AFRL-RI-RS-TR-2022-024

MEASURING AND ANALYZING ONLINE ANONYMOUS (`DARKNET') MARKETPLACES

CARNEGIE MELLON UNIVERSITY

FEBRUARY 2022

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2022-024 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
TODD N. CUSHMAN
Work Unit Manager

/ S /
JAMES S. PERRETTA
Deputy Chief
Information Warfare Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
FEBRUARY 2022	FINAL TECHNICAL REPORT	SEPTEMBER 2020	SEPTEMBER 2021

4. TITLE AND SUBTITLE

MEASURING AND ANALYZING ONLINE ANONYMOUS (`DARKNET') MARKETPLACES

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
N/A	FA8750-20-1-1003	Other
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
DHS2	0C	MU

6. AUTHOR(S)

Nicolas Christin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Carnegie Mellon University
5000 Forbes Ave
Pittsburgh PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Air Force Research Laboratory/RIGA 525 Brooks Road Rome NY 13441-4505	RI	AFRL-RI-RS-TR-2022-024

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We present the results of the project "Measuring and Analyzing Online Anonymous (`Darknet') Marketplaces," carried out at Carnegie Mellon University at the behest of the Department of Homeland Security, Science & Technology Directorate. This project was a follow up to an earlier project ("A Queryable Platform for Online Crime Repositories"). We continued to avail to other researchers data from 12 dark web marketplaces, corresponding to over 22,288 vendors, 348,400 items, and 5,826,115 transactions. The data was availed through 1) a publicly available website (for anonymized data), and 2) a set of databases provisioned through the IMPACT portal (for anonymized and de-anonymized data). For the duration of the project at hand, the IMPACT portal served an additional 19 distinct requests for data from 8 academic institutions (academic, industry, government) in the US, the Netherlands, the UK, and Singapore. Combined with the previous contract, the project has served 69 requests from 25 institutions over six countries (US, Japan, Singapore, Netherlands, UK, and Australia). For the duration of the present contract, serving this data has led in particular to the publication of a paper — by third parties—based on our data, at WEIS 2021. In addition, our research has led to the development of a paper (under revision at the time of writing), to be submitted to a leading computer security conference in 2022.

15. SUBJECT TERMS

Dark web, measurements, IMPACT

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
U	U	U	SAR	15

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
TODD N. CUSHMAN	**N/A**

PREVIOUS EDITION IS OBSOLETE.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 Summary

We present the results of the project "Measuring and Analyzing Online Anonymous (`Darknet') Marketplaces," carried out at Carnegie Mellon University at the behest of the Department of Homeland Security, Science & Technology Directorate. This project was a follow up to an earlier project ("A Queryable Platform for Online Crime Repositories"). We continued to avail to other researchers data from 12 dark web marketplaces, corresponding to over 22,288 vendors, 348,400 items, and 5,826,115 transactions. The data was availed through 1) a publicly available website (for anonymized data), and 2) a set of databases provisioned through the IMPACT portal (for anonymized and de-anonymized data). For the duration of the project at hand, the IMPACT portal served an additional 19 distinct requests for data from 8 academic institutions (academic, industry, government) in the US, the Netherlands, the UK, and Singapore. Combined with the previous contract, the project has served 69 requests from 25 institutions over six countries (US, Japan, Singapore, Netherlands, UK, and Australia). For the duration of the present contract, serving this data has in particular led to the publication of a paper — by third parties—based on our data, at WEIS 2021. In addition, our research has led to the development of a paper (under revision at the time of writing), to be submitted to a leading computer security conference in 2022.

2.0 Introduction

This project stemmed from years of research in collecting large amounts of data from several types of online illicit activity, ranging from illicit sales of pharmaceutical drugs, to online anonymous marketplaces. We have strived to make our data publicly available for others to use, which presents a number of challenges.

First, continuously maintaining these data repositories requires significant software engineering: continuous maintenance of parsers, scrapers, and other data collection primitives in the face of changes in the environment. Second, simple data "dumps" are quite difficult to use, and do not lend themselves to rapid hypothesis testing—this is particularly true for researchers in social sciences (e.g., criminology, economics) who might otherwise have great interest in our data, but not necessarily the ability to write complex, low-level database queries.

The object of this work was to continue the work originated in the project "A Queryable Platform for Online Crime Repositories." Specifically, as part of this original project, we had built and deployed query-able online platforms for our online crime repositories. In this new project, our main tasks were to 1) continue maintaining our infrastructure to be able to continue serving data to potential customers; and 2) to continue fulfilling requests for data from other researchers. These maintenance tasks were the bulk of the work supported in this new project, and led to measurable dissemination of our results.

