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

Niharika V

Sivani Pentapati

Contents

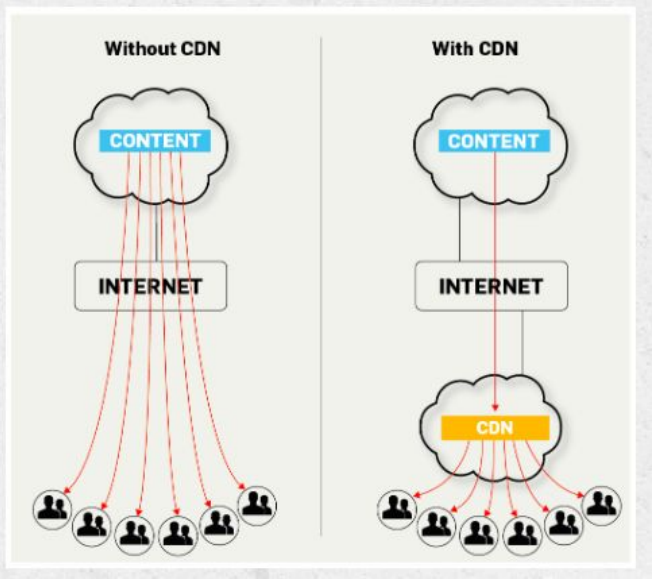
- Introduction : Motivation of the paper
- Concept : Content Delivery Networks
- Netflix Architecture & Analysis
- Hulu Architecture & Analysis
- Measurement Part
- Conclusion

Motivation

1. Netflix and hulu are few of the top content service providers existing in the current trends and technology.
2. Shedding light on the architecture and analysing the performance of then would give us an insight of how state of the art content service providers work and would give us more room to come up with new innovative models or could help us improving efficiency of the existing models
3. In order to achieve this we propose new CDN strategies that can be implemented in real life cases.

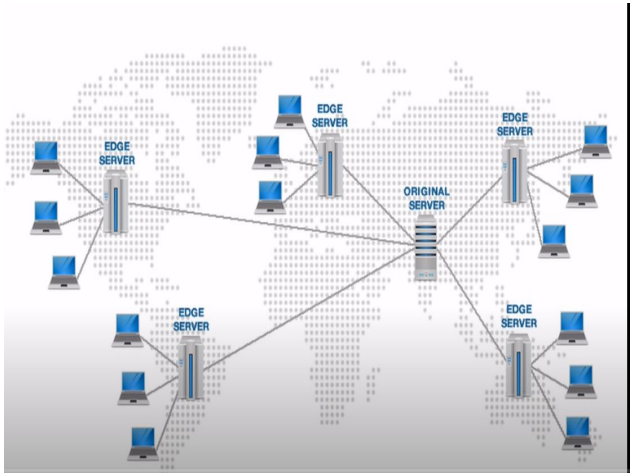
Content Delivery Networks

- A Content Delivery Network (CDN) is a globally distributed web server network or Points of Presence (PoP) whose purpose is to provide faster content delivery.
- The content is replicated and stored throughout the CDN so the user can access the data that is stored at a location that is geographically closest to the user. This is different (and more efficient) than the traditional method of storing content on just one, central server.
- A client accesses a copy of the data near to the client, as opposed to all clients accessing the same central server, in order to avoid bottlenecks near that server.

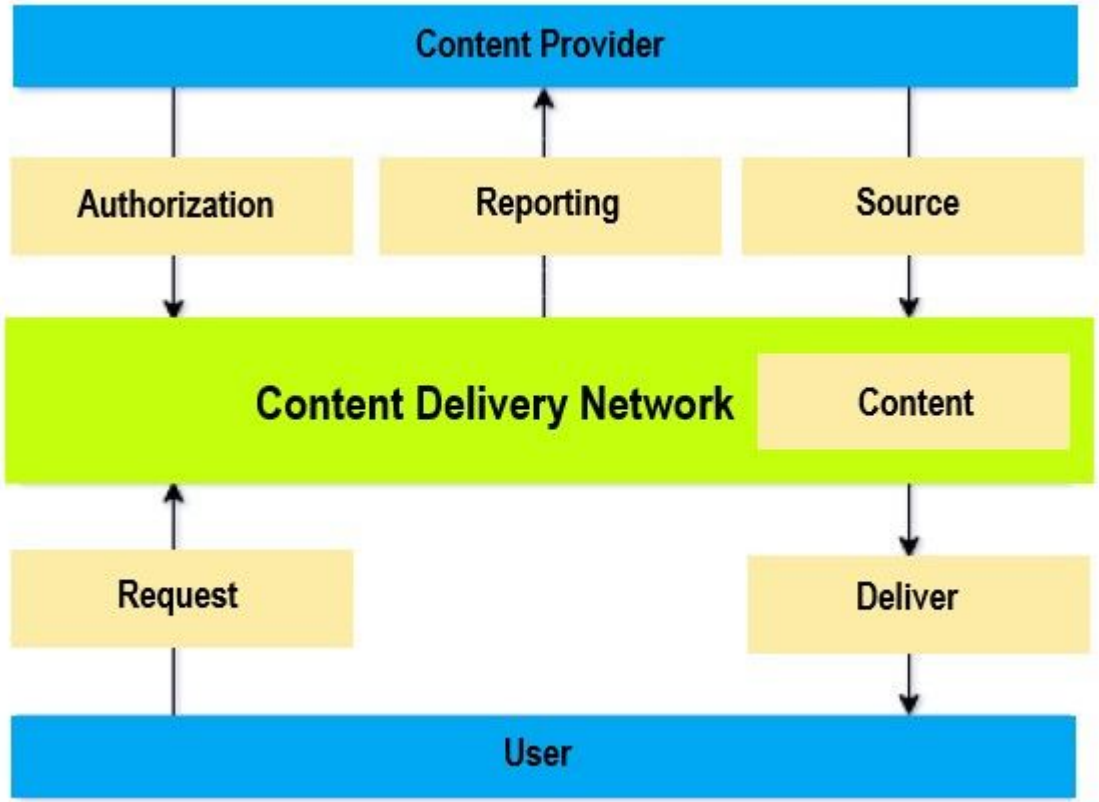


Advantages of CDNs

- High content loading speed = positive User Experience. If all data is located on a central server, the user experience is negatively affected by limited loading speed
- Less Network Congestion: Overcomes Bandwidth & Latency issues.Offload the traffic served directly from the content provider's origin infrastructure.
- Effective against DDoS attacks , since they provide their own architecture.
- Effective Automatic Data Analytics : Data analytics allows companies to improve their business model and see which practices work, which don't, and where there is room for growth.

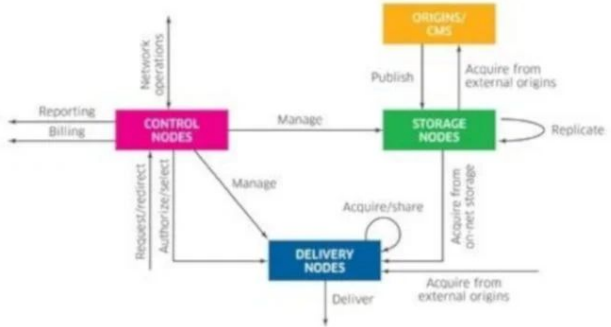


CDNs : Basic Architecture & Functioning



- Content Provider : Eg-Netflix,Amazon,Flipkart,etc.
- Autorisation :Permission to CDN for delivering content of the Content Provider.
- Report : Performance Analytics of CDN
- Source : Original data storage(main server)
- Content:Copy of Source(Cache)

Components of CDN



1 . Delivery Nodes

- Primary purpose is to deliver content to the end user
- Servers that contain caches running one or more content delivery applications. They are typically located as close to the end user as possible.
- Content can be stored manually to these nodes (Push CDNs), or delivery nodes can demand content from origin nodes based on Cache Expiration Rules (Pull CDNs).
- The advantage of Push CDNs is that the content is instantly available to the users demanding it. Disadvantage is that the Content Provider has to proactively “push” content every time it is updated.
- The advantage of Pull CDNs is they automatically demand content from the Content Provider. Its main disadvantage is the initial content delivery speed: when the user demands content for the first time, the delivery speed will be the same as if the Content Provider didn’t use CDN.

Components of CDN

2. Storage Nodes

- Primary purpose is storing copies of original data that is being distributed to Delivery Nodes. Storage Nodes can be deployed in a hierarchical model to allow tiered caching.

3. Origin Nodes

- These are the main sources for content that enable content distribution across the entire network or the content owner's infrastructure.

4.Control Nodes

- Primary purpose is hosting of management, routing and monitoring components of a CDN.

CDNs in Market

- AKAMAI : Serves content to 15% to 30% of overall global traffic as well as for some well respected companies.
- Akamai's CDN infrastructure counts more than 240,000 servers located in more than 130 countries around the world.
- **Amazon CloudFront:** Amazon CloudFront is a CDN service particularly targeted towards developers provided by Amazon Web Services (Amazon AWS), a company established in 2006.

