Anatomy of a Web Connection: A Brief Analysis

Author: **Vasco António Lopes Ramos**

Date: **30/01/2021**

Index

1. Introductory Note

This document provides a basic and plausible identification of the technologies, processes, actors and business models involved in a web connection.

Furthermore. it approaches web connections on a full spectrum of scope retrieving conclusions and ideas about social and economic implications of the web.

2. Summary

In the Linux Operative System, there are two main approaches to solve this kind of problems: **tracepath** and **traceroute**. Both are diagnostic tools, that identify the route hops followed by the connection and measures transit delays of packets across IP networks.

The **tracepath** tool is default one provided by the operating system, but it has much fewer permissions and capabilities, so, provides a simpler answer to the problem.

On the other hand, **traceroute** needs to be installed by the user and requires superuser permissions, but it is more powerful, since it has the option to change the packets that are sent/received, and provides a more detailed answer than the tracepath tool.

With that in mind, the tool proposed to solve the problem is the **traceroute**.

To ensure that we had enough variety of information, the connection to the website `www.cmu.edu` was established through the university (University of Aveiro) network (eduroam) and my home network (Vodafone as ISP). Moreover, the connection was established in different periods to observe if there was any relation to it.

3. Framework

The utilization of the web is currently present in many circumstances of our daily lives. When we sit in front of a computer and open a browser an immense cyberworld is opened in front of us. For that to be possible many technologies, processes and actors are involved. Also, behind of what we see, many other operations take place, some of them not perceived by the ”common user”. These operations might have profound social and economic implications.

4. The Mechanism of a Connection

To get a clearer idea of how a web connection is created, we have made some procedure decisions:

1. The connection will be made to a specific website: `www.cmu.edu`.
2. The connection will be made in two attempts: one from inside the University of Aveiro and another one, outside of the University.
3. It’s going to be used a tool - traceroute [1].

[1] The Linux equivalent to Microsoft’s ”tracert”. Usage can be found at: `https://linux.die.net/man/8/traceroute`

4.1. Creation of the connection using *traceroute*

According to the source [6]:

> "...
> Traceroute tracks the route packets taken from an IP network on their way to a given host. It utilizes the IP protocol's time to live (TTL) field and attempts to elicit an ICMP TIME_EXCEEDED response from each gateway along the path to the host.
> ..."

This means that this tool prints the route that a packet takes to reach the host, allowing us to know about the route and about all the hops that a packet takes.

Therefore, the usage of the traceroute tools in the case of `www.cmu.edu`, presents the following result:

```
traceroute to www.cmu.edu (128.2.42.52), 30 hops max, 60 byte packets
 1  gt1-edu-alunos.core.ua.pt (192.168.63.252)  2.143 ms  3.217 ms  3.224 ms
 2  gt1-vrfinternet-r.core.ua.pt (193.137.173.244)  3.224 ms  3.227 ms  3.229 ms
 3  nx2-ibgp.core.ua.pt (10.0.34.1)  2.123 ms  2.128 ms  2.513 ms
 4  Router42.Porto.fccn.pt (193.136.4.26)  2.530 ms  2.524 ms  2.526 ms
 5  Router43.Porto.fccn.pt (193.137.4.18)  3.182 ms  3.329 ms  3.302 ms
 6  Router60.20GE.DWDM.Backbone2.Lisboa.fccn.pt (193.136.4.1)  7.624 ms  7.050 ms  7.016 ms
 7  ROUTER1.10GE.CR1.Lisboa.fccn.pt (193.137.0.1)  7.178 ms  6.989 ms  6.962 ms
 8  fccn.mx2.lis.pt.geant.net (62.40.124.97)  6.172 ms  6.535 ms  6.543 ms
 9  ae0.mx1.mad.es.geant.net (62.40.98.107)  22.188 ms  22.594 ms  22.606 ms
10  ae3.mx1.par.fr.geant.net (62.40.98.65)  35.232 ms  35.674 ms  35.640 ms
11  et-2-1-5.102.rtsw.newy32aoa.net.internet2.edu (198.71.45.236)  108.631 ms  107.952 ms  107.763 ms
12  et-4-0-0.4079.rtsw.phil.net.internet2.edu (162.252.70.103)  109.409 ms  109.272 ms  110.025 ms
13  204.238.76.33 (204.238.76.33)  110.710 ms  111.647 ms  110.400 ms
14  204.238.76.46 (204.238.76.46)  110.360 ms  110.036 ms  110.094 ms
15  162.223.17.79 (162.223.17.79)  138.442 ms  125.558 ms  126.328 ms
16  CORE255-POD-I-DCNS.GW.CMU.NET (128.2.255.193)  120.677 ms  122.984 ms  123.944 ms
17  POD-D-CYH-CORE255.GW.CMU.NET (128.2.255.202)  119.386 ms  121.551 ms  121.500 ms
18  WWW-CMU-PROD-VIP.ANDREW.CMU.EDU (128.2.42.52)  120.213 ms  119.237 ms  119.228 ms
```

