Inconsistencies in Darknet Researches

Florian Platzer[1,2,*] and Alexandra Lux[3,4]

[1] *Fraunhofer SIT, Germany*
[2] *ATHENE, Germany*
[3] *TU Darmstadt, Germany*
[4] *University of Hohenheim, Germany*
E-mail: florian.platzer@sit.fraunhofer.de; alexandra.lux@sit.fraunhofer.de
[*] *Corresponding Author*

Received 12 December 2022; Accepted 14 December 2022;
Publication 28 April 2023

Abstract

The darknet terminology is not used consistently among scientific research papers. This can lead to difficulties in regards to the applicability and the significance of the results and also facilitates misinterpretation of them. As a consequence, comparisons of the different works are complicated.

In this paper, we conduct a review of previous darknet research papers in order to elaborate the distribution of the inconsistent usage of the darknet terminology. Overall, inconsistencies in darknet terminology in 63 out of 97 papers were observed. The most common statement indicated that the dark web is a part of the deep web. 19 papers equate the terms darknet and dark web. Others do not distinguish between dark web and deep web, or between deep web and darknet.

Keywords: Darknet, dark web, Tor, terminology, review.

Journal of Cyber Security and Mobility, Vol. 12_2, 187–204.
doi: 10.13052/jcsm2245-1439.1223

1 Introduction

In research papers, but also in media reports, various darknet terms such as *darknet*, *dark web* or *deep web* are used inconsistently. Several studies use the term *darknet* and *dark web* identically. Others do not distinguish between *dark web* and *deep web* or describe the *dark web* as a part of the *deep web*. This leads to an incorrect understanding of what the darknet is. The inconsistent use of terminology often makes it difficult to compare the results of research studies. Furthermore, this can lead to deficient conclusions, which are partially reflected in the media.

In a previous work [43], we elaborated six critical aspects that inconsistencies in darknet research may exhibit. This was done by focusing on the terminology used, the methodology of sample collecting and analyzing the data. Through the implications of these aspects, presented descriptions of darknets do not necessarily reflect the actual nature of darknets. This current paper addresses these critical aspects with a focus on the inconsistent use of darknet terminology. In order to gain a better understanding of how widespread the inconsistent use of darknet terminology in research papers is, we conduct a literature review. We analyze past research papers for darknet terms used and examine the context in which they are applied or whether their authors have provided their own definitions.

In total, we found 945 darknet relevant papers. In order to reduce the sample size for a qualitative analysis, we analyzed scholarly papers that conducted a primary focus on research on web services on the Tor network. Of these 227 papers, we identified 97 papers that used darknet terms. Overall, we found inconsistencies in darknet terminology in 63 out of 97 research papers. The most common is the view that the dark web is a part of the deep web. 19 papers equate the terms darknet and dark web.

2 Background

This paper is related to one of our earlier studies [43], which focused on the Tor network. In the following, we briefly introduce Tor. Afterwards, we summarize our previous study in order to understand the issue on which the current work is based on.

2.1 Tor Network

When darknet technologies are considered, the Tor network is often mainly focused on, as the Tor network is the most popular and well-known anonymizing network [2, 12, 45, 46].

Internet services such as SSH, FTP, email, chat or web services can be offered anonymously via the Tor network. Anonymity is achieved through the *onion routing* implemented in Tor. This ensures that traffic is encrypted multiple times and sent over a path consisting of multiple Tor nodes on the Tor network. There are several thousand Tor nodes operated by volunteers distributed around the world. Each Tor node in the path knows only its immediate predecessor and successor. None has knowledge about the entire path. This means that the source node is never directly connected to the destination node. Tor allows users to anonymously access Internet services outside the Tor network, as well as anonymously access or host certain services within the Tor network. [53]. The services within this network are called *onion services* (formerly, hidden services). In order to access such a service, the respective *onion address* is required.

2.2 Critical Aspects for Darknet Research on Related Work

In our previous work [43], we used existing literature to identify critical aspects for darknet research. We identified six aspects which we describe subsequently:

1. Inconsistent use of terminology. In the literature, terms such as *darknet*, *dark web* or *deep web* are not used consistently. Some works do not distinguish between *darknet* and *dark web* and use both terms identically. Others do not distinguish between *dark web* and *deep web*, or use terms that in fact have a different meaning. This leads to misunderstandings of e.g. research papers and makes it difficult to compare results. In addition, it leads to the interpretation of results in inaccurate contexts. Inconsistent use of these terms is also mentioned in other works [21,26,53].

