Measuring onion website discovery and Tor users' interests with honeypots

1st Arttu Paju
Tampere University
Tampere, Finland
arttu.paju@tuni.fi

2nd Waris Abdullah
Tampere University
Tampere, Finland
waris.abdullah@tuni.fi

3rd Juha Nurmi
Tampere University
Tampere, Finland
juha.nurmi@tuni.fi

***Abstract*—Tor enables anonymous web browsing and access to anonymous onion websites. Prior work has focused on crawling and content analysis rather than on what users actually try to access. Our honeypot approach measures engagement across onion-site categories, revealing behavioral interest rather than inferred popularity. In March–April 2025, we deployed honeypot onion websites and seeded neutral-looking links via three channels—the Ahmia Tor search engine, Stronghold paste onion "paste" service, and pastebin.com—to observe discovery and subsequent interaction events (CAPTCHA solves; registration/login attempts). We observe that, almost without exception, human users originate from Ahmia.fi; after removing the honeypot links from the Ahmia.fi search results, visits dropped to nearly zero and no users solved CAPTCHAs. The honeypot landing front pages represent different forums for cybercrime activities—child sexual abuse, violence, malware, stolen goods, illegal firearms, illegal drugs, and forgery items—and, as a baseline comparison, an unclear forum. Within that set, the CSAM-themed honeypot drew markedly higher engagement than the other honeypots. When identical sites were offered in multiple languages, interaction events occurred most often on the English-language versions.**

***Index Terms*—Tor, Search engine, Onion services, Honeypot**

I. Introduction

Tor and Tor Browser are the most widely used tools for ensuring the anonymity of users and websites. Tor enables anonymous hosting of websites using the `.onion` domain, whereas the Tor Browser enables anonymous web browsing and accessing these onion websites. Ahmia (`Ahmia.fi`) Tor search engine is among the most common means by which individuals discover onion websites [1].

Our research is user-oriented and based on their interests: in March–April 2025, we deployed honeypot onion websites to see how human users try to create credentials for different types of onion websites. Our onion honeypot websites appear to have legitimate cybercrime-related content, but they do not. Instead, we monitor user activity to gather information about their behavior. This research provides an overview of how users follow neutral entry descriptions and onion links from three different sources: the Ahmia Tor search engine, the Stronghold paste onion "paste" service, and pastebin.com. To detect humans and robots following the links, the websites are protected with CAPTCHAs. This enables us to measure (i) how users discover onion websites and (ii) which honeypot onion websites human users access and try to create credentials on. Our measurements are novel, as the related *dark web* research focuses on web crawling rather than on studying how users access onion websites. The following are our research questions (RQs).

RQ1. Paste websites vs. Tor search engine: how do users find our honeypot onion websites?

RQ2. What are the most captivating onion websites based on honeypot interaction data?

RQ3. What languages do users prefer on our onion honeypot websites?

Our contributions are as follows: 1) **Discovery channels vs. human engagement:** We shared links via Ahmia, pastebin.com, and Stronghold paste. While all three channels generated visits, most of human engagement (CAPTCHA solves and registration/log in attempts) in our dataset originated from Ahmia. 2) **Theme-level engagement:** Within Ahmia-originated honeypot interactions, the child sexual abuse material (CSAM)-themed honeypot showed disproportionately high engagement relative to other themes. Surprisingly, illegal drugs was among the least engaged, although related research on content available on Tor often concentrates on marketplaces selling illegal drugs. 3) **Language-level engagement:** When we offered the same honeypot website in multiple domains and languages, English-language honeypots received the most interactions.

In Section II, we outline the research gap that our findings address. We provide an ethical rationale in Section III. In Section IV, we explain our methods. We present our results in Section V. The limitations of our work are listed in Section VI. Finally, in Section VII, we collect our findings and conclude our work. In summary, our work provides the first quantitative measurement of how Tor search engines and paste services lead users to attempt interaction with different categories of onion sites.

II. Related work

Tor routes traffic through multiple relays while encrypting the data between each relay to ensure both the anonymity of the Tor user browsing websites and the anonymity of the websites available on Tor [2]. Tor onion websites (formerly known as "hidden services") constitute a substantial portion of the so-called *dark web*. There are roughly 800,000 onion domains [3], but this does not reflect the number of onion

websites, as a website can have several onion domains, and onion domains host all kinds of internet services, not only web services.

Existing research on the dark web has a focus on analyzing content [4–9]. Particularly, a significant portion of research concentrates on illegal marketplaces [10–16], and identifying onion sites, but they lack insight into user interests. While the popularity of various onion sites/categories—that have been previously researched based on the content available [8, 17–19]—can offer some insight into user interests, it may not always accurately reflect user interest.

Our research is the first to directly quantify the types of onion websites that users actually try to log in to, i.e., not clicking and opening by accident or visited by a robot. This has not been measured since 2014, when one study measured requests for resolving onion domain names, but even then, this research method did not make a distinction between humans and automated crawlers accessing onion websites [20]. As such, the process for resolving onion domains contains automated crawlers accessing these websites and does not distinguish if users are actually eager to log in for the content.

While prior work offers limited insight into discovery mechanisms [21], most evidence comes from surveys [22, 23], which introduce self-selection bias. Based on a survey from August 16 to September 11, 2017, Ahmia is among the most common means by which individuals discover onion websites: "The three most popular ways that almost half of our survey participants discovered onion sites by were . . . (ii) search engines such as Ahmia, (46%)" [1].

In 2024, a study combined query logs from Ahmia and presented a survey targeting individuals who sought CSAM on Ahmia to investigate the number of users accessing this material and their motivations [24] finding that 11.1% of users seek CSAM. Furthermore, this study estimated the number of CSAM websites on Tor and found that one-fifth of the onion websites share CSAM.

In another relevant study [25], researchers deployed honeypots to geolocate IP addresses and capture real-time access patterns; this highlighted regional variation on the dark web across different operating systems (e.g., "Windows 10" and "Android 10"). Although the study offers useful information, it does not clarify how users find these honeypot websites or the popularity between categories.

Research into financial transaction statistics on the dark web, specifically investigating which category generated the most revenue, shows that sexual abuse generated the highest revenue (94,241,807 US dollars) in 2021, followed by drugs, cybercrime, and extremism [26].

An article published in 2020 [27] implemented two secure honeypots—a chatroom and a website. Over a period of seven months, the study discovered that CSAM accounted for 50% of chatroom requests, with discussions on pornography and video links following closely behind. The honeypots also faced attacks, including scripting and SQL injections.

III. Ethics

This research complies with relevant laws, regulations, and guidelines. The Ahmia Terms of Service include a statement that Ahmia is used for scientific research and carries out tests with the interface. Particularly, Ahmia's Terms of Service mention A/B tests and edited result views and the goal to carry out tests with different layouts, messages, or redirect options on a randomized basis to understand user preferences to generate scientific knowledge and benefit society. To protect users' privacy, the Ahmia search engine does not record their IP addresses, does not use cookies, and only saves minimal HTTP logs. Similarly, honeypot onion websites only save minimal HTTP logs, and users remain anonymous as they access them via Tor. We did not track users with cookies because doing so requires users' permission, which we did not want to add as an extra step. We do not compromise the users' anonymity and privacy in any way. We do not cause any harm to any users.

