# Analysis of google.com and wits.ac.za websites.

Willie Macharia

Department of Computer Science
University of Cape Town
South Africa

September 2020

## Abstract

This paper contains analysis of two websites: google.com and wits.ac.za. The analysis was done using network tools namely: Wireshark, Traceroute, and curl. The paper starts by analyzing various protocols which are present on the application layer of the two websites. After this, the transport layer security of the two websites is then analyzed. Finally, AS path analysis is performed to determine the internet paths from different continents and countries to where these websites are hosted

***CCS Concepts*** • **Computer systems organization → Embedded systems; Redundancy; Robotics**; • **Network → Network reliability**

***Keywords*** datasets, neural networks, gaze detection, text tagging

## 1. Introduction

Google.com is a website owned by an American company Google. The website is mainly dedicated to provide search functions for information from various web servers hosted on the internet. The website acts as a web search engine. According to [1], the website processes more than 5.4 billion searches everyday and more than 90% searches on the internet. Wits.ac.za is a website owned by University of Witwatersrand which is a public University in Johannesburg, South Africa [8]. The website offers information about the university and also acts as the gateway to access other University online facilities.

This paper will focus on analysing various protocols implemented in various layers of the OSI network model. These protocols includes:

- The Domain Name System (DNS) and various DNS Nameservers which these websites rely on.
- Analysis of whether Internet Protocol version (IPv4 and IPv6) is supported by the two websites.
- The HyperText Transfer Protocol (HTTP).
- The Transport Layer Security (TLS).

The websites will also be analysed on whether they rely on CDNs (Content Delivery Networks) to deliver their content. Since these websites are accessed from various parts of the globe, internet traffic originating from the servers hosting these websites does not follow the same path to serve the different users trying to access content. Therefore, it is important to perform a geographical AS path analysis to analyze the various paths internet traffic follows when users request to access content hosted in these websites. This paper also presents an AS path analysis that shows various paths the internet traffic follow when these websites are accessed from various parts of the world.

## 2. The Domain Name System (DNS)

To perform the DNS analysis, python library dnspython was used to write a python script which performed the analysis. The dns.resolver.query(domain, a) method accepts two parameters namely: 'domain' is the website being analysed and 'a' is the type of DNS record being queried. After running the script, the results that were obtained are shown in figure 1 in the appendix for google.com and figure 2 in the appendix for wits.ac.za.

### 2.1 Nameservers

The DNS record for the name servers are indicated with the initials "NS". Google.com website uses the following 4 names servers:

- ns1.google.com
- ns2.google.com
- ns3.google.com
- ns4.google.com

Wits.ac.za website uses the following 4 names servers:

- ns1.enetworks.co.za
- ns2.enetworks.net
- ns3.enetworks.co.za
- ns4.enetworks.net

### 2.2 Support of IPv4 and IPv6

Support of IPv4 was analysed by the dns record 'A' and IPv6 was analysed by the dns record 'AAA'. Google.com supports both IPv4 and IPv6 as seen in figure 1 above. The IPv4 address of the website is 172.217.170.46 while IPv6 is 2c0f:fb50:4002:805::200e. Wits.ac.za website only supported IPv4 as shown in figure 2 above. The IPv4 address of wits.ac.za is 146.141.13.50. The IPv6 dns query for wits.ac.za returned none and hence we conclude that wits.ac.za does not support IPv6.

### 2.3 MX Records

MX dns records also known as Mail Exchanger records are responsible for specifying the server that accepts mails/emails on behalf of a domain. Google.com has the following MX dns records:

- 10 aspmx.l.google.com.
- 20 alt1.aspmx.l.google.com.
- 30 alt2.aspmx.l.google.com.
- 40 alt3.aspmx.l.google.com.
- 50 alt3.aspmx.l.google.com.

Wits.ac.za has the following MX dns records:

- 0 mg1.wits.ac.za.
- 10 mg2.wits.ac.za.
- 20 mg3.wits.ac.za.

### 2.4 TXT Records and SOA records

TXT dns records are used to associate arbitrarily text with host name or server. Google.com has 4 TXT dns records as shown in figure 3 in the appendix. Wits.ac.za has the following 4 TXT dns records as shown in figure 4 in the appendix:

SOA dns records also known as Start of Authority records record information about administration of the domain zone and also zone transfer. Google.com has the following SOA records as shown in figure 3 in the appendix.

- ns1.google.com. dns-admin.google.com. 330686346 900 900 1800 60

Wits.ac.za has the following SOA dns records as shown in figure 4 in the appendix.

- ns1.enetworks.co.za. admin.wits.ac.za. 2020081603 10800 3600 604800 108

00

### 2.5 Location of where the website have been hosted

The location where the two websites are hosted was determined through geoip2 which is a web service provided by Maxmind geolocation [5]. The webservice API call geoip2.webservice.Client(API ID, API key).insights(ip) returned an object which had a location where the websites were hosted. According to figure 5 in the appendix, Google.com is hosted in the United States and wits.ac.za is hosted in Johannesburg South Africa.

## 3. HTTP Analysis

This section was analysed using chrome developers extension and the Requests python library.