2. Gathering method of onion addresses. In order to solve a research task regarding services offered on a darknet – in the specific case on the Tor network – a valid data gathering strategy is important. In past literature, two methods have been used for this purpose. The first is the use of a web crawler. However, a web crawler can only find services that are listed on the respective web pages. These are mainly links to other web pages. Services such as SSH, FTP, email or IRC chats are not found easily by this method. For a web crawler, a good selection of starting points (seed pages) is important. Many works use as seed a lists compiled by third parties [1,2,6,8,14,18,31,36,50, 54], existing darknet search engines [1,2,18,31,50,54], or onion lists from previous works [2,4].

The second approach is to actively set up Tor nodes. Tor onion services publish their reachability information to particular Tor nodes (called directory servers) and sign their data with their public keys. These keys represent the onion addresses of the servers. This provides the possibility to collect existing onion addresses by reading the published information. However, this method takes some time to get a good overview of all services. Hence, this approach overrepresent short-lived services because they are not all online at a single point in time, but over a period of time.

The way in which a sample is compiled is crucial for research, as it serves as the basis for subsequent analysis. Depending on which method is used to collect onion addresses, this affects the number of the onion addresses found. Therefore, e.g. the size of Tor or the number of onion addresses found must be reflected in terms of the methodology used to compile the sample.

3. Short lifetime of services. Many services exist only for quite a short time. This is observed by many works that collect onion addresses on web pages [1,4,23,54], or by collecting via directory servers [12,35]. A large part of the onion addresses found are often no longer accessible, because the web services no longer exist. Many onion services exist for only a few days, hours or even minutes [36]. When evaluating these collected services, it can lead to an overrepresentation of the short-lived services [36]. This can lead to the total number of existing services being much lower than the number often given in studies.

4. Botnet command and control servers. Due to their anonymity, darknets are popular for botnet infrastructures [3]. Some work show that at certain times up to 50% of all existing onion services belonged to botnet command and control (C&C) services [11,49]. Botnets can therefore represent a large fraction of all darknet services. Thus, the total number of available onion addresses does not necessarily represent the number of unique onion services in the Tor network.

5. Web services with undetermined content. In the past, various works have investigated websites on the Tor network. The content of the websites and their distribution was analyzed. However, many websites possess undetermined content. These include websites that contain less than 20 words, websites that displayed only images, error messages, or a default web page of any service [1, 2, 4, 11, 14, 31, 46, 48]. Also completely empty websites or with unreadable text or words were excluded from the analysis or were assigned to categories such as *none* or *empty*. Additionally, it is common to analyze only websites in English [1, 2, 4, 11, 50]. The exclusion criteria may be conditioned, for example, by the method used to analyze the sample. This should be considered especially in relation to the results and the interpretation of the work.

6. Duplicates of onion services. When evaluating websites, it is important to analyze all found websites for duplicates. Past work has shown that up to 51% of all analyzed websites on the Tor network are duplicate websites [54]. In particular, websites in the category *drug* and *cryptolocker* are faked and operated as phishing websites [2]. But marketplaces are also frequently duplicated. For example, the former Tor darknet marketplace *AlphaBay* was offered 165 times in a fake way under different onion addresses [6].

In a case study we conducted [43], it was shown that darknet marketplaces also have multiple mirror sites. A total of 37 darknet marketplaces were examined. Only 12 marketplaces were accessible under a single authentic onion address. All others had several up to 20 authentic onion addresses.

It is important to consider duplicates of onion services in the initial categorization. Not considering this aspect will lead to biased reporting of e.g. web service content distribution. Furthermore, this can lead to a crime and threat situation that does not accurately reflect the reality of Tor, as the majority of suspicious onion services tend to have duplicated copies [2].

The study described above presents aspects that show inconsistencies in darknet research. These can lead to errors and misunderstandings of the results regarding the Tor network. The inconsistent use of darknet terminology itself can place well-crafted results in the wrong context and lead to misinterpretation. The past work considered only a limited selection of papers. This paper complements the work with a scoping review to gain a more accurate dissemination of the inconsistent use of darknet terminology.

3 Methodology

We are interested in the use of darknet terminology in previous research. We want to find out in which contexts the respective terms are applied and whether they are used correctly. Therefore, we first elaborate the meanings of the darknet terminology and describe how they relate to each other. Subsequently, we describe our process for collecting research for a qualitative analysis.