IV. Methods

We shared honeypot URLs on three sources—(i) the Ahmia Tor search engine; (ii) the Stronghold paste onion service; and (iii) pastebin.com—each with a unique onion domain to separate discovery origins.

Each of our honeypot websites included two pages. A neutral-looking landing page with a CAPTCHA puzzle, and a second page—that could be accessed after solving the CAPTCHA—with descriptions of some type of cybercrime activity as well as registration/login fields. The registration and login fields on these pages had no real functionality, but they did leave a trace in logs that tracked the number of users who attempted to create credentials or log in to the sites. We made no distinction between how many users attempted to create credentials or log in to the sites. Instead, we viewed both actions as attempts to access the content available on the honeypot forum.

Our data collection period started on March 24, 2025 by publishing the honeypot websites and distributing the URLs. The URL distribution happened by randomly showing URLs to the honeypot sites alongside other search results on Ahmia and by publishing the URLs on both pastebin.com and Stronghold paste. We only showed the URLs on pastebin.com and Stronghold paste for the duration of one day on March 24. On April 10, 2025, we removed the URLs from Ahmia. Thereafter, we continued to monitor the activity on all URLs for an additional week. Thus, the total analysis period ran from March 24 to April 17, 2025.

We modified the search engine to randomly add a neutral description of "onion log in page available" to the search results, along with a random link when there were more than 10 search results. As a comparison, in March 2024, Ahmia recorded an average of 137,993 daily searches, equivalent to 95 searches per minute. Furthermore, Ahmia filters CSAM from the search results and prevents CSAM-related queries.

To quantify how users find onion services, we measured the number of visits to our honeypot websites from each source.

This allowed us to compare the popularity of each source in accessing onion services and answer the RQ1.

We used CAPTCHA puzzles to protect the onion websites and distinguish real human users from bots and crawlers that follow URLs on the Tor network. Realistically, only few crawlers/bots are capable of solving CAPTCHAs. This is particularly true when the website is brand new and no human has instructed the crawler on how to solve this new CAPTCHA type. Thus, the number of solved CAPTCHAs indicates the number of real human users who attempted to interact with the honeypot sites.

To quantify what kind of onion websites the users try to access, we created honeypot websites representing eight different categories: (i) CSAM; (ii) violence; (iii) malware; (iv) stolen goods; (v) illegal firearms; (vi) illegal drugs; and (vii) forgery items. We selected these categories because prior content-analysis studies show that they are among the most prevalent categories of illegal activities on Tor network [8, 28]. Additionally, we created a generic (viii) unclear forum front page that acted as a baseline comparison. As an example, Figure 1 shows the honeypot website relating to the CSAM in the English language that we used in the research. Screenshots of other honeypot websites in the English language are in the Appendix A.

By analyzing the number of attempts to create credentials or log in on these sites related to different categories, we can determine which websites the users are actually interested in and trying to access, thus answering the RQ2.

To quantify the disparity between users' preferences for language variants, we implemented identical websites for each category in four languages: English, German, Finnish, and Russian. We selected English, German, Finnish, and Russian since prior studies identify them as prevalent onion-site languages [5, 8]. By comparing the number of attempts to create credentials or log in on different language sites, we can determine the popularity of each language among users. This allowed us to answer the RQ3.

In total, we created honeypot sites for eight categories in four languages, resulting in 32 unique websites. We also created separate domains for each of the three sources, which resulted in 96 unique URLs for the research.

The unit of analysis in our research is honeypot interaction events. Tor's anonymity and our minimal logging prevent reliable de-duplication of individuals; a single person may generate multiple events (e.g., multiple registration attempts after one CAPTCHA). Accordingly, any references to "user interest" should be interpreted as aggregate event-level engagement, not person-level preferences.

V. Results

We tracked the number of visits, number of solved CAPTCHAs, and attempts to register or log in to our honeypot sites from each source (Ahmia, pastebin.com, and Stronghold paste). Comparing the number of visits and solved CAPTCHAs originating from each source allows us to answer the RQ1, and comparing the number of attempts to register/log in for different honeypot sites allows us to answer the RQ2 and the RQ3.

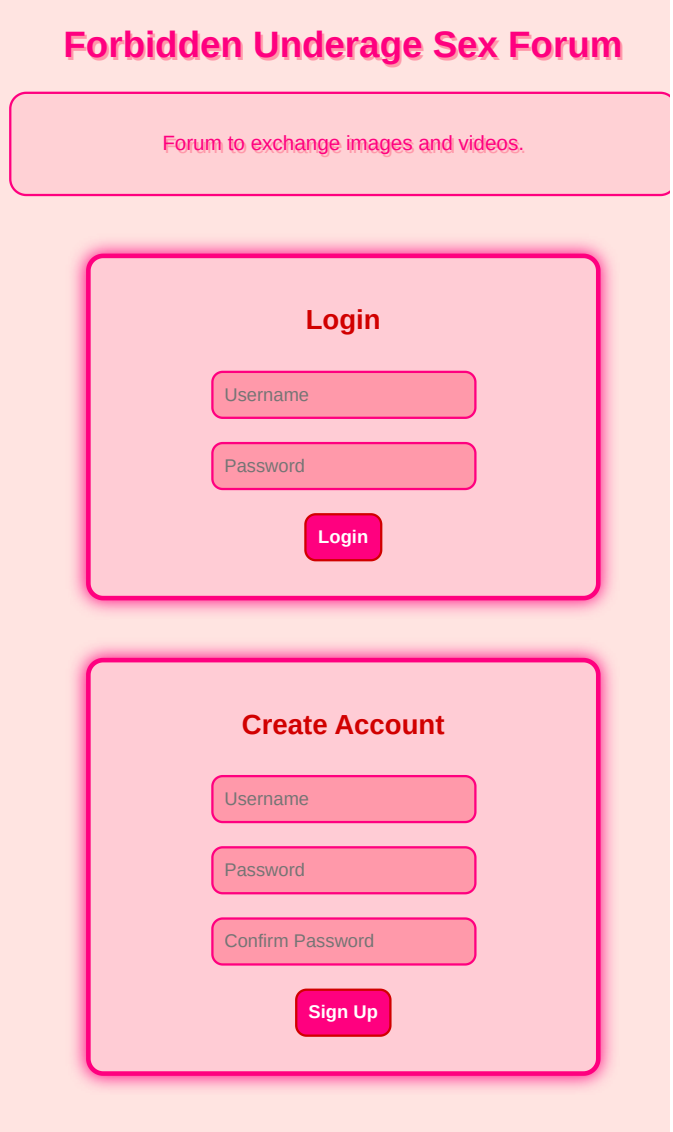


Fig. 1: CSAM honeypot (English version). All honeypot websites contained only text-based hints for users, along with free-to-use legal images, and did not contain any illegal content.