Akamai's PoPs (Point of Presence):

REGION	Number of PoPs
NORTH AMERICA	1000
SOUTH AMERICA	75
EUROPE	700
ASIA	300
MIDDLE EAST	75
AFRICA	50
TOTAL	2200

Amazon CloudFront's PoPs (Point of Presence):

REGION	Number of PoPs
NORTH AMERICA	42
SOUTH AMERICA	5
EUROPE	35
ASIA	31
AUSTRALIA	4
TOTAL	117

CDNs in Market (Contd)

- **Google CDN** : Cloud CDN uses Google's global infrastructure (the same infrastructure that Google uses to deliver their end-user products like Google Search and Youtube) to cache and deliver content for their clients that are also users of the Google Cloud Platform (GCP).
- **Limelight Networks** : A company founded in 2001 that owns and manages a global, private Content Delivery Network, enabling publishers to deliver their digital content (videos, operating system updates, online games, etc.) on any device, anywhere in the world.

Google Cloud PoPs (Point of Presence):

REGION	Number of PoPs
NORTH AMERICA	20
SOUTH AMERICA	5
EUROPE	20
ASIA	15
AUSTRALIA	3
AFRICA	5
TOTAL	68

Limelight's PoPs (Points of Presence):

REGION	Number of PoPs
NORTH AMERICA	19
SOUTH AMERICA	11
EUROPE	26
ASIA	17
AUSTRALIA	4
AFRICA	3
TOTAL	80

NETFLIX ANALYSIS

Netflix Architecture

- The 4 main components of entire Netflix Architecture would be :

  - Netflix data center

  -Amazon cloud

  -CDNs (LimeLight, Level3,Akamai)

  -Players

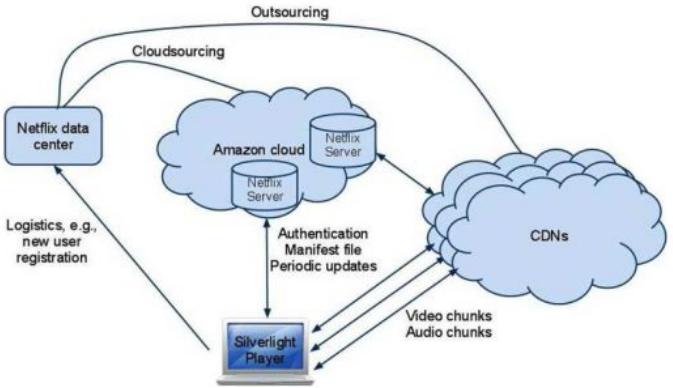


Fig. 1. Netflix architecture.

Netflix Architecture Cont

- Netflix Data Centers:
  - The website : www.netflix.com is hosted at the Netflix Organisation's IP space.
  - Deals with 1.Registration of new accounts 2.Payment activities : PayPal
  - Redirects to other pages based on user.
  - This server DOES NOT interact with user while Video Streaming .
- Amazon Cloud : -Key functions, such as content ingestion, log recording/analysis, DRM, CDN routing, user sign-in, and mobile device support. All queries go here, anything non-streaming related.

TABLE I
KEY NETFLIX HOSTNAMES

Hostname	Organization
www.netflix.com	Netflix
signup.netflix.com	Amazon
movies.netflix.com	Amazon
agmoviecontrol.netflix.com	Amazon
nflx.i.87f50a04.x.lcdn.nflximg.com	Level 3
netflix-753.vo.llnwd.net	Limelight
netflix753.as.nflximg.com.edgesuite.net	Akamai

- The authors performed DNS resolutions to collect the canonical names (CNAMEs) and IP addresses of all the server names that the browser has contacted.
- Based on the IP Address,and WHOIS lookup they were able to figure out the owner.

Netflix Architecture Contd.

- **CDNs** : Multiple CDNs to deliver the video content to end-users. The encoded and DRM protected videos are sourced in the Amazon cloud and copied to CDNs .
- **Players** :
- Netflix uses Silverlight to download, decode, and play Netflix movies on desktop Web browsers. The run-time environment for Silverlight is available as a plug-in for most Web browsers. There are also players for mobile phones and other devices such as Wii, Roku, etc.

TABLE I
KEY NETFLIX HOSTNAMES

Hostname	Organization
www.netflix.com	Netflix
signup.netflix.com	Amazon
movies.netflix.com	Amazon
agmoviecontrol.netflix.com	Amazon
nflx.i.87f50a04.x.lcdn.nflximg.com	Level 3
netflix-753.vo.llnwd.net	Limelight
netflix753.as.nflximg.com.edgesuite.net	Akamai

- The authors performed DNS resolutions to collect the canonical names (CNAMEs) and IP addresses of all the server names that the browser has contacted.
- Based on the IP Address,and WHOIS lookup they were able to figure out the owner.

