An Essay on Anonymization and the Dark Web

Alexandra Candeias, Pedro Araújo, and Tiago Ribeiro

University of Minho, Department of Informatics, 4710-057 Braga, Portugal
e-mail: {a89521,a70699,a76420}@alunos.uminho.pt

Abstract. Anonymization is a highly researched topic in recent times since several surveillance programs have become public knowledge, making this topic of severe importance to the majority of the world population. This essay intends to explore the concept of *Dark Web*, explaining what it is and how it functions and also discuss some Anonymity and Privacy Tools.

1 Introduction

In these last few decades, the internet has grown to unprecedented proportions, which triggered a variety of problems regarding privacy and security. As highlighted by Edward Snowden in 2013 [1], several world Governments have resorted to using the internet against one of the primary fundamental rights of humans: privacy. Anonymity has, as a result of such scandals, became a highly discussed and explored field of research. Similarly, there has been a crescent interest in the *Dark web* and how it functions as well as how to access it. The following sections will discuss what it is, how it works, and how to access it as well as all the software needed to keep ones' anonymity and privacy on the internet.

2 How Search Engines Operate

To explain the many layers of the *Web*, there is the need to primarily explain in a general manner how most of the common search engines work. Nowadays, most, if not all, internet users resort to search engines, which have led to their deep integration on how internet is perceived and used. Search engines - such as *Google, Bing, Yahoo*, among others - obtain their search results by the use of *crawlers*. These programs send *spiders* to crawl through all the internet, following one *hypertext link* to the next, in order to register every possible link they find during this process.

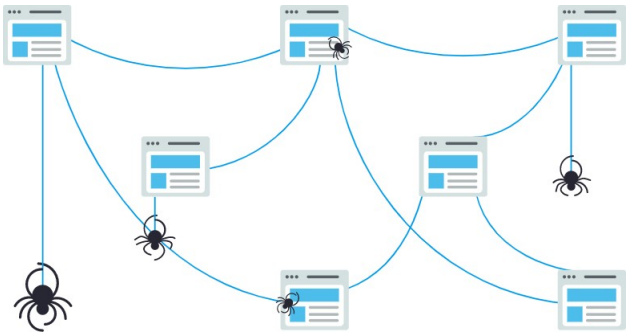

Fig. 1. Crawler sending *spider* through the internet

Upon the *spiders "return"*, a new process takes place where everything that was gathered is stored and organised; after such a process is concluded, the search engine now has a

page ready to be displayed as a result of an enquire. This is what is referred to as *indexing*. Even though such programs exist, there is still some content that can not be searched by our usual *search engines*; this content resides on what its' called the *Deep Web*.

3 The several *Webs*

Nowadays, the term *Web* is divided and convoluted. However, there seems to be a consensus in the scientific community that indicate the existence of three layers that categorise the *Web*, namely *Surface*, *Deep* and *Dark Web*, as illustrated in Figure 2 [2].

3.1 *Surface Web*

Also known as the *Indexable Web*, is the portion of the Internet that is available to the general public through searchable standards when using regular search engines, as *Google, Bing* or *Yahoo*. It is, at its core, made with a collection of public and indexed pages on a server, accessible by any search engine. It is thought that the size of the *Surface Web* is 4000-5000 times smaller than the *Deep Web* [2].

3.2 *Deep Web*

The content in this *Web* layer is not indexed, and therefore, traditional search engines cannot access the information that lies in it. The massive size of this part of the *Web* in comparison with other layers [2] is due to the fact that the vast majority of the information that is contained in this section is composed of private intranets, such as the one *University of Minho* has, as well as commercial and non-commercial databases. As stated in [3], *"You don't surf the Deep Web, you dive into it"*.

3.3 *Dark Web*

The *Dark Web* can either be perceived as the third and last layer of the internet, or as a fully integrated part of the *Deep Web* [2]. This layer can only be accessed through the use of specific software, as discussed in further detail in Section 4. It is believed that nefarious activities take a more prevalent stance on this layer because of the given particularities provided by such software, being the main one anonymization. The majority of this layer is comprised of hidden services that these software are able to reach [2].

