Anatomy of a Web Connection: A Brief Analysis
Author: Tomé Carvalho 97939
Date: 19 March 2022

Index

1. Objectives

The main objectives of this assignment are to provide a plausible identification of the technologies, processes, actors and business models involved in a web connection, as well as some of the possible social and economic implications associated with the identified entities. The Steam Store will be used as an example.

2. What are the Essential Steps Involved in a Connection to a Website?

2.1 Tracing the connection using the *tracert* tool

tracert traces the path data packets take from their source to their destination. Data must hop through multiple devices, such as routers and switches, when transmitted between two points. This tool maps each hop and provides its details: round-trip time (RTT), device name and IP address (when possible).
On Windows, *tracert* sends ICMP (Internet Control Message Protocol) Echo Request packets. Relying on the time-to-live (TTL) mechanism, traceroute uses the returned ICMP Time Exceeded messages to build a list of routers that packets traverse.[1]

```
PS C:\Users\tomec> tracert store.steampowered.com

Tracing route to store.steampowered.com [23.49.245.62]
over a maximum of 30 hops:

  1     8 ms     7 ms     7 ms  gt2-edu-alunos.core.ua.pt [192.168.63.253]
  2     6 ms     4 ms     9 ms  10.1.0.118
  3     4 ms     6 ms     5 ms  gt1-vrfinternet-r.core.ua.pt [193.137.173.244]
  4     3 ms     3 ms     2 ms  nx2-ibgp.core.ua.pt [10.0.34.1]
  5     3 ms     3 ms     7 ms  Router41.Porto.fccn.pt [193.136.4.26]
  6     8 ms    12 ms    11 ms  Router40.Porto.fccn.pt [194.210.7.208]
  7    10 ms    10 ms     7 ms  Router60.Lisboa.fccn.pt [193.136.1.10]
  8     7 ms     8 ms     6 ms  Router6.Lisboa.fccn.pt [194.210.6.205]
  9    21 ms    11 ms    14 ms  Akamai.AS20940.gigapix.pt [193.136.250.70]
 10    55 ms     6 ms    10 ms  a23-49-245-62.deploy.static.akamaitechnologies.com [23.49.245.62]
```

Performed at around 11:30, 2022-03-15 at DETI, University of Aveiro.

2.2 A possible interpretation into some detail

This table shows a possible interpretation of the results of the above command.

Hop	Name	IP Address	Location	Network/Operator/Owner
1	gt2-edu-alunos.core.ua.pt	192.168.63.253	Aveiro	UA Network/STIC/UA
2		10.1.0.118	Aveiro	UA Network/STIC/UA
3	gt1-vrfinternet-r.core.ua.pt	193.137.173.244	Aveiro	UA Network/STIC/UA
4	nx2-ibgp.core.ua.pt	10.0.34.1	Aveiro	UA Network/STIC/UA
5	Router41.Porto.fccn.pt	193.136.4.26	Porto	FCCN
6	Router40.Porto.fccn.pt	194.210.7.208	Porto	FCCN
7	Router60.Lisboa.fccn.pt	193.136.1.10	Lisbon	FCCN
8	Router6.Lisboa.fccn.pt	194.210.6.205	Lisbon	FCCN
9	Akamai.AS20940.gigapix.pt	193.136.250.70	Olival do Basto, Lisbon	GigaPix/FCCN
10	a23-49-245-62.deploy.static.akamaitechnologies.com	23.49.245.62	Lisbon	Akamai

As for the technologies used, the IPv4 communication protocol is used at OSI[12] layer 3 (Network). ICMP is situated at layer 4 (Transport). Since the Steam Store website uses HTTPS, we can infer that port 443 is used[8], an aspect of layer 5 (Session).

```
PS C:\Users\tomec> tracert store.steampowered.com

Tracing route to store.steampowered.com [2.21.168.87]
over a maximum of 30 hops:

  1     3 ms     3 ms     3 ms  vodafonegw [192.168.1.1]
  2     7 ms     7 ms     6 ms  2.0.60.94.rev.vodafone.pt [94.60.0.2]
  3   346 ms    15 ms    12 ms  24.50.174.83.rev.vodafone.pt [83.174.50.24]
  4     *        *        *     Request timed out.
  5    12 ms    11 ms    11 ms  a2-21-168-87.deploy.static.akamaitechnologies.com [2.21.168.87]
```

Performed at around 11:24, 2022-03-16 at Cais da Malhada, Aveiro.

Different locations lead to different logs. The first difference we can see is at the local level. This operation was performed on a Vodafone network. As such, the first hop seen is *vodafeonegw*. This is also the local page where router configuration can be performed. The next two hops are Vodafone servers located in Lisbon. Unfortunately, the fourth request timed out.