Netflix :Service Analysis

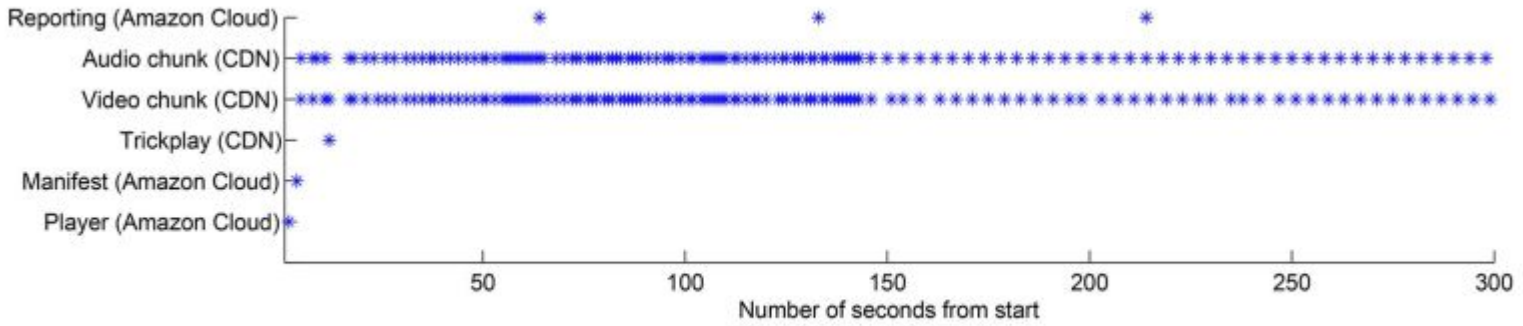


Fig. 2. Timeline in serving a Netflix client.

Netflix Servicing a client.

The major steps involved while user opens Netflix and starts watching a movie :

- Silverlight Player Download and User Authentication: (RIA)
- First, the user gives his credentials and get authentication .

  -Video playback on a desktop computer requires the Microsoft Silverlight browser plug-in to be installed on the computer.

  - When the user clicks on the “Play Now” button, the browser downloads the Silverlight application, and then that application starts downloading and playing the video content. This small

- **Netflix Manifest File** :

  -The Netflix manifest file provides the DASH player metadata to conduct the adaptive video streaming. (Changes with client)

  -For instance, if the user player indicates it is capable of rendering h.264 encoded video, h.264 format video is included in the manifest file. If the player indicates that it can only play back .wmv format, only .wmv format video is included

  -Manifest file - A metadata file transferred over SSL,such that it content of the file cannot be read over the wire using packet capture tools such as **tcpdump or wireshark.**

  -The extracted manifest file is in XML format and contains several key pieces of information including the list of the CDNs, location of the trickplay data, video/audio chunk URLs for multiple quality levels, and timing parameters such as timeout interval, polling interval, and so on.

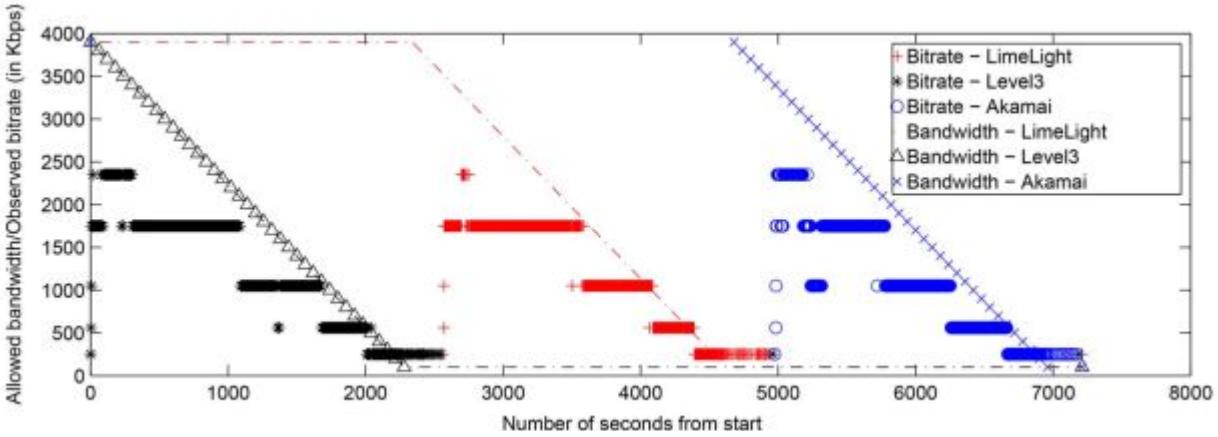
- **Trickplay**: Netflix Silverlight player supports simple trickplay such as pause, rewind, forward, and random seek.
  - Trickplay is achieved by downloading a set of thumbnail images for periodic snapshots. The thumbnail resolution, pixel aspect, trickplay interval, and CDN from where to download the trickplay file are described in the manifest file.
- **Audio and Video Chunk Downloading:**
  - As shown in figure, audio and video contents are downloaded in chunks.Download sessions are more frequent at the beginning so as to build up the player buffer.
  - Once the buffer is sufficiently filled, downloads become periodic.
  - The interval between the beginning of two consecutive downloads is approximately 4 s—the playback length of a typical chunk.
- **User Experience Reporting**: After the playback starts, Netflix player communicates periodically with the control server agmoviecontrol.netflix.com. Based upon the keywords such as “/heartbeat” and “/logblob” in the request URLs and the periodicity of the communication, we conjecture that they are periodic keep-alive messages and log updates.