### 3.1 HTTP Request and Response

The request that was sent to access google.com is shown in figure 6 in the Appendix below. The request that was sent to access wits.ac.za is shown in figure 7 in the Appendix below. Some of the standard HTTP request headers that were present in both requests that were sent to the websites are:

- Method used in both requests is *GET*. However, wits.ac.za uses HTTP version 1.1 while google.com uses https.
- Both requests use a persistent http connection as "keep-alive" is added to the request.
- Both requests have a host name. Google.com uses www.google.com and wits.ac.za uses www.wits.ac.za as the host name.
- Both have the same user-agent as Chrome, language as en-US , Date as 12.09, and accept encoding as gzip, br.

The following are some of the non standard request headers that were identified:

- Both websites requests were accompanied with a cookie. These cookies were the one previously sent from the server to the clarinet and stored by the client.
- Both websites use cache-control : max age = 0 which according to Mozilla, Max age refers to the time a cached copy of a resource takes to expire. Therefore, Max Age =0 means that clients must re-validate cached resources before using the resource. This means that the request is done every time but downloading the HTTP body from the server may be skipped if the cached copy of the resource is valid[3].
- Wits.ac.za request has an upgrade-insecure-request header whose value is 1. This means that when the browser sends a request to the server, the server should redirect the browser to the secure version of the website[3]. This is because the initial request is in the http version.

The response that was received from google.com server is shown in figure 8 in the Appendix below. The

response that was received from wits.ac.za server is shown in figure 9 in the Appendix below.

Some of the standard HTTP responses headers that were present in both responses that were received from to the servers of the two websites are:

- Both responses had response status code 200 meaning that the requests were successful and the response was present.
- Both have the same date which is 12th september, content-type which is text/html, content-length which is 8879 bytes for wits.ac.za and 48527 bytes for google.com, content encoding which is gzip wits.ac.za and br for google.com. Gzip and br works the same which are HTTP response compression algorithms.
- Wits.ac.za has a "last-modified" header which shows the server utilises web caches.

The following are some of the non standard response headers that were identified:

- Both responses had set-cookie header which were new cookies that were issued by the server and sent to the client to be stored by the client.
- Google.com response has an alt-svc which is a header that lists alternative services through which the same resource can be reached. One of the services that can be seen is "quic=":443" which is the QUIC service deployed by Google to implement a quick, reliable and secure UDP service in the transport layer of the website [6].
- Wits.ac.za response has a header X-Cnection': 'close which is a header used to track the client's connections that want to close immediately after receiving a response from the server[3].
- Google.com response has "remote address" and "cache-control:private" headers. Remote address is different from the IP Address and means that the server utilises a private network. Cache-control being private means that part of the response is only intended for a single user and should not be cached by a shared cache which might be a proxy server.

### 3.2 Cookies

Web cookies are used to store and save information about the user such that when the user sends a request again to the same server, the server will generate a customized web page for the user or the server already has login information[4]. Initially, the cookies are generated by the server and sent back to the user embedded in the response. The cookie is stored by the client and it is accompanied by a consecutive request a user makes to the server. Requests may have a header Cookie which means a cookie that had been issued before and responses may have set-cookie which means a cookie originating from server to the client[4]. Google.com had the following cookies:

- Cookie 1PJAR which is a cookie that stores information and preferences about maps and related Google services.
- Cookie:NID which gives a user a Unique ID which is associated with information and preferences about maps and related Google services

Wits.ac.za had the following cookies:

- BIGipServer Web ccms.appccms pool which is a cookie that associates BIGip server web content servers with specific user[4].

## 4. TLS analysis

TLS which means Transport Layer Security is a layer that resides between the application layer and the transport layer. Security analysis was done by using openssl python library, chrome developers tool extensions and SSLLabs[7].

### 4.1 TLS Version used

Google.com uses TLSv1.3, Cipher is TLS-AES-256-GCM-SHA384 which means TLS's encryption is implemented by Advanced Encryption standard and Secure Hashing Algorithm with 384 bits. Wits.ac.za uses TLSv1.2, Cipher is ECDHE-RSA-AES128-GCM-SHA256 which means TLS's encryption is implemented by Elliptic Curve Diffie Hellman Ephemeral and Secure Hashing Algorithm with 256 bits. AES is a symmetric encryption algorithm while RSA is a public-private asymmetric encryption algorithm. Google.com is implemented in QUIC as shown in figure 10 in the appendix.

### 4.2 Sites web certificates

Google.com SSL server certificate had been issued to Google LLC by Google Trust Services and the root issuer is GlobalSign Root. Wits.ac.za SSL server certificate had been issued to University of Witwatersrand by Thawte RSA CA 2018. and the root issuer is DigiCert Global Root .Google.com certificate had a short validity period which was 26 August 2020 to 18 November 2020 while wits.ac.za had a longer validity period which was 23 July 2020 to 27 March 2021.