In addition, we proposed two research tasks. First, we offered to verify the quality of the measurements we have obtained, to give assurance to IMPACT customers the data provided is useful.

Second, we were interested in investigating behaviors at a vendor level. We wanted to understand how specific vendors move across marketplaces, whether they diversify their offerings or, on the other hand, simply list all their wares across several marketplaces. Ultimately, we were interested in characterizing various vendor "types," with different economic activity.

While the work was primarily infrastructural, we have seized the opportunity provided by this grant to make scientific advances. A paper based on our scientific work is currently in preparation.

The next sections describe our data sharing and research (Section 3), a summary of the results of this grant (Section 4), as well as plans for future work (Section 5).

3.0 Methods, Assumptions and Procedures

3.1 Web interface

A critical aspect of this overall work (dating back to our original project on "A Queryable Platform for Online Crime Repositories") was to provide access to the larger community to our datasets of online crime, specifically, dark web marketplace data. After consultation with our Institutional Review Board, we provided access through two media: a publicly accessible website, which only featured *anonymized* data, and a set of databases available both in anonymized or de-anonymized format to researchers; the latter (de-anonymized) required signature of a Memorandum of Understanding between the interested researchers and Carnegie Mellon University. This process was facilitated through the IMPACT cyberportal (https://www.impactcybertrust.com).

Our data sharing eschews potential thorny human-subject issues, as all the data we collect are already publicly available, and should not contain private identifiers (e.g., no IP addresses). We obfuscate textual content that could plausibly contain (unidentifiable) contact information such as "throw-away" email addresses used by questionable businesses. We worked with our IRB and general counsel to ensure the soundness of our proposed approach. The IRB confirmed this was not human subject research; legal indicated specific language to be used on our website.

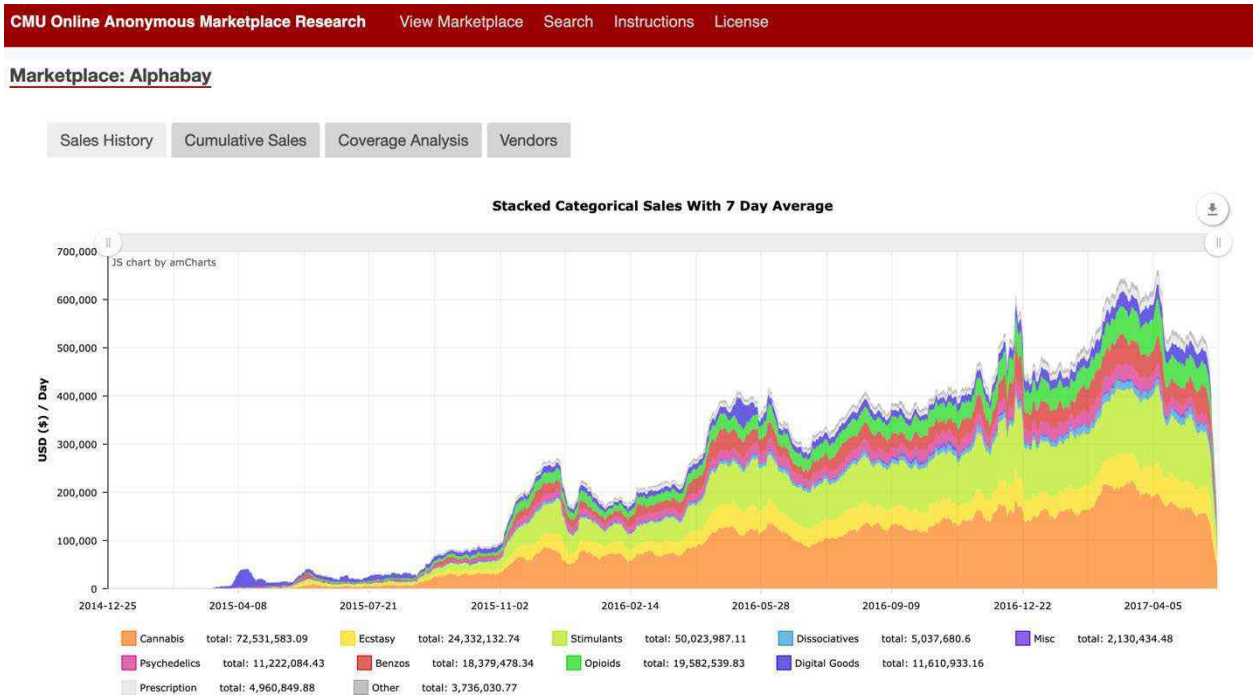

Figure 1: Online portal (Example of marketplace aggregate data)

Figure 1 gives a glimpse of the online portal we developed.

For 12 different dark web marketplaces (Agora, Evolution, Silk Road, Silk Road 2, Black Market Reloaded, Pandora, Hydra, Alphabay, Dream Market, Traderoute, Berlusconi Market, and Valhalla), we provide panels that allow users to quickly plot the amount of sales (broken by category, e.g. "Cannabis") over time. Figure 1 shows the entire timeline for the Alphabay marketplace, in which we plot the amount of daily sales on the market, over two years. Users can move the slider to "zoom in" on specific time periods.

By clicking on "Vendors," users are redirected to a page listing all of Alphabay's vendors.

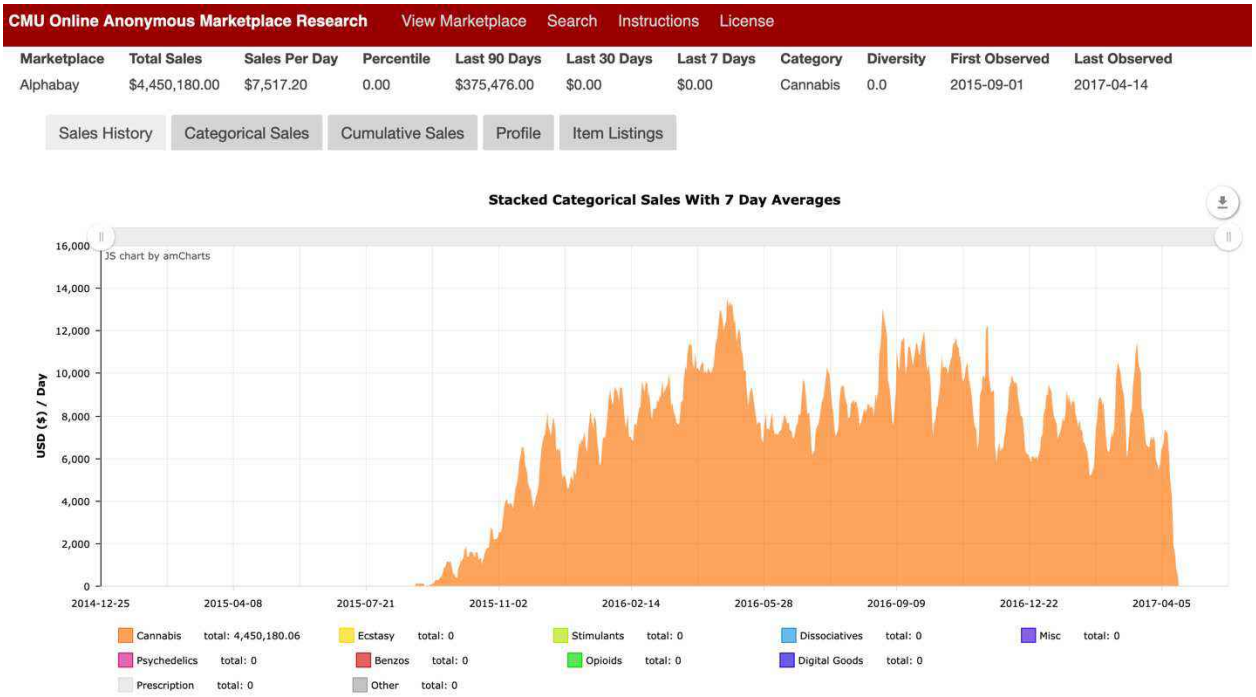

Figure 2: Online portal (Example of vendor data)

Subsequently, users can obtain more detailed information simply by clicking on a vendor's record. This, in turn, leads them to a page as depicted in Figure 2. In this case, the vendor was specializing fully in cannabis, and sold approximately 4 million USD of such goods over their lifetime on Alphabay.

The data and software were previously provided to AFRL at the conclusion of the original project ("A Queryable Platform for Online Crime Repositories"). The main task in this iteration of the project was to 1) maintain the infrastructure – which was successfully done, as the website continues to operate and serve users and 2) continue provision data for the IMPACT portal.

