TOR for Dark and Deep Web

Survya Singh | UID: U00803205 | singh.143@wright.edu

Abstract

“Tor is a real-world, circuit-based low-latency anonymous communication network, supporting TCP applications over the Internet” [2]. The paper discusses how Tor is implemented in Deep and Dark web and describe the detail explain of how Tor works towards providing anonymity to user in network. It also explains the basic difference between surface web, deep web and dark web with respect to search engine indexing and content anonymity. Paper also describe various methods of disrupting the anonymity of Tor users in network and explain the reason behind 3 famous incidents where FBI was able to caught Tor users involved in illegal activity. Tor is a network and a software package that helps us anonymously use the Internet, but if used in wrong way it can easily reveal your identity, thus paper also discuss safe way of using Tor.

1. Overview

Privacy is an important aspect of human life and same applies for internet user’s. In this digital era where everything is available online, more and more people are getting connected to internet, which imposes a threat on internet user’s privacy. Most of us do not want to reveals the things that we do online, like posting against political parties or protesting against someone or something, and our personal browsing habit to get connected to our real identity.

Anonymity does not only means hiding of one’s personal information, in internet world it means not revealing one’s network information such as system IP address and MAC address to rest of the users on internet. Tor one of the freely available software, when used correctly hides system IP address and MAC address in TCP network connection. It is considered as one of the best privacy tools currently available, and is widely used for deep and dark web. Deep web is a that part of the internet which cannot be crawled by link-crawler used in most of the search engine’s like Yahoo and Google. But that does not mean that deep web always uses anonymity tools(Tor) to hide their data from link crawlers, any content that cannot be indexes by search engines fall into the category of deep web for e.g. data present on website, but not in the form of link. Dark web is a part of deep web which uses tools like Tor to conceal websites or servers from surface web and is inaccessible by normal search engines and standard web browser. Detail implementation of Tor in dark web and some infamous dark web websites is briefly discussed.

Paper discuss how Tor encrypts data and anonymize data by wrapping it into multiple layer like an onion, and the concepts of relay, cells, directories and bridges used in Tor. It discusses “Protocol level attack” and “Traffic Analysis attack “and their effectiveness in disrupting the anonymity of Tor network. It discusses the “Harvard Bomb threat”, “Freedom Hosting” and “Silk Road” cases as an example of misusing Tor and the loop holes in their Tor network which led them to reveal their identity and finally getting caught by FBI. If someone is using Tor to login into their social network website like Facebook or Gmail, then this can end up in destroying your own anonymity, thus paper also discuss the safe ways of using Tor and some cases where one should not use Tor.

2. Tor Design and Working [1]

2.1 Components of Tor:

Tor a freeware software easily available and widely used for connecting anonymously with public internet and does not require root privilege or kernel modification to work. Tor works by forming the circuit of relay or nodes and uses symmetric encryption to encrypt the data before sending it to destination. By default, Tor uses 3 relay node to form chain of circuit/node and each node in the circuit knows only the address of its adjacent node. Below are the details of components used in Tor.

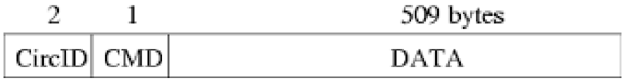
i. **Directory authorities/Servers:** Tor has a hardcoded list of 10 nodes which are called Directory servers which stores the relay nodes from all over the world. These authority servers are maintained by Tor volunteers and have responsibility of updating the DA. Even though Tor is used for providing anonymity to user on internet, DA are publicly available on internet and anyone can see the list of relay node used for Tor connection. Out of this 10 DA, 9 are used for updating relay/node list and 1 is used for bridge node. Update of DA server is done hourly and requires 5 DA server to vote in favor of update, which means to keep Tor network functional 5 DA server must be operational always.

ii. **Tor Relay/Node:** Node in Tor networks are the volunteered system/computers, who are willing to share their bandwidth to make Tor connection more secure. There are around 7500 tor relay node available and anyone in the world can make their system as a relay node. Higher the number of relay node available in Tor network more secure it becomes. By default, Tor client uses 3 nodes to form the circuit before sending data to destination, and Tor client are free to modify the number of nodes for circuit formation. Based on the position of nodes in circuit relay node are divided into 3 types. Entry Node – This is entry point of connection in Tor circuit and have a high bandwidth available. Middle Node – as name suggest it is the middle node in the Tor circuit and prevent entry and exit node from knowing each other. Exit node – this is the last node of Tor circuit which send data to the destination and share its IP address with destination server.