#### 4.2.1 TLS Negotiation

TLS negotiation occurs when a client and the server have different TLS versions and they have to negotiate which version to use. Websites attracting high traffic such as google needs to implement TLS negotiation since not every client will have the same TLS version. According to analysis done via SSL Labs[7], Google.com

and wits.ac.za implements TLS-FALLBACK-SCSV which is used mainly by clients such web browsers to fall back to a lower version of TLS if the first attempt to access the server with a higher version of TLS fails. It is also used to guard against protocol downgrades if a request is sent from malicious users. The trade-offs that these websites have to establish is to either enable TLS-FALLBACK-SCSV so that to prevent malicious users from taking advantage of vulnerabilities that lower versions have or enable it so that to serve many users.

#### 4.2.2 Page-load times

This analysis was done using the curl tool. For accuracy, 5 curl commands were issued to each website and average were taken. After averaging, the following results were obtained: Screenshots can be seen in figure 11 and 12 in the appendix.

**Table 1.** Data showing loading times for the two websites

| Website | wits.ac.za | google.com |
|---|---|---|
| **Name Lookup Time:** | 0.004747 | 0.002997 |
| **Connect Time:** | 0.024721 | 0.023354 |
| **Pre-transfer Time:** | 0.024895 | 0.023402 |
| **Start-transfer Time** | 0.269779 | 0.306406 |
| **Total Time:** | 0.269652 | 0.318679 |

Google.com domain name gets resolved faster with 3ms compared to wits.ac.za with 4ms. Connect Time which is the time it takes for TCP to connect to the remote server is faster in google.com than in wits.ac.za. This might be probably that Chrome uses QUIC to connect to google servers hence faster connection. The pre-transfer time is fast in google.com compared to wits.ac.za but transfer time is faster in wits.ac.za than in google.com. This can be attributed to wits.ac.za server being hosted locally compared to Google.com server which is hosted in the USA.

## 5. AS-Path Analysis and CDN Presence

The AS-Path analysis was done using the RIPE Atlas [2] and Maxmind geolocation database.

### 5.1 Vantage points

To perform the analysis, a python script was written with 24( 8 countries each with 3 probes) vantages points. Measurements were taken after 20 minutes and the traceroute was saved locally. The vantage points chosen were from Kenya, Senegal, Canada, United Kingdom, USA, Russia, Brazil and China.

### 5.2 AS Paths

#### 5.2.1 AS Path analysis - wits.ac.za

Wits.ac.za is connected to the internet via the tenet.ac.za ASN as the last five hops of all the traces from probes to wits.ac.za are routed via tenet.ac.za ASN Number (2018). Tenet.ac.za ASN is present at Johannesburg IXP but many ASNs can be seen not to be present at the IXP. Tenet.ac.za can also be seen in Europe and USA IXPs. This causes the internet traffic from Europe and USA to have few hops as compared to other traffic from other continents such as Asia where traffic from China probes takes a route with many hops. The traffic from the USA and Canada utilises IXPs as it can be seen from the traces. Some of these IXPs are Miami IXP which is initiated as “Ampath.net”.

Internet traffic from probes hosted in African countries such as Senegal is routed to either Europe or USA in order to get to tenet.ac.za which is present in Europe and USA IXPs. The trace from Senegal ASN(8346) is routed to Cogent ASN (174) which is present in Europe and USA. Cogent then routes the traffic to tenet.ac.za which is connected to wits.ac.za. This analysis shows that Africa’s intra-continental internet traffic is forced to be routed via IXPs present in Europe and USA. Traffic sent from the probe hosted in Kenya has few hops since the probe is hosted by Tespok ASN 37578 and Tespok has presence in Johannesburg IXP where tenet.ac.za has presence too. The trace from Senegal could have taken few hops if ASN(8346) was present at Johannesburg IXP.

#### 5.2.2 AS Path analysis - google.com

Google.com server is connected to the internet via Google ASN 15169. Internet traffic that was from probes hosted in the USA and Canada could only hop once and get to the target server. Traffic from probes hosted in Europe utilised Cogent ASN(174) to get to Google.com. Internet traffic from probes in Kenya could connect to Google.com as there was presence of Google ASN in Kenya. However, traffic from probes in Senegal was getting routed to Europe to connect to Cogent ASN(174) so as to connect to Google ASN.

#### 5.2.3 CDN Reliance

CDN(Content Delivery Network) are content caches that are distributed geographically to reduce the time the user takes to get response from the server hosting the content[9]. The caches are mainly deployed by content providers such as Facebook, Google, Netflix. Most CDNs are private networks operated privately by the content provider.

Cogent communications ASN 174 has CDN deployed mostly in Europe and North America. Cogen runs a CDN called CDN 77 which has a presence in Europe.

Google also operates its cloud CDNs which are deployed all over the globe with presence in the five continents. Google.com trace shows the probes' internet traffic utilises google CDNs to reach google.com content. This is evident as the last three hops for the traces from probes to google.com are within Google ASN. This means the content are being searched from different caches hosted in different geographical regions. No CDN support seen for the traces from the probes to wits.ac.za.

## References


[1] 10 google search statistics you need to know [april 2020]. https://www.oberlo.co.za/blog/google-search-statistics#:~:text=Google%20processes%20over%203.5%20billion,queries%20every%20second%20on%20average. (Accessed on 09/14/2020).

[2] atlas.ripe.net. https://atlas.ripe.net/probes/. (Accessed on 09/14/2020).

