

Tuesday 17 May 2022

BASIC PRINCIPLES IN
NETWORKING
ELEC-C7420

Replacement Assignment

How YouTube works?

PSALTAKIS GEORGIOS



Section 1: Goals of the experiments , design criteria and description of test cases

The goals of the experiment are of course to understand how such a large-scale video streaming provider, in this case, Youtube is designed to work so seamlessly and of course the main magic behind its service which is its CDN model.

The models and test case scenarios I performed are of course repeated and cover a long amount of available options to test everything. The first test I initiated was the traceroute test on youtube.com. This was done to understand how many waypoints my request to open youtube will pass through. The results are mapped and explained below analytically for this first test scenario.

The second experiment includes the use of Wireshark to capture the youtube packages of video playback of youtube (since youtube slices off the video to multiple small videos) to identify the processes and the protocols that are in place in the video playback such as which transfer protocol is in place what HTTP it uses as well as the streaming protocol.

The final and most important experiment was of course to identify how the CDN of youtube works and how it affects the video you play depending on your location. To do this I chose specific locations and videos to try out at various times. The locations I chose were Finland in comparison with Cyprus with local videos of Finland and local Greek Videos. But I didn't stop there since I took a look at Japan, Serbia as well as Mexico. This was done to verify that it was not a single-case scenario and to make it scientifically sound.