4.2. Analysis of the results provided by *traceroute*

Even using traceroute, the tool can only do so much for us. The responsibility of analysing and conclude the results is ours. Thus, the first thing we need to properly understand is the concept of the hop.
In the subsection above, agreeing with the source referenced, I used the term "hop" without any clarification: each hop represents a jump between pairs of equipment. The successive hops transport the data from the source of the request to the requested (final) destination, in our example: `www.cmu.edu`.

4.2.1. Interpretation of traceroute results

Using the following table, it'll be presented a somewhat detailed analysis and interpretation of each hop listed in the result of traceroute.

Hop	Device or Media	Local	Network/ Operator/Owner	Technologies/ Protocols	OSI Layer
0	Personal Computer (192.168.1.71)	GSBL UA	UA Ethernet Network/ STIC/UA	HTTP Port: XXXX TCP IPV4 WiFI-IEEE802.11x Free-Space Radio	7 - Application 5 - Session 4 - Transport 3 - Network 2 - Data Link 1 - Physical
Transport		UA	Free-Space Radio		
1	Router (192.168.63.252)	DETI UA	UA Ethernet Network/ STIC/UA	IPV4 FastEthernet (802.2; 802.3) 100BASE-T (IEEE 802.3)	3 - Network 2 - Data Link 1 - Physical
Transport		UA	OPTICAL FIBRE (Gigabit Ethernet)		
2	Router (193.137.173.244)	DETI UA	UA Ethernet Network/ STIC/UA	IPV4 FastEthernet (802.2; 802.3) 100BASE-T (802.3)	3 - Network 2 - Data Link 1 - Physical
Transport		UA	OPTICAL FIBRE CAMPUS BACKBONE		
3	Router (10.0.34.1)	Exterior Gateway UA	UA Ethernet Network/ STIC / UA	BGP IPV4 FastEthernet (802.2; 802.3) 100BASE-T (IEEE 802.3)	7 - Application 3 - Network 2 - Data Link 1 - Physical
Transport		Linha do Norte	OPTICAL FIBRE FCCN BACKBONE (GIGABIT INTERNET)		
4	Router (193.136.4.26)	Estação Porto Campanhã	RCTS IP/ FCCN / REFER	IPV4 PPP (??) GE, OTN, SDH, SONET, ...	3 - Network 2 - Data Link 1 - Physical
Transport		Porto	UTP/ Optic Fibre Building-Cable (Ethernet)		
5	Aggregator Router Porto (193.137.4.18)	Porto	RCTS IP/ FCCN / FCCN	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		Linha do Norte	OPTICAL FIBRE FCCN BACKBONE		
6	Aggregator Router Lisboa (193.137.4.18)	Lisboa	RCTS IP/ FCCN / REFER	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical

Hop	Device or Media	Local	Network/ Operator/Owner	Technologies/ Protocols	OSI Layer
Transport		Linha do Norte	OPTICAL FIBRE FCCN BACKBONE		
7	Aggregator Router Lisboa (193.137.0.1)	Lisboa	RCTS IP/ FCCN / FCCN	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		Lisboa	OPTICAL FIBRE FCCN BACKBONE		
8	Gateway GEANT Router Lisboa (62.40.124.97)	Lisboa	GEANT/ FCCN / NREN	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		Lisbon-Espanha	GEANT EUROPEAN BACKBONE		
9	Gateway GEANT Router Madrid (62.40.98.107)	Madrid	GEANT/ RedIRIS / RED.ES	IPV4 20 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		Espanha-França	30 GIGABIT GEANT LINE		
10	Gateway GEANT Router Paris (62.40.98.65)	Paris	GEANT/ RENATER / RENATER	IPV4 30 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		França-USA	100 GIGABIT GEANT LINE		
11	Internet 2 Router Proxy (198.71.45.236)	New York	Internet2/ Internet2 / Internet2	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		USA	OPTICAL FIBER (Gigabit Ethernet)		
12	Internet 2 Router Proxy (162.252.70.103)	Philadelphia	Internet2/ Internet2 / Internet2	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
Transport		USA	OPTICAL FIBER (Gigabit Ethernet)		

Hop	Device or Media	Local	Network/ Operator/Owner	Technologies/ Protocols	OSI Layer
13	Drexel University Router (Entry Point) (204.238.76.33)	Philadelphia	DREXEL-N3/ ??? / Drexel University	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
	Transport	Drexel University	OPTICAL FIBER (Gigabit Ethernet)		
14	Drexel University Router (Entry Point) (204.238.76.46)	Philadelphia	DREXEL-N3/ ??? / Drexel University	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
	Transport	Philadelphia	OPTICAL FIBER (Gigabit Ethernet)		
15	Kinber Router/Gateway (162.223.17.79)	Philadelphia	PennREN/ Kinber / Kinber	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
	Transport	US	OPTICAL FIBRE CMU (Gigabit Ethernet)		
16	Carnegie Mellon Uiversity (128.2.255.193)	Pittsburgh	CMU/ CMU CMU Hostmaster	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
	Transport	Pittsburgh	OPTICAL FIBRE CMU (Gigabit Ethernet)		
17	Carnegie Mellon Uiversity (128.2.255.202)	Pittsburgh	CMU/ CMU CMU Hostmaster	IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	3 - Network 2 - Data Link 1 - Physical
	Transport	Pittsburgh	OPTICAL FIBRE CMU (Gigabit Ethernet)		
18	Carnegie Mellon Uiversity (128.2.42.52)	Pittsburgh	CMU/ CMU CMU Hostmaster	HTTP Port : 80 ICMP IPV4 10 Gigabit Internet 10GBase (IEEE 802.3aX)	7 - Application 5 - Session 4 - Transport 3 - Network 2 - Data Link 1 - Physical

Table 1: Analysis/Interpretation of traceroute results

4.2.2. The Possible Situation of Request TimeOut

In the traceroute result presented above, there were no situations of **Request TimeOut**. However, this situation can occur in multiple scenarios:

- The destination's firewall or other security devices (e.g: proxy or protected gateway) is blocking the request. This case doesn't mean that the destination is unreachable by the applications commonly used (web/HTTP); it just means that the firewall is preventing the final hops at the destination from showing up in traceroute output.
- There may be a connection problem at that particular system or the next system.
- There could be a problem on the return path from the target system: the forward route and the return route often follow different paths, therefore, if there is a problem on the return route, it may not be evident in the command output.

4.2.3. Traceroute Logs and Origin Request Location

As referenced in the *Summary*, the traceroute tool was used to analyse the path of `www.cmu.edu` in different networks and times of the day. One of the results was already shown in the **sub-section 4.1**. The other one, originated from my home network is presented below:

```
traceroute to www.cmu.edu (128.2.42.52), 30 hops max, 60 byte packets
 1  vodafonegw (192.168.1.1)  1.935 ms  5.440 ms  5.439 ms
 2  2.64.54.77.rev.vodafone.pt (77.54.64.2)  6.574 ms  7.308 ms  8.586 ms
 3  113.41.30.213.rev.vodafone.pt (213.30.41.113)  8.583 ms  10.262 ms  10.264 ms
 4  195.10.48.9 (195.10.48.9)  13.571 ms  14.363 ms  14.946 ms
 5  ae6-xcr1.mat.cw.net (195.2.30.238)  20.681 ms  22.362 ms  22.364 ms
 6  ae0-xcr1.mtt.cw.net (195.2.8.149)  22.736 ms  21.263 ms  19.961 ms
 7  cogent-gw-xcr1.mtt.cw.net (195.2.19.6)  19.232 ms  20.911 ms  19.376 ms
 8  be3480.agr21.mad05.atlas.cogentco.com (154.25.1.17)  22.824 ms  22.834 ms  22.829 ms
 9  be3378.ccr32.mad05.atlas.cogentco.com (154.54.36.173)  22.806 ms  25.426 ms  25.433 ms
10  be2325.ccr32.bio02.atlas.cogentco.com (154.54.61.133)  26.614 ms  25.712 ms  25.673 ms
11  be2332.ccr41.iad02.atlas.cogentco.com (154.54.85.245)  110.333 ms  110.350 ms  110.350 ms
12  be3084.ccr42.dca01.atlas.cogentco.com (154.54.30.65)  107.563 ms  110.333 ms  111.171 ms
13  be2820.rcr21.pit02.atlas.cogentco.com (154.54.83.54)  116.300 ms  119.503 ms  119.514 ms
14  te0-0-2-3.nr11.b015486-0.pit02.atlas.cogentco.com (154.24.50.198)  119.514 ms  119.515 ms  122.013 ms
15  38.140.44.154 (38.140.44.154)  119.509 ms  120.056 ms  120.616 ms
16  CORE0-POD-I-CYH.GW.CMU.NET (128.2.0.249)  121.976 ms  118.303 ms  116.228 ms
17  POD-D-DCNS-CORE0.GW.CMU.NET (128.2.0.210)  117.001 ms  117.733 ms  118.312 ms
18  WWW-CMU-PROD-VIP.ANDREW.CMU.EDU (128.2.42.52)  200.411 ms  201.419 ms  201.993 ms
```

Comparing both results, it is clear that the path taken from the Univesity of Aveiro to CMU's website is completely different from the one originated from my home. If we pay attention to detail, it is clear that, apart from the last 3 hops (located at the Carnegie Mellon University), all remaining hops are substantially different. The reason for this is that each operator has its network (graph) of connections to the Internet which can or cannot be the same. In this particular case, they're not.

Related to the times of the day, these operations were performed, the only spotted difference is regarding the response time, in ms, that revealed to be lower at night.

4.3. Protocols, Mechanisms and Tools

4.3.1. Web Browser

Nowadays browsers are the face of the World Wide Web (from now on plainly designated **the web**) since just opening a web page is enough to access it and benefit from a huge repository of information (images, videos, documents, among others), scattered in computers around the world.

So, how do this works?

1. The web browser (which is a piece of software capable of fetching and rendering information resources from different locations throughout the web and also of travelling across the web looking for the desired source determined by the appropriate address, URL) goes to DNS server and finds the real address of the server where the website lives on.
2. The web browser sends an HTTP request to the server, asking for a copy of the website. This message and all the data sent between the server and the client is sent across an internet connection, using TCP/IP.
3. If the server accepts the client's request, it sends the client a **200 OK** message and starts sending the website's files to the browser as a series of small chunks of data named packets.
4. The browser assembles the small chunks into a complete document and displays it to the end-user.

4.3.2. DNS

"Domain Name System (DNS) is a hierarchical decentralized naming system for computers, services or other resources connected to the Internet or a private network." [9]

Taking into account the quote from above, Domain Name Servers are like an address book for websites. When it's inserted a web address in the browser, the browser looks at the DNS to find the website's real address so it can fetch the actual website data, because it needs to find out which server the website lives on.

DNS servers allows humans to use the internet without having to remember IP addresses such as 193.137.168.32 (IPv4) or 2001:db8:85a3::8a2e:370:7334 (IPv6).

4.3.3. TCP/IP

Transmission Control Protocol and Internet Protocol are communication protocols that define how data should travel across the web. Its purpose is to ensure that all received bytes at one end of a communication system are identical to the bytes that are sent from the other end and are in the correct order. This involves the following aspects:

1. Data handling and processing (streams, segments and sequence numbers)
2. Data transport, reliability and flow control
3. Management of ports, connections and connection identification

4.3.4. HTTP

Hypertext Transfer Protocol (HTTP) is an application-layer protocol that defines a language for clients and servers to speak to each other and it is responsible for transmitting hypermedia documents, such as HTML.

HTTP is a stateless protocol, which means that the connection is closed as soon as the request has been made. The three main HTTP messages are: **GET**, **POST** and **HEAD**.

4.3.5. ICMP

Internet Control Message Protocol (ICMP) is an Internet-layer protocol used by network devices to diagnose network communication issues.

ICMP is mainly used to determine whether or not data is reaching its intended destination promptly. Commonly, the ICMP protocol is used on network devices, such as routers.

The ICMP message contains certain fields such as: **TYPE**, **CODE** and **CHECKSUM** to help identify a certain response message.

5. Social and Economic Implications

Without a doubt, one of the greatest achievements of mankind was the ability to live in a society. By definition, living in a society means that everyone can change one's life. What this means is that every action always has an impact, either positive or negative, on other people's lives. That's the reason why concepts like **social networking and data privacy** and also **net neutrality** are so immeasurably important.

Nowadays, the internet is dominated by big multinational corporations, so, the actions taken by these companies affect millions of people every day and, unfortunately, the decisions taken by these multinationals are always driven by greedy and unethical intentions.