3.1 Darknet Terminology

Within the Internet, various services such as SSH, FTP, email, chat or web services are offered. Each service is accessible under a specific port number. These can vary, but there are standardized default port numbers for each type of service. For example, FTP services are accessible under port number 20 and 21, and web services are accessible under port number 80 or 443. All web resources are referred to as the ***World Wide Web*** (WWW or Web). It comprises a huge collection of different documents typically linked with each other and are accessible with web browsers using HTTP(S) protocol [53]. Both users and servers are easily identified by their IP address [5].

The World Wide Web is divided into *surface web* and *deep web*. The ***surface web*** is the part of the World Wide Web that can be found and thus indexed by conventional web search engines such as Google, Bind or Yahoo. The remaining part is referred to as the ***deep web***. Contents in the *deep web* are therefore web resources to which web search engines do not have access, e.g. contents of web pages that require passwords or that are protected by other security measures [50]. All corporate web resources that are only provided internally are also part of the *deep web* [42]. Furthermore, there are websites that belong to the *surface web* from a technical point of view, but are not indexed by web search engines due to their poor findability, and thus belong to the *deep web* (e.g. unlinked and standalone websites [26]). There is no hard line between *surface web* and *deep web*. Individual web pages can belong to the *surface web*, but their contents belong to the *deep web*. An example of this are websites that present both static and dynamic content. Due to the fact that search engines are constantly expanding and changing their indexing capabilities, the amount of web content that belongs to both the *deep web* and the *surface web* is constantly changing [53].

Darknets are networks that use the infrastructure of the Internet and are therefore referred to as *overlay networks*. Those networks are organized decentrally and enable anonymous participation and communication [53]. Example technologies for a darknet are I2P, Freenet or Tor. Within these networks, Internet services or the sharing of files can be provided anonymously. All these services are available on the Internet but only accessible through the respective network [32]. Darknets ensure both a technically anonymous provision and a technically anonymous use of information and services.

The ***dark web*** is the part of a *darknet* that offers all web services i.e. HTTP(S) services [26] and are exclusively accessible via the respective darknet [37]. The relationship between the darknet and the dark web is comparable to the Internet and the World Wide Web.

Finally, the part of the Internet that does not belong to any darknet is called ***clearnet***.

3.2 Scoping Review

We selected research papers from the databases *Scopus*, *Web of Science* and *Ebsco*. We performed queries with the following search terms:

"Tor" or "Tor-Netzwerk" or "Tor-Network" or "Tor Netzwerk" or "Tor Network" or "Tor-Browser" or "Tor Browser" or "Darknet" or "Dark Net" or "Darkweb" or "Dark Web" or "Onion Router" or "Onion Routing". In order to also account for articles that used terminology inconsistently we further included "Deep Web" or "Deepweb".

In total, we received a set of 6,445 items from the three databases.

After removing duplicates, we excluded the following non-scholarly items:

- Proceedings summaries
- Conference reviews
- News items
- Books
- Book reviews
- Monographs
- Anthologies
- Film reviews
- Meeting abstracts
- Homonyms
- Editorials
- Erratum

The remaining 3,468 papers were examined for non-darknet-specific work. These are papers that exclusively address issues of onion routing (beyond the darknet context), or consider research related to the deep web. For this, we analyzed the title and abstract of each paper. The term *darknet* is also used in other areas. For example, in computer science in the area of network security, unused IP spaces of a local network are referred to as *darknet*. In image processing, a framework called *darknet* exists. These works have also been removed.

This results in a total number of 945 scholarly papers.

In order to perform a qualitative analysis, we reduced the sample. For this purpose, we analyzed all entries with respect to title, abstract and keywords for web-specific works in the Tor network. Papers that do not explicitly mention web content in the Tor network were removed. This also applies to papers that consider darknet web content but do not explicitly state that the subject is the Tor network. However, works whose descriptions imply the Tor network, such as the Tor darknet marketplace SilkRoad, have been retained. This results in a sample of 227 papers.

Figure 1 shows the workflow of the scoping review.