2.3 What happens during a typical web session

2.3.1 Browser

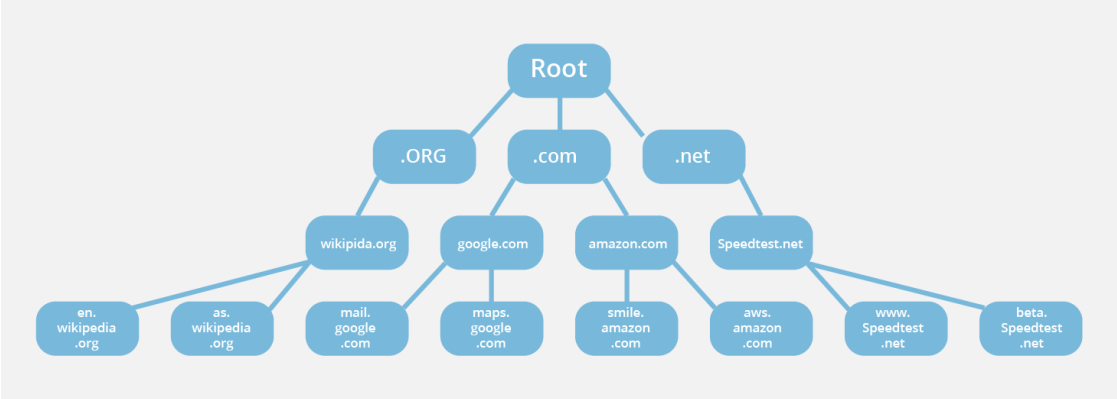
To begin a web session, a user typically opens a browser. When a page from a website is requested, the browser fetches the necessary content from a web server and then displays the page on the device. This process starts when a URL (Uniform Resource Locator) is input into the browser.

The typical format of a URL is *https://www.google.com/*. The first part (*https*, in this case), specifies the protocol to be used. Virtually all URLs start with either *http* (Hypertext Transfer Protocol) or *https* (HTTP Secure). The hostname (*www.google.com*) follows this. Lastly, there's the name of the resource to be accessed.[2] In this example, we don't specify anything, so Google will give us what they have set up for the root (Google Search). We could, however, for example, input *https://www.google.com/maps*, which would take us to Google Maps.

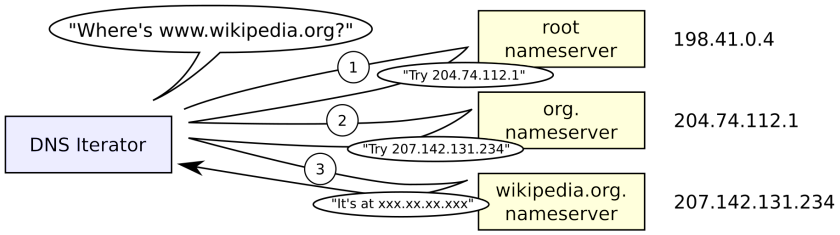
2.3.2 DNS

URLs are human-friendly, but machines must convert them into numerical IP addresses (unique identifiers), such as *172.217.168.174* for *https://www.google.com/*. This conversion is done through a Domain Name System (DNS).

There are a multitude of DNS servers. Each ISP features its own default server, but users may configure their system to use a different one, such as Cloudflare's 1.1.1.1, for reasons such as speed, privacy and circumventing the censoring of websites by authorities (for example, when the government of Turkey blocked Twitter after recordings showing a government corruption scandal leaked online).[3]



DNS Hierarchy[4]



DNS lookup example[5]

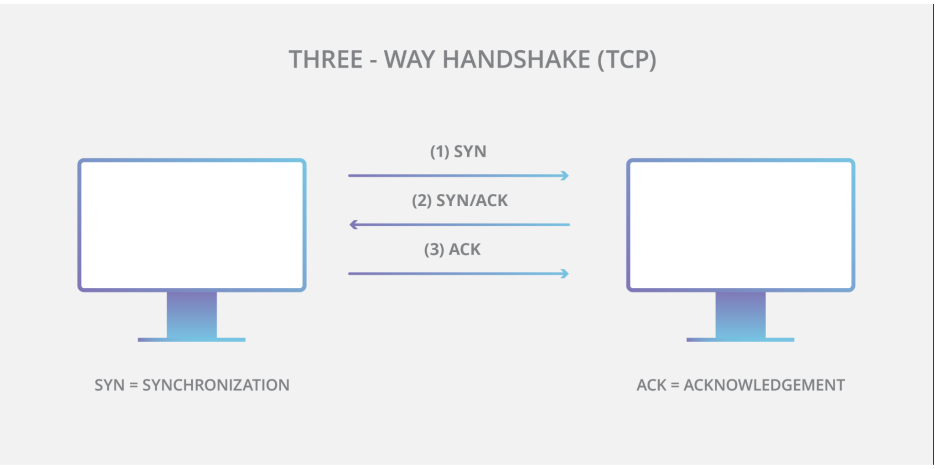
2.3.3 TCP/IP

The Internet Protocol (IP) is the address system of the Internet. Its purpose is to deliver packets from a source device to a target device. It is the primary way of making network connections, and establishing the basis of the Internet. The main version of IP used today is IP Version 4 (IPv4). A newer protocol, IPv6, was developed because of size constraints with the total number of possible addresses in IPv4. It is increasing in adoption.

IP does not handle packet ordering or error checking. For this purpose, another protocol is required, usually TCP.

IP is a connectionless protocol. This means each unit of data is individually addressed and routed from the source to the destination, and the latter does not send an acknowledgement

back to the source. TCP is used in conjunction with IP to maintain a connection between the sender and the target and to ensure packet ordering.[6]



2.3.4 HTTP/HTTPS

HTTP is an application layer protocol designed to transfer information between networked devices. It runs on top of other layers of the network protocol stack. A typical flow over HTTP involves a client sending a request (the way platforms like browsers ask for the information they need to load a website) to a server, which then sends a response message to the former.

Each request contains encoded data with different types of information: HTTP version type, URL, HTTP method, HTTP request headers and, optionally, HTTP body. HTTP methods indicate the action a request expects from the server. The two most common methods are GET (resource fetching) and POST (submission of information). Responses contain an HTTP status code, HTTP response headers and, optionally, HTTP body.

In client-server protocols, the client establishes the connection. Opening a connection in HTTP means initiating a connection in the underlying transport layer, usually TCP.

With TCP, port 80 is the default port for an HTTP server. Others can also be used, like 8000 or 8080. The URL of a page to fetch contains both the domain name and the port number, though the latter can be omitted if it is 80.[7]

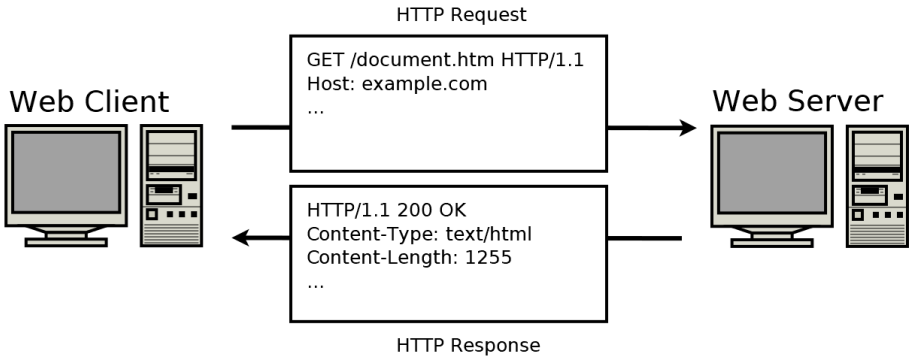


Image source: O'Reilly

HTTPS is an extension of HTTP, where the communication protocol is encrypted using Transport Layer Security (TLS). It's used for the authentication of the accessed website and protection of the privacy and integrity of the exchanged data while in transit. It provides protection against man-in-the-middle attacks, eavesdropping and tampering.[8]

