Measurement Study of Netflix, Hulu, and a Tale of Three CDNs

Final Project Report

Pentapati Sivani
Niharika Vadlamudi

April 27 2020

1 Introduction

Netflix and hulu are leading over the top content providers in most of the countries. Analysing the architecture and working of such complex systems can provide good information on how such platforms work and could potentially help us in developing a better version and can also shed light on innovative possibilities to enhance working on current systems.As a part of that, we are going to analyse the architecture and service strategies adopted by over the top (OTT) content service providers Netflix and hulu and also perform measurement studies of the three content distribution networks (CDN's) used by these platforms in this project.

2 Netflix Streaming Platform

The entire architecture of Netflix can be divided into following parts :

- **Data Centers**

  The Data Center hosts the website "www.netflix.com" which is the opening website/ the first site the user is presented with. The main functionalities of data centers include :

  - Registration of new users and their payment information signing up.
  - Redirection based on the user it them to login(movies dedicated for streaming) or sign up page .

  It performs only the above two functions and doesn't interact with user during streaming.

- **Amazon Cloud**

  All other sites for streaming which include Movies.Netflix, Agmovie etc are hosted by Amazon Cloud. The servers perform **content ingestion**, **log recording**, **DRM**,**CDN Routing**, **User Sign-in**, and **Mobile Device Support**.

- **Content Distribution Networks**

  The encrypted streaming content is copied from amazon cloud into CDN's which is then delivered to the end user. Netflix employs 3 cdn's which include :

  - Akamai
  - Limelight
  - Level 3

- **Players**

  Netflix uses silver light player on the client side to download, decode and play movies on browsers. The run-time environment for silver-light is available as plugin for most of the browsers.

Architecture of netfix:

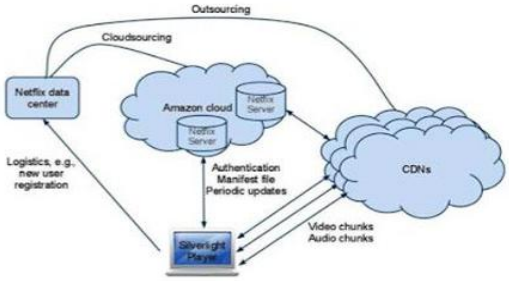


Fig. 1. Netflix architecture.

2.1 Netflix in Action

- Netflix uses DASH protocol, Dynamic streaming over HTTP for streaming. In DASH, the video stream is delivered in chunks of various quality levels. Each chuck is requested via HTTP. A rate determination algorithm determines the quality of next video chunk to be delivered and this makes DASH freely switch between various video quality levels depending on the constraints
- **Netflix manifest file installation and execution**: Silver light client downloads manifest file which contains metadata for DASH player to perform adaptive video streaming. Manifest files are user specific and are generated depending on the user requirements and the video types the user player can support.(i.e different manifest files are generated for .mkv,.h264, .wmv video formats) The manifest file is in XML format and contains the following information:
  - List of CDN's in the order of their ranking
  - Location of trickplay data
  - For each quality level multiple Video/Audio chunk urls corresponding to each CDN .
  - Timing parameters like timing interval, polling interval etc
- **Trickplay**: This feature enables the user to pause,fast forward ,rewind the videos. All the data to perform trickplay, The thumbnail resolution, pixel aspect, trickplay interval, and CDN from where to download the trickplay file is available in the manifest file.
- **Audio and Video chunk downloading**:The audio and video data is downloaded in chunks to build up the player buffer. After the buffer is built, data is downloaded in periodic fashion where the interval between 2 consecutive downloads is 4s, typically the length of each playback.

2.2 Analysis of Manifest File

Effect of environmental constraints like geographic locations, client capabilities, and content type Eg: popular versus unpopular, movies versus TV shows on streaming parameters.