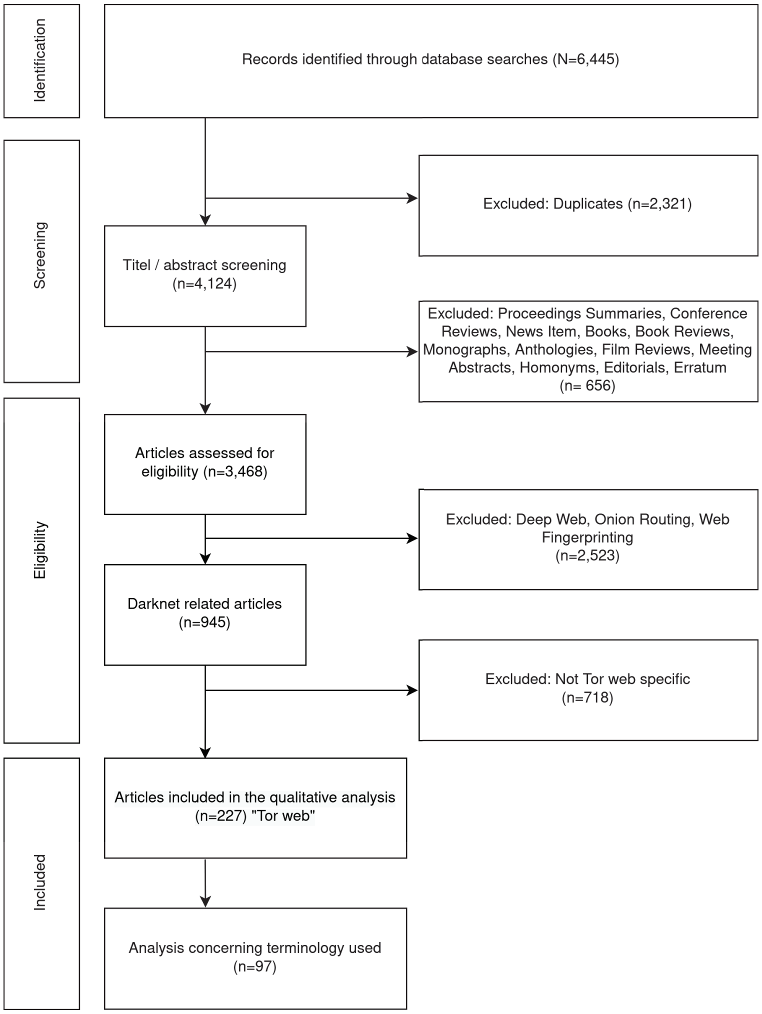

Identification
Records identified through database searches (N=6,445)
Screening
Excluded: Duplicates (n=2,321)
Titel / abstract screening (n=4,124)
Excluded: Proceedings Summaries, Conference Reviews, News Item, Books, Book Reviews, Monographs, Anthologies, Film Reviews, Meeting Abstracts, Homonyms, Editorials, Erratum (n= 656)
Eligibility
Articles assessed for eligibility (n=3,468)
Excluded: Deep Web, Onion Routing, Web Fingerprinting (n=2,523)
Darknet related articles (n=945)
Excluded: Not Tor web specific (n=718)
Included
Articles included in the qualitative analysis (n=227) "Tor web"
Analysis concerning terminology used (n=97)

Figure 1 Workflow of scoping review.

4 Results

Of 227 analyzed papers, we identified 97 that described darknet terms used in their work. All others either do not use any of these terms, or have not previously described or defined them explicitly. Table 1 shows that many papers (1) equate multiple terms and do not distinguish between them (=), (2) describe that one term is a subset of another term, or (3) use one term but actually describe another (→). It is possible that a paper may contain several discrepancies at the same time.

Some of the remaining 34 papers use inaccurate or superficial descriptions that are not necessarily incorrect. For example, that web pages in the Tor network (dark web) are known as hidden services [14, 15, 28]. But not all services in the Tor network have to be web services.

Many do not explicitly describe the difference between darknet and dark web. Since the dark web is a part of a darknet, many descriptions about the darknet also apply to the dark web. Thus, the dark web is described as

- "a part of the Internet that cannot be accessed by mainstream software" [17],
- "a part of the internet that cannot be reached through conventional means as it depends upon encryption techniques" [33],
- "an encrypted part of the Internet only accessible through specialized software such as the Tor web browser" [52] or
- "a part of the Internet that can only be reached through anonymisation software" [9].

The same applies to descriptions such as "Such web applications are called onion sites in the Dark Net" [27]. Which is not incorrect, since such web

Table 1 Inconsistent use of terminology

Inconsistent Use	Amount
Dark web = Darknet	19
Dark web = Deep web	6
Darknet = Deep web	3
Dark web part of Deep web	27
Darknet part of Deep web	6
Darknet part of Dark web	3
Deep web part of Darknet	1
Dark web → Deep web	2
Deep web → Dark web	5
Darknet → Dark web	2
Dark web → Darknet	1

applications represent the dark web, and the dark web is located in the darknet.