3.2 Maintenance activities

We continuously maintained the https://arima.cylab.cmu.edu/markets website throughout the project, and we did not experience any significant downtime during the course of this project.

In particular, we managed to migrate to a new hardware server and a new co-location facility in early 2021 with very limited downtime – we kept the old machine online until the new machine was perfectly stable, and then performed a simple DNS switch. In the process we upgraded the operating system on which the server was running, as well as the versions of the software (php, Apache, mysql) powering the website.

The new machine is hosted at Carnegie Mellon's main server room, and as such, benefits from significant upgrades (See: https://www.cmu.edu/computing/services/infrastructure/server/physical-colo/index.html):

- Emergency power by generator and Uninterruptible Power Supply (UPS)
- Redundant power distribution in the racks
- Environmentally-controlled space
- Fire suppression
- Restricted access for authorized users

3.3 Ongoing research activities

As discussed above, this grant has primarily been used for infrastructural maintenance, and to assist other researchers willing to use our data. We have, however, complemented this infrastructural work with research activities using our existing data. This research is unpublished and currently under submission; we thus outline it in broad strokes to maintain the confidentiality of the submission and review process.

We have established a partnership with colleagues at TU Delft (unfunded by this grant), who themselves have partnered with the National Dutch Police. The National Dutch Police has consented to grant our partners (and ourselves) access to an image of a seized dark web marketplace. This is invaluable as it provides us with unfettered access to "ground truth," against which the external measurements we have been collecting can be compared.

Specifically, we build a framework to reason about online anonymous marketplace data collection and projections. We mathematically define a model to express possible sources of inaccuracies in online anonymous market measurements: scraping errors, missing data, estimation errors (e.g., statistical projection error ranges).

Building this framework is important to understand data collection limitations. Importantly, even police forces that seize a market server do not necessarily have perfect information about all the activities of the marketplace: they have merely a snapshot of what the server looks like at seizure time, but they may not necessarily have access to historical data. For instance, if the server operators routinely purge older data (as was the case with Dream Market, in which user reviews older than six months were deleted), a server seizure would provide an incomplete picture of the data available.

We then use this mathematical model in the context of a case study with the market data provided to us by our partners. We also simulate marketplaces based on the data provided to generalize our findings beyond that of the case study.

4.0 Results and Discussion

4.1 Website use and data sharing

All in all, our website provides data about:

- 12 different marketplaces as noted earlier
- 22,288 vendors,
- 348,400 items,
- and 5,826,115 transactions.

We believe this is the most comprehensive publicly available archive of dark web data. The website is available at https://arima.cylab.cmu.edu/markets/. Code to generate the website was provided as part of the software bundle included in the final report for of the original project ("A Queryable Platform for Online Crime Repositories").

In addition to this website, we provided data through the IMPACT cyberportal. In this project period, we served 19 distinct requests for data from 8 academic institutions in the US, the Netherlands, the UK, and Singapore. Combined with the original, the project has, in total, served 69 requests from 25 institutions over six countries (US, Japan, Singapore, Netherlands, UK, and Australia). Figure 3 shows a partial list of three requests that were honored.

Besides IMPACT data, at the time of this writing, the website https://arima.cylab.cmu.edu/markets, which provides anonymized data in the interface described above, at no cost, serves on average 6 to 7 pages per day, excluding bots and automated collection engines.

DSR-6910	Dataset Request	[Restricted] Carnegie Mellon University record 'Dream, Traderoute, Berlusconi and Valhalla marketplaces, 2017-2018: Non-anonymized datasets' requested by [redacted] from Maastricht University, Faculty of Law	Approved (receipt not confirmed)	2021-02-25	2021-03-26	TBD
DSR-6939	Dataset Request	[Unrestricted] Carnegie Mellon University record 'Dream, Traderoute, Berlusconi and Valhalla marketplaces, 2017-2018: Anonymized datasets' requested by [redacted] from Temple University	Approved	2021-04-18	2021-04-21	TBD
DSR-6958	Dataset Request	[Unrestricted] Carnegie Mellon University record 'Dream, Traderoute, Berlusconi and Valhalla marketplaces, 2017-2018: Anonymized datasets' requested by [redacted] from National University Of Singapore	Approved (receipt not confirmed)	2021-06-02	2021-06-04	TBD

Figure 3: IMPACT Cyberportal request example (administrator view)

This has yielded a number of additional important publications *by other researchers not affiliated with our group*. In particular, in this project period, Scott Lee Chua presented at WEIS 2021 a paper entitled "Measuring the Deterioration of Trust on the Dark Web: Evidence from Operation Bayonet" (https://weis2021.econinfosec.org/wp-

content/uploads/sites/9/2021/06/weis21-chua.pdf), which extensively uses the data made available on IMPACT.

