Títle: **The Onion Router: A in-depth analysis**

Author: **Ricardo Manuel Batista Rodriguez**

Date: **29/06/2022**

Index

Introductory Note:

This report was made for the *Aspetos Profissionais e Sociais da Engenharia Informática* course and aims to explain a project called *The Onion Router*, its purposes and how it operates on the internet.

1. Summary / Abstract

Nowadays, browsing on the internet raises security and privacy problems, with popular browsers like *Google Chrome* or *Mozilla Firefox* gathering large amounts of personal information to present targeted advertisement and with people having difficulty in hiding their own identity and online fingerprint. [1]

The Onion Router is a free software which allows users to surf the internet in an anonymous and secure way by redirecting traffic to a web of repeaters, preventing surveillance techniques and encrypting data in order to browse freely.

The safety provided by TOR may bring different kinds of usages, with legal and illegal ends, but it also presents some weaknesses and performance issues.

2. Framework

Over the last decades, technology has been growing exponentially and more people use the internet every day for daily routine tasks, for work, entertainment and educational purposes.

Current modern browsers like *Google Chrome, Mozilla Firefox* and *Microsoft Edge* provide a bridge between the user and internet, presenting many functional tools that improve the way we see information in our screens. Internet data is transferred using the Hypertext Transfer Protocol, or HTTP, which is a standard that defines how different types of data are shared in the web, and browsers can choose how to display that information, by translating the information written in HyperText Markup Language, or HTML, and rendering the information to be visually represented to the user. [2]

Nevertheless, these browsers give users a “free” service in return for personal data extraction. Browsers often store information about users likes and patterns, stored in cookies, hardware and software, search history and many other types of private data, and sell the information to advertisers in order to trace a user profile and present personalized advertisements. [3]

Thus, internet privacy is a must, and many internet users feel the need to have the right kind of software that can prevent data leakage, data encryption and assure anonymity in the digital world.

3. The Onion Router

The Onion Router is a free open-source software created in the 1990s by David Goldschlag, Mike Reed, and Paul Syverson, employees at the United States Naval Research Laboratory, with the purpose of creating a secure and anonymous environment for internet users around the world. [4]

This being said, the main idea behind the TOR project is to allow the use of the internet without the monitoring and mass surveillance of external entities like governmental agencies and network eavesdroppers.

More than 3 million users connect to the TOR network each day and the number keeps rising, since the urge to privacy is more wanted. [5]

3.1. Operation

The TOR network consists of an overlay network with thousands of intermediate relay servers, called onion routers, whose job is to redirect encrypted internet traffic through itself 3 hops until reaching the destination server the user wants to connect to. With this in mind, Onion Routing is like a complex proxy server with a unique redirect protocol.

When a user wants to establish a TOR connection, an immutable path is generated constituted by the client, an entry node, a middle node, an exit node and the destination server. However, instead of a normal connection, the data sent between these actors is always encrypted, except for the last hop between the exit node and the destination server.

Initially, three symmetric keys are shared between the client and the onion routers using the Diffie-Hellman protocol [6]. The client shares the symmetric key K1 with the input node, the key K2 with the middle node and, finally, key K3 with the exit node. Each one of these nodes only have knowledge of the key shared between themselves and the client, which means no node knows how all the communication is done in its whole, only a fraction of it.

After establishing the symmetric keys between the network nodes, the message going to be sent by the client is encrypted with all the symmetric keys in a direction contrary to the communication path. First, the message is encrypted with key K3, known by the exit node, then with the key K2, known by the middle node, and by the key K1, known by the input node.

Therefore, the initial message is encrypted three times with three different keys, meaning that we have three layers of encryption in that message, just like the onion’s layers. When the encryption is done, the client starts the communication by sending the data to the first node of the path: the input node. This node, having knowledge of the outer layer of the encrypted message’s key K1, decrypts the message, resulting in an unreadable and still encrypted message with two layers of encryption. The same happens between the remaining nodes of the path, with the middle node decrypting the received data with the key K2 and with the exit node decrypting the message received by the middle node. When reaching the exit node, the last layer of the message’s encryption gets "peeled of" by the key K3 and the message, initially encrypted, is sent to the server decrypted.