A. RQ1: Paste websites vs. Tor search engine: how do users find our honeypot onion websites?

Between March 26, 2025 and April 10, 2025—while distributing links to our honeypot sites—we received a combined total of 219,173 visits on all honeypot sites, with 87.65 % (n = 192,101) coming from Ahmia, 6.42 % (n = 14,066) from pastebin.com, and 5.93 % (n = 13,006) from Stronghold paste, respectively. The total number of visits over time for each source is presented in Figure 2.

During the first day of the trial, the majority of visits originated from the Ahmia search engine, while pastebin.com and Stronghold paste received less traffic by an order of magnitude.

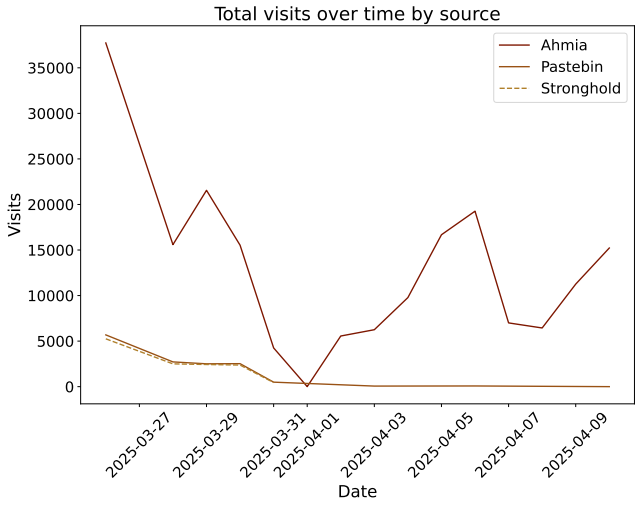


Fig. 2: Visits to our honeypot websites from all three distribution sources.

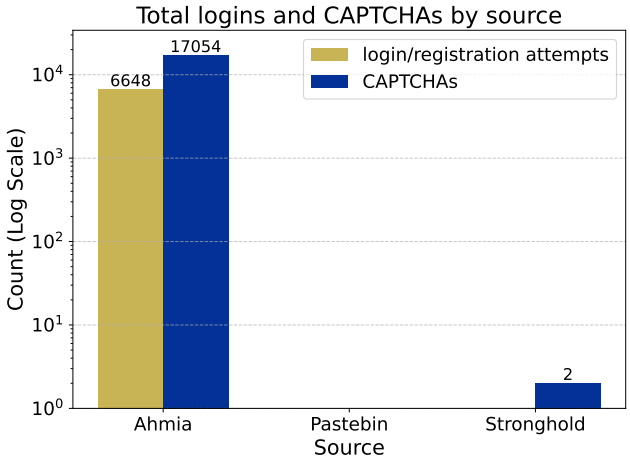


Fig. 3: Resolved CAPTCHAs and log in attempts from all three sources.

The difference in the number of visits from pastebin.com and Stronghold paste was negligible. On the first day of the analysis (March 26), we had significantly more visits from all three sources than on any other day. This is to be expected, as whenever new onion sites are indexed on Tor, active crawlers send massive amounts of traffic to those sites. After the first day, we received fluctuating visits from Ahmia, but at least thousands of daily visits throughout the analysis period, except on April 1 when the search functionality was not working properly.

Two "paste" websites, pastebin.com and Stronghold paste, displayed the links for 24 hours to compare their popularity on the first trial day. Please note that these paste websites only highlight the most recently added content, and Stronghold paste even automatically removes the paste after 24 hours. Thus, we compare the popularity of Ahmia users and paste site users on the first day. We see some visiting activity after the first day, which reveals that some automated systems collected the onion links from these paste websites and continued to visit them. Nevertheless, these visits are not mainly from human users, as only two of them actually led to resolving the CAPTCHA, and there have been no log in attempts; we will discuss these findings further next.

During our active URL sharing period between March 26, 2025 and April 10, 2025, our dataset recorded a combined total of 17,056 solved CAPTCHAs (7.78 % of visits leading to solved CAPTCHAs) and 6,648 attempts to register or log in (3.03 % of visits leading to registration/log in attempts). The total count of solved CAPTCHAs and attempts to register/log in originating from each source is presented in Figure 3.

We observed a continuous stream of solved CAPTCHAs and log in attempts exclusively from Ahmia's search results. In total, 17,054 solved CAPTCHAs and 6,648 registration/log in attempts originated from Ahmia. Only two completed CAPTCHAs originated from Stronghold paste. No registration/log in attempts originated from either pastebin.com or Stronghold paste. *These results show that real users prefer Ahmia search engine over paste sites for accessing dark web onion websites*. Stronghold paste and pastebin.com do not represent actual user traffic, but rather bots and crawlers.

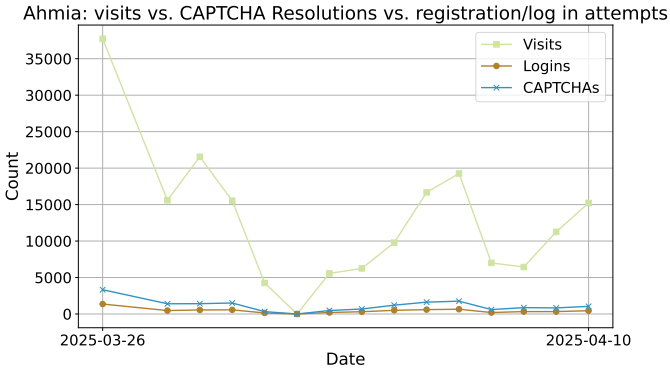


Fig. 4: Visits, solved CAPTCHAs, and registration/log in attempts originating from Ahmia's results.

Figure 4 shows the total number of visits, solved CAPTCHAs, and registration/log in attempts from Ahmia. The solved CAPTCHAs and registration/log in attempts follow the trend of overall visits quite consistently. On March 26—the first day of the analysis—there were 3,329 CAPTCHAs solved and 1,362 registration/log in attempts, which correspond to a 6.84 % CAPTCHA-solving rate and a 38.98 % follow-up registration rate among those who solved the CAPTCHA. These figures had dropped to 1,045 and 444, respectively, by the end of the URL sharing period on April 10. Only a small fraction of the visits resulted in CAPTCHA resolutions, and an even smaller number led to registration/log in attempts.

After removing the URLs from Ahmia on April 10, we continued to monitor the activity on all URLs for an additional week. We did not detect any attempts to create credentials or

log in. In fact, the CAPTCHA was solved only 20 times, and there were only 1,053 visits during this seven-day period. This again implies that Ahmia was the primary source for human users accessing the onion addresses during the investigation.