[3] Http cache headers - a complete guide - keycdn. https://www.keycdn.com/blog/http-cache-headers#:~:text=The%20max%2Dage%20directive%20states,for%20the%20next%2090%20seconds. (Accessed on 09/14/2020).

[4] Http cookie - wikipedia. https://en.wikipedia.org/wiki/HTTP_cookie. (Accessed on 09/14/2020).

[5] Ip geolocation and online fraud prevention | maxmind. https://www.maxmind.com/en/home. (Accessed on 09/14/2020).

[6] Quic - wikipedia. https://en.wikipedia.org/wiki/QUIC. (Accessed on 09/14/2020).

[7] Ssl server test (powered by qualys ssl labs). https://www.ssllabs.com/ssltest/. (Accessed on 09/14/2020).

[8] University of the witwatersrand - wikipedia. https://en.wikipedia.org/wiki/University_of_the_Witwatersrand. (Accessed on 09/14/2020).

[9] What is a cdn? | how do cdns work? | cloudflare. https://www.cloudflare.com/learning/cdn/what-is-a-cdn/. (Accessed on 09/14/2020).


# Appendices

### .1 Github Code Repo

Python scripts that were used to perform the analysis can be found here:https://github.com/willie84/CSC5032Z--Assignment

```
/usr/bin/python3.7 /home/willie-macharia/PycharmProjects/Assignemnt1/main.py
The Website being tested is:  google.com
The DNS Record:  A : 172.217.170.46
The DNS Record:  AAAA : 2c0f:fb50:4002:805::200e
The DNS Record:  NS : ns1.google.com.
The DNS Record:  NS : ns4.google.com.
The DNS Record:  NS : ns3.google.com.
The DNS Record:  NS : ns2.google.com.
The DNS Record:  MX : 50 alt4.aspmx.l.google.com.
The DNS Record:  MX : 30 alt2.aspmx.l.google.com.
The DNS Record:  MX : 20 alt1.aspmx.l.google.com.
The DNS Record:  MX : 10 aspmx.l.google.com.
The DNS Record:  MX : 40 alt3.aspmx.l.google.com.
The DNS response does not contain an answer to the question: google.com. IN CNAME
The canonical name of the analytics.google.com is www3.l.google.com.
```

**Figure 1.** Showing google.com dns records.

```
The Website being tested is:  wits.ac.za
The DNS Record:  A : 146.141.13.50
The DNS response does not contain an answer to the question: wits.ac.za. IN AAAA
The DNS Record:  NS : ns1.enetworks.co.za.
The DNS Record:  NS : ns2.enetworks.net.
The DNS Record:  NS : ns3.enetworks.co.za.
The DNS Record:  NS : ns4.enetworks.net.
The DNS Record:  MX : 20 mg3.wits.ac.za.
The DNS Record:  MX : 10 mg2.wits.ac.za.
The DNS Record:  MX : 0 mg1.wits.ac.za.
The DNS response does not contain an answer to the question: wits.ac.za. IN CNAME
The canonical name of the ftp.wits.ac.za is sunsite.wits.ac.za.
```

**Figure 2.** Showing wits.ac.za dns records.

```
The DNS Record:  TXT : "docusign=05958488-4752-4ef2-95eb-aa7ba8a3bd0e"
The DNS Record:  TXT : "docusign=1b0a6754-49b1-4db5-8540-d2c12664b289"
The DNS Record:  TXT : "facebook-domain-verification=22rm551cu4k0ab0bxsw536tlds4h95"
The DNS Record:  TXT : "globalsign-smime-dv=CDYX+XFHUw2wml6/Gb8+59BsH31KzUr6c1l2BPvqKX8="
The DNS Record:  TXT : "v=spf1 include:_spf.google.com ~all"
The DNS Record:  SOA : ns1.google.com. dns-admin.google.com. 330686346 900 900 1800 60
```

**Figure 3.** Showing google.com txt dns records.

```
The DNS Record:  TXT : "google-site-verification=cQQjTdb5QMfOog-Dh2PJPjnH6P-_hPWLXaHKneL0nFA"
The DNS Record:  TXT : "MS=ms99071949"
The DNS Record:  TXT : "v=spf1 include:spf.protection.outlook.com ip4:146.141.11.0/24 ip4:146.141.9.6 ip4:146.141.14.0/24 ip4:146.1
The DNS Record:  TXT : "ufO1sh3mohMMqJBoKSWr0J8ITf4uYBRycvlJ/Rp+/n7dr5dM5hNcnVFosYHH7xdFbp/LxHoX1yIS4habPmYEXw=="
The DNS Record:  TXT : "ciscocidomainverification=63c04492f024d2480ae4156c1fedb4a8286c6ebbfee25eadbf13195807c8f7fb"
The DNS Record:  SOA : ns1.enetworks.co.za. admin.wits.ac.za. 2020081603 10800 3600 604800 10800
```

**Figure 4.** Showing wits.ac.za txt dns records.

```
The IPv4 of this website is  google.com 172.217.170.46
The Location where hosting of this website: Latitude google.com -97.822 Longitude:  37.751
The city of this website server google.com None
The country name where the server of this google.com United States
The IPv4 of this website is  wits.ac.za 146.141.13.50
The Location where hosting of this website: Latitude wits.ac.za 28.0583 Longitude:  -26.2309
The city of this website server wits.ac.za Johannesburg
The country name where the server of this wits.ac.za South Africa
```