This happens because exit nodes are responsible for connecting with the requested destination server and receiving the response from the server, thus the last hop can't be encrypted by TOR, only the remaining hops can be encrypted. Encryption in the last hop of the connection can only be achieved if the remote server supports an encryption mechanism like SSL or TLS.

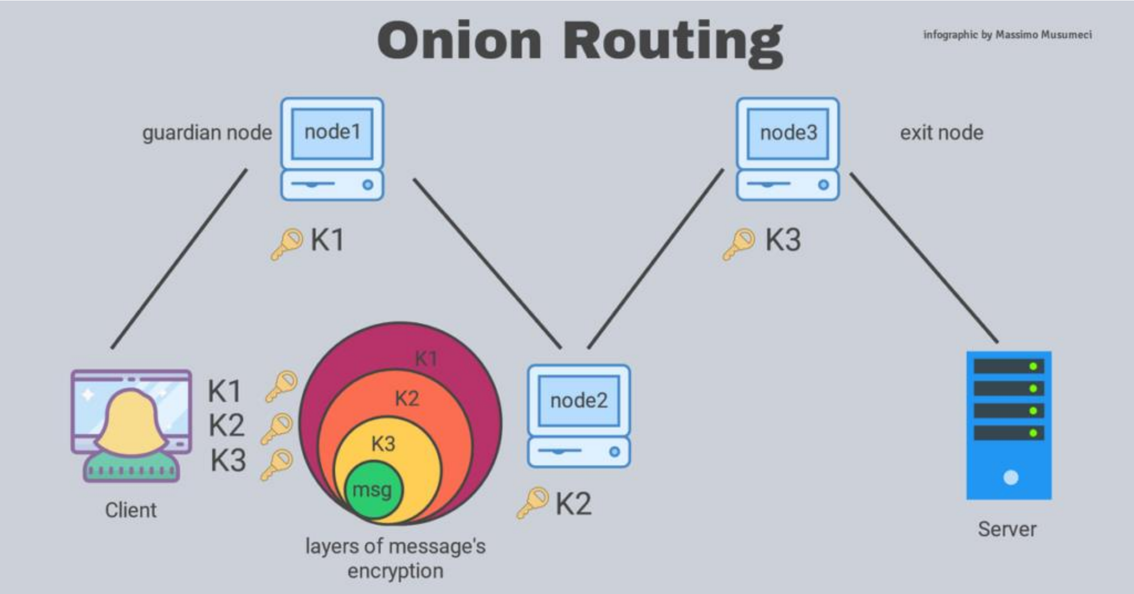


Figure 1 – How Onion Routing works – connection's actors and encryption [7]

The exit node of the connection knows the content of the message and the destination server of the client. However, he doesn't know who the client is. The only knowledge he has about the connection path are the nodes adjacent to it.

It's important to note that not only the content of the message is encrypted but also the IP address of the sender, at any point of the communication path. If an eavesdropper is listening, for example, between the exit node and the remote server, he cannot identify the original sender, since the last node appears to be the originator of the connection. This encryption mechanism, alongside with a series of onion routers working like a proxy, assure users anonymity on the internet.

TOR's anonymity is not determined only by its encrypted transmission mechanism. Although the destination server doesn't know the client's IP address, the client also doesn't know the IP address of the server he wants to connect to. This bidirectional privacy is due to TOR's hidden services, servers configured to receive traffic from the TOR network, exclusively. These hidden services, also known as onion services, are accessed through their onion address (.*onion*), instead of the common IP address or DNS [8]. This address corresponds to the onion service's public key, which is exposed to the public.

When a server is set up online as an onion service, he starts asking three random TOR relays if they want to be their introductory points. These brokers, responsible for introducing clients to the server, establish a three-hop circuit between themselves and the server.