Section 2: Traceroute of Youtube from Finnish ISP

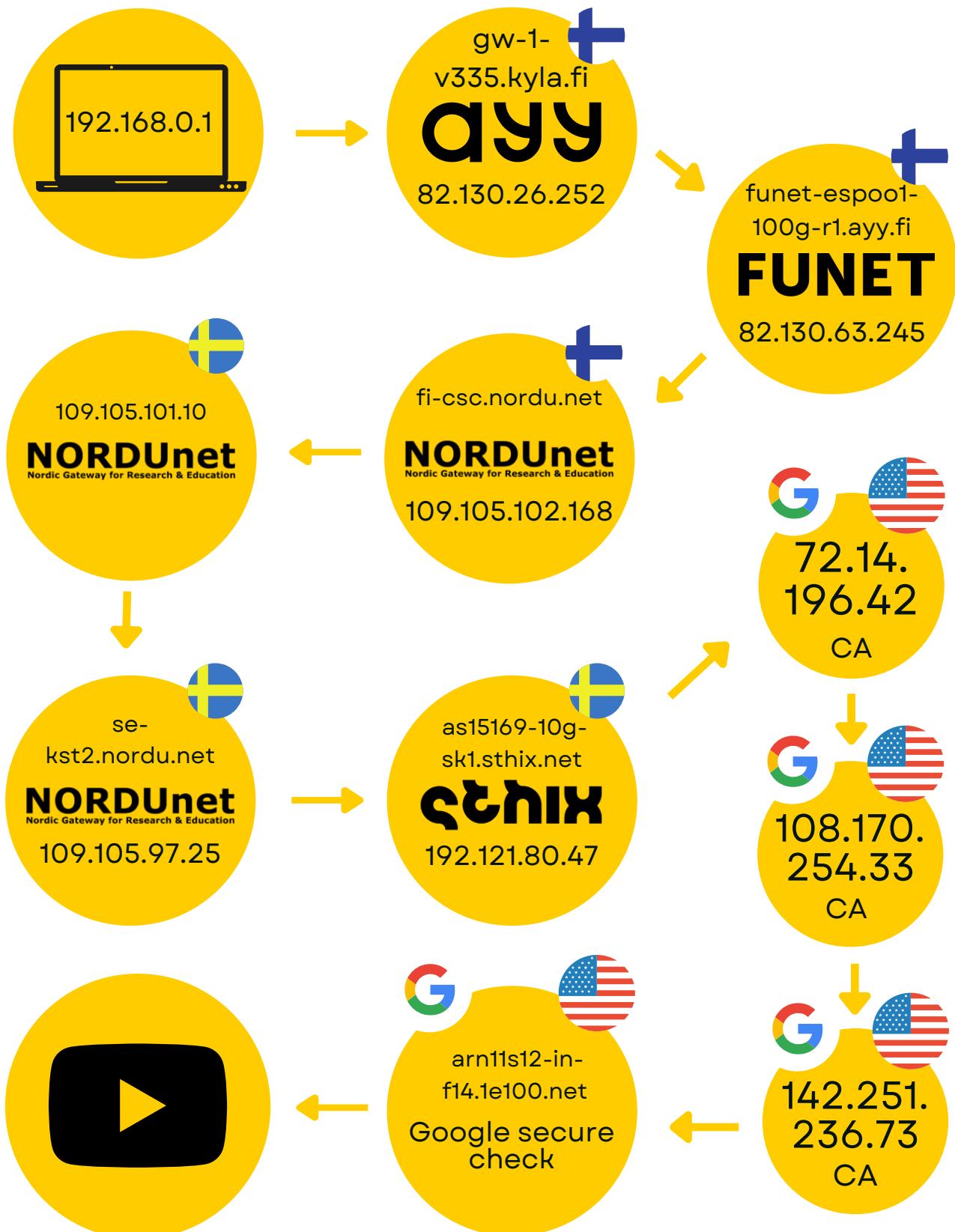
I Begin the assignment with Traceroute and what "directions" my pc takes to connect to youtube.com. One serious topic to take notice of would also be that by doing that Youtube identifies my IP address as Finnish and pushes me to the Finnish Youtube as well as the Nordic CACHE of youtube videos first.

```
georgepsaltakis@Georges-MacBook-Air ~ % traceroute youtube.com
traceroute to youtube.com (142.250.74.174), 64 hops max, 52 byte packets
 1  192.168.0.1 (192.168.0.1)  3.614 ms  3.090 ms  3.457 ms
 2  gw-1-v335.kyla.fi (82.130.26.252)  3.624 ms  3.662 ms  3.137 ms
 3  funet-espool1-100g-r1.ayy.fi (82.130.63.245)  46.643 ms  9.310 ms  3.660 ms
 4  fi-csc.nordu.net (109.105.102.168)  3.323 ms  7.364 ms  3.340 ms
 5  109.105.101.10 (109.105.101.10)  40.885 ms  32.389 ms  14.796 ms
 6  se-kst2.nordu.net (109.105.97.25)  14.533 ms  44.679 ms  14.892 ms
 7  as15169-10g-sk1.sthix.net (192.121.80.47)  33.607 ms
    72.14.196.42 (72.14.196.42)  16.148 ms  15.321 ms
 8  108.170.254.33 (108.170.254.33)  36.468 ms  15.700 ms
    108.170.254.49 (108.170.254.49)  19.352 ms
 9  142.251.236.73 (142.251.236.73)  15.537 ms
    142.251.236.71 (142.251.236.71)  15.341 ms  17.430 ms
10  arn11s12-in-f14.1e100.net (142.250.74.174)  15.512 ms  15.394 ms  36.549 ms
georgepsaltakis@Georges-MacBook-Air ~ %
```

A brief explanation of the Route

Firstly my computer (me) puts a request to connect to youtube.com this is of course firstly leaving through my ISPs which in this case is AYY it then continues to the University Network of Finland FUNET before going to the Finnish side of the Nordic Internet Association NORDUnet. It then moves to the Swedish part and ends in the STHIX servers before going through multiple google IPs in the United States in Californian Data Centers and then before granting entry to my request it goes through the googles security check. This can all be achieved thanks to the DNS which translated my request to visit Youtube into an Address. Below, I created a map of that request.