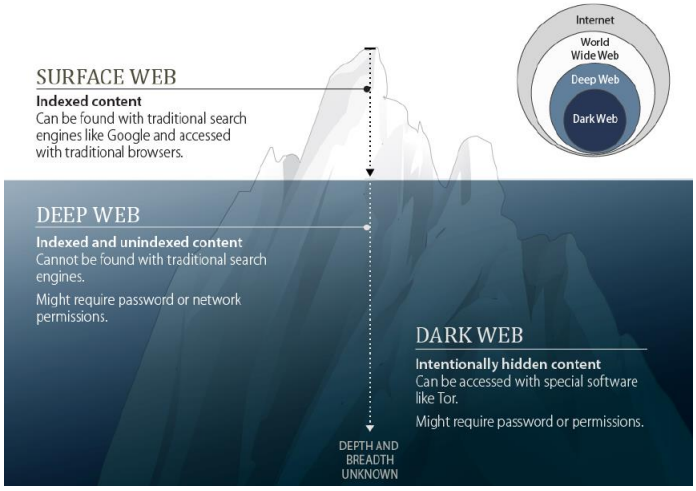

Fig. 2. Modern representation of the internet layers [2]

4 Accessing *Dark Web*

As mentioned above, the *Dark Web* is a portion of the *Deep Web*, but while we can access the latter without needing any form of anonymization or special software, the same cannot be said if there is the need to use the *Dark Web*. An example of such software is *The Onion Router* (*TOR*). The Project responsible for the creation of *TOR* browser developed a revolutionary network overlay technique called *Onion Routing*, which relies on four components - Sender, Receiver, Onion Routers (ORs) and Directory Servers -, that interact as shown in Figure 3 [4].

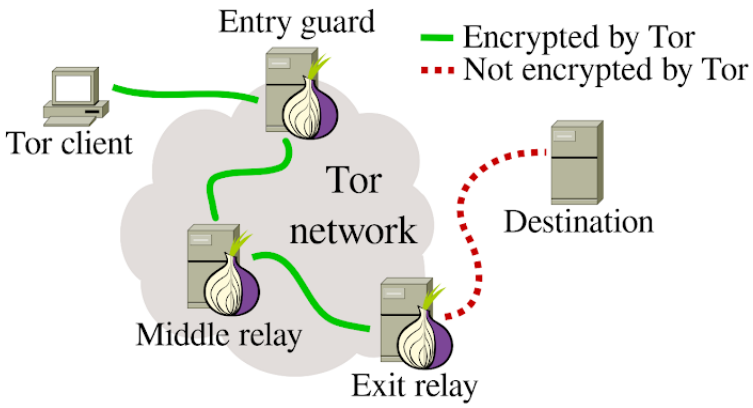

Fig. 3. Functioning of TOR network

With the intent of transmitting a message through *TOR*, an onion is created. The Sender, represented in the Figure 3 by the *TOR Client*, selects a set of *nodes/relays* from a list of public *OR*s provided by the *Directory Servers*. The chosen *nodes/relays* are arranged into a path, represented by the *green line*, through which the message will be transmitted. To preserve the anonymity of the *Sender*, no *nodes/relays* in the path can tell whether the others before itself are the originator or just simply another intermediary. This also means that none of *nodes/relays* in the path can know how many others exist, and only the final *node/relay*, represented in the figure by the *Exit relay*, can determine its own location within the path [4].