4.1 Evaluation

Table 1 shows that several papers describe the dark web as a part of the deep web. This could be due to the fact that these works do not consider darknets to be independent networks in which their own services can be offered. Consequently, there is no separate web to which the definition of the deep web does not apply. *Deep web* is a definition of a division of the World Wide Web. However, they deal with all web resources that can be offered by the infrastructure of the Internet as a single set. The definition of the deep web is consequently applied to this set of all web resources.

Some works classify the Internet into several layers. The top layer is attributed to the surface web, the layer below to the deep web, and the bottom layer to the dark web [13, 19, 24, 29, 39]. Another metaphor often used is an iceberg in the ocean [16,25,30,47]. The surface web is described as the tip of the iceberg visible above sea level. Below the water surface is the deep web, and at the bottom of the iceberg is the dark web or sometimes the darknet.

The lack of clarity of the relationship between the deep web and the dark web is further increased by the fact that web resources of the dark web cannot be found by conventional web search engines. Web resources of the dark web are provided by a separate network, and are not in the scope of the World Wide Web. Conventional web search engines, however, only index the area of the World Wide Web. Due to these natural conditions, web search engines cannot find web resources of the dark web. Due to the lack of a clear differentiation, the argument of web search engines is often used in relation to the dark web. As a consequence, this leads to a correlation with the definition of the deep web. This is shown by the following examples:

- "It has been recognized that most of the Internet is not accessible through regular search engines and web browsers. This part of the web is known as dark web" [38]
- "Darkweb refers to the portion of the internet that is not indexed by search engines and hence cannot be accessed by standard browsers." [34]
- "[...] the Surface Web, which is the portion of content on the World Wide Web that may be indexed by popular engines, and lately in the Dark Web, a portion that is not indexed by conventional search engines and is accessed through network overlays such as the Tor network." [7]

- "With 'dark web' we mean parts of the Internet the use of which requires specifically configured browsers or specialized software, such as Tor or I2P. Such parts tend to be pseudo-anonymous and are not usually indexed by search engines such as Google." [20]
- "The Open Web is generally understood as the segment of the web that can be accessed using ordinary web browsers (e.g., Chrome, Firefox, Safari) and search engines (e.g., Google, Yahoo, Bing), whereas the Dark Web is a portion of the web that can only be accessed using specialized browsers that have added layers of encryption technology, such as The Onion Router (Tor) network" [22]

Due to the described conception, some studies also equate dark web and deep web.

Furthermore, additional terms such as *hidden web* or *invisible web* are used [24, 41, 44, 51]. However, these terms are also used inconsistently and are indicated as synonyms for darknet, deep web or dark web. Some works do not distinguish between several of these terms, as the following example shows:

- "Hidden web, also known as deep web, deep net, dark net, invisible web or undernet." [44]
- "This underbelly or underground version of the internet is referred to as the deep web (or deepnet, dark web, invisible web)" [41]

The following example "A Study on Analytical Visualization of Deep Web" [40] shows how challenging it can be to correctly interpret results when terminology is used inconsistently. The author describes that in the paper "the analysis visualization through the Deep Web crime data and the number of Deep Web users in each country" has been studied. However, the following definition is given: "The Deep Web is another range of the Internet called the Deep Internet. 'Deep Web refers to an encrypted network that is not detected on search engine like Google etc. Users must use Tor to visit sites on the dark web"'. No distinction is made between deep web and dark web, but both terms are used. Thus, it is not known whether the results actually refer to the deep web or to the dark web.

5 Conclusion

In our previous paper [43], we elaborated six critical aspects in darknet research. These aspects can lead to misinterpretation of results or inconsistency. They can also lead to misrepresentation of elaborated research results. One aspect is the inconsistent use of darknet terminology. Inconsistent use of darknet terms such as *darknet*, *dark web*, or *deep web* leads to large discrepancies and challenges in making comparisons between multiple studies or results. In this work, we investigated the inconsistent use of darknet terminology in past research. For this purpose, we conducted a scoping review and analyzed how darknet terms have been used or defined. In total, we examined 227 papers, of which 97 listed descriptions or definitions of darknet terms. We found that 28 papers equated multiple terms and did not distinguish between them. Ten papers were identified as having passages of description, which were each a description of a different term. In 36 papers, the relationships between the individual terms were misrepresented.