- **CDN ranking and user accounts**: Manifest file ranks CDN in the order of their preference. Experimental studies show that CDN ranking/assignment solely depends on the user account and is independent of geographic factors, movie types, computers, time etc. It has also been observed that CDN ranking for each user remains unchanged for several days. It is also independent of the available bandwidth from each CDN which implies that the cdn which is ranked 1 still remains 1 even if other CDN can offer a better video quality.
- **Data bit rates**: As mentioned above the user requests for a manifest file depending on the formats it can support and then the manifest file is sent to the user from the Netflix server. Manifest file contains urls for multiple bit rates like 100-1750 kbps for normal videos(12 different bit rates) and 2350 or 3600 kbps for HD video streaming (14 different bit rates).

2.3 CDN Selection Strategy

We know that the user has flexibility to choose between various bit rates and to select CDN for streaming and in order to test that, we congested bandwidth for a particular CDN using a dummy net which virtually created traffic and throttled the bandwidth. As a part of the experiment we lowered the bandwidth of the CDN in the order of their ranking one after the other, from 3900 kbps to 100 kbps.As expected, the client started downloading video chunks from the first CDN. As we lower the available bandwidth for the first CDN while leaving the other CDNs ,something interesting has been noticed. Instead of switching to a different CDN, which is not throttled, the client keeps lowering the bit rate and stays with the first CDN. Only when it can no longer support even the very low quality level (i.e when the available bandwidth for the first CDN reaches 100 kbps), it switches to the second CDN. It repeats almost the same behavior as we leave the first CDN at 100 kbps and gradually lower the available bandwidth for the second CDN while leaving the third CDN intact. In general, the Netflix clients stay with the same CDN as long as possible even if it has to degrade the playback quality level.

3 Hulu Steaming Platform

Unlike Netflix Hulu provides a subscription based subscription based plan and a free plan for its users. Monetization through advertisements is also an additional feature observed in Hulu which was absent in Netflix. Hulu has similar architecture as that of Netflix as even Hulu uses third party commercial data centers and CDN's for delivering content to its users The client gets the HTML pages for video from Hulu's front end web server at the main website www.hulu.com and the manifest file is obtained from - s.hulu.com which also provides instructions to contact a video server to download the video. As a part of user experience reporting, feedback is periodically sent to t.hulu.com .Unlike Netflix, HULU users can switch between bit rates during playback. Videos are streamed at 480kbps,700kbps and HD videos are streamed at 1000kbps.

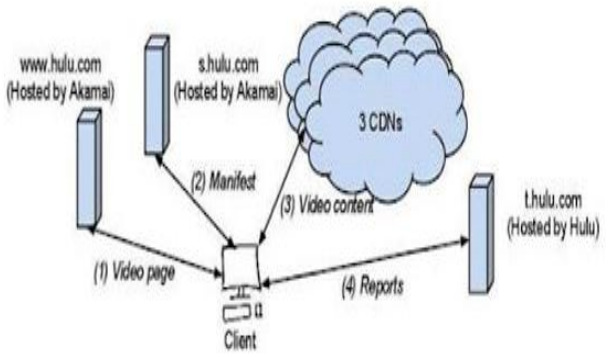


Fig. 6. High-level Hulu architecture.

3.1 Streaming Protocol

Hulu uses Real time messaging protocol (RTMP). Video are delivered in two ways, RTMP over port 35 or RTMP tunneled over HTTP (RTMPT). Level3 prefers raw RTMP, whereas Akamai and Limelight prefer RTMPT. All three CDNs use RTMP when TCP port 1935 is blocked (by a firewall for example).The same protocols are also adopted by Hulu plus. On mobile devices, Hulu Plus uses adaptive streaming over HTTP. For instance, on iPhone and iPad, Hulu Plus content is delivered over HTTP live streaming technology.