Following the brief explanation of TOR network functions, a step-by-step guide of common practices before its use is provided [5]:

1. The user has to download of the most recent released version of *TOR*, in order to ensure that the most up-to-date patches and fixes of possible vulnerabilities have been corrected;
2. Upon a successful installation, the user must *"ensure JavaScript, Flash, and other platforms that can execute code on your computer without your knowledge are disabled"* [5];
3. In case the user is not *tech-savy* and is using the default configuration, they shall change the "DNS requests to be by proxy as this is not set by default. This means that otherwise any DNS lookup will see your IP." [5];
4. If possible, the user should connect TOR to a *VPN* in order to add an extra layer of protection;

5. The user is now set to navigate the *Dark Web*. Due to possible vulnerabilities of *exit nodes*, it is common practice to only visit hidden services sites, since venturing to the *Surface Web* might lead to the user being "caught" in case of monitoring. But in case the user lives in a country that has strict internet policy or censorship there are, fortunately, some *Surface Web* services that offer a *Dark Web* equivalent, such as the case of *Facebook* and *DuckDuckGo*.

5 Anonymity and Privacy Tools

As discussed above, where it was explained the difference between *Deep* and *Dark Web* and how these terms are used interchangeably, also anonymity and privacy are two terms whose usage is commonly mistaken. A simple example to explain the difference between such terms would be the following: Sending an encrypted message may protect ones' privacy, since, assuming the proper procedures were followed, no one besides ones' recipient and themselves will be able to view the content of the message. However, encryption does not protect the metadata and thus does not provides anonymity.

To mitigate such confusion, the following subsections will attempt to explain the most prevalent and used privacy and anonymity tools.

5.1 Privacy Tools

- *Virtual Private Network* (VPN): creates a private network using a public internet access, this means that a VPN will shift the traffic from the users' *Internet Service Provider* (ISP) to the VPN's servers in question [6, 7] .
- *Proxy*: A server application that establishes communication, for example between two machines. *Proxy* servers act as an intermediary. Following this example we can explain the *modus operandi* as such: Machine A asks for a specific request, the *Proxy* acting as an intermediary sends the request to Machine B. After it is processed and concluded, it is sent back. Upon receiving it, the *Proxy* sends it to Machine A. Through this whole process, Machine B does not know who asked for the request [8].
- *Zero-knowledge services*: To explain this concept, the *Cloud Storage* services will be used as an example. Several *Cloud Storage* services on the market are considered *Zero-knowledge services* [9–11]. This services differ from the rest by relying on either *end-to-end encryption* that uses *keys* to encrypt the users' files or *Advanced Encryption Standards* (AES). This means that only the user has access to their encrypted files, and whoever they decide to share their keys with. The company that provides the service does not know nor has the power to know what its users have stored on their servers. There are other *zero-knowledge services* besides *Cloud Storage*, such as *ProtonMail*, an email provider that relies on *PGP keys* to encrypt their users' emails [12].

5.2 Anonimity Tools

- *The Onion Router* (*TOR*): As previously explained it is a software that privatises and anonymizes the connection to the internet as well as giving access to hidden services [5].
- *JonDonym*: Formerly known as *Java Anonymous Proxy* (*JAP*), this network functions by connecting its users' not directly to a web server they intend to visit/use but instead establishing an encrypted connection through several Proxies that are denominated by *mixes* [4, 13].
- *Internet Invisible Project* (*I2P*): Uses a variant of *TOR*s' revolutionary network overlay technique denominating it the *Garlic Routing*. The main difference between these techniques is that while *TOR* sends a single encrypted message, *I2P* chooses to send multiple encrypted messages together to make it extremely difficult for any kind of traffic analysis. It also increases the speeds of data transfer.[4, 14].

– *Freenet*: Decentralised *peer-to-peer* (*P2P*) network that uses a technique denominated *distributed data store*. This manages to keeps the anonymity of its publishers and users [4, 15].
– *Ad-Blockers*: Using *uBlock Origin* as an example, *"It is a wide-spectrum blocker – which happens to be able to function as a mere "ad blocker". The default behavior of uBlock Origin when newly installed is to block ads, trackers and malware sites – through EasyList, EasyPrivacy, Peter Lowe's ad/tracking/malware servers, Online Malicious URL Blocklist, and uBlock Origin's own filter lists"* [16].
– *Tails*: A *GNU/Linux* distribution, based on *Debian*, with a security focus. Revolutionary for blocking all non-anonymous connection and for not leaving a digital footprint on the machine where it is ran. It was designed to be mainly used as a live boot [17].
– *WhoniS*: Like *Tails*, it is a *GNU/Linux* distribution, based on *Debian*, with a security focus. Its main particularity being that the system is composed by two Virtual Machines (VM) that are connected through *TOR* [18].

