Case Study on Deep Dark Web

And just deep Web

Difference between deep Web vs. Dark Web

The Surface

To start on our journey of the different aspects of the web, we'll begin with the surface; the parts you're most familiar with. **The Surface Web is anything that can be indexed by a typical search engine like Google, Bing or Yahoo.** Google has a great interactive story explaining how they index and search the web in depth.
To help you understand how search engines work, I want you to open a traditional news or blog site (CNN, BBC, etc.) and begin clicking different links to new article pages. Once you have finished doing that, come back to the blog posting.

If you're done clicking links, you've just behaved how search engines' crawling technology finds and identifies websites. Search engines rely on pages that contain links to find and identify content. You'll find that this is a great way for finding new content on the web that most of the people generally care about (blogs, news, etc.). But this technique of navigating links also misses a lot of content. Let's go a little deeper to find out exactly what type of content is missed.

Moving a little deeper

From a purist's definition standpoint, the Surface Web is anything that a search engine can find while **the Deep Web is anything that a search engine can't find**. The *Forbes* article that we mentioned previously used BrightPlanet's definition for the Deep Web as the definition for the Dark Web. There are a number of reasons that a search engine can't find data on the web, today we plan on covering the most common one.

Getting a little darker

Continuing with our definitions, we've learned that the Surface Web is anything that a search engine can access and the Deep Web is anything that a search engine can't access. **The Dark Web then is classified as a small portion of the Deep Web that has been intentionally hidden and is inaccessible through standard web browsers.**

The most famous content that resides on the Dark Web is found in the TOR network. The TOR network is an anonymous network that can only be accessed with a special web browser, called the TOR browser. This is the portion of the Internet most widely known for illicit activities because of the anonymity associated with the TOR network.

The key thing to keep in mind is the Dark Web is a small portion of the Deep Web. Some media is inaccurately defining both and we want to do our best to clear up the confusion.

Source : https://brightplanet.com/2014/03/clearing-confusion-deep-web-vs-dark-web/

What is dark Web?

The Dark Web is a term that refers specifically to a collection of websites that are publicly visible, but hide the IP addresses of the servers that run them. Thus they can be visited by any

web user, but it is very difficult to work out who is behind the sites. And you cannot find these sites using search engines.

Almost all sites on the so-called Dark Web hide their identity using the Tor encryption tool. You may know Tor for its end-user-hiding properties. You can use Tor to hide your identity, and spoof your location. When a website is run through Tor it has much the same effect. Indeed, it multiplies the effect. To visit a site on the Dark Web that is using Tor encryption, the web user needs to be using Tor. Just as the end user's IP is bounced through several layers of encryption to appear to be at another IP address on the Tor network, so is that of the website. So there are several layers of magnitude more secrecy than the already secret act of using Tor to visit a website on the open internet - for both parties (See also: How to delete your Google location history).

