## *[9217] Examining the trends and operations of modern Dark-web marketplaces*

PRELIMINARY DECISION: accept

AUTHORS: Víctor Labrador Ortiga and Sergio Pastrana Portillo

# Summary of Reviews

- Review 1: 1 (3)
- Review 2: 3 (4)
- Review 3: 2 (5)
- Review 4: -2 (4)

# Reviews

***Review 1***

**SCORE:** 1 (weak accept)

I read this paper with great interest, and overall I think it is a nice addiction to the existing literature. The approach used is original, and could be of use to researcher in both computer science, criminology, and other disciplinary areas. The organization of the paper is clear, and the empirical results provide an up-to-dated understanding of the situation in CERTAIN markets.

My main concern is that the limitations of the methodology used are scarcely addressed. In particular, I would like to see many more details as regards how the markets were sampled, how this affects the results and their interpretation, and how data were (in practice) analysed. There are a lot of details on the software architecture and crawler operation, but that is only a part of the story. Without further details, it is difficult to assess the solidity of the results presented, and their interpretation.

***Review 2***

**SCORE:** 0 (borderline paper)

This paper examines 123k listings across 6 darkweb markets containing information about the volume and type of products sold, and origin/destination of the sale, and overall find that drugs are the most prevalent category of item sold and that some vendors did not ship internationally. The paper is generally well written, though a proper proofreading from a native english speaker would be worthwhile.

One of the innovations presented by the authors, beyond analysis of darkweb markets, is an updated tool that allows for this kind of data collection. Ostensibly, the illicit marketplace owners are becoming more sophisticated at tracking and blocking automated crawlers, and so these techniques to crawl and collect data also require advancement.

Are the author certain that lost trust was the reason for Cannazon's termination, and did they find any actual evidence of this? Otherwise, this statement, "the attack suffered by the market lead to a lack of trusts from sellers and buyers, ultimately leading to the decision of the operators to close the market" may be

overstated. This is one place where the authors could bolster their research and provide more novel and interesting findings. But absent more evidence, this is all speculation.

The authors write how they conduct "longitudinal analysis, we select one productive and dynamic market to conduct periodic crawls and get different snapshots over time, in order to understand the evolution of such market." -- however, even given the authors' later admission, there were not enough data to properly perform longitudinal analysis. One might suggest reframing the initial claim that the paper will be performing longitudinal analysis.

By now there is a considerable body of literature examining darkweb and illicit marketplaces. In the beginning, any information was novel and useful, but now there is a higher bar to demonstrating a novel contribution to the field. Yes, most current data works in the authors' favor, yet the field demands more analysis and insight than just descriptive summaries of top products and sellers. And so overall the contribution is modest. It does appear that the authors spend considerable work developing the crawler and adapting it to new security mechanisms by the darkweb sites. And while the analysis and insights were somewhat interesting, there did not appear to be any novel techniques used, hypotheses tested, or insights gained.

***Review 3***

**SCORE:** -2 (reject)

This paper presents new datasets (and a new crawler to obtain such data) and some analysis on the new datasets. Especially interesting is the analysis of the “death” of one of the marketplaces. Overall, I believe that this work can be improved and could be on the path to an acceptance, but considerable changes are needed, especially in terms of writing, including comparing the results to related work, clarifying the novelty, differentiating between the markets, and cleaning up/clarifying the descriptions in the text.

* Missing related work:

First, a very relevant related report[a] is missing, which looks at Cannazon (and 2 other marketplaces) at the beginning of the COVID-19 pandemic. A comparison to this work is necessary.

* Novelty

The paper states that one thing this paper addresses is that increased "security measures have been added such as CAPTCHAs that are more difficult to resolve and other similar mechanisms that make crawling more difficult." However, the crawler presented does not address this. CAPTCHAS and logins are resolved manually. The paper could more clearly state the advantages of their crawler and how it differs form others, because the novelty of the crawler is not clear.

* Analysis

The markets analyzed are quite distinct, especially in terms of products and geography, but there is no per market analysis.

* Textual issues

Overall, the paper is quite poorly written, and should receive a heavy copy-edit.

There are some factual issues with the descriptions of the dark-web, deep web, and the tools used to access the "dark-web". The paper starts by stating that almost the entire Internet is the deep web, which is technically true, but not relevant and written in a way that may be misleading. The "dark-web," as the focus of this paper, only makes up a small percentage of the deep web, which is dominated by more mundane content: university web mail, banking, government or corporate intranet, etc. Even then, the "dark-web" is not completely dominated by illegal activity [b, c, d]. As written, a naive reader may assume that 96% of the internet is illegal activity. A few sentences later, again, the paper states that the Deep Web itself provides anonymity. The authors are conflating a tool like Tor which allows anonymous access to the service-web as well as the "dark-web" (recently more often referred to as ".onion sites", previously "hidden services"). Tor does also provide .onion se

rvices themselves with anonymity, hiding the IP from the visiting user, but that is independent. Tor provides the same anonymity to a user accesses Facebook.com as a user visiting their onion "dark" address: facebookwkhpilnemxj7asaniu7vnjjbiltxjqhye3mhbshg7kx5tfyd.onion. It may be that these issues arise from trying to make the descriptions simple, but it is important to remain on the accurate side of the thin line and not oversimplify to the point of saying something that is incorrect or misleading.