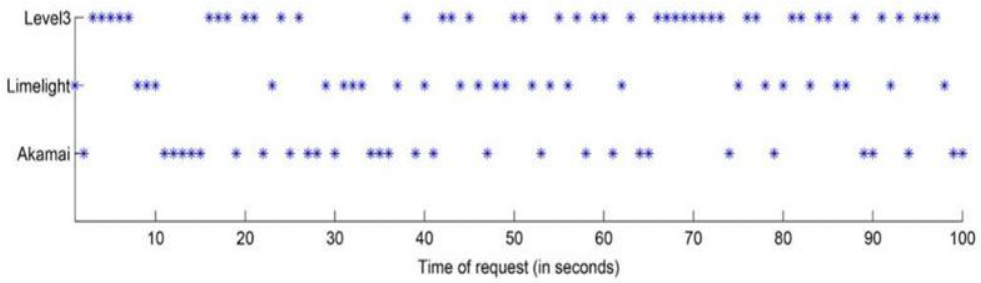
3.2 User Experience Reporting

Periodic status report from client which contains the information about video bit rate, current video playback position, total amount of memory the client is using, the current bandwidth, number of buffer under-runs etc is

sent to t.hulu.com where reports from all the locations in US are mapped to a single IP: 208.91.157.68.

3.3 Manifest files and CDN selection strategy

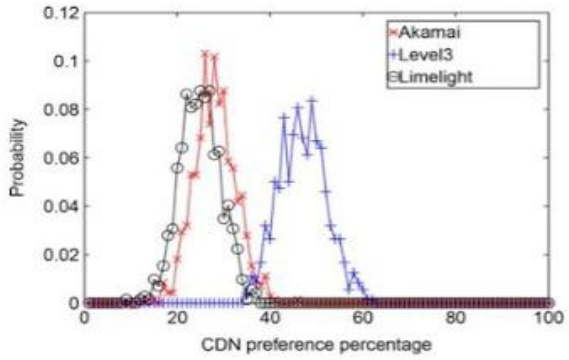
It has been experimentally observed that hulu sticks to a single cdn server throughout a video and switches to another cdn for the next video. The below figure shows a use case use case of how hulu selects different CDN's. The manifest file is requested for every 100s seconds from the same computer. "*" in the figure for a given CDN shows that it has been selected for that window of 100s. As these requests are made from the same computer with same networking conditions, it can be understood that CDN selection is not based on instantaneous network conditions.



3.4 CDN Usage Analysis

To understand CDN usage, an "experiment" has been performed where we download the manifest file 100 times with 1s interval between them. Each download has a selected CDN as it's assigned CDN. We define following metrics for each CDN.

- Preference Percentage: Number of times a particular CDN has been selected among all the three CDN's. Below figure shows us that each CDN has a distribution which is very close to the Gaussian and Level 3 has highest preference percentage with 47 percent .



Overall CDN preference distribution.

- CDN preference over different locations: We observe that there is a heterogeneous distribution of all three CDN's across various locations. The experiment is conducted over 13 locations which span across 8 states in the US. The below figure shows variation of average preference percentage of each cdn with respect to all locations. It can be observed that level-3 has significantly higher percentage than rest of two.

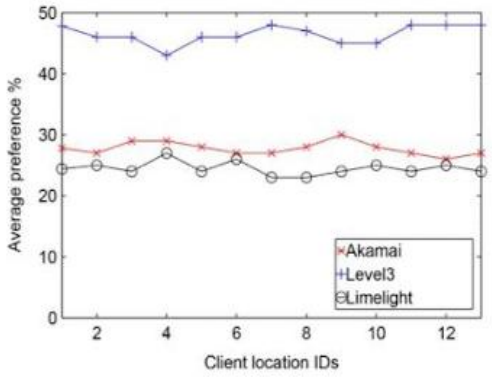
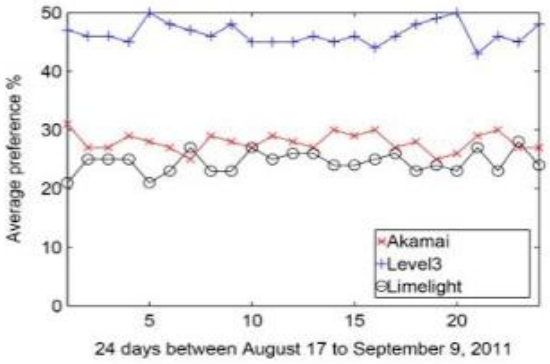


Fig. 10. CDN preference from geographic regions.