6 Conclusions

This essay started with the intent to explain how the *Web* works and what tools users have at their disposal for maintaining their anonymity and privacy while using the Internet. Several of the most reliable and used application have been reviewed and the authors tried to give a summarised explanation to each of them in order to foment the interest of the non-*tech-savy* in the use of such software. It is of the utmost importance that users become aware of the dangers that the lack of knowledge about internet privacy and anonymity may result in. The authors hope that this essay may have helped shed some light on the topics discussed.

References

1. Pohle, J., Van Audenhove, L.: Post-snowden internet policy: Between public outrage, resistance and policy change. Media and Communication **5**(1) (2017) 1–6
2. Finklea, K.: Dark Web Kristin Finklea Specialist in Domestic Security. Dark Web (2017)
3. Gabriela González: What is the Deep Web. https://blogthinkbig.com/what-is-the-deep-web (2014) Online; accessed 16 October 2020.
4. Ali, A., Khan, M., Saddique, M., Pirzada, U., Zohaib, M., Ahmad, I., Debnath, N.: TOR vs I2P: A comparative study. Proceedings of the IEEE International Conference on Industrial Technology **2016-May** (2016) 1748–1751
5. Corianna, J.: The Onion Router and the Darkweb. Technical report (2016) 1–14
6. Mason, A.G.: Cisco secure virtual private networks. Cisco Press (2001)
7. J.M. Porup: 8 steps to being (almost) completely anonymous online . https://www.csoonline.com/article/2975193/9-steps-completely-anonymous-online.html (2020) Online; accessed 16 October 2020.
8. Luotonen, A., Altis, K.: World-wide web proxies. Computer Networks and ISDN systems **27**(2) (1994) 147–154
9. Sync: Sync protects your privacy with end-to-end encryption — ensuring that your data in the cloud is safe, secure and 100% private. https://www.sync.com/your-privacy/ (2020) Online; accessed 16 October 2020.
10. Mega: Secure Cloud Storage and Communication. Privacy by Design. https://mega.nz/security (2020) Online; accessed 16 October 2020.
11. SpiderOak: SpiderOak products leverage a unique combination of Zero-Trust encryption and private Blockchain, delivering solutions with a level of confidentiality, integrity, and availability unavailable until today. https://spideroak.com/about/ (2020) Online; accessed 16 October 2020.
12. ProtonMail: Secure Email . https://protonmail.com/ (2020) Online; accessed 16 October 2020.
13. Technische Universität Dresden, the Universität Regensburg and Privacy Commissioner of the state of Schleswig-Holstein: The anonymisation service. https://anonymous-proxy-servers.net/ (2020) Online; accessed 16 October 2020.

14. Invisible Internet Project: The I2P network provides strong privacy protections for communication over the Internet. https://geti2p.net/ (2003) Online; accessed 16 October 2020.
15. Ian Clarke: Freenet is a peer-to-peer platform for censorship-resistant communication and publishing. https://freenetproject.org (2000) Online; accessed 16 October 2020.
16. Raymond Hill: An efficient blocker add-on for various browsers. Fast, potent, and lean. https://github.com/gorhill/uBlock (2015) Online; accessed 16 October 2020.
17. Tails: Tails is a portable operating system that protects against surveillance and censorship. https://tails.boum.org/ (2020) Online; accessed 16 October 2020.
18. Whonix: Software That Can Anonymize Everything You Do Online. https://www.whonix.org/ (2020) Online; accessed 16 October 2020.