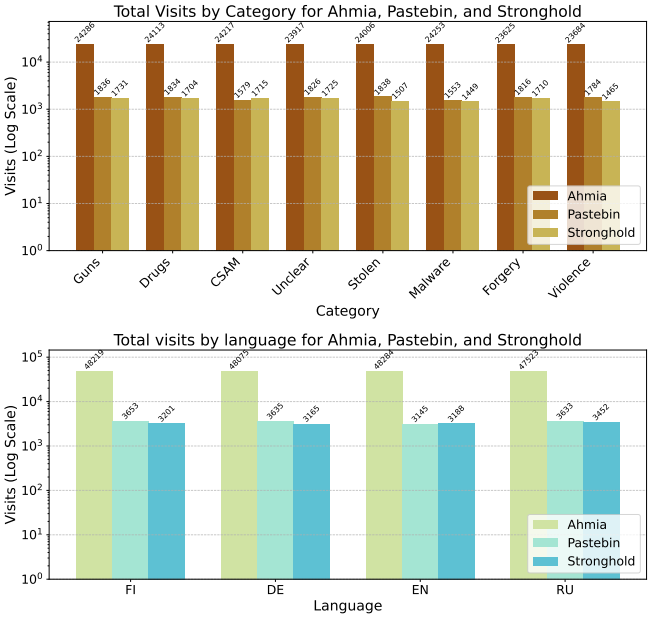
B. RQ2: What are the most captivating onion websites based on honeypot interaction data?

We analyzed traffic to our honeypot sites in eight distinct categories (CSAM, violence, malware, stolen goods, illegal firearms, illegal drugs, forgery items, and unclear) and four distinct languages (English, German, Finnish, and Russian). Different categories and languages have different numbers of registration/log in attempts, indicating user preferences. Thus, comparing the total number of registration/log in attempts on different categories and languages allows us to determine the type of content users seek on the dark web as well as their language preferences.

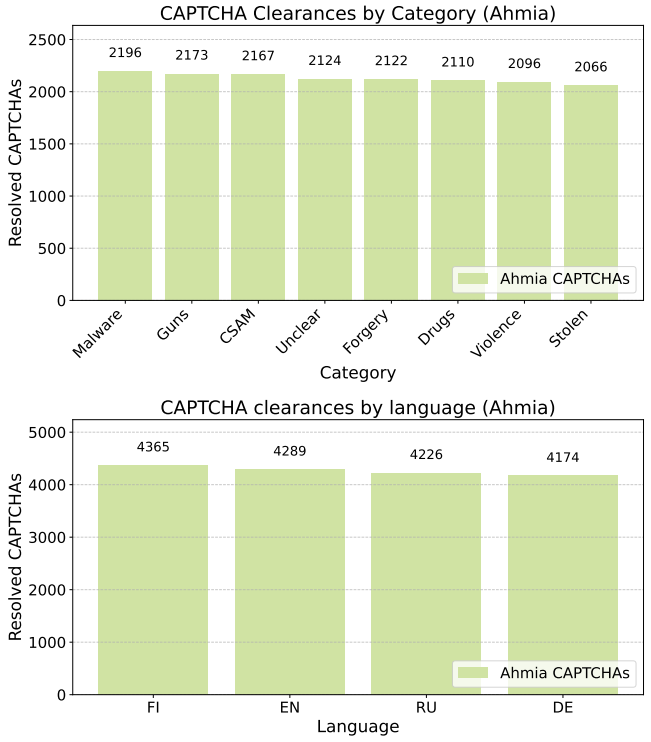
Because Ahmia displays a neutral "Onion login page" and URL for all websites without content information before CAPTCHA clearance, and URLs posted in pastebin.com and Stronghold paste do not show the context of honeypot sites, all categories and languages receive equal traffic and CAPTCHA clearances, as expected. Figure 5a shows a consistent distribution of visits across categories and languages, while Figure 5b shows a similar consistent distribution of CAPTCHA clearances across categories and languages. Thus, the number of visits or resolved CAPTCHAs cannot be used for determining the user preference for different categories or languages. However, the number of registration/log in attempts indicates genuine user interest, as only authentic users typically bypass the CAPTCHA and gain access to the site's content, and users will only attempt to create credentials or log in if they are interested in the material.

We recorded a total of 6,648 registration/log in attempts originating from Ahmia. Figure 6 displays the number of attempts for each category. CSAM was by far the most popular category, with 37.67 % (n = 2,504) of all registration/log in attempts. Violence was the second most popular category by a significant margin, with 20.47 % (n = 1,361) of all attempts. Following these, the baseline unclear and malware categories received roughly equal interest, with 10.51 % (n = 699) and 10.36 % (n = 689) of all attempts, respectively. Following that, the stolen goods and illegal firearms categories also saw roughly equal popularity, with 7.31 % (n = 486) and 6.86 % (n = 456) of all attempts, respectively. Surprisingly, illegal drugs was the second least popular category, with only 3.43 % (n = 228) of all attempts. The least popular category was forgery items, which had 3.38 % (n = 225) of all attempts—just barely less than illegal drugs. There were no registration/log in attempts from pastebin.com or Stronghold paste, again indicating that the majority of visits from these sources were bots or crawlers.

The statistics in Table I show the percentage of users who tried to register/log in to each category of websites after resolving the CAPTCHAs. We only included the traffic from



(a) Total visits by category and language for Ahmia, pastebin.com, and Stronghold paste.



(b) Resolved CAPTCHAs for each category and language of websites for traffic coming from Ahmia.

Fig. 5: All honeypot websites—as expected—received roughly equal amount of visits and resolved CAPTCHAs.

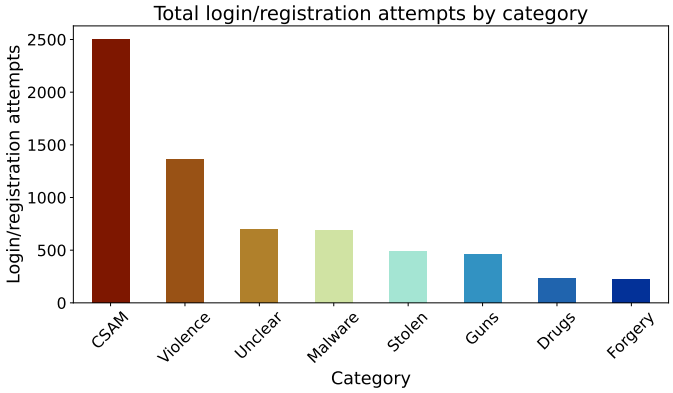


Fig. 6: Total registration/log in attempts for each category of websites.

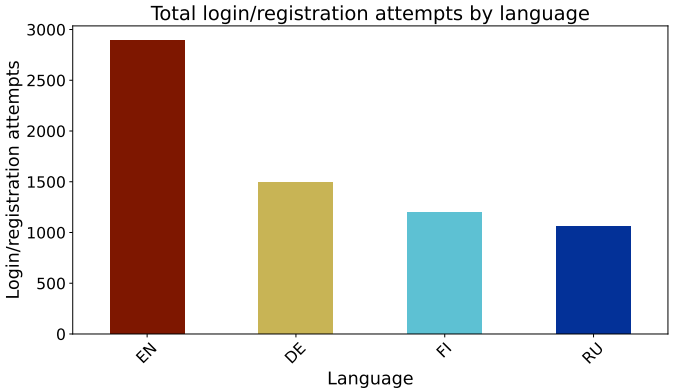


Fig. 7: Through all categories, the log in attempts by language.