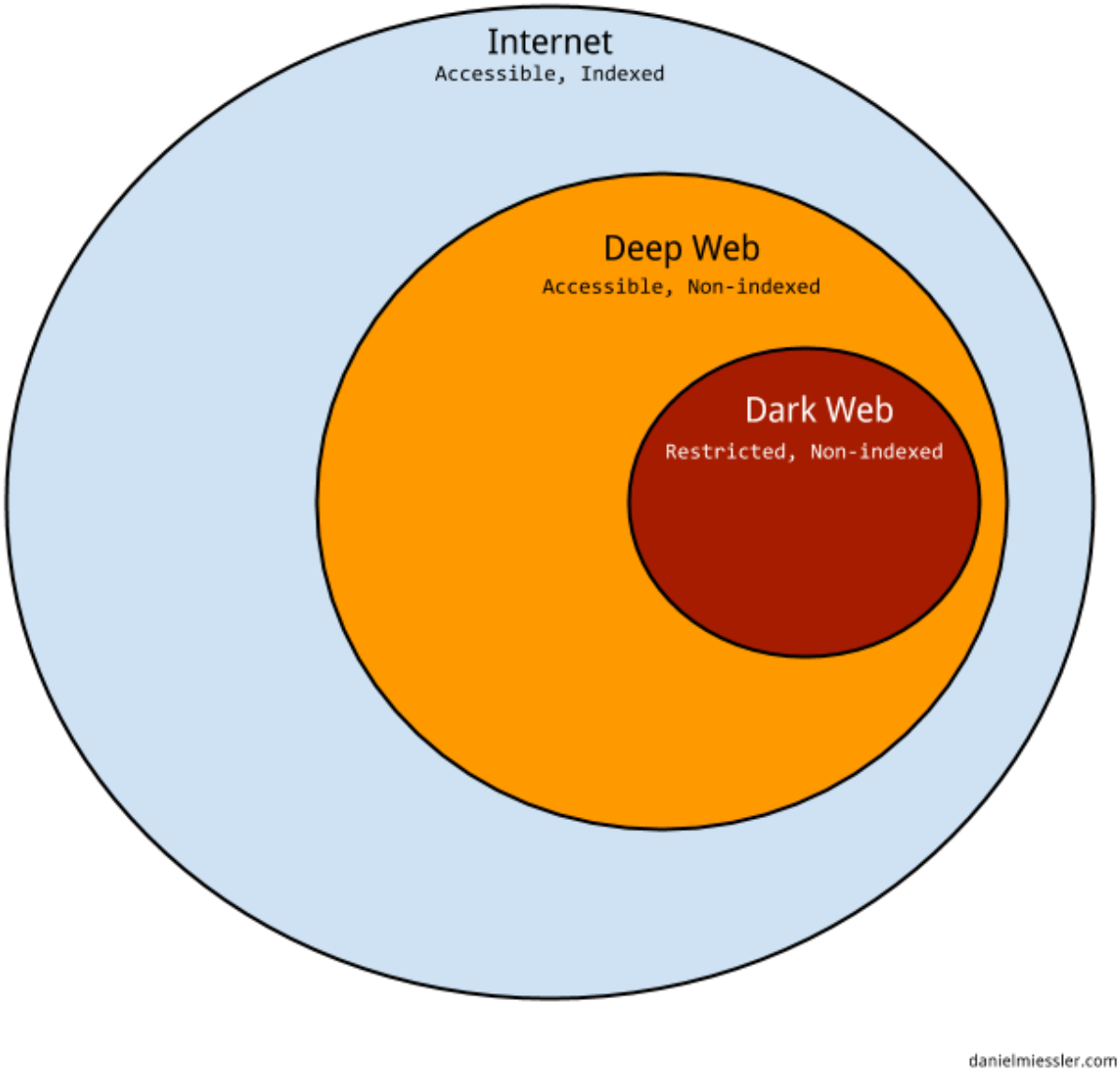
Not all Dark Web sites use Tor. Some use similar services such as I2P - indeed the all new Silk Road Reloaded uses this service. But the principle remains the same. The visitor has to use the same encryption tool as the site and - crucially - know where to find the site, in order to type in the URL and visit.

Infamous examples of Dark Web sites include the Silk Road and its offspring. The Silk Road was (and maybe still is) a website for the buying and selling of recreational drugs. But there are legitimate uses for the Dark Web. People operating within closed, totalitarian societies can use the Dark Web to communicate with the outside world. And given recent revelations about US- and UK government snooping on web use, you may feel it is sensible to take your communication on to the Dark Web.