Section 2: Map of the TraceRoute



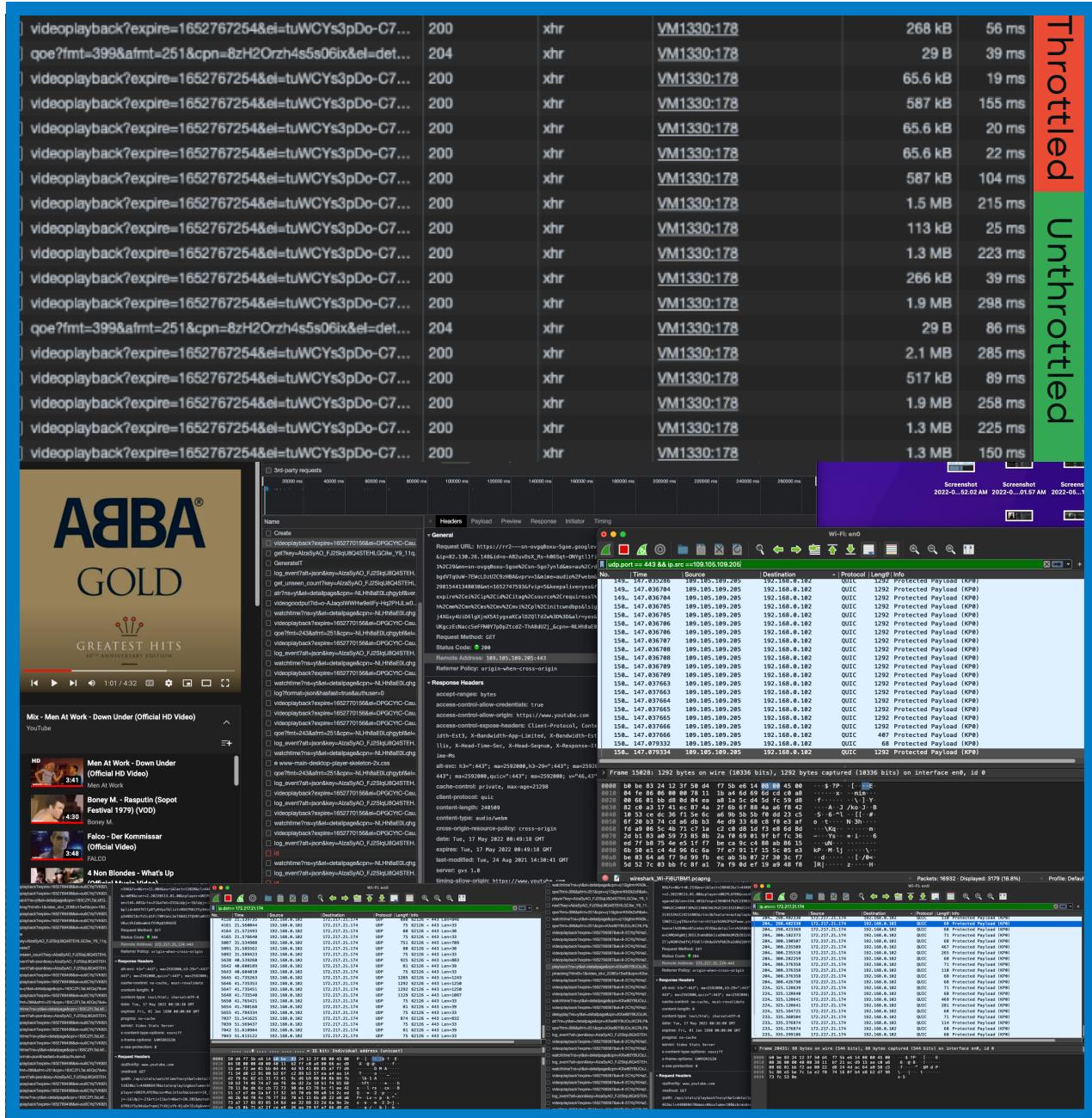
Section 3: Architecture of Youtube and Protocols

Youtube is a large scale Video Streaming Service which includes a giant database with many many videos in place with each one a unique identifying code. The videos can be searched quite easily by the user who is also the one who uploads the content. How Youtube then handles that video is indeed quite magical.

Youtube takes the original uploaded videos and breaks it to many many pieces that are quite small. Then it also downgrades the quality to multiple levels of video quality and has everything available in order. When a user is requesting that video the actual video player does the stitching of the clips to be seen as seamless as possible. Another incredible feature implemented is the automatic video quality upgrade or downgrade completely depended on your available traffic and internet speeds. So if you have for example slow speed it will push you a 360p video. If your internet can seem to handle that quite easily the packets it will bump you up immediately to a higher quality without stopping thanks to the hundreds of pieces of video available all in all the available qualities. The popular content is also served through Caches by the CDN to avoid connection to the main database but more in detail regarding that in the next session.