Similarly, the description of Tor (please also note the capitalization) is very brief and gives an overly simplistic view of the system. It should either be correctly described or omitted.

There are some normative statements about drug sales vs other activities, which are out of place in the paper.

Should "good learning" be "Deep learning" in this sentence: "even crawlers supported by artificial intelligence for good learning"

The statement: "A similarity between all these markets is that the most prevalent category is drugs, representing 80% of the total number of products for sale" does not line up with the results presented later in the paper (table III: 61k/123k = 49.5%). Is this 80% true for some forums and not others?

* Ethics

Was IRB (or similar) approval obtained before this research was conducted? Was the Tor Research Safety Board consulted? What ethical considerations were taken into account when conducting this research? Will the datasets be released? How?

* Minor comments

The data is figure 2 would better be displayed as a bar chat. The pie chart implies that there are no other products aside from these. I'm also not entirely sure what data it is showing: "the number of sellers willing to offer their products, and subsequently, in the number of products in each market." Is the number of products the raw number of things for sale, the variety ("wider catalog"), or the sellers?

* Bib:

[a] https://www.emcdda.europa.eu/system/files/publications/13042/EMCDDA-report_COVID19-darknet-final.pdf

[b] Mirea, Mihnea, Victoria Wang, and Jeyong Jung. "The not so dark side of the darknet: A qualitative study." Security Journal 32.2 (2019): 102-118.

[c] Biryukov, Alex, et al. "Content and popularity analysis of Tor hidden services." 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, 2014.

[d] https://blog.torproject.org/tor-80-percent-percent-1-2-percent-abusive/

***Review 4***

**SCORE:** 0 (borderline paper)

This paper presents a collection of up-to-date datasets of six illicit marketplaces on dark webs and then uses the collected data to draw characteristics and trends of those marketplaces. The authors also conduct a longitudinal analysis specifically on a productive market, Cannazon, to provide interesting insights on the last days of the market and its evolution, notably through a DDoS attack which leads to the market closure, explained by the loss of trust from buyers on the market. Overall, I found the dataset itself and the paper quite interesting and easy to follow; and it nicely fits the scope of the workshop. The dataset is great; there are a lot of analyses and insights, but they miss a number of important work in the literature review which also analyses large and up-to-date datasets of dark web marketplaces. I also have some concerns about the reliability of the methodology they used; ethics; and the way they share the data with other researchers.

I would probably push for an "Accept" if the author could appropriately address my concerns and questions below during the rebuttal phase, otherwise, I will go for a "Borderline".

1. The scraper:

+ That is great to see around one page describing how their scraper works in detail. There are two separate modules in the scraper: (1) to index the URLs and put them into a list (2) to visit the list, crawl the page and remove the collected URL. However, the delay gap between the executions of these two functions is not clear: Ideally, they should be run in parallel to avoid some products being removed during the waiting time,

but to my understanding, the authors run it in turn (i.e., function 1 then function 2), so how many hours to wait for the function 1 to finish before function 2 starts? Please explain that, because it will clarify how complete the dataset is and to make sure the scraper can capture all of the listings.

+ The authors mention that the crawler is able to deal with modern anti-scraping techniques, but it is not clear what they are. I understand CAPTCHA puzzles on dark webs can be very frustrating (maybe even more challenging than reCAPTCHA or hCAPTCHA as we can't use cookies and PrivacyPass on dark webs), but can the authors give some examples of different types of CAPTCHA that they have encountered, and how often it appears for each marketplace during the data collection process? it appears that the authors solve the CAPTCHAs manually, so it is particularly important for audiences to estimate the human effect that needs to be paid for collection and to what extent the process can be automated, which affects long term maintenance (i.e., if it involves a lot of human interventions, it is fairly expensive to have a man just to sit in front of the computer to solve CAPTCHA). It has been known that some kinds of text CAPTCHA can be machine-solvable (for example, see [1]), so it may

be worth examining those solvers to make the effort more convincing, for example, "the CAPTCHA we have seen on dark webs cannot be solved by any popular methods that we have tried, so we have to do it manually" would strengthen the author's contribution.

+ The authors run the scraper in parallel for 5 over 6 marketplaces, but it is not clear how many processes they used. An ethical issue that the authors should bear in mind when running things in parallel is that: don't flood the targeted servers and don't put lots of pressure on the Tor network (which is already very slow).

+ Does the scraper automatically resume upon crashes? What is the mechanism e.g., storing the progress as the scraper goes along? What is the plan to maintain the dataset? Would it be for a long time?

2. The dataset:

+ The dataset itself is great, and the details of fields collected are good and clear explained. However, the collection period is not clear, which is essential to tell how up-to-date the dataset is. The authors claim it is more up-to-date than the existing ones, but without the collection period, I could not say how up-to-date it is. Please see [3] and [4], they also collect an up-to-date dataset and they do seem to maintain it regularly. How do you compete with them in terms of "up-to-date"? That saids, there is a lack of a survey of available dark web marketplace datasets available, and comparisons between the authors' data and existing data to highlight what is the advantages, what is similar, and what is distinct.