In the same way, the server creates a descriptor containing its public key and the IP addresses of each of his introductory points. After defining the server's descriptor, it will be sent to TOR's distributed hash table. Then, when a client wants to connect to a hidden service using its onion address, the distributed hash table gets the IP addresses of the introductory points of the server with the requested public key, allowing the user to establish a connection with the server's introductory points. [9]

When a person wants to connect to a hidden service, he first needs to know the server's public key. Since an onion address consists of a string of 56 characters, followed by the ".onion" suffix, users typically can't memorize it or simply don't know the public key to a specific server. Because of this, users often turn to websites with curated lists of onion websites or by search engines.

After knowing the server's public key, the distributed hash table maps the onion address to all the server's introduction points IP address, if a server with that address exists. Then, the client chooses a random TOR relay node to act as a *rendezvous* point and establishes a circuit between themselves. This circuit has two-hops, considering that the third hop is the *rendezvous* point. The client asks the *rendezvous* node to connect to a hidden service by a specific introduction point with an introduction message, consisting of the *rendezvous point*'s address and a one-time cypher. This introduction message is encrypted using the hidden service's public key.

After the client's message is received by the introduction point, the last one forwards the message to the onion service which decrypts the final message. Only the server can decrypt the message since it was previously encrypted with its public key, meaning it can only be decrypted by its private key. In analogy, this asymmetric encryption mechanism works like a postal service. Everyone knows the location of my mailbox (public key) and they can send letters to it (messages), but only I can open the mailbox with its key (private key) and read the content of the message.

At this point, the onion service can decide if it's going to accept the client's request or to do nothing. If it accepts the client's request, it creates a circuit to the *rendezvous* point and it sends a one-time cypher to it.

Receiving the one-time cypher, the *rendezvous* point informs the client that a connection has been established between himself and the onion service and they can interact with each one using the chosen relay node as an intermediary.

Afterwards, messages get sent back and forth encrypted with the one-time cypher shared between the *rendezvous* point and the other connection entity, which can be the client or the onion service. It's important to note that both of them share different one-time cyphers with the relay node and, since there's a circuit between the client and the *rendezvous* point and between the last and the onion service, all the circuit has six hops (3 hops each).

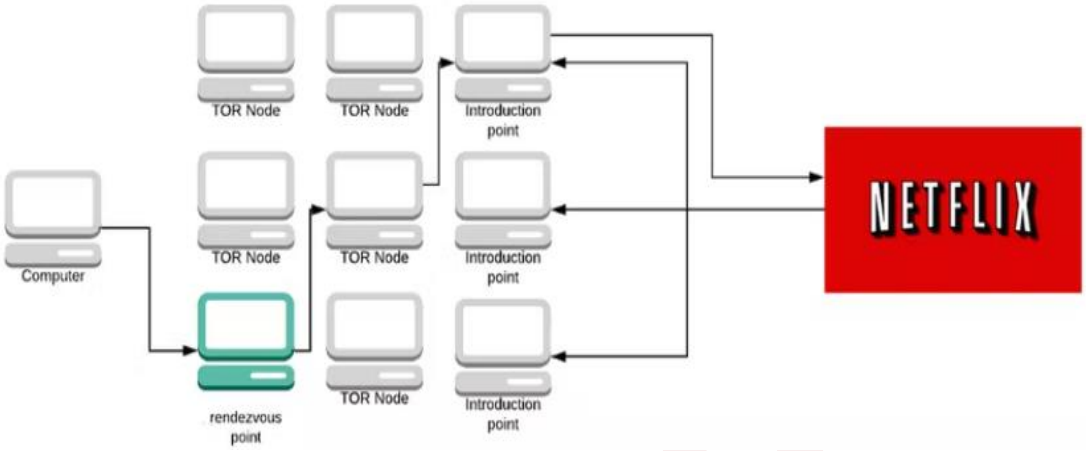


Figure 2 – Connection between the client and an onion service (Netflix) [9]

Traffic sent along the TOR network doesn't have variations in length, data is sent in fixed-size 514 bytes cells. The reason behind this protocol choice is that varying the length of the cells provides an additional security to traffic analysis, since it's difficult for eavesdroppers to identify the real size of a file transferred in different connections and prevents them from correlate timing information between two points of the network.