Ahmia, since no attempts to register/log in originated from either Stronghold paste or pastebin.com.

TABLE I: Percentage of registration/log in attempts per resolved CAPTCHA for each category.

Category	CAPTCHAs	Log./cred.creat. attempts	Percentage
CSAM	2,167	2,504	115.55
Violence	2,096	1,361	64.93
Unclear	2,124	699	32.91
Malware	2,196	689	31.38
Stolen goods	2,066	486	23.52
Illegal firearms	2,173	456	20.98
Illegal drugs	2,110	228	10.81
Forgery	2,122	225	10.60
Total	17,054	6,648	38.98

After resolving the CAPTCHA, an average of 38.98 % of users attempted to register/log in to the website. This number varied greatly depending on the category. The CSAM category had an unusually high rate of user engagement at 115.55 %, indicating multiple registration/log in attempts per user who solved the CAPTCHA, underscoring that results are event-level rather than person-level. The violence category also had significant user engagement at 64.93 %. On the other hand, illegal drugs and forgery had significantly lower user engagement rates than the average at 10.81 % and 10.60 %, respectively. Overall—as expected—the user engagement rate closely matches the number of registration/log in attempts per website category (see Figure 6). *These interactions show disproportionately high user engagement with the CSAM- and violence-themed honeypots*. The unexpectedly high CSAM engagement—despite Ahmia's filtering—may indicate that search-engine-referred visitors are a less experienced subset of Tor users who engage with first-accessible content.

C. RQ3: What languages do users prefer on our onion honeypot websites?

When the same category of a website is available in multiple languages, the number of registration/log in attempts reveals the popularity between them. Figure 7 displays the number of registration/log in attempts per language. The most popular language was English, with 43.49 % (n = 2,891 of 6,648) of all attempts, followed by German and Finnish (22.47 % (n = 1,494) and 18.07 % (n = 1,201) of all attempts, respectively). Russian was the least-popular language with 15.97 % (n = 1,062) of all attempts. The percentage of resolved CAPTCHAs leading to log in attempts is detailed in Table II. English shows the highest conversion rate at 67.40 %, followed by German at 35.8 %, Finnish at 27.5 %, and Russian at 25.1 %.

TABLE II: Percentage of log in attempts per resolved CAPTCHAs by language on Ahmia.

Language	CAPTCHAs	Log./cred.creat. attempts	Percentage
English	4,289	2,891	67.40
German	4,174	1,494	35.79
Finnish	4,365	1,201	27.51
Russian	4,226	1,062	25.13
Total	17,054	6,648	38.98

Despite being generally more prominent on onion websites [5, 8] and having more global speakers [29], Russian was less popular in our dataset than Finnish. The Finnish language overrepresentation in our dataset might be explained by the fact that Ahmia is a Finnish search engine, so it may have a higher popularity among the Finnish-speaking population compared to other websites. Similarly, Russian Tor users may not use Ahmia as frequently as English-, German-, or Finnish-speaking users, thus resulting in underrepresentation of the language in our dataset.

We also analyzed registration/log in attempts in the most popular category—CSAM—for each language in Figure 8. The distribution was similar, with English being the most popular language with 40.26 % (n = 1,006 of 2,499) of all log in attempts, followed by German (24.93 % or n = 623 of 2,499), Finnish (18.81 % or n = 470 of 2,499), and Russian (16.01 % or n = 400 of 2,499). *Our results clearly indicate that users prefer English over other languages.*

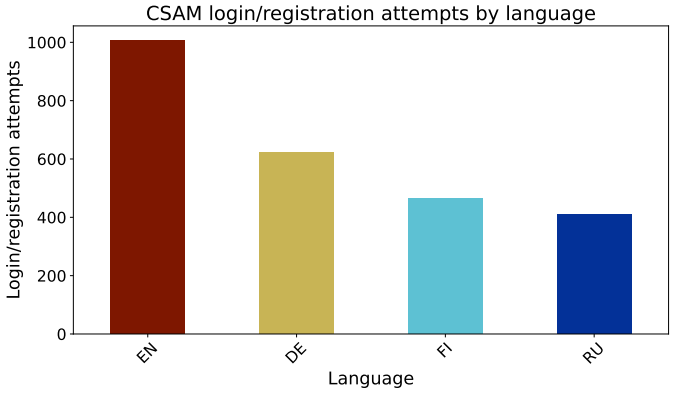


Fig. 8: For the most popular category, CSAM, log in attempts per language.

VI. Limitations

Our dataset captures only a subset of Tor activity: Not all Tor users access onion websites. In 2015, the Tor Project estimated that onion service traffic accounts for about 3.4% of total Tor traffic [30]. Moreover, Ahmia users represent only a subset of Tor search engine users, introducing sampling bias.

Ahmia filters CSAM from the search results and prevents CSAM-related queries: A major limitation is that human interactions in our dataset stem almost entirely from Ahmia, which blocks CSAM-related queries. Consequently, our RQ2 interpretations are conditional on a filtered, general-purpose search-engine discovery channel and may not reflect behaviors of users who rely on unfiltered/private link-sharing or specialized forums. Paradoxically, the high CSAM engagement observed despite filtering is therefore surprising, potentially reflecting a less-experienced audience arriving via search rather than seasoned offenders using alternative discovery paths. Similarly, the low engagement with drug sites could be due to serious buyers already knowing the major marketplaces and having no reason to search for them on Ahmia or paste services.

Honeypot realism: To allow for controlled comparisons, all of our honeypots had a minimalist and similar design. This improves internal validity while decreasing realism in comparison to authentic illicit forums, potentially altering observed engagement tendencies.

Language coverage: Our honeypots were available only in English, German, Finnish, and Russian. Prior analyzes of Tor hidden-service content indicate that French and Italian also appear among the most common languages on onion websites. Consequently, our dataset does not capture engagement from those linguistic communities. Finnish was included primarily because we could generate accurate content easily, even though Finnish has a comparatively smaller presence on Tor than French and Italian. Future work should expand honeypot deployments to include French and Italian to test whether the observed engagement patterns generalize across these additional languages.

VII. Conclusion

Methodologically, the honeypot design and the CAPTCHA test allowed us to distinguish between human and automated traffic. We measured honeypot discovery and interaction across three channels: the Ahmia Tor search engine, pastebin.com, and Stronghold paste. All three channels generated traffic, but human interactions originated almost entirely from Ahmia.

Analysis of registration/log in attempts—our proxy for genuine user interest—reveals that interaction events were disproportionately concentrated on abusive material. The CSAM-themed honeypot attracted most user engagement, with more than twice the engagement of the next most popular category (violence). Other categories with moderate interest included malware, stolen goods, and illegal firearms, while illegal drugs and forgery received the least engagement. Notably, the drug category's low engagement contrasts with its prominence in the larger dark web discourse.

In terms of language, English-language variants had the most interactions. English was the language with both the highest overall and CSAM-specific engagement, followed by German and Finnish. Despite its large global speaker base, the Russian language had the lowest user engagement. Language preferences remained consistent across categories. This over-representation of Finnish may be due to Ahmia's origin and user base, rather than general dark web trends. Overall, our findings reflect user engagement patterns that are specific to the discovery channel (public, general-purpose, filtered search engine) rather than the entire dark-web ecosystem.