iii. **Onion Proxy (OP):** Once Tor is installed on a system OP decide the relay node to be used as entry/middle/exit node and communicate with each node, by hiding the IP address of client system. OP also fetches the list of updated DA and Tor nodes.

iv. **Bridges:** We have already discussed that list of all the relay node are publicly available, so what if ISP or government blocks all the relay node available to disrupt the Tor network. Bridges are the non-published relay node which helps in providing the Tor network in censored network. Out of 10 DA, 1 DA named as “Tonga DA” (82.94.251.203:443) contains the information about bridges node.

v. **Cells:** Data in the Tor circuit travels in the fixed size packet called as cells. The size of each cell is 512 bytes and consist of heard and payload. Two bytes of data are used as circuit identifier(CircID) which specify the circuit cells belong too, one byte is used for command(CMD) which describe what has to be done with 509 bytes of payload.



2	1	509 bytes
CircID	CMD	DATA

Fig: Control Cells

Based on the command type cells are divided into 2 types Controls cells and Relay cells.

Controls cells – it is the data packet which is used by node to control the Tor circuit and is not used for data transmission. There are 3 types of command in the control cells *padding* – used for keeping the circuit alive, *create or created*- used for creating the new circuit and *destroy*- used for tearing down the circuit.

Relay cells – it is used for transmission of payload in circuit or for extending the number of node in the circuit. It has additional header with streamID, digest (for integrity check of payload), length of relay payload and command.

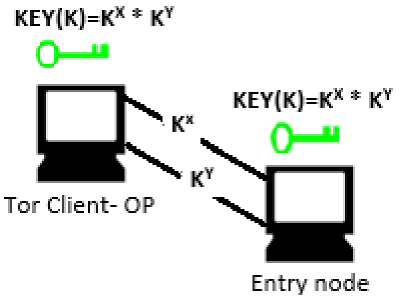
2	1	2	6	2	1	498
CircID	Relay	StreamID	Digest	Len	CMD	DATA

Fig: Relay Cells

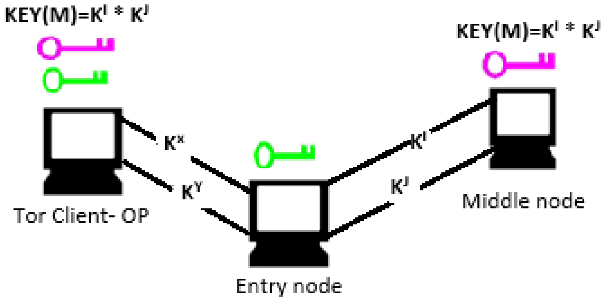
Relays cells are encrypted at each node, before forwarding it to another node. Key is shared between OP and node, no other node in the circuit can have information about symmetric key of other nodes. Command that are present in relay cells are *relay data* – for relay of payload from node to node, *relay begin* – for opening the relay stream, *relay end* – to close the relay stream, *relay connected* – for notifying the OP that relay stream has be opened successfully, relay *extend* – for extending the Tor circuit by increasing the number of node, and *relay truncate* – used to tear down the part of circuit.

2.2 Construction of Tor Circuit:

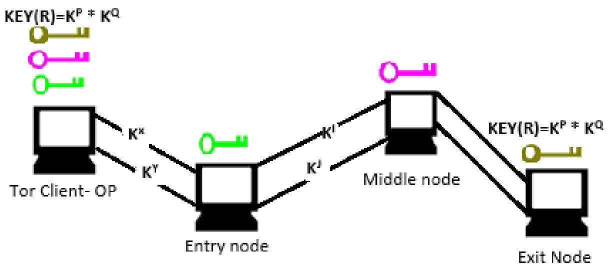
After the Tor bundle software is installed on system, OP start fetching the list of updated relay node. Based on the bandwidth available it will choose an entry relay node. OP will then send a control cell with create command, the payload of control cell will have first half of Diffie-Hellman key(K^x) and entry node will respond with other half of the key(K^y) along with the hash of key. After the key negotiation is done between OP and entry node, OP will have a symmetric key for encryption between OP and entry node.