Because of the extra hops a connection does in the TOR network, alongside with the encryption mechanism and the relay nodes' overload, a typical connection takes way longer than a simple connection on the internet, sometimes in the order of seconds. This also depends on the geographic location of the relay nodes of the circuit, because they can be opposites sites of the globe.

TOR is a decentralized network, to some extent. There are 10 directory authorities, since there's no form a client can find nodes inside the network with these authorities involved. The list of the current authority relays can be found in this website.

3.2. Weaknesses

Although TOR is an awesome software for anonymous communications, retransmitting data along different relay nodes, encrypting data and having mechanisms that allow the presence of unidentified entities like the onion services and the client, it also has some weaknesses.

3.2.1. Consensus Blocking

As previously mentioned, TOR is a decentralized mechanism. It needs a consensus mechanism between the authority nodes, which are publicly listed with their IP address, to update the way the network operates: updating relays, entry and exit nodes, etc.

Every hour, authority nodes communicate with themselves to update the state of the network and reach for a consensus on how everything should be. However, this communication can be blocked when some attacker congests the network, with a DDoS attack, for example, making the authority nodes unavailable for communication and preventing votes to update the consensus. [10]

3.2.2. Traffic-analysis attack

This attack can be divided into two different ones: active and passive. In an active traffic-analysis attack, the attacker, listening on the two opposite sites of the network, modifies timing flow of the sent packets by the client following a specific pattern and observes timing correlation on the other side of the network.

In a passive traffic-analysis attack, the intruder doesn't alter the timing flow of the packets. Instead, he tries to correlate the timing flow of the sent packets on one point of the network and compares it to the other point he is listening to. [10]

3.2.3. Eavesdropping

Like the previous topic, the eavesdropping technique can be divided into different categories: exit node eavesdropping and an internal communication attack.

In exit node eavesdropping, since the last hop of a connection made between an exit node and the hidden service is not encrypted by TOR, attackers can listen to traffic on these points and intercept it, if the remote server doesn't use end-to-end encryption like TLS or SSL. Although this technique doesn't break the sender's anonymity, it can expose partial information about him due to payload or the protocol data. [10]

On the other hand, internal communication attacks, discovered by a research team from ESIEA, consists of creating a map of TOR relay nodes, one-third of them being controlled by the same entity, and then gather their encryption keys and algorithm seeds. Afterwards, using the collected keys and seeds, they can decrypt two of the three TOR encryption layers, breaking the last one using a statistical attack. Redirection of the traffic to the entity's owned relay nodes can be done through a DDoS attack.

3.2.4. Bad apple attack

In 2011, a document made by INRIA researchers described an attack that can reveal the IP addresses of *BitTorrent* users on the TOR network, a popular software used for peer-to-peer file sharing. Researchers ran the attack for 23 days, listening in 6 exit nodes, and collected 10000 IP addresses of TOR users.

Since the IP address of the user is sent in the datagram payload of *BitTorrent*, attackers can exploit insecure application use and map the use of secure application with the IP address of the TOR's user.

3.2.5. Tor exit node block

Internet sites' operators can simply prevent traffic from Tor exit nodes or offer less functionalities for users coming from these points. Websites can block all traffic incoming from TOR exit nodes using IP tables, since the all exit nodes IP addresses is publicly listed and operators can create a firewall that blocks communication with these relays.

3.3. The TOR browser

The TOR browser is an internet browser that allows users to surf the web using the TOR software. Instead of being a normal browser which supports normal internet traffic, it operates using the overlay network and encryption methods that onion routing uses.

This is the most popular way to access the TOR network, it's like the Mozilla Firefox browser but with additional security and privacy features. First, HTTPS Everywhere is included by default in the TOR browser. This extension allows switching insecure "http" websites into secure "https" ones, protecting users against eavesdroppers. Using this feature, all communications done using the TOR browsers assure encryption, with the browser blocking every unsafe connection.