**Figure 5.** Showing geolocation of google.com and wits.ac.za servers.

▼ **Request Headers**

**:authority:** www.google.com

**:method:** GET

**:path:** /

**:scheme:** https

**accept:** text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/

ion/signed-exchange;v=b3;q=0.9

**accept-encoding:** gzip, deflate, br

**accept-language:** en-US,en;q=0.9

**cache-control:** max-age=0

**cookie:** OTZ=5593600_48_48__48_; SEARCH_SAMESITE=CgQIwJAB; ANID=AHWqTUmCpgdjRsLqjVd_Le

**Figure 6.** Showing google.com HTTP Request Headers.


**Figure 7.** Showing wits.ac.za HTTP Request Header.

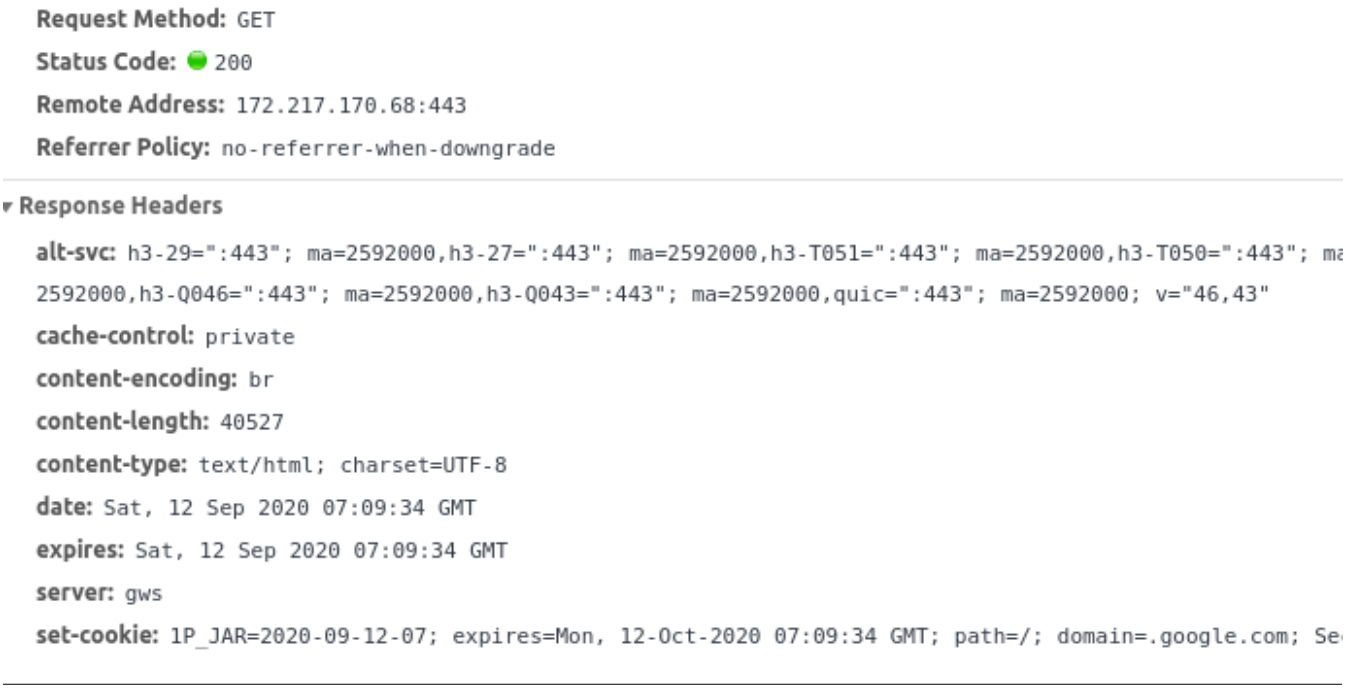


**Figure 8.** Showing google.com HTTP Response Header.

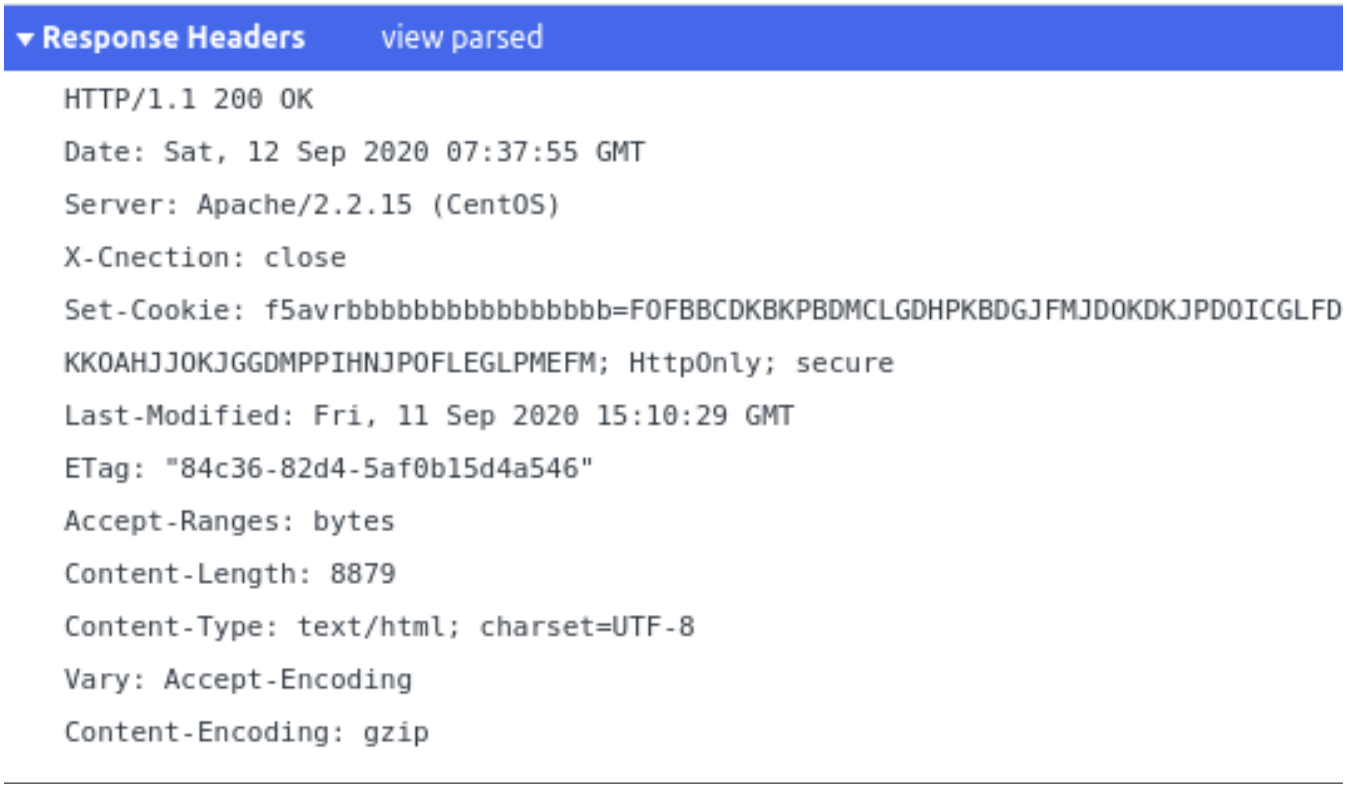


**Figure 9.** Showing wits.ac.za HTTP Response.

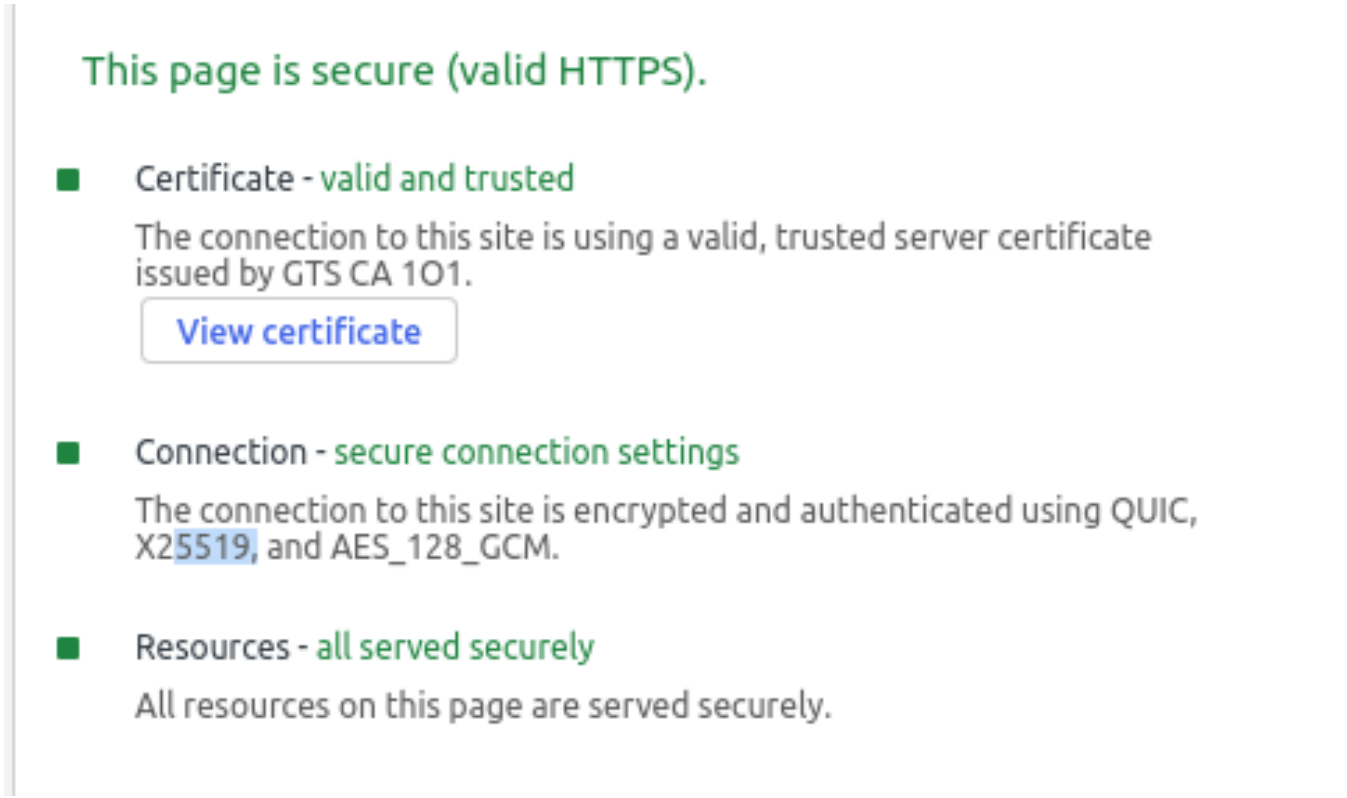


**Figure 10.** Showing indication of QUIC in google.com.

```
 curl -s -w 'Testing Website Response Time for :%{url_effective}\n\nLookup Time:
\t\t%{time_namelookup}\nConnect Time:\t\t%{time_connect}\nPre-transfer Time:\t%{
time_pretransfer}\nStart-transfer Time:\t%{time_starttransfer}\n\nTotal Time:\t\
t%{time_total}\n' -o /dev/null http://www.google.com
Testing Website Response Time for :http://www.google.com

Lookup Time:            0.002995
Connect Time:           0.023054
Pre-transfer Time:      0.023102
Start-transfer Time:    0.302406

Total Time:             0.308679
```

**Figure 11.** Showing google.com *curl* request.

```
 curl -s -w 'Testing Website Response Time for :%{url_effective}\n\nLookup Time:
\t\t%{time_namelookup}\nConnect Time:\t\t%{time_connect}\nPre-transfer Time:\t%{
time_pretransfer}\nStart-transfer Time:\t%{time_starttransfer}\n\nTotal Time:\t\
t%{time_total}\n' -o /dev/null http://www.wits.ac.za
Testing Website Response Time for :http://www.wits.ac.za

Lookup Time:            0.004247
Connect Time:           0.024521
Pre-transfer Time:      0.024695
Start-transfer Time:    0.269279

Total Time:             0.269352
```