- Below figures show representation of distribution of CDN's over a range of time (24 day period in this case) and for different videos :



. CDN preference over time.

As a whole, we can conclude that Hulu selects CDN randomly following a fixed latent distribution for each of the playback requests. On average it has been observed that Level3 is preferred more than others. It can also be understood that CDN selection is not affected by client location or networking conditions.

3.5 CDN Performance Measurement

In both cases we have observed that Netflix ties CDN to user accounts and Hulu chooses CDN for video based on a latent distribution. In both cases geographic factors or other constraints do not affect the CDN selection strategy and both attempt to balance video traffic based on certain latent distribution. These observations suggest that the CDN preference and selection strategies employed by Netflix and Hulu are plausibly due to business considerations such as business relations etc between the content providers and CDN's. Therefore, we can clearly see that there is a trade off between CDN selection strategy, business constraints versus end user QOE.In order to evaluate performance of CDN we consider various metrics evaluated under different kinds of scenarios

3.6 Bandwidth Throughput

We measure bandwidth throughput of all the three CDN's across 95 vantage points. Video files URLs are collected for all the three CDN's from Netflix manifest files. We take advantage of the fact that the URL remains active for several hours after the manifest file is generated. We download data in terms of rounds where each round lasts for 96 s. Further each round is divided into 4 chunks where the first three chunks corresponding to three CDN's and then we download from all the three CDN;s in the last chunk. Each round lasts for 96s of size 1.8Mb and each chunk lasts for 24s. The above values for length and size are decided based on empirical observations. **Bandwidth Throughput = data downloaded/time taken**

The above experiment is performed on 12 residential sites and 83 Planetlab nodes for a duration of 2 hours

everyday for 19 days. We consider one day average bandwidth for each day and average bandwidth over all the 19 days. Let us assume the CDN's Akamai, Limelight, Level 3 as A,B,C, respectively for proper convenience.

Fig. 13. Best CDN at each vantage point.

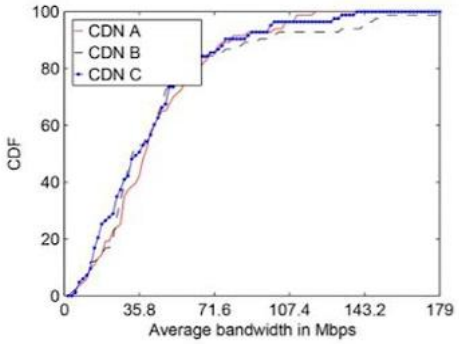


Fig. 14. CDF of average bandwidth at PlanetLab nodes.

Overall CDN performance throughout the experiment: The below figure shows the top CDN across all the vantage points during the measurement period.

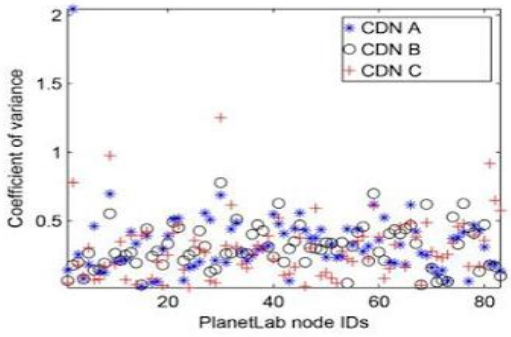
From the above CDF graph of average bandwidth of all the three CDN's it can be seen that no CDN outperforms the other as three of them have similar distribution.

3.7 Further Analysis

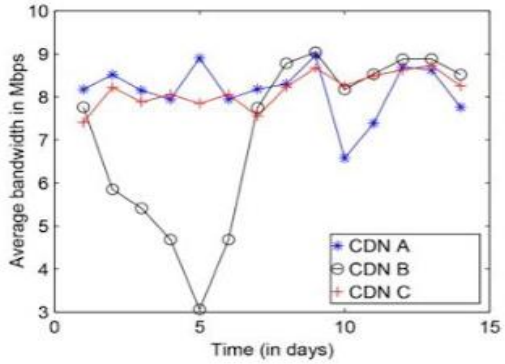
The above two graphs represent variation of average band width across all the Planetlab nodes and residential areas.We can observe that the bandwidth range in planet lab node spanned for a higher range varying from 3 to 200 mbps whereas residential areas had limited range.

- In few residential ID's 4,9,11 we see that CDN B significantly outperforms the other two CDN's Daily Bandwidth variations.
- We examine bandwidth variations at different sites over various time scales over all the three different CDN's. We consider coefficient of variance of daily bandwidth as an evaluation metric where : **COV= Standard Deviation/Mean** (Calculated for each location)

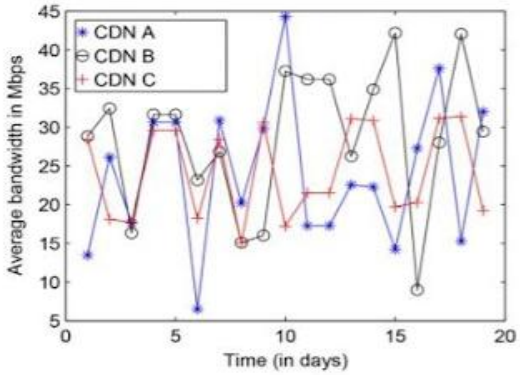
Average values of COV: CDN A: 0.33 CDN B: 0.3 CDN C: 0.3



Coefficient of variance for the one-day average at PlanetLab nodes.



. One-day average bandwidth over time at residential site 7.

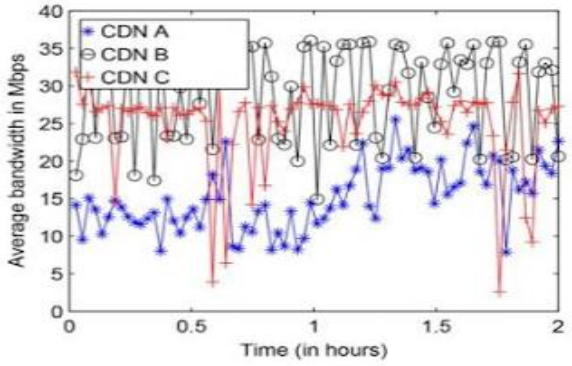


. One-day average bandwidth at a PlanetLab node over time. . C

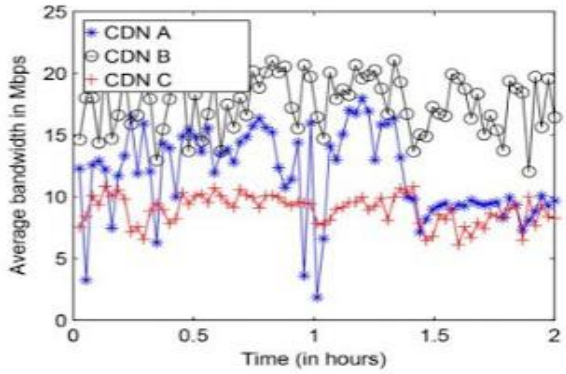
The above two graphs show the variation of one day average bandwidth across a Planet Lab node and across a residential site. Significant variations in bandwidth are also observed which suggest further improvements in CDN selection strategies

3.8 Instantaneous Bandwidth Variations

The following figures show the variation of instantaneous bandwidth for a duration of 2 hours. As DASH player periodically monitors instantaneous bandwidth and requests for next chunk depending on the current speed, looking into variations in bandwidth can help to track how DASH player works. The small timescale bandwidth may significantly impact the Netflix users viewing experience as 2 hours is a typical length of a movie. The below figure shows variation of instantaneous across planetlab nodes and residential site .