In addition, the NoScript extension is also present on the browser, which allows content only from trusted domains in order to prevent exploitation that can be done using cross-site scripting attacks (XSS), cross-site request forgery (CSRF) and other security vulnerabilities with no performance loss. [11]

Besides the default extensions, the browser makes sure all cookies are deleted at startup, including the browser's history, and websites can't use any information that can be used to identify the person, breaking the concept of anonymity TOR provides. The browser has cross-platform availability. [12]

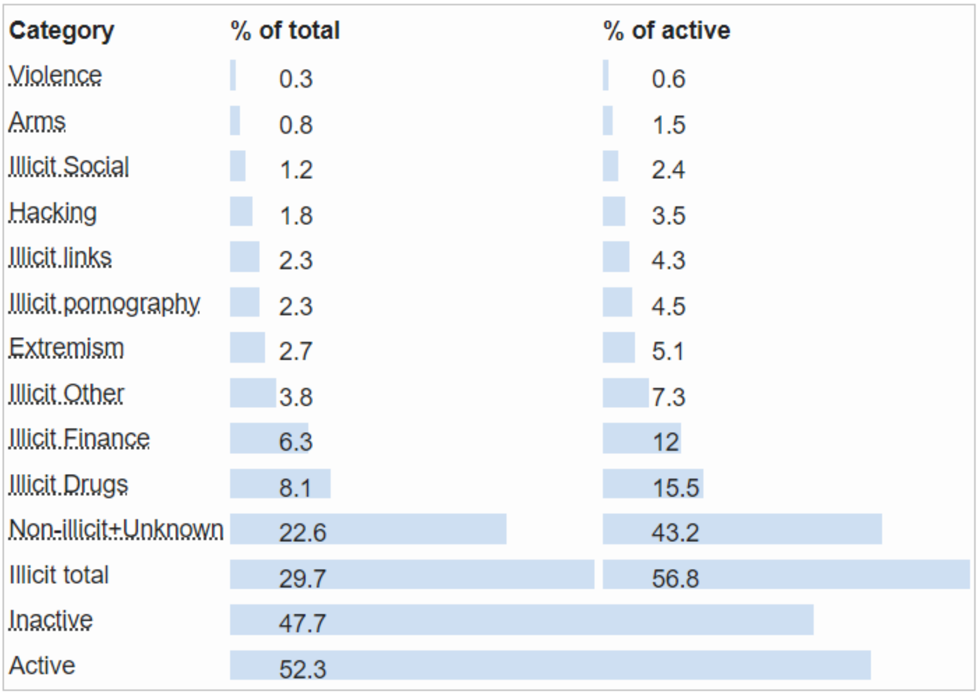
4. Socioeconomic Consequences

The onion routing protocol is one of the most important inventions in a modern world where technology is present in the daily routine of billions of users. Its privacy mechanism, explained before, allows users to surf on the internet without being profiled by companies or other entities.

Because of this, more and more users are joining the TOR network every day on behalf of freedom. This privacy toolkit offers privacy and security when doing simple tasks like making a post on an online forum about your point of view, without fearing the big eye watching us.

However, when onion routing allows users to look all the same, making difficult for eavesdroppers or entities to identify who is behind the screen, unregulated websites or clients with illegal purposes start rising in the network.

Typically, not only everyone who uses TOR is a criminal but almost every cyber criminal uses TOR for its illegal activities. As described in the figure 3, TOR is often used for illicit activities like, for example, drugs, violence, extremism, etc. A study estimated that, on average, nearly 6.7% of TOR users connect to the network for illegal purposes. [10]



Category	% of total	% of active
Violence	0.3	0.6
Arms	0.8	1.5
Illicit Social	1.2	2.4
Hacking	1.8	3.5
Illicit links	2.3	4.3
Illicit pornography	2.3	4.5
Extremism	2.7	5.1
Illicit Other	3.8	7.3
Illicit Finance	6.3	12
Illicit Drugs	8.1	15.5
Non-illicit+Unknown	22.6	43.2
Illicit total	29.7	56.8
Inactive	47.7	
Active	52.3	

Figure 3 – Web based onion services by category/purpose in February 2016 [10]

The dark web, a popular term in the internet world, is used to describe the websites that can only be accessed by specialized browsers. In this case, it refers to the hidden services present on the TOR network, which can be accessed by any user who has a browser that supports onion routing like the TOR browser.