Now OP will select middle relay node and will send the IP address of middle relay node and first half of key (K^I) in encrypted relay extend packet to entry node for circuit extension. Entry node will copy the payload and will create a control cell with create command and send it to middle node. Middle node will respond with created cell and other half of the key(K^J) in encrypted packet. Entry node will pass this information to OP, without the knowledge of symmetric key that is shared between OP and middle node.



Same circuit extension process is repeated for exit node and once circuit is formed, tor client will have 3 symmetric key 1 from each node, but nodes in the circuit are unaware of symmetric key of other nodes in the circuit.



Exit node is the one which send data packets to destination address, thus destination server is only able to know the IP address of exit node. Middle node prevents communication between entry and exit nodes.

2.3 Tor Data Encryption in Circuit:

Once circuit is created between Tor client and destination server, client will start sending relay cells for data transmission, but before sending the data packets are encrypted with 3 layers of symmetric encryption. Packet is first wrapped with destination address information and encrypted with exit node key, second layer of wrapping contains information of exit node and is encrypted with middle node key, finally whole packet is then wrapped with information of second node and encrypted with the entry node key.

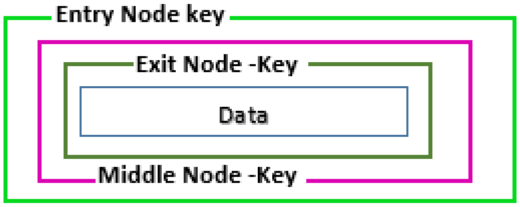


Fig: Tor Data encryption layer

After encryption is done, packet is send to entry node, where entry node decrypts the first layer of wrapping by his symmetric key. Packet is then passed to middle node where is it again decrypted by the middle node and finally packet reaches the third layer where packet get decrypted. If Tor user is sending non- encrypted or using Http, then data at exit node will be in plaintext which poses threat to user anonymity. Packet is then send to destination server, and same circuit is used for sending data from destination to Tor client. And again if destination server is sending data in plaintext then exit node will receive data in plaintext imposing threat to user anonymity.

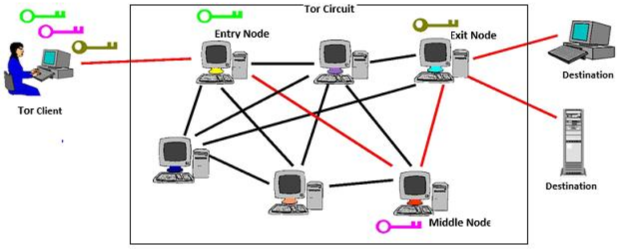


Fig: Complete Flow of Tor circuit.

3. Tor implementation in Dark Web[9]

Based on the anonymity of content or data, internet is divided into 3 categories.

Surface web – Anything on internet which can be indexed by search engines like Google, Yahoo etc. fall under this category.

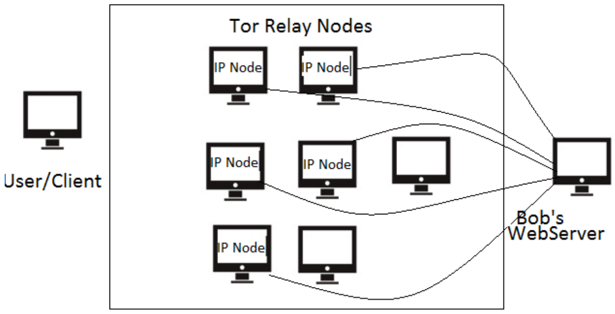
Deep web –Search engines work by crawling the links present on website for indexing of webpages present on internet. Anything which fails to get crawled by link crawler's fall into the category of Deep web. One of the best example of deep web can be a website without any links, but with a search box and allow user to search content of website using search box. Link browser fails to use search box for indexing.

Dark web – "It is classified as a small portion of the deep web that has been intentionally hidden and is inaccessible through standard web browsers" [3]. Dark web is implemented in such a way that normal web user cannot access it from public internet. Tor is mostly used for implementing such website and the user has a special way of accessing it.