Best CDN Selection Strategy

- Authors of this paper have followed the procedure :
- Given , a client he is allowed to stream the movie in different bit rates and resolutions .
- Hence ,the authors conduct experiments to help understand how Netflix makes such choices when bandwidth is dynamic, given its huge array of choices with the CDNs.
- Once the playback starts, we gradually throttle the available bandwidth of the top-ranked CDN in the manifest file.
- At the beginning, servers from each CDN are allowed to send data at 3900 kb/s. After every minute, we reduce the available bandwidth for the current CDN by 100 kb/s until it reaches 100 kb/s.
- In this instance, Level3, Limelight, and Akamai CDNs are ranked first, second, and third, respectively.

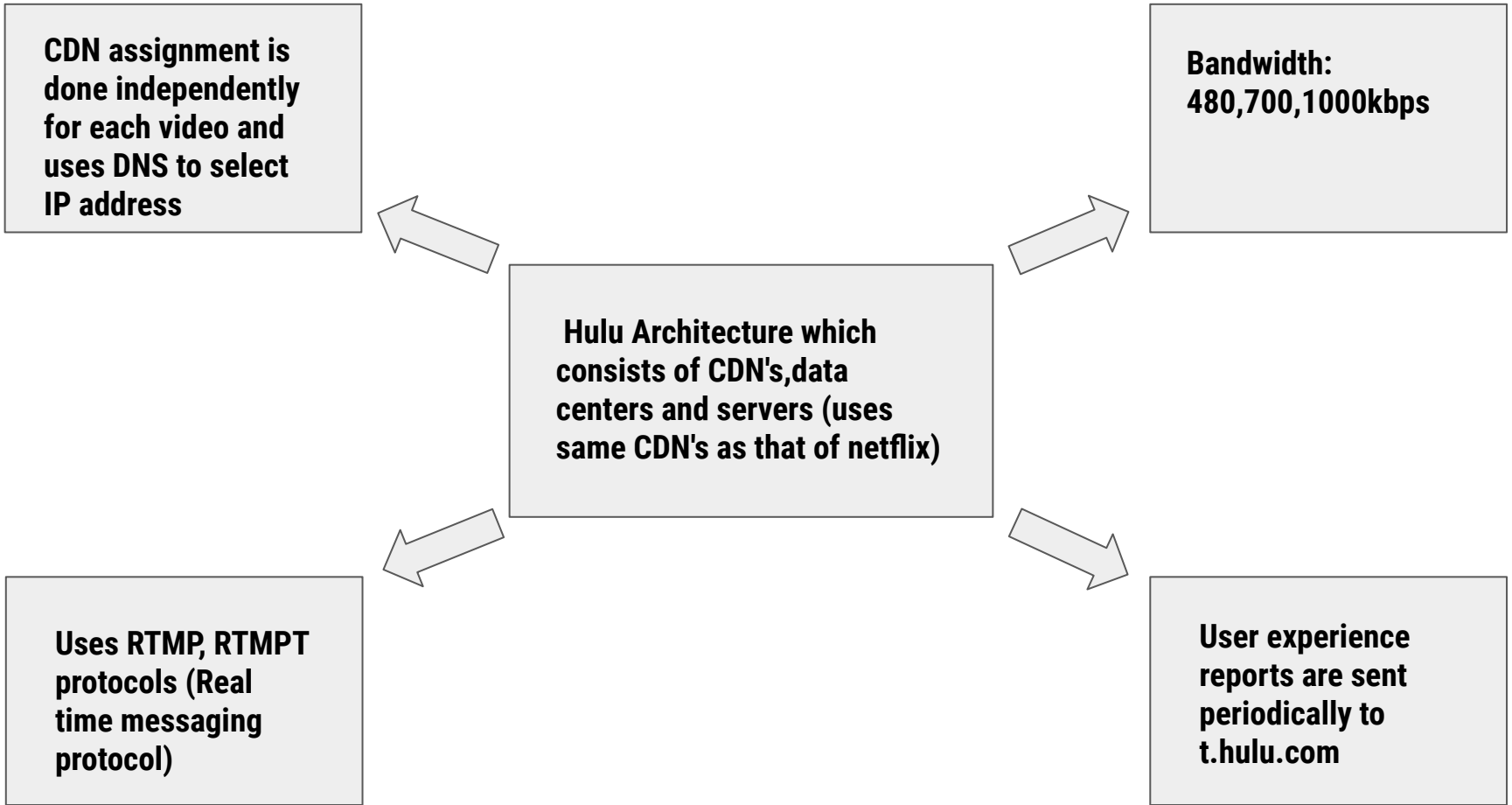
CDN Experiment

- 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 kb/s), it switches to the second CDN. It repeats almost the same behavior as we leave the first CDN at 100 kb/s 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.