These inconsistent representations of the individual terms are possibly due to the different metaphorical illustrations. Some divide the Internet into three layers, others use an iceberg or ocean metaphor to describe terms such as *surface web*, *deep web*, and *dark web*. An additional complicating factor is that the term *darknet* has changed over time. In the first uses of the term, a darknet was described as any network in which objects protected by licensing law (e.g., movies, music, software, etc.) are distributed without authorization [10]. It was only later that the term darknet was detached from its original meaning as file-sharing networks and became the designation for overlay networks such as Tor, whose main goal is to preserve the anonymity of its participants [26, 42].

It is essential that the darknet terminology is used consistently. By applying metaphorical illustrations, different levels of abstraction are created which do not result in a detailed description of the terms. Accordingly, an increased number of inconsistencies and contradictions can be observed. A conceivable approach would be to ensure that definitions are based on their respective technical implementation. This way, consistency in terminology usage can be achieved.

Acknowledgment

The joint project PANDA on which this publication is based was funded by the Federal Ministry of Education and Research under the funding codes 13N14355 and 13N14356. The authors are responsible for the content of this publication.

References

[1] Mhd Wesam Al Nabki, Eduardo Fidalgo, Enrique Alegre, and Ivan 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: Volume 1, Long Papers*, pages 35–43, 2017.

[2] Mhd Wesam Al-Nabki, Eduardo Fidalgo, Enrique Alegre, and Laura Fernández-Robles. Torank: Identifying the most influential suspicious domains in the tor network. *Expert Systems with Applications*, 123: 212–226, 2019.

[3] Marios Anagnostopoulos, Georgios Kambourakis, Panagiotis Drakatos, Michail Karavolos, Sarantis Kotsilitis, and David KY Yau. Botnet command and control architectures revisited: Tor hidden services and fluxing. In *International Conference on Web Information Systems Engineering*, pages 517–527. Springer, 2017.

[4] Georgia Avarikioti, Roman Brunner, Aggelos Kiayias, Roger Wattenhofer, and Dionysis Zindros. Structure and content of the visible darknet. *arXiv preprint arXiv:1811.01348*, 2018.

[5] Andres Baravalle and Sin Wee Lee. Dark web markets: turning the lights on alphabay. In *International Conference on Web Information Systems Engineering*, pages 502–514. Springer, 2018.

[6] Frederick Barr-Smith and Joss Wright. Phishing with a darknet: Imitation of onion services. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–13, 2020.

[7] Dario Adriano Bermudez Villalva, Jeremiah Onaolapo, Gianluca Stringhini, and Mirco Musolesi. Under and over the surface: a comparison of the use of leaked account credentials in the dark and surface web. *Crime Science*, 7(1):1–11, 2018.

[8] Massimo Bernaschi, Alessandro Celestini, Stefano Guarino, and Flavio Lombardi. Exploring and analyzing the tor hidden services graph. *ACM Transactions on the Web (TWEB)*, 11(4):1–26, 2017.

[9] Venkataraman Bhaskar, Robin Linacre, and Stephen Machin. The economic functioning of online drugs markets. *Journal of Economic Behavior & Organization*, 159:426–441, 2019.

[10] Peter Biddle, Paul England, Marcus Peinado, Bryan Willman, et al. The darknet and the future of content distribution. In *ACM Workshop on digital rights management*, volume 6, page 54, 2002.

[11] Alex Biryukov, Ivan Pustogarov, Fabrice Thill, and Ralf-Philipp Weinmann. Content and popularity analysis of tor hidden services. In *2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 188–193. IEEE, 2014.

[12] Alex Biryukov, Ivan Pustogarov, and Ralf-Philipp Weinmann. Trawling for tor hidden services: Detection, measurement, deanonymization. In *2013 IEEE Symposium on Security and Privacy*, pages 80–94. IEEE, 2013.

[13] Bruno Requião da Cunha, Pádraig MacCarron, Jean Fernando Passold, Luiz Walmocyr dos Santos, Kleber A Oliveira, and James P Gleeson. Assessing police topological efficiency in a major sting operation on the dark web. *Scientific reports*, 10(1):1–10, 2020.

[14] Mohd Faizan and Raees Ahmad Khan. Exploring and analyzing the dark web: A new alchemy. *First Monday*, 2019.

[15] Mohd Faizan, Raees Ahmad Khan, and Alka Agrawal. Ranking potentially harmful tor hidden services: Illicit drugs perspective. *Applied Computing and Informatics*, 2020.