+ Without completeness guaranteed, we can't say anything about trends and longitudinal analyses. So how do the authors ensure the completeness of the data? Is there any sanity check for completeness? The scale of the data collection here means that we never know what will be going on, and even when you watch and see that the crawler is working just fine, it is still hard to guarantee completeness because the scraper may have bugs at any time due to an unknown reason.

+ Do they share the data with other researchers? The data itself may contain sensitive information, but it is still worth sharing for reproducibility purposes. If the authors plan to share the data, they have to pay attention to the data-sharing regime to make it works across different jurisdictions such as the US, the EU, and China. The sharing agreement must be comprehensive enough to avoid misconduct.

+ The authors store the data in an SQLite database, but I wonder if the database is stored on an encrypted hard disk? It is important to make sure the sensitive dataset won't be leaked even when the hard disk is stolen, for example.

3. The literature review

+ When the authors talk about the crawler architecture, it is worth mentioning this paper [7] which details a very good picture of how they have built a robust and sustainable cybercrime data collection system.

+ There is a number of important related work missing e.g., recent work from Damon McCoy [3] and [4], where they also use lots of data from dark web marketplaces to analyse the effect of the pandemic. And these data are also up-to-date. Please have a look.

+ When you look at the evolution of marketplaces over time, it is also worth looking at [5], where they analyse the evolution of a different type of marketplace on HackForums.

+ When you talk about how the market got ruined by the lack of trust between the users, you should appropriately cite this paper [6], which is about the market's quality uncertainty and is one of the fundamental theories of the market mechanism.

4. The analysis:

+ For the dataset, did you conduct some statistical tests to see if any noise or if a vendor creates fake products and leaves fake ratings to gain reputation? i.e. if they fake the numbers, the dataset will not be in a normal distribution.

+ In IV B, using the regex for preprocessing sounds reasonable, but when they convert from dollar to euros, what conversion rates are used?

+ In figure 2, IV C, how do you count the number of products? it's not as straightforward as normal counting: some products may appear to have a similar name e.g., two vendors advertise the same product but with different descriptions. How can the authors map them all together? Or maybe they just count the pure number?

+ Type of products: grouping by product type is good, and breaking them into sub-categories is also interesting. The authors use the 'category' field from the database, but it is unclear how the 'category' is identified. If they collect it from the marketplace, different marketplaces may have different category names. So how did you unify them? It is interesting to see fake covid certificates and vaccination proof, which is also found in existing work [3] and [4].

+ They analyse the cross-market presence of vendors by comparing nicknames. Yes I agree that it is likely users use the same nickname, but to make it more robust, they will need to add another constraint to the nickname e.g., the n-gram of that nickname is rare enough, see the paper here [2].

+ Information about ratings of sellers and the number of ratings in Table IV can be unreliable e.g., a vendor creates a bunch of fake accounts to buy stuff themself and give good ratings in order to get more reputation. I think this problem should not be overlooked and should be addressed in the paper.

+ Geography of the sales: Some vendors indicate "Ships from" and "ships to" information. "Ship to" may be credible, but "ship from" may not be always reliable because criminals will tend to hide their actual location. This is important and should be addressed because using that unreliable information may be cause serious inaccurate measurements. Maybe interested to break them by countries as well?

+ I really like the analysis of the effect of the DDoS and the last days and how the market structure changed and lead to a closure, the relationship between different variations in the stock and prices of the products with the attacks and the close of the site. which may be the most interesting part of the paper. The finding that the number of products added was affected by the attacks is interesting. However, they only "extended the data collection by the crawler for one particular case" on a weekly basis", why don't they do it with the other marketplaces? Is that because a lot of human effort needs to be done to solve CAPTCHAs, as I mentioned above? If you could collect it more frequently, you may have captured the marketplace right before the DDoS attacks and after the DDoS attack, which will be very useful. Please clarify that point. In

section V, when comparing the number of products between different snapshots, they said, "it is reasonable to conclude that the products

were removed due to them being sold". It is not convincing to me. Vendor can always take down their listings without any reason, and it does not necessarily need to be referred to as "being sold".

5. Ethical considerations

+ There is a huge lack of ethics discussion here. They do mention in Table IV that they anonymise the seller's name for ethical reasons, it is good, but to me, it is not enough. Dealing with that sensitive data means that the author will need to seek approval from their institution for both data collection and data usage before carrying out the research. Please pay attention to it.

[1] Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach.

[2] What's in a name? An unsupervised approach to link users across communities.

[3] Dark Web Marketplaces and COVID-19: before the vaccine.

[4] Dark web marketplaces and COVID-19: the vaccines.

[5] Turning Up the Dial: The Evolution of a Cybercrime Market Through Set-up, Stable, and Covid-19 Eras.

[6] The market for "lemons": Quality uncertainty and the market mechanism.

[7] Crimebb: Enabling cybercrime research on underground forums at scale.