4.2 Results in scientific papers in preparation as part of this grant

As discussed above, this grant has primarily been used for infrastructural maintenance, but we have been working on a mathematical model to evaluate the accuracy of external data collection. As noted, the paper is currently under submission (and potential revisions), which prevents us from discussing the results in great depth.
However, preliminary findings show that:

1. Scraping often and repeatedly is necessary to achieve sound coverage. On the other hand, "one-shot" or "two-shot" scrapes result in missing out on a majority of the data. This echoes the findings of our previous work [7].

2. Abundance estimators – e.g., Jolly-Seber or Schnabel estimators – rest on a number of assumptions about population growth, that may not hold in practice and greatly bias the projections. For instance, the Schnabel estimator assumes that population is stationary: no new births, and no new deaths. This is unlikely with dark web marketplaces where pages, reviews, items, and vendors come and go, and results in estimates whose quality decreases over time.

   The Jolly-Seber estimator removes that assumption, but requires a model for births (i.e., arrival of new data) and deaths (i.e., deletion of data), which may not be readily available. Empirically, using multiple estimators in parallel is a potentially viable strategy to get a sense of actual coverage; hard bounds are practically difficult to obtain.

We are currently working toward publishing these findings. The paper is expected to be finalized in 2022, post-completion of the grant.

5.0 Conclusions

As we have shown, in addition to being a lynchpin of a number of research efforts at Carnegie Mellon, our work under IMPACT has facilitated considerable additional research *by others*, as evidenced by the numerous downloads of our data and their actual use in publications. In that respect, the main objective set forth by the proposal was met.

We were also able to investigate important research questions on statistical analysis of our data, notably with respect to better understanding the reliability of estimates based on external (scraping) measurements. Those have not been fully completed during the period of performance of this contract but will likely be published later in 2022.

6.0 References

[1] European Monitoring Centre for Drugs and Drug Addiction and Europol. Drugs and the Darknet: Perspectives for Enforcement, Research and Policy. EMCDDA–Europol Joint publications, Publications Office of the European Union, Luxembourg. November 2017.

[2] Malte Möser, Kyle Soska, Ethan Heilman, Kevin Lee, Henry Heffan, Shashvat Srivastava, Kyle Hogan, Jason Hennessey, Andrew Miller, Arvind Narayanan, and Nicolas Christin. An Empirical Analysis of Traceability in the Monero Blockchain. To appear in Proceedings of the Privacy Enhancing Technology Symposium (PETS 2018), volume 3. Barcelona, Spain. July 2018.

[3] Rolf van Wegberg, Samaneh Tajalizadehkhoob, Kyle Soska, Ugur Akyazi, Carlos Gañán, Bram Klievink, Nicolas Christin, and Michel van Eeten. Plug and Prey? Measuring the Commoditization of Cybercrime via Online Anonymous Markets. In Proceedings of the 27th USENIX Security Symposium (USENIX Security'18). Baltimore, MD. August 2018.

[4] Xiao Hui Tai, Kyle Soska, and Nicolas Christin. Adversarial Matching of Dark Net Market Vendor Accounts. To appear in Proceedings of the 25th ACM SIGKDD Conference of Knowledge, Discovery, and Data Mining (KDD'19). Anchorage, AK. August 2019.

[5] James E. Arps and Nicolas Christin. Open Market or Ghost Town? The Curious Case of OpenBazaar. To appear in Proceedings of the 24th International Conference on Financial Cryptography and Data Security (FC'20). Kota Kinabalu, Malaysia. February 2020.

[6] Nicolas Christin. After the Breach: The Monetization and Illicit Use of Stolen Data. Testimony before US Congress, March 2018.
http://www.andrew.cmu.edu/user/nicolasc/publications/20180315-testimony-christin.pdf

[7] Kyle Soska and Nicolas Christin. Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem. In Proceedings of the 24th USENIX Security Symposium (USENIX Security'15), pages 33-48. Washington, DC. August 2015.

LIST OF ACRONYMS

EMCDDA	European Monitoring Centre for Drugs and Drug Addiction
IMPACT	Information Marketplace for Policy and Analysis of Cyber-risk & Trust
IP	Internet Protocol
IRB	Internal Review Board