Although TOR is constituted by thousands of onion services that operate for illegal purposes, it's not illegal to access the TOR network or to use the TOR browser, since it also houses thousands of legal websites that need users' anonymity to do simple tasks without surveillance or censorship.

One of the most popular hidden services in the TOR network is Silk Road. It was the biggest online market for buying and selling illegal products, mainly drugs. Since 2012, one year after the launch of the website, the annual sales would reach an estimated 22 million dollars, which is just a fraction of all the money sent back and forth using the TOR and Blockchain technology together. [13]

While using TOR is much more secure than using the normal internet because of the additional procedures and features, TOR users are not completely safe. This was explained in the previous chapter, where we exposed some of the many weaknesses TOR presents. TOR's main flaw is the fact that eavesdroppers can identify users if they're listening for traffic between the exit node and the onion service, if there's no encryption mechanism like TLS or SSL and if the client sends personal information, like a username or a password, in the payload.

5. Conclusions

With this report, we can conclude that the TOR network is not only a simple free open-source software to reach the internet, but it’s also a tool that allows users and services to keep anonymity in a digital world full of mass surveillance and data gathering. This is due to its overlayed network, its encryption mechanism present in each circuit hop.

However, these privacy implementations have some performance and security flaws explained in this document, which can break TOR’s main principle: anonymity.

Furthermore, there’s a dichotomy in the use of the TOR protocol, with persons using it legally for political or mediatic purposes and criminals carrying out illegal activities on the network like selling drugs, distributing explicit content and other tasks that tarnish TOR’s reputation every single day.

~

6. References

[1] Wired, "It's time to ditch Chrome," [Online]. Available: https://www.wired.co.uk/article/google-chrome-browser-data. [Accessed June 2022].

[2] Mozilla, "What is a web browser," [Online]. Available: https://www.mozilla.org/en-US/firefox/browsers/what-is-a-browser/. [Accessed June 2022].

[3] CNN, "Your browser history is for sale, here's what you need to know," [Online]. Available: https://money.cnn.com/2017/04/05/technology/online-privacy-faq/. [Accessed June 2022].

[4] Tor Project, "History," [Online]. Available: https://www.torproject.org/about/history/. [Accessed June 2022].

[5] TorMetrics, "Users," [Online]. Available: https://metrics.torproject.org/userstats-relay-country.html.. [Accessed June 2022].

[6] Skerritt, "How to Share a Secret (Diffie-Hellman-Merkle)," [Online]. Available: https://skerritt.blog/diffie-hellman-merkle/. [Accessed June 2022].

[7] Massmux, "HOW TOR REALLY WORKS FOR PROTECTING IDENTITY," [Online]. Available: https://www.massmux.com/how-tor-really-works-for-protecting-identity/. [Accessed June 2022].

[8] Tor, "How Do Onion Services Work?," [Online]. Available: https://community.torproject.org/onion-services/overview/. [Accessed June 2022].

[9] Hackernoon, "How does Tor actually work?," [Online]. Available: https://hackernoon.com/how-does-tor-really-work-5909b9bd232c. [Accessed June 2022].

[10] Wikipedia, "Tor (network)," [Online]. Available: https://en.wikipedia.org/wiki/Tor_(network). [Accessed June 2022].

[11] ExpressVPN, "All about Tor, the Tor Browser, and the Dark Web," [Online]. Available: https://www.expressvpn.com/blog/tor/. [Accessed June 2022].

[12] Tecmint, "Tor Browser: An Ultimate Web Browser for Anonymous Web Browsing in Linux," [Online]. Available: https://www.tecmint.com/tor-browser-for-anonymous-web-browsing/. [Accessed June 2022].

[13] Wikipedia, "Silk Road (marketplace)," [Online]. Available: https://en.wikipedia.org/wiki/Silk_Road_(marketplace). [Accessed 2022 June].