2017. A public policy perspective of the Dark Web. *Journal of Cyber Policy* 2, 1 (2017), 26–38.
[7] Nicolas Christin. 2013. Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. In *WWW*. 213–224.
[8] Roger Dingledine, Nick Mathewson, Paul F Syverson, et al. 2004. Tor: The second-generation onion router.. In *USENIX Security Symposium*, Vol. 4. 303–320.
[9] Abeer ElBahrawy, Laura Alessandretti, Leonid Rusnac, Daniel Goldsmith, Alexander Teytelboym, and Andrea Baronchelli. 2020. Collective dynamics of dark web marketplaces. *Scientific reports* 10, 1 (2020), 18827.
[10] Kristin M Finklea. 2017. Dark Web.
[11] Sean Foley, Jonathan R Karlsen, and Tālis J Putniņš. 2019. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies* 32, 5 (2019), 1798–1853.
[12] Google. 2024. Cloud Computing Services. https://cloud.google.com/
[13] Danny Yuxing Huang, Maxwell Matthaios Aliapoulios, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C Snoeren, and Damon McCoy. 2018. Tracking ransomware end-to-end. In *IEEE Symposium on Security and Privacy*. 618–631.
[14] Seunghyeon Lee, Changhoon Yoon, Heedo Kang, Yeonkeun Kim, Yongdae Kim, Dongsu Han, Sooel Son, and Seungwon Shin. 2019. Cybercriminal minds: an investigative study of cryptocurrency abuses in the dark web. In *NDSS*. 1–15.
[15] Malte Möser, Rainer Böhme, and Dominic Breuker. 2013. An inquiry into money laundering tools in the Bitcoin ecosystem. In *APWG eCrime Researchers Summit*.
[16] Satoshi Nakamoto. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. https://bitcoin.org/bitcoin.pdf White paper.
[17] Masarah Paquet-Clouston, Matteo Romiti, Bernhard Haslhofer, and Thomas Charvat. 2019. Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem. In *ACM Conference on Advances in Financial Technologies*. 76–88.
[18] Hampus Rosenquist, David Hasselquist, Martin Arlitt, and Niklas Carlsson. 2024. On the Dark Side of the Coin: Characterizing Bitcoin Use for Illicit Activities. In *PAM*. 37–66.
[19] Kyle Soska and Nicolas Christin. 2015. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *USENIX Security Symposium*.