What is the Deep Web?
A first trip into the abyss

From: www.SecurityAffairs.co

Article written by: Pierluigi Paganini: Chief Information Security Officer at Bit4Id[1]

The Deep Web (or Invisible web) is the set of information resources on the World Wide Web not reported by normal search engines.

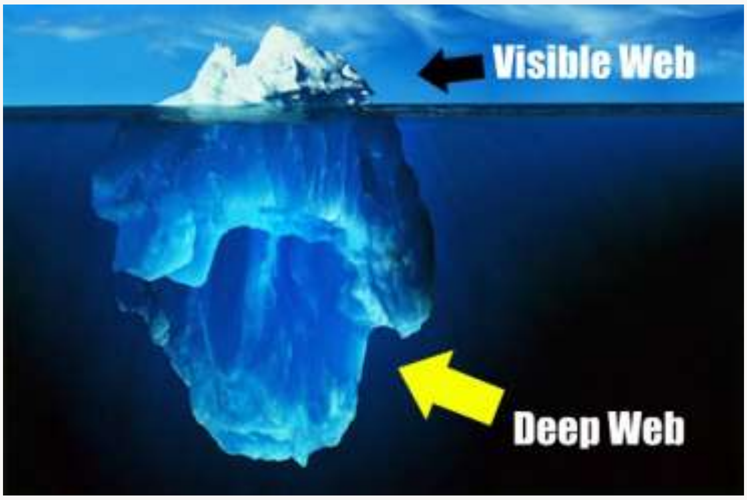
According several researches the principal search engines index only a small portion of the overall web content, the remaining part is unknown to the majority of web users.

What do you think if you were told that under our feet, there is a world larger than ours and much more crowded? We will literally be shocked, and this is the reaction of those individual who can understand the existence of the Deep Web, a network of interconnected systems, are not indexed, having a size hundreds of times higher than the current web, around 500 times.

Very exhaustive is the definition provided by the founder of BrightPlanet, Mike Bergman, that compared searching on the Internet today to dragging a net across the surface of the

[1] Company Director, Researcher, Security Evangelist, Security Analyst and Freelance Writer. Security expert with over 20 years experience in the field. Certified Ethical Hacker at EC Council in London. The passion for writing and a strong belief that security is founded on sharing and awareness led him to found the security blog "Security Affairs". Today he is Chief information security officer for Bit4id company, firm leader in identity management, and he works as a writer with some major publications in the field such as Cyber War Zone, Infosec Island, The Hacker News.

ocean: a great deal may be caught in the net, but there is a wealth of information that is deep and therefore missed.



Ordinary search engines to find content on the web using software called “crawlers”. This technique is ineffective for finding the hidden resources of 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.

A parallel web that has a much wider number of information represents an invaluable resource for private companies, governments, and especially **cybercrime**. In the imagination of many persons, the DeepWeb term is associated with the concept of anonymity that goes with criminal intents the cannot be pursued because submerged in an inaccessible world.

As we will see this interpretation of the Deep Web is deeply wrong, we are facing with a network definitely different from the usual web but in many ways repeats the same issues in a different sense.

What is a Tor? How to preserve the anonymity?

Tor is the acronym of “The onion router”, a system implemented to enable online anonymity. Tor client software routes Internet traffic through a worldwide volunteer network of servers hiding user’s information eluding any activities of monitoring.

As usually happen, the project was born in military sector, sponsored the US Naval Research Laboratory and from 2004 to 2005 it was supported by the Electronic Frontier Foundation.

Actually the software is under development and maintenance of Tor Project. A user that navigate using Tor it’s difficult to trace ensuring his **privacy** because the data are encrypted multiple times passing through nodes, Tor relays, of the network.

Connecting to the Tor network

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.