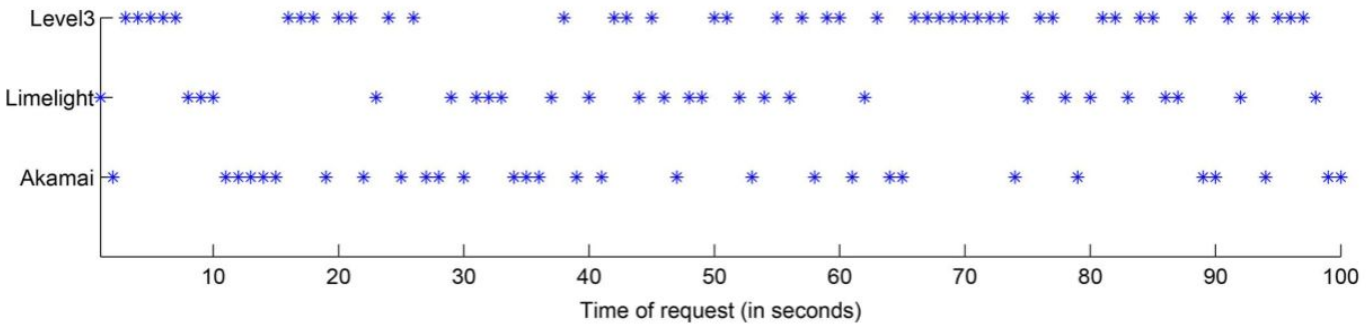
HULU ANALYSIS

Hulu Architecture



CDN selection strategy

1. It has been observed that hulu uses a single cdn throughout a single video and switches to other CDN in the next video
2. An experiment has been performed where we throttle the bandwidth of the selected CDN and observe how the system to responding to the changes
3. It has been observed that the client sticks to the same cdn even if other CDN's can offer a better quality videos and shifts to another CDN only if the quality of the video supported by this CDN is very low



CDN usage analysis

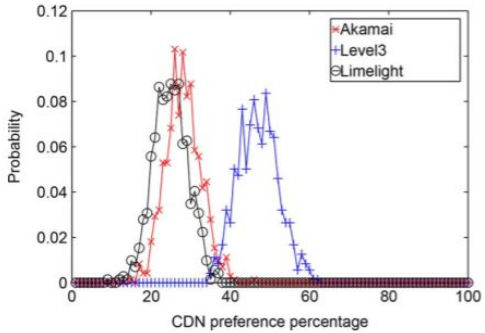
Objective: To understand impact of various factors such as client location,video, time on CDN selection

Experiment: For a given video we download manifest files with 1s interval for 100 times.

Preference percentage: Number of times a particular CDN is chosen out of all three in the 100 files

Observation: We observed that curves representing each CDN is very close to Gaussian Distribution

and level 3 had highest preference percentage 47% ,while limelight and akamai has 25% and 28%



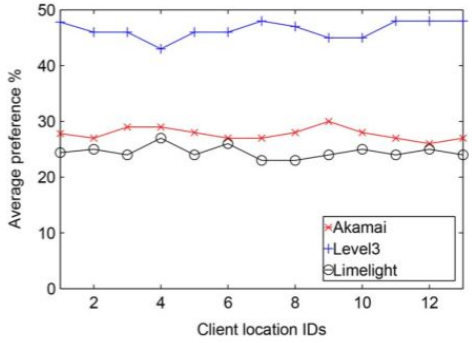
Overall CDN preference distribution.

