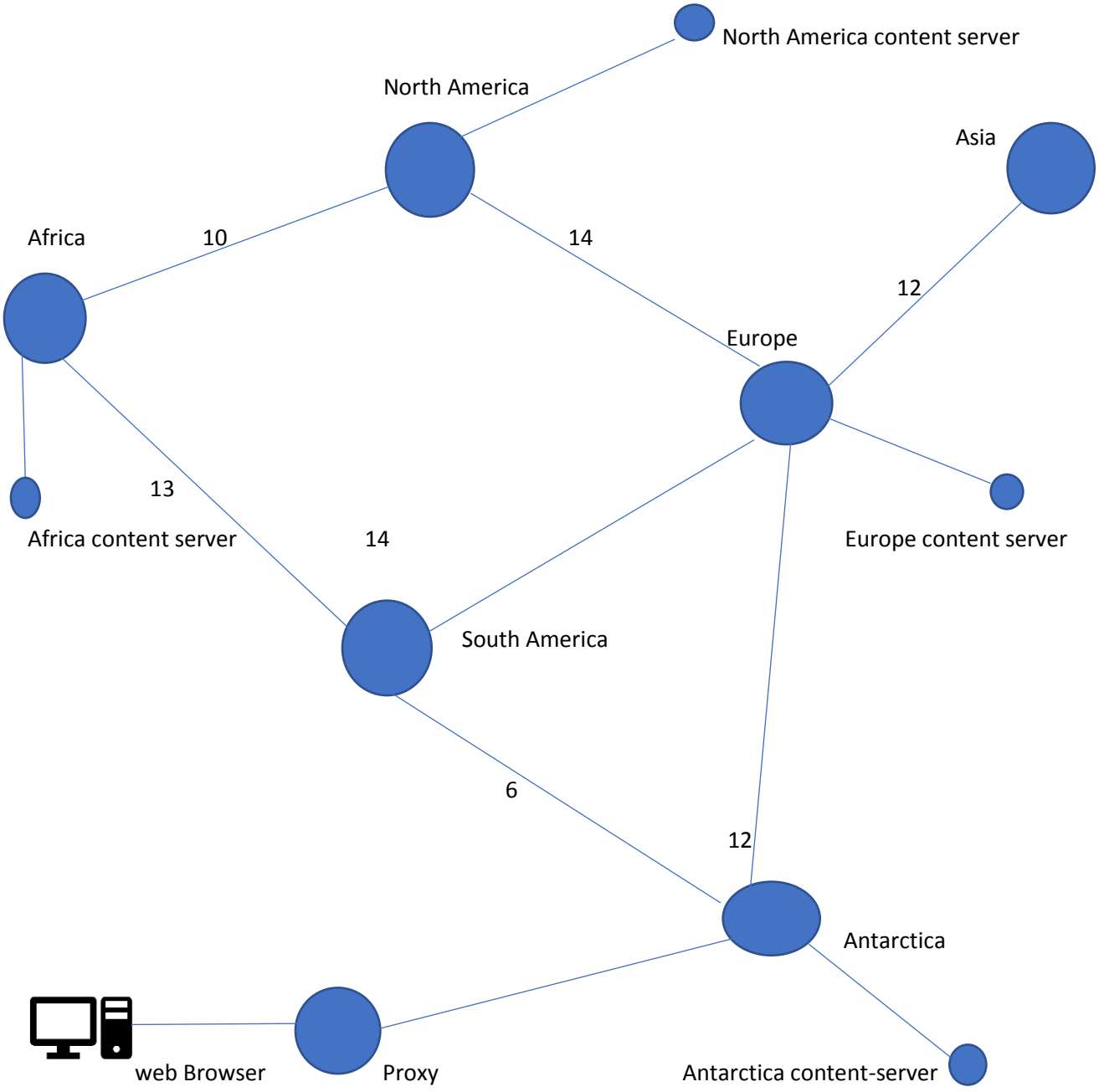
Content Distribution Network

Introduction-The goal of content distribution network (also known as a content delivery network)is to accelerate the delivery of web content to internet-connected devices. The world's largest content distribution network, owned and operated by Akamai, spans more than 216,000 servers in over 120 countries and within more than 1,500 networks around the world.(
https://www.akamai.com/us/en/cdn/)

Design-To build Content distribution network

Consider any network topology. For example-We take below topology



1.Make CDN program which will read its configuration file, which includes it CDN IP address CDN port, and its neighbors information (ip , port and delay time to reach port)
2.Measure total delays from each node with some intervals and create routing table (delays are used as cost to reach from one node to another node).
3. send routing table to neighbors on some intervals (I have taken 2 seconds) .
CDN uses caching scheme so that it can provide file to user with high performance or we can see that less delay time.
4.This CDN network uses distance vector algorithm to compute its path to other node and update its table accordingly.
5.Whenever CDN get request its serves file using its routing table.
6.CDN uses cache scheme to provide data to end user with less delay time .
7.Below we will discuss caching policy using examples.

Here in CDN project we have implemented two cache schemes

1.Cache policy-1

Let us discuss scenario where user (who is in Antarctica) needs a file which is located in asia. .(Let us assume Antarctica , Europe and Asia does not have file in its CDN cache.)