The quality check was also done by throttling the speed to slow 3G on the browser developer tools and then unthrottling it to full speed to compare the two received packages together which happened automatically the moment I lifted the throttle as seen below.

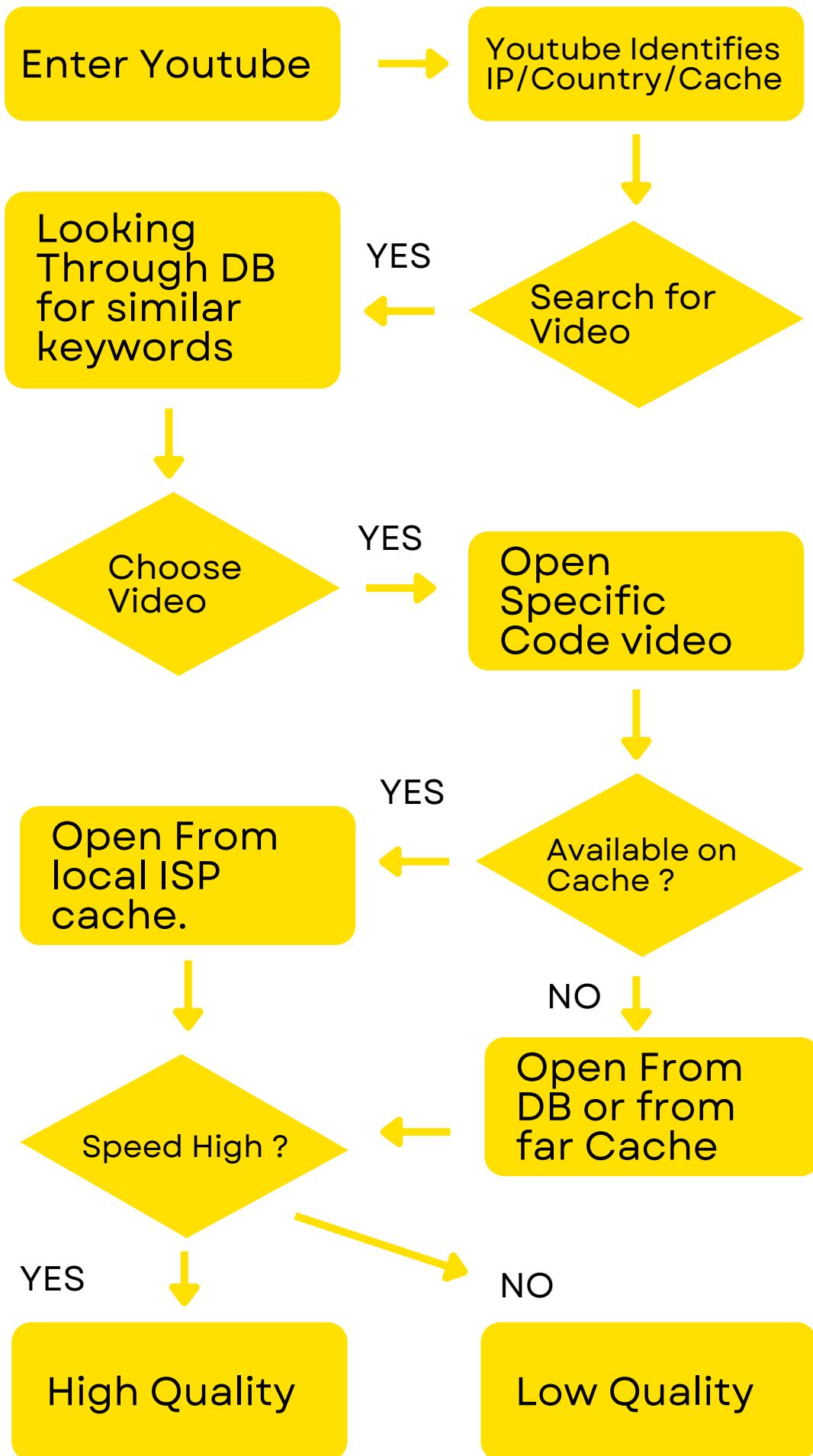
Section 3: Architecture of Youtube and Protocols