CDN preference over different locations and videos and over a range of time

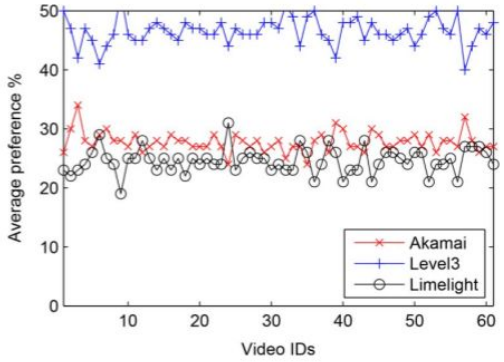
Objective: To check how above factors would affect CDN selection strategy

Experiment: The experiment as defined before is performed for multiple times over multiple locations and with different videos we also perform the same experiment in the same location for 24days

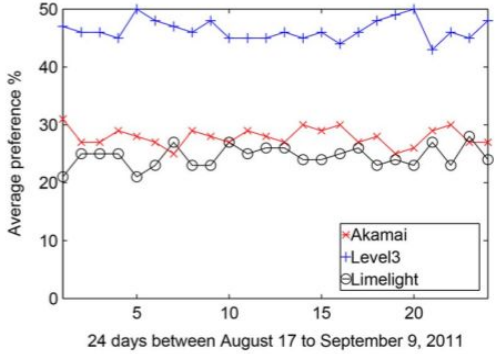
Observations: All the observations are made with the average preference percentage (the mean of all values obtained in multiple experiments across different locations and for different videos. We observe that different CDNs have different popularity, but the popularity does not change over different locations as Level 3 always had the highest score. It has also been observed that CDN preference does not change with time either



CDN preference from geographic regions.



. CDN preference for different videos.



. CDN preference over time.

Inference

1. From the results of the above experiment we can conclude that Hulu selects the preferred CDN randomly following a fixed latent distribution for each of the playback requests. On average, one CDN (Level3) is preferred more than others, but such selection preference does not seem to depend on instantaneous network conditions or location or time
2. We therefore deduced that such CDN preference is most likely based on pricing and business arrangements and is not dependent upon instantaneous bandwidth or past performance history of the CDNs.

CDN performance measurement

Objective: To see how performance metrics vary for each CDN over external constraints and other factors.

Experiment:

1. Extract urls from manifest files
2. Send get requests to all the CDN and use DNS to obtain the IP
3. Download data from each cdn in round robin order and 3 rounds and in the last round, download data from all the cdn's

We consider how the following performance metric vary:

1. Bandwidth throughput= data downloaded/time taken
2. Daily Bandwidth Variations
3. Instantaneous Bandwidth Variations

The above experiment is done at 95 locations with 12 residential sites and 83 planetlab nodes

Overall CDN performance

Best CDN at each vantage point.

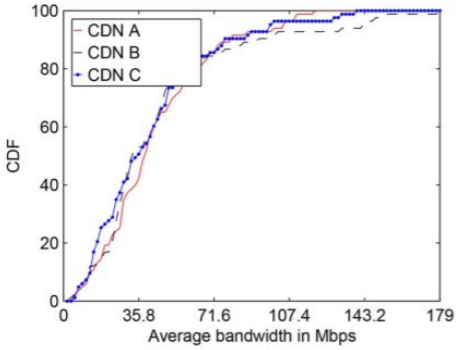
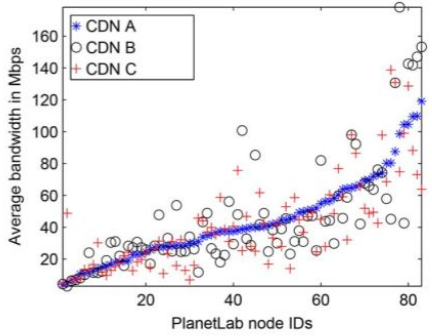
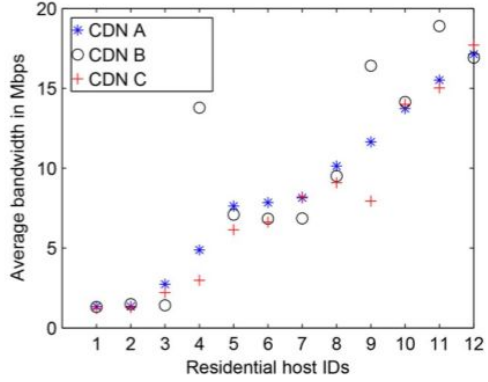


Fig. 14. CDF of average bandwidth at PlanetLab nodes.