Acknowledgment

The authors declare no competing interests. This work was supported by the European Commission under the Horizon Europe funding programme, as part of the project SafeHorizon (Grant Agreement 101168562). The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

Appendix

Contributor Roles Taxonomy (CRediT) statement of authorship contribution.

Contributions	A.P.	W.A.	J.N.
Conceptualization	●	○	●
Data curation	●	○	●
Formal Analysis	●	●	●
Funding acquisition	○	○	●
Investigation	●	●	●
Methodology	●	○	●
Project administration	●	○	●
Resources	○	○	●
Software	○	○	●
Supervision	●	○	●
Validation	●	●	●
Visualization	●	●	●
Writing—original draft	●	●	●
Writing—review & editing	●	○	●

References

[1] P. Winter, A. Edmundson, L. M. Roberts, A. Dutkowska-Zuk, M. Chetty, and N. Feamster, "How Do Tor Users Interact With Onion Services?" in 27th USENIX Security Symposium, USENIX Security 2018,

Fig. 9: Screenshots of the honeypot websites (English versions): malware

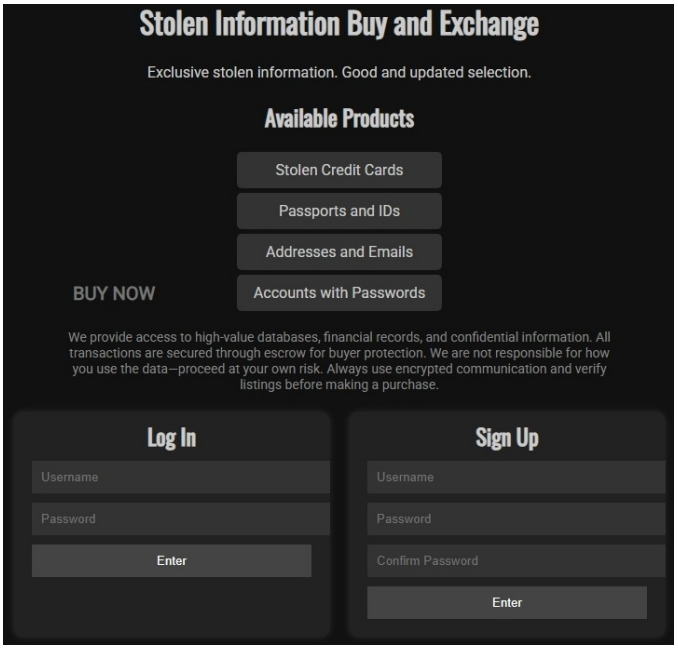


Fig. 10: Stolen goods honeypot (English version)

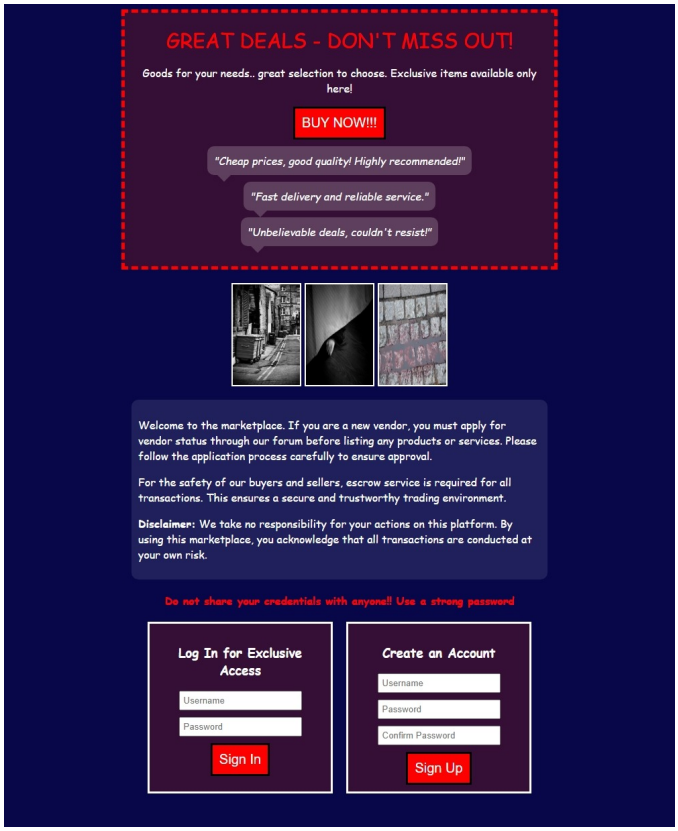


Fig. 11: Unclear honeypot (English version)

Baltimore, MD, USA, August 15–17, 2018, W. Enck and A. P. Felt, Eds. USENIX Association, 2018, pp. 411–428. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/winter

[2] R. Dingledine, N. Mathewson, and P. F. Syverson, "Tor: The second-generation onion router," in Proceedings of the 13th USENIX Security Symposium, August 9–13, 2004, San Diego, CA, USA, M. Blaze, Ed. USENIX, 2004, pp. 303–320. [Online]. Available: http://www.usenix.org/publications/library/proceedings/sec04/tech/dingledine.html

[3] The Tor Project, "Tor metrics – onion services between March 2024 and May 2025," https://metrics.torproject.org/hidserv-dir-v3-onions-seen.html?start=2024-03-01&end=2025-05-20, 2025, accessed: 2025-05-20.

[4] S. Aked, "An Investigation Into Darknets and the Content Available via Anonymous Peer-to-Peer File Sharing," in Proceedings of the 9th Australian Information Security Management Conference, Perth, Western Australia, December 5 – 7, 2011. secau Security Research Centre, Edith Cowan University, Perth, Western Australia, 2011. [Online]. Available: https://doi.org/10.4225/75/57b52857cd8b3

[5] M. Spitters, S. Verbruggen, and M. van Staalduinen, "Towards a Comprehensive Insight into the Thematic Organization of the Tor Hidden Services," in IEEE Joint Intelligence and Security Informatics Conference, JISIC 2014, The Hague, The Netherlands, 24–26 September, 2014. IEEE, 2014, pp. 220–223. [Online]. Available: https://doi.org/10.1109/JISIC.2014.40

[6] M. Zabihimayvan, R. Sadeghi, D. Doran, and M. Allahyari, "A Broad Evaluation of the Tor English Content Ecosystem," in Proceedings of the 11th ACM Conference on Web Science, WebSci 2019, Boston, MA, USA, June 30 – July 03, 2019, P. Boldi, B. F. Welles, K. Kinder-Kurlanda, C. Wilson, I. Peters, and W. M. Jr., Eds. ACM, 2019, pp. 333–342. [Online]. Available: https://doi.org/10.1145/3292522.3326031

[7] J. Dalins, C. Wilson, and M. J. Carman, "Criminal Motivation on the Dark Web: A Categorisation Model for Law Enforcement," Digit. Investig., vol. 24, pp. 62–71, 2018. [Online]. Available: https://doi.org/10.1016/j.diin.2017.12.003