[16] Nicolas Ferry, Thomas Hackenheimer, Francine Herrmann, and Alexandre Tourette. Methodology of dark web monitoring. In *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–7. IEEE, 2019.

[17] Robert W Gehl. Power/freedom on the dark web: A digital ethnography of the dark web social network. *new media & society*, 18(7):1219–1235, 2016.

[18] Clement Guitton. A review of the available content on tor hidden services: The case against further development. *Computers in Human Behavior*, 29(6):2805–2815, 2013.

[19] Hardik Gulati, Aman Saxena, Neerav Pawar, Poonam Tanwar, and Shweta Sharma. Dark web in modern world theoretical perspective: A survey. In *2022 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–10. IEEE, 2022.

[20] J Tuomas Harviainen, Ari Haasio, and Lasse Hämäläinen. Drug traders on a local dark web marketplace. In *Proceedings of the 23rd International Conference on Academic Mindtrek*, pages 20–26, 2020.

[21] Masayuki Hatta. Deep web, dark web, dark net a taxonomy of “hidden” internet. *Annals of Business Administrative Science*, page 0200908a, 2020.

[22] Thomas J Holt and Jin Ree Lee. A crime script model of dark web firearms purchasing. *American Journal of Criminal Justice*, pages 1–21, 2022.

[23] Intelliagg. Deeplight: Shining a light on the dark web, 2016.

[24] Shubhdeep Kaur and Sukhchandan Randhawa. Dark web: a web of crimes. *Wireless Personal Communications*, 112(4):2131–2158, 2020.

[25] Dimitrios Kavallieros, Dimitrios Myttas, Emmanouil Kermitsis, Euthimios Lissaris, Georgios Giataganas, and Eleni Darra. Understanding the dark web. In *Dark Web Investigation*, pages 3–26. Springer, 2021.

[26] Robert Koch. Hidden in the shadow: The dark web-a growing risk for military operations? In *2019 11th International Conference on Cyber Conflict (CyCon)*, volume 900, pages 1–24. IEEE, 2019.

[27] Denis Korolev, Alexey Frolov, and Irina Babalova. Classification of websites based on the content and features of sites in onion space. In *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pages 1680–1683. IEEE, 2020.

[28] Massimo La Morgia, Alessandro Mei, Simone Raponi, and Julinda Stefa. Time-zone geolocation of crowds in the dark web. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pages 445–455. IEEE, 2018.

[29] Junyan Li. Threats and data trading detection methods in the dark web. In *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, pages 1–9. IEEE, 2021.

[30] Sergio Mauricio Martínez Monterrubio, Joseph Eduardo Armas Naranjo, Lorena Isabel Barona López, and Ángel Leonardo Valdivieso Caraguay. Black widow crawler for tor network to search for criminal patterns. In *2021 Second International Conference on Information Systems and Software Technologies (ICI2ST)*, pages 108–113. IEEE, 2021.

[31] Daniel Moore and Thomas Rid. Cryptopolitik and the darknet. *Survival*, 58(1):7–38, 2016.

[32] Kathleen Moore. Dark web, black markets: Crisis as opportunity. In *ISCRAM 2019. Proceedings*. 2019.

[33] Joanna Moubarak and Carole Bassil. On darknet honeybots. In *2020 4th Cyber Security in Networking Conference (CSNet)*, pages 1–3. IEEE, 2020.

[34] Eric Nunes, Paulo Shakarian, and Gerardo I Simari. At-risk system identification via analysis of discussions on the darkweb. In *2018 APWG symposium on electronic crime research (eCrime)*, pages 1–12. IEEE, 2018.

[35] Gareth Owen and Nick Savage. Empirical analysis of tor hidden services. *IET Information Security*, 10(3):113–118, 2016.

[36] Gareth Owenson, Sarah Cortes, and Andrew Lewman. The darknet's smaller than we thought: The life cycle of tor hidden services. *Digital Investigation*, 27:17–22, 2018.

[37] Jonathan Pace. Exchange relations on the dark web. *Critical Studies in Media Communication*, 34(1):1–13, 2017.

[38] Mandeep Pannu, Iain Kay, and Daniel Harris. Using dark web crawler to uncover suspicious and malicious websites. In *International Conference on Applied Human Factors and Ergonomics*, pages 108–115. Springer, 2018.

[39] George Pantelis, Petros Petrou, Sophia Karagiorgou, and Dimitrios Alexandrou. On strengthening smes and mes threat intelligence and awareness by identifying data breaches, stolen credentials and illegal activities on the dark web. In *The 16th International Conference on Availability, Reliability and Security*, pages 1–7, 2021.