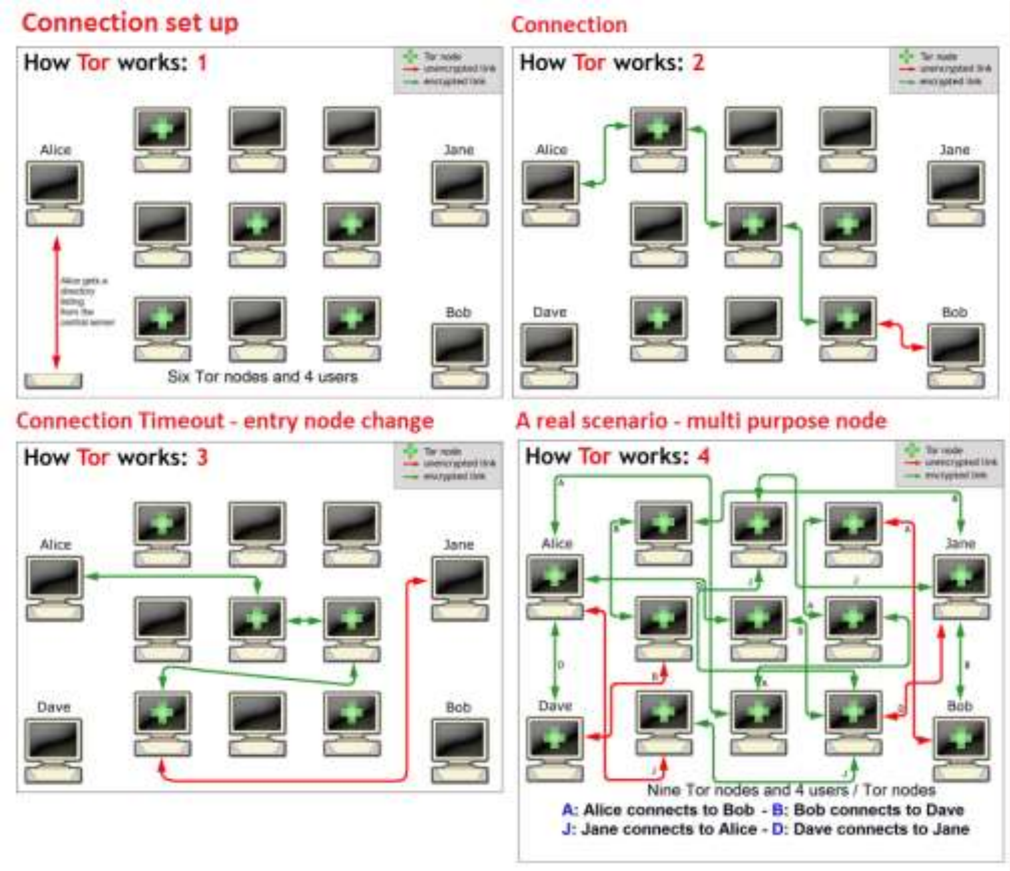
She makes an **unencrypted** connection to a centralized directory server containing the addresses of Tor nodes. After receiving the address list from the directory server the Tor client software will connect to a random node (the entry node), through an encrypted connection. The entry node would make 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 are 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.


Connection set up
How Tor works: 1
Tor node
unencrypted link
encrypted link
Alice
Jane
Alice gets a directory listing from the central server
Bob
Six Tor nodes and 4 users
Connection
How Tor works: 2
Alice
Jane
Dave
Bob
Connection Timeout - entry node change
How Tor works: 3
Alice
Jane
Dave
Bob
A real scenario - multi purpose node
How Tor works: 4
Alice
Jane
Dave
Bob
Nine Tor nodes and 4 users / Tor nodes
A: Alice connects to Bob - B: Bob connects to Dave
J: Jane connects to Alice - D: Dave connects to Jane

After this necessary parenthesis on Tor network routing we are ready to enter the Deep Web simply using the Tor software from the official web site of the project. Tor is able to work on all the existing platforms and many add-ons make simple they integration in existing applications, including web browsers. Despite the network has been projected to protect user's privacy, to be really anonymous it's suggested to go though a VPN.

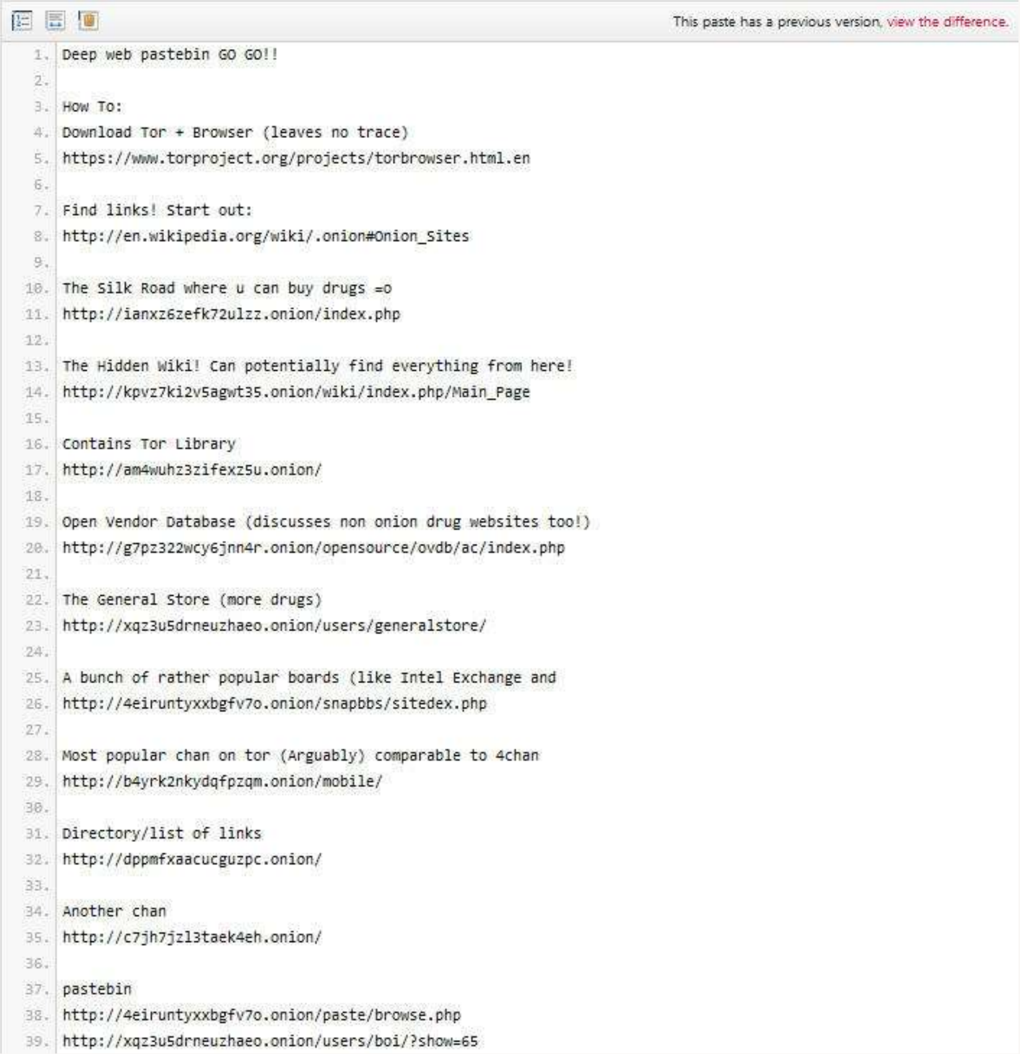
A better mode to navigate inside the deep web is to use the **Tails OS distribution**which is bootable from any machine don't leaving a trace on the host. Once the Tor Bundle is installed it comes with its own portable Firefox version, ideal for anonymous navigation due an appropriate control of installed plugins, in the commercial version in fact common plugins could expose our identity.