**Figure 12.** Showing wits.ac.za *curl* request

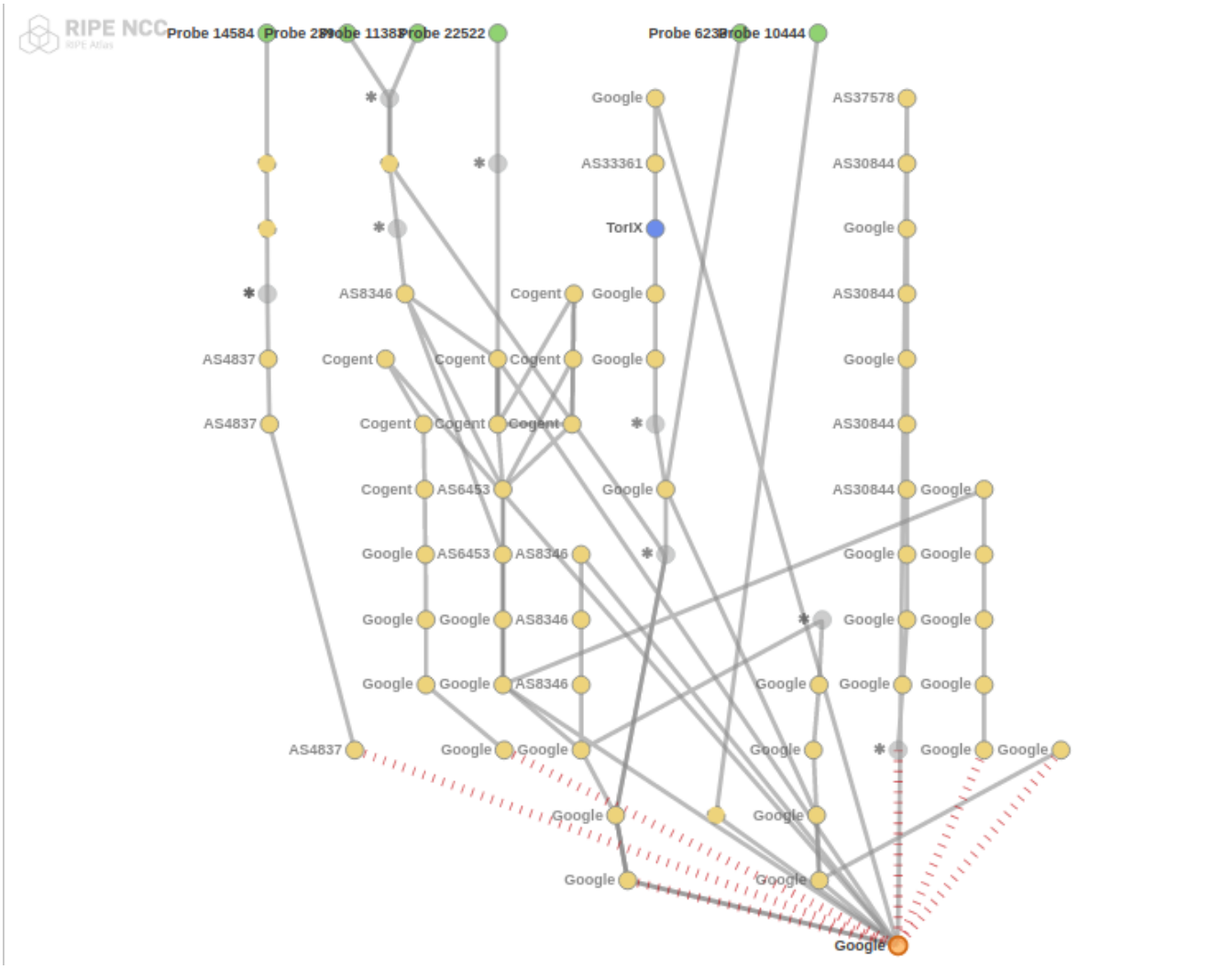


**Figure 13.