An Extract from www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp_below_the_surface.pdf:-

I n te rest in the Deep Web peaked in 2013 when the FBI took down the Silk Road marketplace and exposed the Internet’s notorious drugtrafficking underbelly. Ross Ulbricht, aka Dread Pirate Roberts, was charged for narcotics trafficking, computer hacking conspiracy, and money laundering. While news reports were technically referring to the Dark Web—that portion of the Internet that can only be accessed using special browsing software, the most popular of which is TOR [1]—negative stereotypes about the Deep Web spread. The Deep Web is the vast section of the Internet that isn’t accessible via search engines, only a portion of which accounts for the criminal operations revealed in the FBI complaint [2]. The Dark Web, meanwhile, wasn’t originally designed to enable anonymous criminal activities. In fact, TOR was created to secure communications and escape censorship as a way to guarantee free speech. The Dark Web, for

example, helped mobilize the Arab Spring protests. But just like any tool, its impact can change, depending on a user's intent. In our 2013 paper, "Deep Web and Cybercrime [3],"and subsequent updates [4, 5, 6], we sought to analyze the different networks that guarantee anonymous access in the Deep Web in the context of cybercrime. In the process, we discovered that much more happens in the murkier portions of the Deep Web than just the sale of recreational drugs. It has also become a safe haven that harbors criminal activity both in the digital and physical realms. This paper presents some relevant statistics derived from our collection of Deep Web URLs and takes an even closer look at how criminal elements navigate and take advantage of the Deep Web. It provides vivid examples that prove that people go there to not only anonymously purchase contraband but also to launch cybercrime operations, steal identities, dox high-profile personalities, trade firearms, and, in more depraved scenarios, hire contract killers.



How is possible that resources located on the web are not visible and which are the content of the hidden web?

Ordinary search engines use software called “crawlers” to find content on the web, they are computer programs that browse the World Wide Web in a methodical, automated manner and are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches.

This technique is ineffective in finding the hidden resources on the Web that could be classified into the following categories:

- ***Dynamic content***: dynamic pages which are returned in response to a submitted query or accessed only through a form, especially if open-domain input elements (such as text fields) are used; such fields are hard to navigate without domain knowledge.
- ***Unlinked content***: pages which are not linked to by other pages, which may prevent Web crawling programs from accessing the content. This content is referred to as pages without backlinks (or inlinks).
- ***Private Web***: sites that require registration and login (password-protected resources).
- Contextual Web: pages with content varying for different access contexts (e.g., ranges of client IP addresses or previous navigation sequence).
- ***Limited access content***: sites that limit access to their pages in a technical way (e.g., using the Robots Exclusion Standard, CAPTCHAs, or no-cache Pragma HTTP headers which prohibit search engines from browsing them and creating cached copies).
- ***Scripted content***: pages that are only accessible through links produced by JavaScript as well as content dynamically downloaded from Web servers via Flash or Ajax solutions.
- ***Non-HTML/text content***: textual content encoded in multimedia (image or video) files or specific file formats not handled by search engines.
- ***Text content using the Gopher protocol and files hosted on FTP that are not indexed by most search engines***. Engines such as Google do not index pages outside of HTTP or HTTPS.

How does Tor network work?

Imagine a typical scenario where Alice desire to be connected with Bob using the Tor network. Let’s see step by step how it is possible.

She makes an unencrypted connection to a centralized directory server containing the addresses of Tor nodes. After receiving the address list from the directory server the Tor client software will connect to a random node (the entry node), through an encrypted connection. The entry node would make an encrypted connection to a random second node which would in turn do the same to connect to a random third Tor node. The process goes on until it involves a node (exit node) connected to the destination.

Consider that during Tor routing, in each connection, the Tor node is randomly chosen and the same node cannot be used twice in the same path. To ensure anonymity the connections have a fixed duration. Every ten minutes to avoid statistical analysis that could compromise the user’s privacy, the client software changes the entry node

Up to now we have considered an ideal situation in which a user accesses the network only to connect to another. To further complicate the discussion, in a real scenario, the node Alice could in turn be used as a node for routing purposes with other established connections between other users.

A malevolent third-party would not be able to know which connection is initiated as a user and which as node making impossible the monitoring of the communications.