The main Transfer protocol for the video playback packets was QUIC for almost all the times. This transport protocol is implemented on UDP and on HTTP/3 plus QUIC is encrypted out of the box as seen by the captured packages. There were also captures on TCP (HTTP) as well as UDP but on rare occasions as it was mostly on QUIC at least 95% of the time. The captured packages were H.264/MPEG-4 AVC but mostly on WebM VP9 as shown in the developer toolkit of chrome video captures.

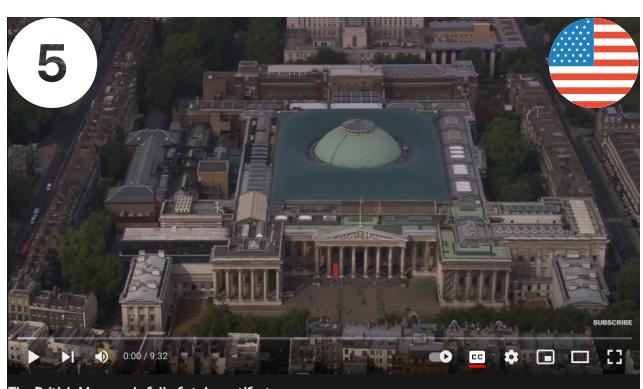
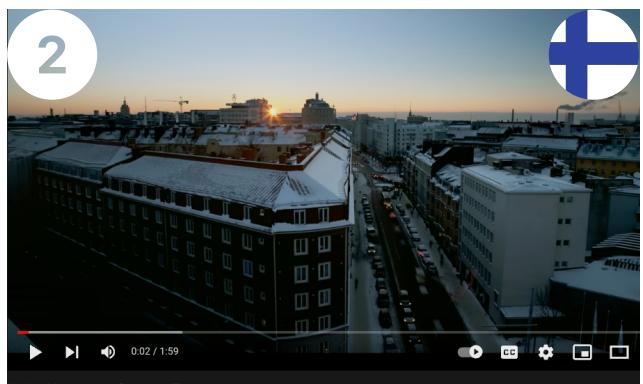
content-type: video/webm

Section 4: Workflow of video search and delivery



Section 5: YouTube CDN and Video Delivery.

With this, the main experiment Begins. I chose five videos. Two Finnish Two Greek and One international. as listed below.



As well as some extras from Japan Serbia and Mexico as it would be listed later in the analysis.

By using the developer options in the browser i was able to trace the location where each song was streamed from and developed a scientific scenario. I used VPN and repeated the experiment. (ill provide a link to a pdf of all the results)



Section 5: YouTube CDN and Video Delivery.



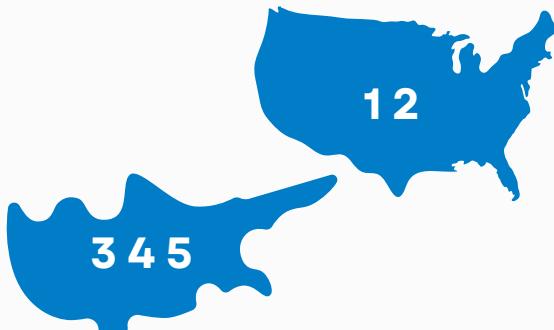
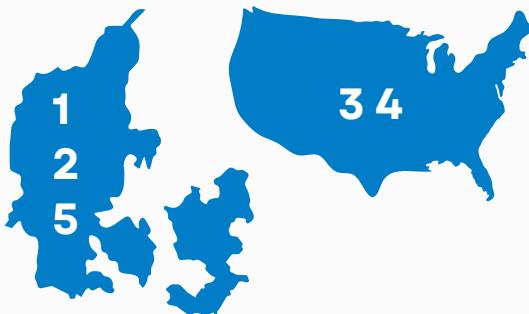
1st Time

1. **109.105.109.205**
2. **109.105.109.205**
3. **74.125.8.201**
4. **172.217.132.134**
5. **109.105.109.204**



1st Time

1. **74.125.100.71**
2. **172.217.132.10**
3. **213.140.213.206**
4. **213.140.213.205**
5. **213.140.213.205**

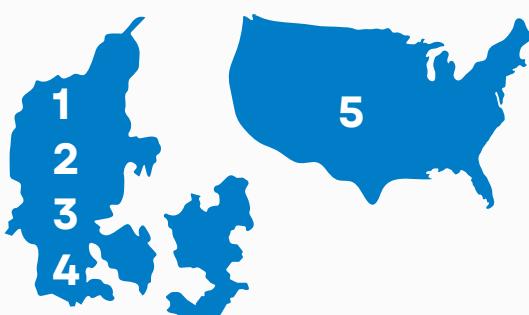


2nd Time

1. **109.105.109.205**
2. **109.105.109.205**
3. **109.105.109.206**
4. **109.105.109.206**
5. **74.125.110.138**

2nd Time

1. **213.140.213.206**
2. **213.140.213.205**
3. **213.140.213.206**
4. **213.140.213.205**
5. **213.140.213.205**



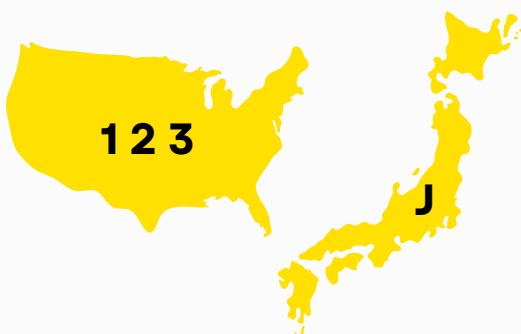
Section 5: YouTube CDN and Video Delivery.



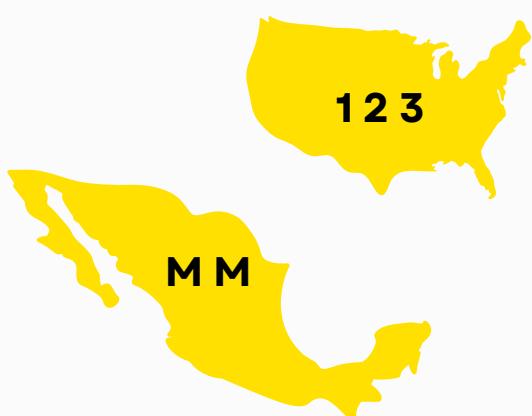
1. **174.125.100.71**
2. **172.217.132.10**
3. **172.217.132.105**
J. **109.105.109.207**