[40] WooHyun Park. A study on analytical visualization of deep web. In *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, pages 81–83. IEEE, 2020.

[41] Amy Phelps and Allan Watt. I shop online–recreationally! internet anonymity and silk road enabling drug use in australia. *Digital Investigation*, 11(4):261–272, 2014.

[42] Florian Platzer, Robert Landwirth, Sandra Wittmer, and Yannikos York. White paper: Was ist das darknet? White paper, Fraunhofer SIT, 2020.

[43] Florian Platzer and Alexandra Lux. A synopsis of critical aspects for darknet research. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, pages 1–8, 2022.

[44] I Gede Surya Rahayuda and Ni Putu Linda Santiari. Crawling and cluster hidden web using crawler framework and fuzzy-knn. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, pages 1–7. IEEE, 2017.

[45] Saad Saleh, Junaid Qadir, and Muhammad U Ilyas. Shedding light on the dark corners of the internet: A survey of tor research. *Journal of Network and Computer Applications*, 114:1–28, 2018.

[46] Iskander Sanchez-Rola, Davide Balzarotti, and Igor Santos. The onions have eyes: a comprehensive structure and privacy analysis of tor hidden services. In *Proceedings of the 26th international conference on world wide web*, pages 1251–1260, 2017.

[47] Vidyesh Shinde, Shahil Dhotre, Vedant Gavde, Ashwini Dalvi, Faruk Kazi, and SG Bhirud. Crawlbot: A domain-specific pseudonymous crawler. In *International Conference on Cybersecurity in Emerging Digital Era*, pages 89–101. Springer, 2020.

[48] Martijn Spitters, Stefan Verbruggen, and Mark Van Staalduinen. Towards a comprehensive insight into the thematic organization of the tor hidden services. In *2014 IEEE Joint Intelligence and Security Informatics Conference*, pages 220–223. IEEE, 2014.

[49] Martin Steinebach, Marcel Schäfer, Alexander Karakuz, Katharina Brandl, and York Yannikos. Detection and analysis of tor onion services. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, pages 1–10, 2019.

[50] Sugiu Takaaki and Inomata Atsuo. Dark web content analysis and visualization. In *Proceedings of the ACM International Workshop on Security and Privacy Analytics*, pages 53–59, 2019.

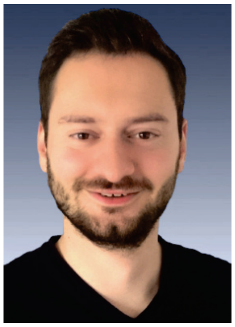
[51] Tarun Trivedi, Vinod Parihar, Manas Khatua, and BM Mehtre. Threat intelligence analysis of onion websites using sublinks and keywords. In *Emerging Technologies in Data Mining and Information Security*, pages 567–578. Springer, 2019.

[52] Madeleine van der Bruggen and Arjan Blokland. Profiling darkweb child sexual exploitation material forum members using longitudinal posting history data. *Social Science Computer Review*, 40(4):865–891, 2022.

[53] York Yannikos, Quang Anh Dang, and Martin Steinebach. Comparison of cyber attacks on services in the clearnet and darknet. In *IFIP International Conference on Digital Forensics*, pages 39–61. Springer, 2021.

[54] Changhoon Yoon, Kwanwoo Kim, Yongdae Kim, Seungwon Shin, and Sooel Son. Doppelgängers on the dark web: A large-scale assessment on phishing hidden web services. In *The World Wide Web Conference*, pages 2225–2235, 2019.

Biographies



Florian Platzer is a research assistant at the Fraunhofer Institute for Secure Information Technology. He is part of the PANDA project at Fraunhofer SIT. The PANDA project is an interdisciplinary project researching the darknet. Within this project he is responsible for the computer science part. Florian studied IT security at the Technical University of Darmstadt, Germany. He wrote his master thesis about deanonymization of Tor hidden services.

Alexandra Lux has been a research assistant in the interdisciplinary project PANDA (Parallel Structures, Activity Forms and User Behavior in the Darknet) since November 2017, and is an associated doctoral student in Sabine Trepte's Team at the Department of Media Psychology at the University of Hohenheim. In her dissertation, she is investigating communication on social networking sites on the darknet. She studied Mass Media and Communication Science with minors in Psychology and Sociology at the University of Vienna and the University of Ottawa.