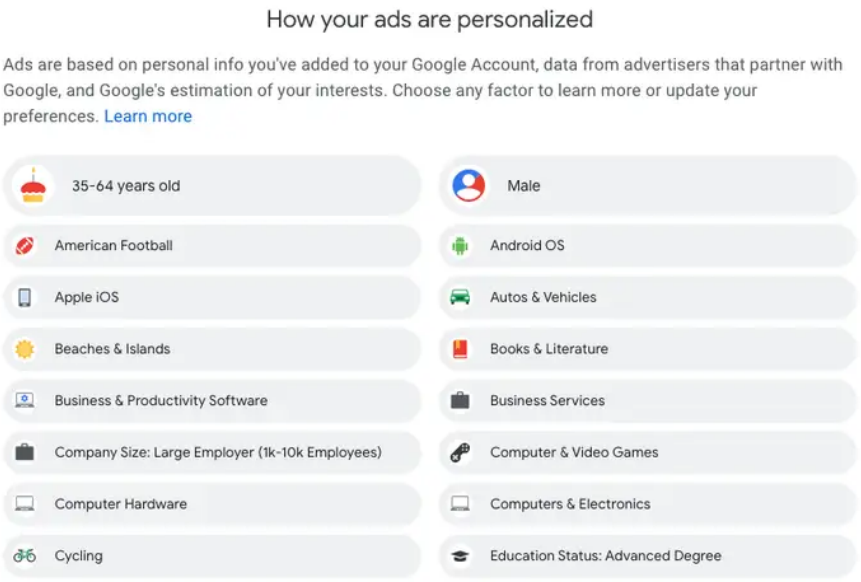
2.3.5 Advertising and Recommendations

Quite often, when users use services for free, they are in reality paying not with money, but with their personal information and attention. Ad networks and websites collect users' data for targeted advertising, that is, advertising directed towards an audience with certain traits, based on what the advertiser is promoting. To afford keeping their services free, these sites profit by selling ads. Advertisers pay websites or ad networks to display their ads, in hope users will be drawn to click on them, which is more probable if targeted advertising is used.

There are multiple types of online advertising.[9]

Display ads are visual ads that appear on web pages in the form of banners or pop-ups, for example. Paid search advertising allows advertisers to place their content at the top of the results page of search engines such as Google Search (through Google Ads, in this concrete example). It is now the largest segment of online advertising. Social media advertising consists in the targeting of a specific audience based on information a social media platform knows about its users (like interests, age, location, etc.). Many social media ads allow the users to interact with the content (liking or commenting, for instance). Ads of this type can take multiple shapes, such as in-feed ads ("disguised" as posts), message ads (in the user's inbox), story ads (ads in the middle of a stream of videos, in apps like Instagram and TikTok) and sponsored posts (where advertisers pay an influencer to use their product on social media, as opposed to purchasing ad space).

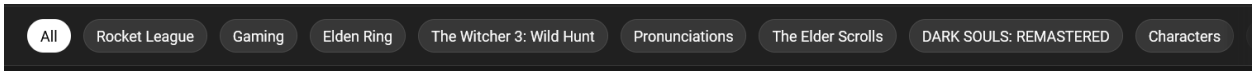
The pros of targeted ads are quite obvious for companies, as they are more effective by appealing to the user's needs and interests, leading to an increase in ROI (return on investment). There are pros for users too. Mainly, they are more likely to discover a product of their interest through targeted advertising, as opposed to seeing. On the other hand, one of the biggest drawbacks for companies is that it can be difficult to get around privacy policies that keep them from acquiring the necessary information for targeting. As for users, ads can sometimes get too personal, making them aware of how much information they are actually offering these companies. They can also be annoying and repetitive (for example, a user might keep receiving ads for phones after purchasing one, due to their search history, even though they are no longer of interest to them). Additionally, users who value their privacy are less likely to visit websites known for employing these data acquisition practices, which, needless to say, is bad for the company the website belongs to.[10]



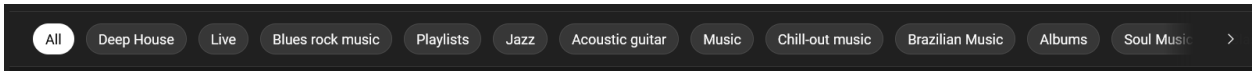
Example of a user's information acquired by Google, including demographic information and estimation of interests.[11]

The widespread adoption of these advertising practices has led to the development of many tools with the objective of enhancing users' privacy and combating excessive advertising. The image on the left, taken from Privacy Tools, shows multiple browser extensions with such purpose.

Targeted advertising and recommendations go hand in hand, as they both make use of the same type of information about the users. In the case of recommendations, instead of it being used for ads, it's used to provide users with more relevant suggestions in cases like searching and content discovery.



YouTube recommendations (logged in)



YouTube recommendations (Private Firefox window, signed out)

As demonstrated by the two images above, Google uses users' data to suggest video categories on YouTube's home page. In the first image, the categories are related to the interests Google has estimated I have through my searches and history on their websites. In the second image, the categories are the most popular ones among all YouTube users (music videos are among the most watched).

Just like target ads, recommendations have pros and cons. If correctly implemented, they can help the user discover content that interests them. Otherwise, they might even worsen the user's experience, compared to a recommendation-less experience. Just as with personalised ads, this requires the users' data to be collected, reducing their privacy.

2. References

1. traceroute - Wikipedia
2. Web browser - Wikipedia
3. What is 1.1.1.1? | Cloudflare
4. DNS root server | Cloudflare
5. Domain Name System - Wikipedia
6. What is TCP/IP? | Cloudflare
7. What is HTTP? | Cloudflare
8. HTTPS - Wikipedia
9. How Does Online Advertising Work? - Neeva
10. Targeted Advertising: The Good and The Bad | by Maya Frai | ART + marketing
11. What Does Google Know About Me? Click This Link to Find Out.
12. What is the OSI Model? | Cloudflare