Once inside the network, where it possible to go and what is it possible to find?

Well once inside the deep web we must understand that the navigation is quite different from ordinary web, every research is more complex due the absence of indexing of the content.

A user that start it's navigation in the Deep Web have to know that a common way to list the content is to adopt collection of Wikis and BBS-like sites which have the main purpose to aggregate links categorizing them in more suitable groups of consulting. Another difference that user has to take in mind is that instead of classic extensions (e.g. .com, .gov) the domains in the Deep Web generally end with the .onion suffix.

Following a short list of links that have made famous the Deep Web published on**Pastebin**



```
Deep web pastebin GO GO!!

How To:
Download Tor + Browser (leaves no trace)
https://www.torproject.org/projects/torbrowser.html.en

Find links! Start out:
http://en.wikipedia.org/wiki/.onion#Onion_Sites

The Silk Road where u can buy drugs =o
http://ianxz6zefk72ulzz.onion/index.php

The Hidden Wiki! Can potentially find everything from here!
http://kpvz7ki2v5agwt35.onion/wiki/index.php/Main_Page

Contains Tor Library
http://am4wuhz3zifexz5u.onion/

Open Vendor Database (discusses non onion drug websites too!)
http://g7pz322wcy6jnn4r.onion/opensource/ovdb/ac/index.php

The General Store (more drugs)
http://xqz3u5drneuzhaeo.onion/users/generalstore/

A bunch of rather popular boards (like Intel Exchange and
http://4eiruntyxxbgfv7o.onion/snapbbs/sitedex.php

Most popular chan on tor (Arguably) comparable to 4chan
http://b4yrk2nkydqfpzqm.onion/mobile/

Directory/list of links
http://dppmfxaacucguzpc.onion/

Another chan
http://c7jh7jzl3taek4eh.onion/

pastebin
http://4eiruntyxxbgfv7o.onion/paste/browse.php
http://xqz3u5drneuzhaeo.onion/users/boi/?show=65
```

Cleaned Hidden Wiki should be a also a good starting point for the first navigations

http://3suaolltfj2xjksb.onion/hiddenwiki/index.php/Main_Page

Be careful, some content are labeled with common used tag such as CP= child porn, PD is pedophile, stay far from them.

The Deep Web is considered the place where every thing is possible, you can find every kind of material and services for sale, most of them illegal. The hidden web offers to cybercrime great business opportunity, hacking **services**, malware, stolen credit cards, weapons.

We all know the potentiality of the e-commerce in ordinary web and its impressive growth in last couple of years, well now imagine the Deep Web market that is more that 500 times bigger and where there is no legal limits on the odds to sell. We are facing with amazing business controlled by ciber criminal organizations.

Speaking of dark market we cannot avoid to mention Silk Road web site, an online marketplace located in the Deep Web, the majority of its products are derived from illegal activities. Of course it's not the only one, many other markets are managed to address specify products, believe me, many of them are terrifying.

Most transactions on the Deep Web accept **BitCoin**system for payments allowing the purchase of any kind of products preserving the anonymity of the transaction, encouraging the development of trade in respect to any kind of illegal activities. We are facing with a with an autonomous system that advantage the exercise of criminal activities while ensuring the anonymity of transactions and the inability to track down the criminals.

But is it really all anonymous? Is it possible to be traced in the Deep Web? What is the position of the governments towards the Deep Web?

I will provide more information on the topic in next articles … in meantime let me thank a great expert of the Deep Web, "**The gAtOmAlO**" with whom I collaborate on a project which we will present you soon.

Pierluigi Paganini

(Security Affairs – Deep Web)