[8] S. He, Y. He, and M. Li, "Classification of illegal activities on the dark web," in Proceedings of the 2nd International Conference on

Fig. 12: Forgery honeypot (English version)

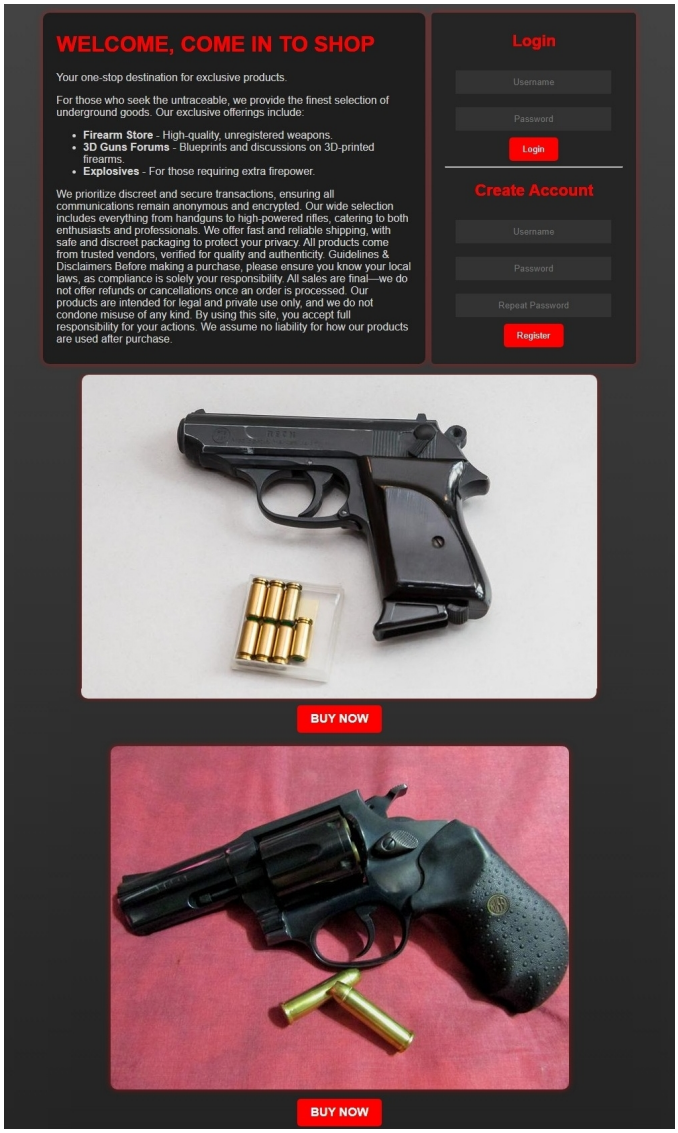


Fig. 13: Firearms honeypot (English version)

Information Science and Systems, ser. ICISS '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 73–78. [Online]. Available: https://doi-org.libproxy.tuni.fi/10.1145/3322645.3322691

[9] Y. Boshmaf, I. Perera, U. Kumarasinghe, S. Liyanage, and H. A. Jawaheri, "Dizzy: Large-Scale Crawling and Analysis of Onion Services," in Proceedings of the 18th International Conference on Availability, Reliability and Security, ARES 2023, Benevento, Italy, 29 August 2023 – 1 September 2023. ACM, 2023, pp. 9:1–9:11. [Online]. Available: https://doi.org/10.1145/3600160.3600167

[10] N. Christin, "Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace," in 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013, D. Schwabe, V. A. F. Almeida, H. Glaser, R. Baeza-Yates, and S. B. Moon, Eds. International World Wide Web Conferences Steering Committee / ACM, 2013, pp. 213–224. [Online]. Available: https://doi.org/10.1145/2488388.2488408

[11] R. van Wegberg, S. Tajalizadehkhoob, K. Soska, U. Akyazi, C. H. Gañán, B. Klievink, N. Christin, and M. van Eeten, "Plug and prey? measuring the commoditization of cybercrime via online anonymous markets," in 27th USENIX Security Symposium, USENIX Security 2018, Baltimore, MD, USA, August 15–17, 2018, W. Enck and A. P. Felt, Eds. USENIX Association, 2018, pp. 1009–1026. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/van-wegberg

[12] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem," in 24th USENIX Security Symposium, USENIX Security 15, Washington, D.C., USA, August 12–14, 2015, J. Jung and T. Holz, Eds. USENIX Association, 2015, pp. 33–48. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/soska

[13] J. Nurmi, M. S. Niemelä, and B. B. Brumley, "Malware finances and operations: a data-driven study of the value chain for infections and compromised access," in Proceedings of the 18th International Conference on Availability, Reliability and Security, ARES 2023, Benevento, Italy, 29 August 2023 – 1 September 2023. ACM, 2023, pp. 108:1–108:12. [Online]. Available: https://doi.org/10.1145/3600160.3605047

[14] K. Oosthoek, J. Cable, and G. Smaragdakis, "A Tale of Two Markets: Investigating the Ransomware Payments Economy," Communications of the ACM, vol. 66, no. 8, pp. 74–83, 2023. [Online]. Available: https://doi.org/10.1145/3582489

[15] A. Baravalle, M. S. Lopez, and S. W. Lee, "Mining the Dark Web: Drugs and Fake Ids," in IEEE International Conference on Data Mining Workshops, ICDM Workshops 2016, December 12–15, 2016, Barcelona, Spain, C. Domeniconi, F. Gullo, F. Bonchi, J. Domingo-Ferrer, R. Baeza-Yates, Z. Zhou, and X. Wu, Eds. IEEE Computer Society, 2016, pp. 350–356. [Online]. Available: https://doi.org/10.1109/ICDMW.2016.0056

[16] R. S. Portnoff, S. Afroz, G. Durrett, J. K. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson, "Tools for automated analysis of cybercriminal markets," in Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3–7, 2017, R. Barrett, R. Cummings, E. Agichtein, and E. Gabrilovich, Eds. ACM, 2017, pp. 657–666. [Online]. Available: https://doi.org/10.1145/3038912.3052600