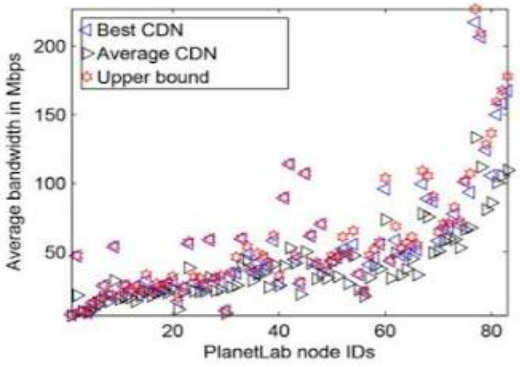
Instantaneous bandwidth at a PlanetLab node.



Instantaneous bandwidth at residential site 7.

3.9 Alternate video delivery strategies

Let upper-bound average bandwidth denote the maximum potential bandwidth which can be achieved by optimal and dynamic selection of CDN instead of sticking to one for a given video.

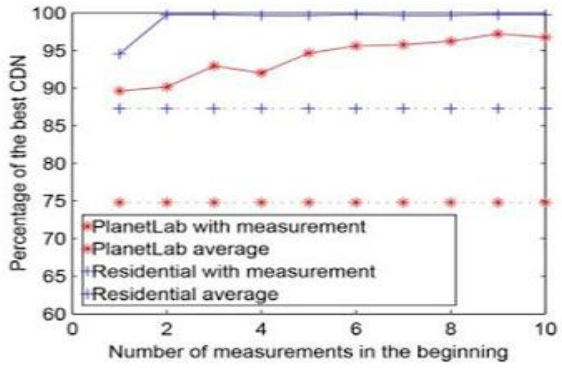


Average bandwidth and the upper bound at PlanetLab nodes.

We can see that the upper bound is 33 percent better than the average CDN in the above case.

3.10 Measurement based CDN selection

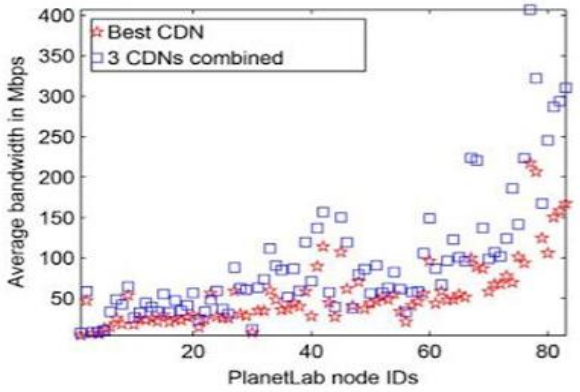
The player can conduct multiple bandwidth measurements in the beginning and can assign a cdn for the movie based on them. From the below figure it can be observed that as the number of measurements increases the probability that the selected CDN is close to the best CDN also increases.



Effect of number of measurements.

3.11 Using multiple CDN's simultaneously

The player can download three different chunks from 3 different CDN's in order to obtain larger bandwidth. The below figure shows that we can obtain promising results by using all the three CDN's as there has been good increment in average bandwidth.



4 Conclusion

We have provided a detailed analysis about each component present in the architectures of Netflix and hulu and along with a detailed series of experiments with quantitative results,hence giving more insight about the functionality of each block. Various CDN's and the selection strategies employed by NETFLIX and HULU are also studied and also a few adaptable strategies to improve bandwidth and efficiency of CDN's.

5 References

- https://en.wikipedia.org/wiki/Content_delivery_network
- MeasurementStudyofNetflix,Hulu,andaTaleofThreeCDNs-OfficialPaper
- https://www.cs.princeton.edu/courses/archive/fall14/cos561/papers/NetFlix12.pdf
- https://en.wikipedia.org/wiki/Over-the-top_media_service