3.1 Hidden Web-Services implementation with Tor [1]:

The goal behind hidden web-service is to deploy a server on internet, that anyone can have access too, without knowing the IP address of the server or who owns that webserver and is accessible around the whole world and is resistance to censorship or denial of service attack. Below are the steps involve in creating such a server on internet and letting anyone to access it.

i. Let's assume Bob wants to create a hidden web service on internet, so he creates his web server and choose few entry nodes (for e.g. 5 nodes) called as "Introduction Point Node (IP Node)" available on Tor network and connects his web server to all the chosen entry node.



ii. Bob then generated public key for each IP node and publish his website fake name along with some description and valid public key for one of the IP node to public database, which is easily accessible to normal internet users.

iii. User then access this public database and ask more information about website on public database in encrypted message. User also has to create one rendezvous point, which is tor relay node selected by user.

iv. User shares his rendezvous point address with Bob on public database, and the message is encrypted with Bob public key.

v. Bob, then connects to user rendezvous point and provides one of the entry point IP address, which connect directly to web server. By this way Bob has a control over his webserver in terms of user access.

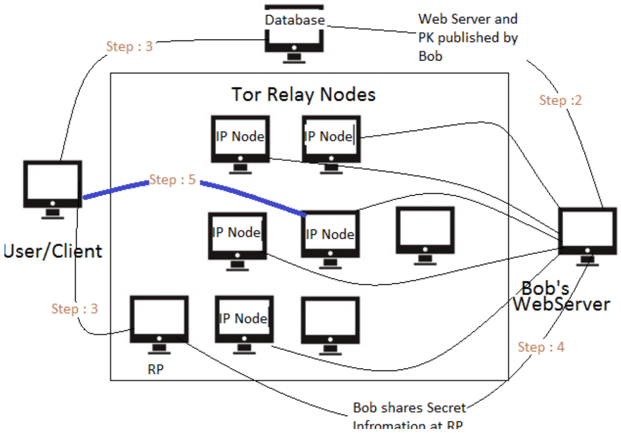


Fig: Steps of user connection to dark web

In this implementation neither the client has any information about Bob, neither Bob has any information about client using his web service. Dark web is mostly associated with illegal activity for e.g. "Silk Road" a website which was based on dark web and was used for selling illegal items to user. It was even called as "Amazon" of illegal products. But that is not always the case, one can use dark web to store their family's photos providing access only to their family members.

4. Breaking of Tor Anonymity

There has been long debate, whether Tor anonymity can be broken and various research paper has been published on the same topic. Recently researchers from MIT publish that they were able to break the Tor anonymity, but there is still no proof whether they were able to break the anonymity of Tor. Below is the detail explanation of 2 attacks published in research papers about disruption to Tor anonymity.

4.1 Traffic Analysis Attack [2]:

This method of attack allow attacker to guess/infer the nodes which are being used for anonymous data transfer, thus reducing the effectiveness of Tor anonymity. Traffic analysis means extracting the information from sniff network meta data, including network overload volume and timings of packets. Attacker will use this information to gain knowledge and trace the network originator or destination address. Now let assume that attacker controls some of the relay node of Tor network and the destination server which Tor user is trying to connect is also controlled by attacker. Let's assume that Tor user is sending some data to attacker website, which is getting sniffed by attacker. Attacker will analyses that packet size and timing of receiving each packet on destination address. Attacker will send the same packet he received from the relay node which he controls to all other node, which he thinks was used in transmission data from user to destination. Assuming here that attacker has the ability to monitor the CPU process of the tor node to which he is sending data.

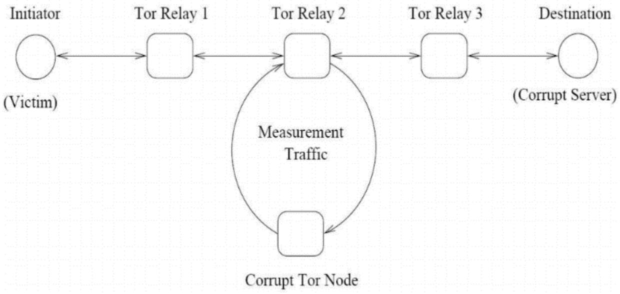


Fig: “Traffic Analysis Attack” [2].

Attacker can user timing characteristic to observer which tor node was used for data traffic and degrade the anonymity of tor to simple proxy servers. But if there is higher volume of traffic on Tor node then this attack will fail to infer any knowledge about timing or network overload. Even though this attack can infer some knowledge about the Tor node, but it clearly fails to fetch the information about Tor client.

The loop holes in Tor, which make this attack possible on tor connection are:

- No explicit delay or reorder in sending the cell or packet.
- Load on the Tor node affects the latency of all the connections traffic.

In my opinion this attack if difficult to implement in real Tor network.

4.2 Protocol- Level attack [3]:

For this attack to be possible the attacker should have a control on entry relay node and exit relay node. This attack works by modifying or deleting some part of the data at the entry node.

User when sends data to entry node, the attacker at entry node will modify the packet and will forward that packet to middle node. Middle node will work as usual and will forward the packet to exit node, which is also under the control of attacker. As Tor uses the counter mode AES for encryption of packets, the modifies data packets will encounter the error during decryption of packets at exit node. By carefully analyzing the error at node, attacker can successfully trace the tor circuit, thus degrading the anonymity of Tor network. This attack can also be used to launch the denial of service attack on middle node.

4.3 Effectiveness of both Attack:

In my opinion protocol level attack can achieve better result than traffic analysis attack, as protocol level attack can determine the circuit of tor network on successful attack. But both of these attack seems highly dependent on knowledge of some part of Tor circuit or controlling of some of the Tor node, thus making this type of attack highly infeasible in real world Tor network connection. By default, Tor uses 3 nodes to from circuit, but user can easily modify Tor bundle to use more of the Tor relay node to form circuit and in that case both of this attack will fail to disrupt the anonymity of Tor network. There are various other way/tricks which can break the Tor anonymity, but it requires some form of mistake form Tor user, otherwise breaking Tor at present seem highly tough task.

There is various real incident where FBI caught users involved in illegal activity using Tor and below is the detail exploitation and case studies of three such incidents.

4.3.1 Harvard bomb threat Hoax using Tor [5]:

Student named Eldo Kim was caught by FBI, for sending an email about “Bomb Threat” at Harvard university building. Intention behind this hoax was to delay the exam at university. Email was sent at around 8:30 am on Monday. He was using Tor network to send email from email service provider called as “GuerillaMail” by using campus Wi-Fi network. So now question arises, how FBI was able to caught him? was FBI able to break Tor network?

Answer is No, Harvard university maintains the logs of all the recent activity on their network and their campus Wi-Fi requires one to sign with their campus ID. GuerillaMail was able to provide all the IP address who were connected around 8:30 am on Monday and we know that exit node communicates with destination server, so one of the IP address was of exit node. As all of the relay node are available on public internet, FBI was able to find the exit node IP address and thus had information that someone was using Tor to send email. Now it was easy for FBI to find out who was connected to Tor network at 8:30 am by campus Wi-Fi from the recent log activity. Unfortunately, there was only one student who was campus ID was used for connecting to Tor at that time and was caught by FBI.

4.3.2 Freedom Hosting Service with Tor [6,7]:

Freedom hosting was one of the dark net most popular hosting service and was run by "Eric Eoin Marques". Some of the most infamous website of Tor like TorMail, HackBB, and Hidden Wiki was hosted on Freedom hosting server. And it was believed that half of the dark net website were using Freedom Hosting server for hosting their website's. Freedom hosting was also accused for hosting some of the most famous child pornography websites like Lolita City, PedoEmpire and the Love Zone.

In 2011 some Anonymous group issued warning against Freedom hosting to remove all of the CP websites from there server, but Freedom hosting refused to remove those websites. Two and half hours later anonymous attacked the webserver of Freedom hosting and taking down all the websites. Later Freedom hosting webserver was under the control of US and Irish government.

After taking the server under control, US and Irish government injected some java script code to all the website of Freedom hosting and made it operation again. The java script code would run automatically by the browser of users who were accessing those website using Tor network. Java script code would first check, whether the Tor browser which is being used if having Firefox 17 version. If Tor user browser is matched, then code start downloading some exploit on user machine by taking the advantage of Firefox 17 vulnerability. Java Script code was able to send Tor user IP address and Mac address to FBI, thus totally disrupting the anonymity of Tor network. The reason behind how anonymous was able to attack Freedom hosting webserver is still unknown and thus raise question whether anonymous group was able to break Tor anonymity. FBI was able to identify all those Tor users whose browser was not updated.

4.3.3 Silk Road – The Amazon of Dark Net [8]:

Silk road was considered as one of the most sophisticated and criminal friendly market place available on Internet, and was taken down by FBI recently. Silk road was designed like Amazon.com, where user was able to buy illegal drugs, fake passports, fake driver license and illegal service provide such as hit men, hackers and forgers. Silk worm had $1.2 billion in sales and million customers making it one of the most famous dark net website based on Tor. Bitcoin was used for buying the products available on Silk road.

"Ross Ulbricht" a former student of Pennsylvania State university was accused as an owner of Silk road and is under FBI custody. It is not clear how FBI was able to compromise the server of Silk road, and it is believed that FBI found some weakness in the computer code that used to operate the Silk road website and was able to exploit those vulnerability and forced server to reveal their IP address and Mac address.

In all of the 3 cases, user was caught or anonymity was broken due to Tor user's mistake or by vulnerability present in website. It is still not clear whether FBI or anonymous or any other group were able to break Tor anonymity, thus making Tor a most powerful software for providing anonymity to internet user if implemented properly.

5. How to Use Tor Safely

If Tor is used properly and correctly it is hard to disrupt the anonymity of Tor network, and one can hide their identity on global internet very easily.

5.1 Using Tor browser [4]:
"Tor bundle does not protect all of your system internet traffic when you run it" [4]. Tor will protect only that application which are configured to use Tor network. Sometimes even by using Tor browser can reveal your identity for e.g. suppose you are trying to connect to xyz.com and your system does not know the IP address of xyz.com, then in that case your system will send DNS query to ISP which may not be using the Tor network. There are many Tor browser plugins which can be used to get the IP address of any website thus resolving this issue.

5.2 Never Use Torrent over Tor [4]:
"Torrent file-sharing applications have been observed to ignore proxy settings and make direct connections even when they are told to use Tor" [4]. Not only torrent reveals your identity it will also slow down the tor network for other user of Tor.

5.3 Always Using HTTPS version of Websites [4]: If Http version of website is used, then it is useless to use Tor network because data will be received in plaintext at entry and exit node and if entry and exit node are malicious Tor node, then anonymity of user is broken. "Tor will encrypt your traffic to and within the Tor network, but the encryption of your traffic to the final destination website depends upon on that website.[4]"

5.4 Never open documents by Tor browser while online [4]:
Documents lie PDF, Doc may contain some script which may run once you open that document and Tor network is not configured to hide the anonymity of traffic originating from other application like Microsoft Word, Adobe Reader etc. "The Tor Browser will warn you before automatically opening documents that are handled by external applications. DO NOT IGNORE THIS WARNING. [4]"

5.5 Using Bridge Relay nodes [4]:
Bridge relay nodes are the node which are not publish by Tor on public internet. Using bridge Tor user can avoid censorship of network and make it hard for anyone analyzing traffic to reveal your identity. In case of "Harvard Bomb Threat" if student had used bridge node, then it was highly impossible for FBI to catch him.

6. Conclusion

Tor, a circuit-based low-latency anonymous communication service available for free, Tor is considered as one of the most powerful software used to hide one's identity on global internet. Many research paper have been published to disrupt the anonymity of Tor network, but still there is no such method which can really harm the anonymity of Tor connection if user has implemented Tor properly. Even though many cases are there ,where FBI was able to caught Tor user's involve in illegal activity , but not by breaking Tor network anonymity. Attack like "Traffic analysis" can guess few nodes which are involve in Tor anonymous traffic connection but fails to completely reveal the Tor user identity. Thus we can conclude that Tor is still the most potent freeware software used by internet users to hide their identity.

References:

[1] Tor: The Second-Generation Onion Router. (2004).
http://www.onion-router.net/Publications/tor-design.pdf

[2] Murdoch, S., & Danezis, G. (2005). Low-cost traffic analysis of Tor. 2005 IEEE Symposium On Security & Privacy (S&P'05), 183. doi:10.1109/SP.2005.12
http://sec.cs.ucl.ac.uk/users/smurdoch/papers/oakland05torta.pdf

[3] One Cell is Enough to Break Tor's Anonymity. (2009).
https://www.blackhat.com/presentations/bh-dc-09/Fu/BlackHat-DC-09-Fu-Break-Tors-Anonymity.pdf

[4] https://www.torproject.org/download/download-easy.html.en#warning

[5] http://www.dailydot.com/crime/tor-harvard-bomb-suspect/

[6]https://bitcoinmagazine.com/articles/freedom-hosting-taken-down-founder-arrestes-users-fed-javascript-exploits-1375664047

[7] http://www.dailydot.com/news/eric-marques-tor-freedom-hosting-child-porn-arrest/

[8] http://www.usatoday.com/story/news/nation/2013/10/21/fbi-cracks-silk-road/2984921/

[9] http://www.brightplanet.com/2014/03/clearing-confusion-deep-web-vs-dark-web/