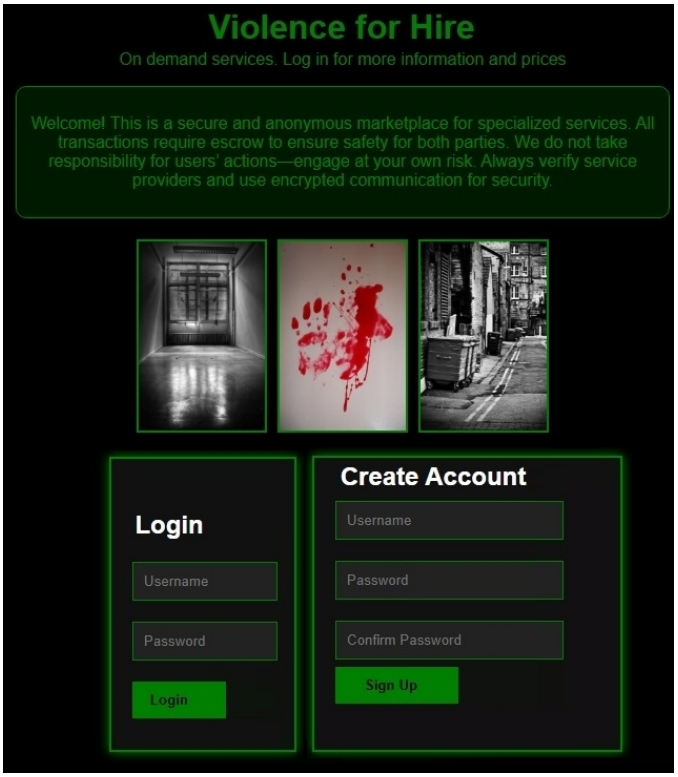


Fig. 14: Violence honeypot (English version)

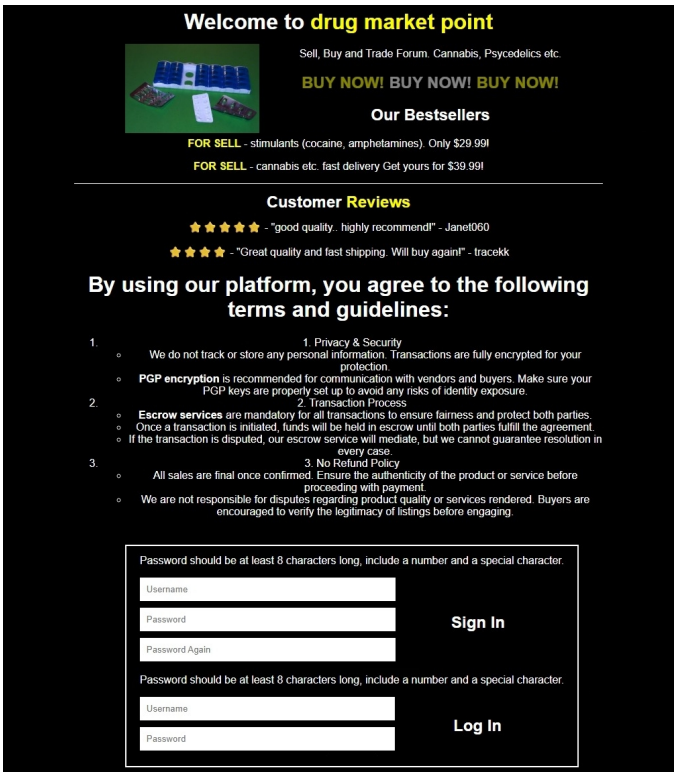


Fig. 15: Drugs honeypot (English version)

[17] N. Tavabi, N. Bartley, A. Abeliuk, S. Soni, E. Ferrara, and K. Lerman, "Characterizing Activity on the Deep and Dark Web," in Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13–17, 2019, S. Amer-Yahia, M. Mahdian, A. Goel, G. Houben, K. Lerman, J. J. McAuley, R. Baeza-Yates, and L. Zia, Eds. ACM, 2019, pp. 206–213. [Online]. Available: https://doi.org/10.1145/3308560.3316502

[18] M. Campobasso and L. Allodi, "Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale," in 32nd USENIX Security Symposium, USENIX Security 2023, Anaheim, CA, USA, August 9–11, 2023, J. A. Calandrino and C. Troncoso, Eds. USENIX Association, 2023, pp. 553–570. [Online]. Available: https://www.usenix.org/conference/usenixsecurity23/presentation/campobasso

[19] M. W. A. Nabki, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "ToRank: Identifying the Most Influential Suspicious Domains in the Tor Network," Expert Syst. Appl., vol. 123, pp. 212–226, 2019. [Online]. Available: https://doi.org/10.1016/j.eswa.2019.01.029

[20] G. Owen and N. Savage, "Empirical analysis of Tor Hidden Services," IET Information Security, vol. 10, no. 3, pp. 113–118, 2016. [Online]. Available: https://doi.org/10.1049/iet-ifs.2015.0121

[21] K. Gallagher, S. Patil, and N. D. Memon, "New Me: Understanding Expert and Non-Expert Perceptions and Usage of the Tor Anonymity Network," in Thirteenth Symposium on Usable Privacy and Security, SOUPS 2017, Santa Clara, CA, USA, July 12–14, 2017. USENIX Association, 2017, pp. 385–398. [Online]. Available: https://www.usenix.org/conference/soups2017/technical-sessions/presentation/gallagher

[22] J. Lusthaus, "Beneath the Dark Web: Excavating the Layers of Cybercrime's Underground Economy," in 2019 IEEE European Symposium on Security and Privacy Workshops, EuroS&P Workshops 2019, Stockholm, Sweden, June 17–19, 2019. IEEE, 2019, pp. 474–480. [Online]. Available: https://doi.org/10.1109/EuroSPW.2019.00059

[23] H. Zhang and F. Zou, "A Survey of the Dark Web and Dark Market Research," in 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1694–1705. [Online]. Available: https://doi.org/10.1109/ICCC51575.2020.9345271

[24] J. Nurmi, A. Paju, B. B. Brumley, T. Insoll, A. K. Ovaska, V. Soloveva, N. Vaaranen-Valkonen, M. Aaltonen, and D. Arroyo, "Investigating Child Sexual Abuse Material Availability, Searches, and Users on the Anonymous Tor Network for a Public Health Intervention Strategy," Scientific Reports, vol. 14, p. 7849, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-58346-7

[25] F. Gallo-Serpillo and J. Valls-Prieto, "Analysis of CSEM offenders on the dark web using honeypots to geolocate IP addresses from spain," Comput. Hum. Behav., vol. 154, p. 108137, 2024. [Online]. Available: https://doi.org/10.1016/j.chb.2024.108137

[26] K. Oosthoek, M. van Staalduinen, and G. Smaragdakis, "Quantifying Dark Web Shops' Illicit Revenue," IEEE Access, vol. 11, pp. 4794–4808, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3235409

[27] R. B. Zeid, J. Moubarak, and C. Bassil, "Investigating the darknet," in 16th International Wireless Communications and Mobile Computing Conference, IWCMC 2020, Limassol, Cyprus, June 15–19, 2020. IEEE, 2020, pp. 727–732. [Online]. Available: https://doi.org/10.1109/IWCMC48107.2020.9148422

[28] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and I. de Paz, "Classifying illegal activities on tor network based on web textual contents," in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers, M. Lapata, P. Blunsom, and A. Koller, Eds. Association for Computational Linguistics, 2017, pp. 35–43. [Online]. Available: https://doi.org/10.18653/v1/e17-1004

[29] Ethnologue, "What are the top 200 most spoken languages?" https://www.ethnologue.com/insights/ethnologue200/, 2025, accessed: 2025-07-23.

[30] R. Dingledine, "Some statistics about onions," https://blog.torproject.org/some-statistics-about-onions/, 2015, accessed: 2025-05-18.