** Showing google.com RIPE Atlas ASN path.

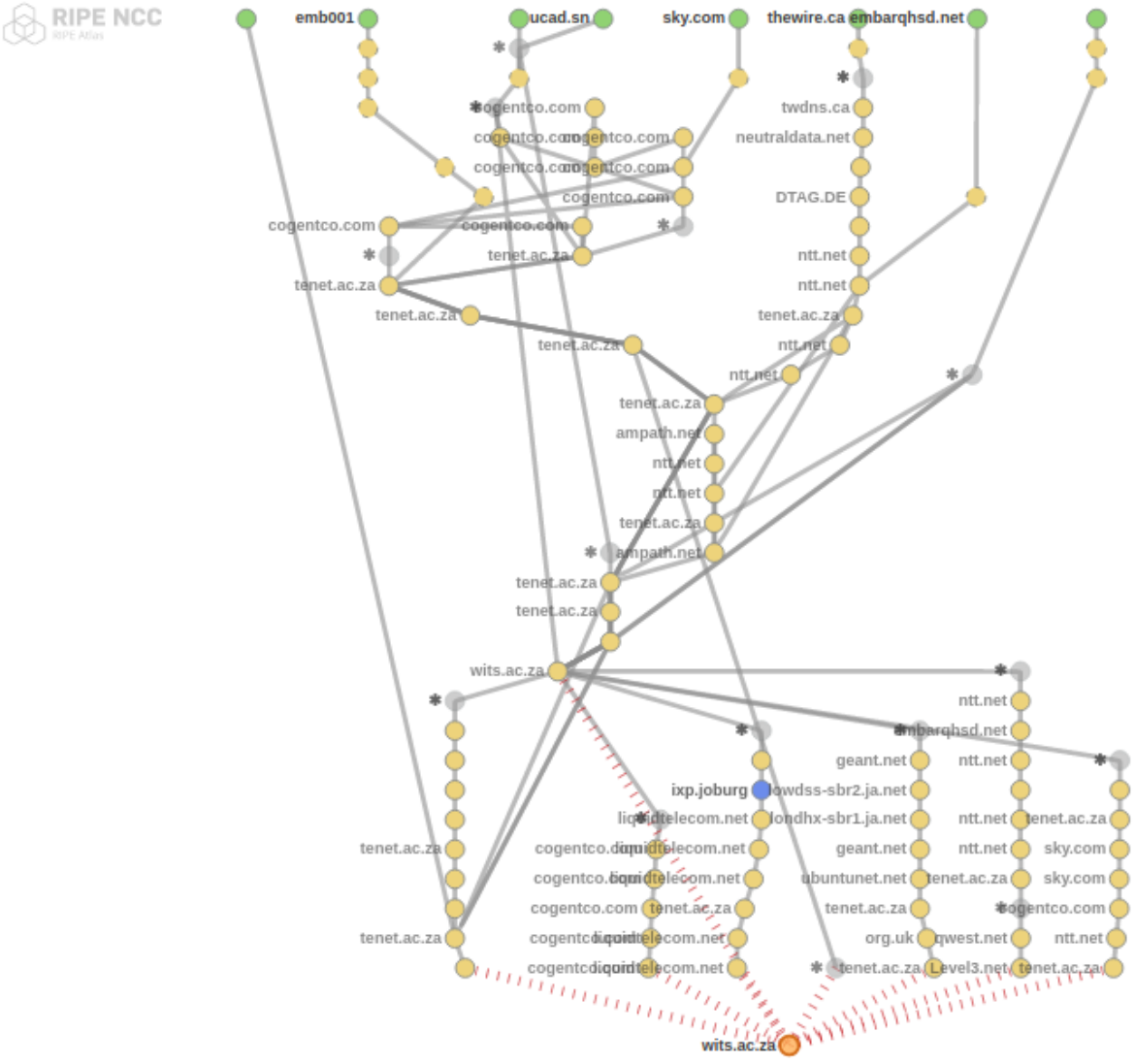


**Figure 14.** Showing wits.ac.za RIPE Atlas ASN path.