1.User request a file which will go to proxy and proxy forwards request to CDN(which is located in user's region or a nearest CDN)
2.CDN search file in its cache which in Antarctica CDN, if Antarctica CDN doesn't have file then it forwards request to another CDN(Europe).
3. Before forwarding request to Europe. CDN looks into its routing table for the shortest path for reaching Europe
4.Then Europe's CDN look into its cache if it has file and if not, then forwards request to Asia's CDN.
5. Asia's CDN will look into its cache if it has and if not then it will get file from its content server and cache in its CDN and send back file to CDN(Europe) and Europe's CDN will send back file to Antarctica's CDN and Antarctica CDN will serve file to user.
6. According to cache Policy-1 -Each node will cache file while returning to the node where the request originated from. Hence according to cache policy-1 – Asia's CDN, Europe's CDN , Antarctica's CDN will cache the file.
7. Next time when user (who is in Antarctica) request same file will get the file in less time as compared to last time when he requested the file, as now Antarctica CDN have that file.

As all Nodes will cache file while writing response back. So here source will be the nearest CDN to user and second time it will get file from Source CDN

Performance measurement for Cache Policy-1

Source	destination	Time1	Time2
North America	Antarctica	36 seconds	Less than 1 second
Europe	Africa	33 seconds	Less than 1 second
Europe	Asia	12.38 sec	Less than 1 second

Here if we see first time the delay time is more as compare to second time and can see CDN network provide data with high performance.

Time-1-First time when file come from Destination caches (Antarctica)
Time-2- Second time when file come from source(North America or Last CDN).

2. Cache Policy-2:

Let us discuss scenario where user (who is in Antarctica) needs a file which is located in asia. .(Let us assume Antarctica , Europe and Asia does not have file in its CDN cache.)

1.Similar Cache policy-1 steps (1 to 5)
6. According to cache Policy-2 - Asia's CDN node will cache file and forwards file to the node where the request originated from (Europe's CDN). Now Europe's CDN forwards and caches file to Antarctica CDN and Antarctica CDN forwards file to user (via proxy).
7.Second time when user (who is in Antarctica) requests same file ,he will get that file from Asia cache. And Europe's CDN will cache file
8.Third time when same user requests file, users will get file from Europe CDN's cache and while returning request to Antarctica, Antarctica will cache the file.
9.Fourth time Antarctica will serve file to user.
10. Hence according to cache policy-2 – Fourth time when user requested file takes less time compare to first, second and third

Performance measurement for cache policy-2

source	via	destination	Time-1	Time-2	Time-3	Time-4
Europe	North America	Africa	33sec	Approx. 33sec	23sec	Less than 1 second
Asia	Europe	Antarctica	35.41 sec	35.37 sec	24 sec	Less than 1seconds
North America	Europe	Antarctica	37sec	36.78 sec	30 sec	Less than 1 second

Time-1 – First Time when content server will provide file to destination cache (for example (case-1)Africa's content server)
Time-2- Second Time when destination cache will serve file to user (Africa's CDN).
Time-3 – Third Time when North America's CDN will serve file.
Time-4- Fourth Time when Africa's CDN will serve file.
Here if we see above analysis. We can see cache policy-2 also improve performance in providing file to network as fourth time when users requests file it get data in us few seconds as compare to Time-1, Time-2, Time-3. As closer is CDN, the faster the content will be delivered to the user.

Implementation-Developed code using python(python-3) and
This CDN network provide **html, jpg, jpeg files**

New feature in CDN- New feature in CDN is that it log the information in log file about the final path or path CDN took to return file to user .Additionally in while requesting to other CDN nodes it will provide information in headers that from which last CDN it came from and how many nodes file came from with respect to current node.

Here are few screenshots from the above performance in Table(northamerica-antarctica)

```
requesting....http://127.0.0.1:20304/antarctica/file.html

2018-11-25 15:33:37.605925
headers---Accept-Encoding: identity
Host: 127.0.0.1:20304
User-Agent: Python-urllib/3.6
From: northamerica
Mypath: northamerica
Connection: close


2018-11-25 15:33:37.606885
requesting....http://127.0.0.1:20311/antarctica/file.html

2018-11-25 15:33:37.608693
headers---Accept-Encoding: identity
Host: 127.0.0.1:20311
User-Agent: Python-urllib/3.6
From: europe
Mypath: northamerica-europe
Connection: close


2018-11-25 15:33:37.609364
finalPath....northamerica-europe-antarctica
```

Figure 1-First Time when request made in Cache policy-2

Here In above image we can see in headers two field From and Mypath
From-last node from which request came
MyPath-Path indicates last route covered by file w.r.t current node

```
2018-11-25 15:56:04.096226
requesting....http://127.0.0.1:20302/africa/minion.jpg

2018-11-25 15:56:04.097954
headers---Accept-Encoding: identity
Host: 127.0.0.1:20302
User-Agent: Python-urllib/3.6
From: europe
Mypath: europe
Connection: close


2018-11-25 15:56:04.098506
finalPath....europe-northamerica
```

Figure 2-Third time when request made in cache policy-2

Conclusions- We can conclude CDN, a system of distributed servers network that deliver content to a user, based on the geographic locations of the user. This service is effective in speeding the delivery of content of websites with high traffic. The closer the CDN server is to the user geographically, the faster the content will be delivered to the user.

Resources-
https://www.akamai.com/us/en/cdn/
https://www.webopedia.com/TERM/C/CDN.html

Testing on Amazon EC2-

Created one instance but got few issues in creating other instances

mv Downloads/new-aws.pem ~/ .ssh

scp -i .ssh/new-aws.pem cp1.py ubuntu@ec2-18-234-72-140.compute-1.amazonaws.com

scp -i .ssh/new-aws.pem test.json ubuntu@ec2-18-234-72-140.compute-1.amazonaws.com: