

# Monitoring the most active sites on the Tor-Network

Degree programme : BSc in Computer Science | Specialisation : IT Security  
Thesis advisor : Prof. Dr. Emmanuel Benoist  
Expert : Daniel Voisard

Anyone who surfs the Internet or offers web services is usually not anonymous. The onion router (Tor) allows providers and consumers to protect their privacy by using the onion routing protocol. The web services offered in this way are summarized under the term the Dark Web and only accessible using the Tor software. Many rumors exist about the Dark Web and its content. Therefore, we statistically analyzed the public part of the Dark Web.

### Public Onions

In order to crawl addresses, a list of public Dark Web addresses (xyz.onion) is created. For this purpose, we deploy an open-source project called Ahmia Search Engine. This search engine, which is specialized on the Dark Web, browses through Dark Web pages and stores the newly encountered links. With this method we collect a list of unique and public Dark Web addresses. At the time of this writing, we had collected over 18,061 onion addresses.

### Crawling and Categorization

Our self-developed crawler visits the list of public Dark Web addresses automatically and analyzes them. On each page it extracts various characteristics of the site. This includes, for example, the language, the category, the type of web server, if CAPTCHA are in use or if there is a member area.

Additionally, we created a machine learning program that determines the corresponding category for each site, primarily based on the textual content of the site. In total twenty-two different categories are in use. Some examples are Forums, Markets, Search, Bitcoin, Drugs, Abuse, Porn and others. The software is capable of detecting the category with an accuracy of 89% and a deviation of three percent.

Table-1 illustrates the distribution of Dark Web sites in the different categories. The category Markets leads the ranking significantly. This is not surprising, since only the public area of the Dark Web is considered. Markets assumably have the biggest motivation to be found, whereas other dark websites want to remain hidden. On a market, mostly illegal goods like drugs, fake money and credit cards and even books can be purchased. The Abuse category ranks second. It includes pornography, which is clearly in an illegal (western jurisdiction) context. The category Directory contains, for example, websites that list Dark Web

links, but is a very general and broad category, which explains it being in third place. The table confirms the assumptions that the majority of the Dark Web is dominated by illegal activities.

### Popularity of v2 Onions

In order to estimate how popular Dark Web sites are compared to each other, a bug in version two of the onion protocol is used to count when a client looked up an onion address. This bug is fixed in version three and therefore only statements about the popularity of v2 addresses can be made. It can be shown that a few Dark Web sites make up the majority of the lookups. However, distributed denial of service attacks against Dark Web sites can be observed, which distort the picture a bit.

### Conclusion

Summarized, there are a few v2 addresses that make up a large part of the traffic. These are placed in legally questionable to illegal corners. The whole range of dark websites is vast, but concentrates mainly in the area of Directories, Markets and Abuse, concerning their popularity.



Jan Friedli  
jan.friedli@immerda.ch

Unique Address Count In Each Category

Category	Percent	Category	Percent
Market	40.40	Counterfeit	0.63
Abuse	14.19	Hosting	0.39
Directory	13.37	Mail	0.33
Fraud	12.21	Forum	0.28
Wiki	7.23	Guns	0.27
Blog	2.70	Hacking	0.27
Search	2.44	Gambling	0.17
Bitcoin	1.93	Anonymity	0.16
Porn	1.47	News	0.15
Drugs	0.66	Books	0.08
Whistleblower	0.65	Chat	0.03

Table-1