1. **173.194.188.167**
2. **74.125.153.198**
3. **173.194.182.232**
S. **188.120.127.140**
S. **188.120.127.140**



1. **173.194.60.199**
2. **173.194.186.103**
3. **173.194.57.105**
M. **201.174.77.144**
M. **201.174.230.13**



J. https://www.youtube.com/watch?v=p_2F2IKV9uA&ab_channel=NaotaroMoriyamaVEVO

S. https://www.youtube.com/watch?v=UqOL7LOR6ko&ab_channel=LagoonaLoire

S. shorturl.at/qvOXO

M. https://www.youtube.com/watch?v=E3JTcbJGCoc&ab_channel=Mira%C3%A7E%C5%9E

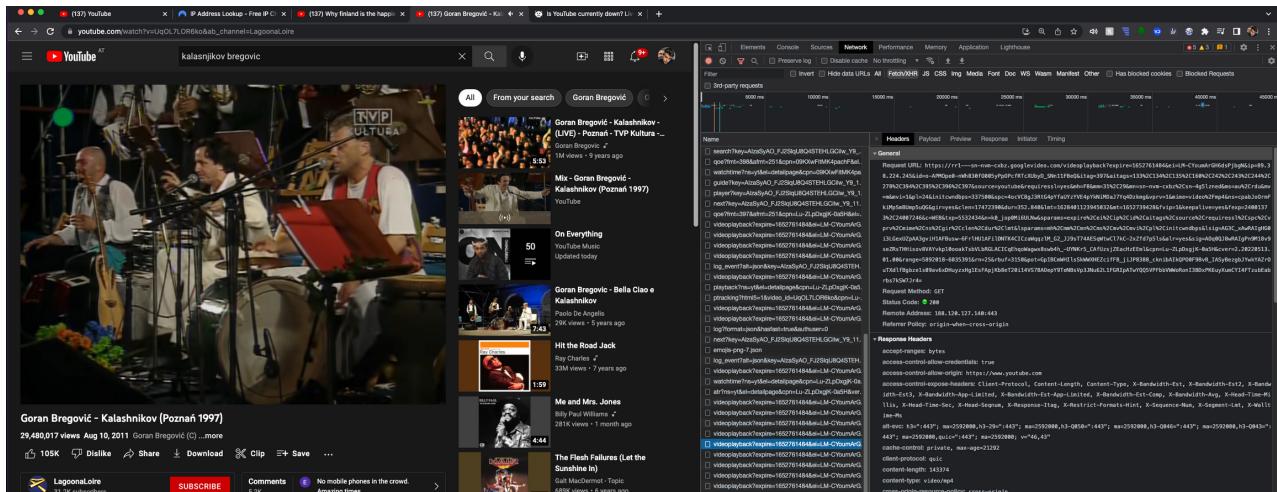
M. https://www.youtube.com/watch?v=rs6Y4kZ8qtw&ab_channel=ManuChao

Section 5: YouTube CDN and Video Delivery.

Let's explain now. We have 5 videos 2 that are local Finnish 2 local greek and 1 from the US. The greek videos are almost impossible to be stored in the NORDIC ISP Caches of Youtube due to the low demand for such content in the Nordic Countries. Same Thing on the Cyprian Side. The Finnish videos are almost impossible to be in demand in Greece/Cyprus. So what we see is that each country's videos are available locally. So for Cyprus, the Greek songs are available in the Cache of the ISP due to the demand for such songs therefore it's faster to access such content. On the other side, the one Finnish comedy video and the Aalto introduction is of course available in the Nordic Caches due to the demand for these videos. The videos that are not accessed before are accessed directly either from a data center cache in the US or the main database. This happens for the greek videos in Finland and the Finnish Videos in Cyprus. On the other hand, an international Video from Vox with millions of views is stored locally in the ISP cache for both the countries due to the demand and being seen recently by someone in those countries. Now we repeat the same experiment after several hours. We can see that all these videos are now available in the Local ISP's Cache in both countries. This happened because of my request to view them a couple of hours ago which therefore saved them in Cache. We can also see that the vox video is no longer available locally in the Nordics which could mean that the demand was low and it got pushed away or that it was faster to access from the main data center/cache in the US. I also repeated the same experiment with three other countries where we could see that videos that could be popular there or were locally created are available directly from each local ISP respectively as Cache, and the Finnish/Greek Videos were accessed by the main server/cache in the US due to low demand of such videos in these countries.

Section 5: YouTube CDN and Video Delivery.

Therefore we can understand easily by the experiment results that YouTube CDN (Content Delivery Network) is indeed very intricate. YouTube has its main databases in the US where all content is stored. It also duplicates popular content (either by expecting it to become popular in a country or by this content getting accessed by someone in that country leaving a copy behind that way it lets the people choose what they want to see) in Local servers or ISP's Cache ready to be accessed way faster than the "long way back to the US". This cache is getting filled by the content that people want as well by getting pushed due to popularity. YouTube also knows which Cache you are closest to due to the DNS server which translates your request giving your IP address to Google along with the Request. Google has a giant database with IP addresses which are bound to the location making it easy to get assigned to multiple cache servers before the main server. That's the general idea of how it works. It identifies your location and based on that the popular content on your location which is already available for you on the local cache without you even knowing about it.



Here are the screenshots of the gathered IP Addresses.