This constantly puts at risk the important concept of **net neutrality**. One of the best examples is the dissemination of online **advertisement** and **"recommendation algorithms"**.

Let's start with the online advertisement case. Today, we have advertisement everywhere: on our music (Spotify), on our videos (YouTube), on our online research (common webpages) and much more. However, these types of advertisement aren't always ethical. For example, when we are on YouTube watching some movie trailer or an info-graphic-learning video, we are already paying the internet bill that lets us access every content on the internet, however, we are faced with advertisement because "we don't pay for that service so we have to watch ads". That is not true, I paid for the internet and I am using the internet, I shouldn't need to "pay" anything more. In that sense, those ads were imposed on me without my consent. Is this ethical? Or fair? Maybe not. One similar example is the Spotify approach to the advertisement.

Still on this subject about advertisement, one other example is Google. Google's core business is advertisement, some results are artificially inserted in every page result because the owners of those pages paid for it, this part may be accepted as ethical. On the other hand, Google has developed a whole ecosystem around the web (Google chrome, Google Web Services, Android, etc) and uses this ecosystem to impose its services in every platform to every user. If we remember, one company did a similar thing that originated a bunch of lawsuits in the 90s and early 2000s: Microsoft and the Antitrust Lawsuit. This is what needs to happen to Google and other big tech companies globally (it already started to happen in Europe): force these companies to give the user a choice. Force these companies to split down because enormous tech giants don't make good things to the web. They polarize the internet and impose themselves to a point where there is no other choice, disturbing completely the so important net neutrality responsible for assuring that the internet is fair and everyone gets the same opportunity to show the world their business, ideas e thoughts.

The last thing I would like to talk about is the recommendation systems implemented everywhere these days because they jeopardize two important things: **education** and **privacy**. These kinds of systems are presented to the users as the holy grail of innovation, but all they do is collect data, create patterns and feed information taking those patterns into account. And this is bad because, on one hand, those broad pattern analysis can make a lot of damage if shared with the wrong organisations (for example, the usage of information and

patterns about one country's health could be devastating in the sense that those patterns provide accurate information about how to effectively kill every person on that country). On the other hand, **education**: highly customised feeds of news and information (intelligent news providers - Google News - or Facebook social network) narrow the length and diversity of information everyone gets, which threatens multicultural societies and promotes extremist behaviours and thoughts.

These are serious concerns because if no action is taken or people are not aware of this problems/dangers, we could end up jeopardising our civil and humanitarian way of living and destroying the culture of acceptance and respect that so many people died fighting for.

6. Conclusions

Considering the analysis performed in this document, it is reasonable to conclude that although traceroute is a powerful tool, it's not complete. If we want to know more information about each hop, as it was in this assignment, we need to do an additional online search for that information.

Furthermore, it was possible to understand that each web connections flows from the source to the destination hopping through multiple companies, countries and policies, which is a very interesting concept because it raises a lot of questions and concerns about politics, economics and, as it was mentioned above, a risk to net neutrality.

References

[1] A. Manuel de Oliveira Duarte, *Network Challenges*, [Online]. Available: eLearning.

[2] A. Manuel de Oliveira Duarte *Anatomy of a Network Connection - Web and OTT cases*, DETI-UA, 2018.

[3] A. Manuel de Oliveira Duarte, *Networks Overview*, [Online]. Available: eLearning.

[4] M. Newman *Networks: An Introduction*, Oxford University Press, 2010.

[5] *Layers of OSI Model*, [Online]. Available: `https://www.geeksforgeeks.org/layers-of-osi-model`.

[6] *Traceroute - Linux Man Page*, [Online]. Available: `https://linux.die.net/man/8/traceroute`.

[7] *RCTS IP Connectivity Service - FCCN*, [Online]. Available: `https://www.fccn.pt/en/connectivity/rcts-ip/`.

[8] *How the Web Works - Learn web development*, [Online]. Availabe: `https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/How_the_Web_works`.

[9] *Domain Name System - Wikipedia*, [Online]. Availabe: `https://en.wikipedia.org/wiki/Domain_Name_System`.

[10] D. Easley, J. Kleinberg *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

[11] Paul DiMaggio, Eszter Hargittai, W. Russell Neuman, John P. Robinson, *Social Implications of the Internet*, Annual Review of Sociology 27:1, 307-336, 2001.

[12] George Reynolds, *Ethics in Information Technology*, Cengage Learning, ISBN-13: 97885197159 / ISBN-10: 1285197151, 2015.

[Acedido em 30/01/2021].