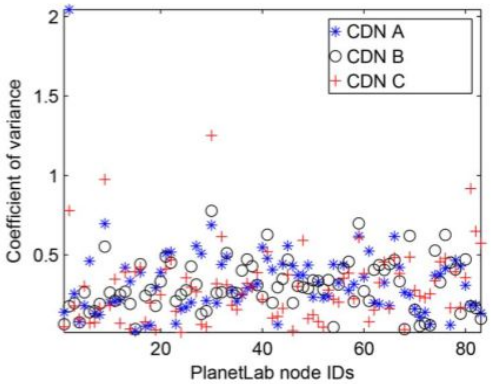


Inference

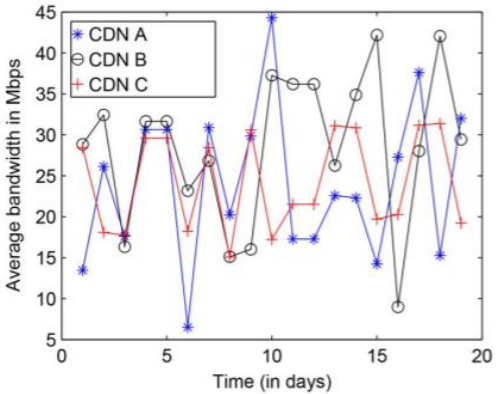
1. All the three CDN's have similar CDF graph for planetlab nodes
2. Level 3 has significantly high average bandwidth at residential host IDs 4,9,11
3. All the CDN's have similar average bandwidth at planet lab nodes
4. We can observe heterogeneous distribution of first ranked CDN across all vantage points
5. Also planetlab nodes have relatively higher bandwidth than residential nodes

Daily Bandwidth variations

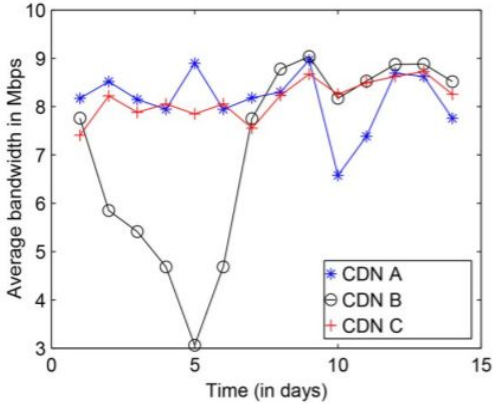
1. Metric : Coefficient of variance = standard deviation /mean (of average bandwidth)
2. We observed that bandwidth varied between 3mbps to 200mbps
3. The average COV for three CDNs was 0.33,0.3,0.3 respectively



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



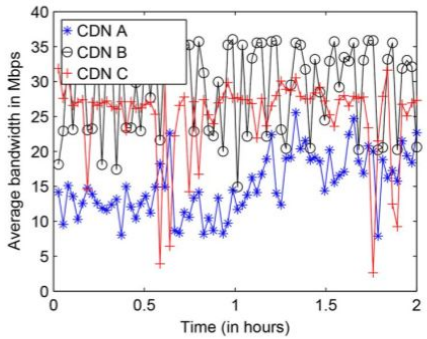
One-day average bandwidth at a PlanetLab node over time.



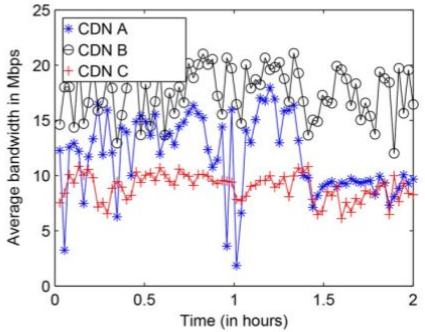
One-day average bandwidth over time at residential site 7.

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 .



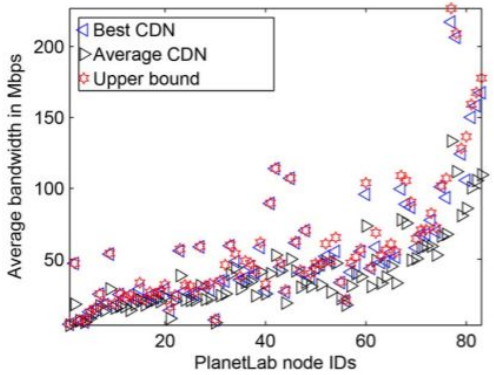
Instantaneous bandwidth at a PlanetLab node.



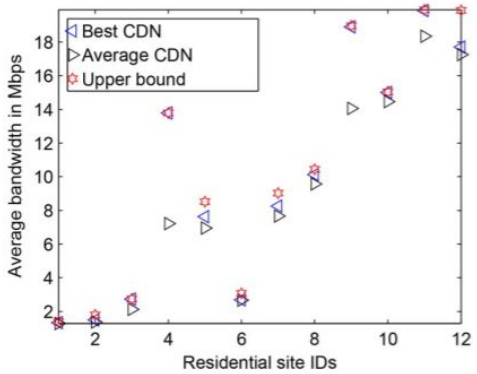
Instantaneous bandwidth at residential site 7.

Room for improvement

Upperbound average bandwidth: The bandwidth obtained by considering the best CDN all the time instead of sticking to a single CDN (as how netflix and hulu does stick to a single CDN for a video). The below graph shows that upab is always higher than bandwidth with one single CDN suggesting that there is room for improvement



Average bandwidth and the upper bound at PlanetLab nodes.



Average bandwidth and the upper bound at residential sites.

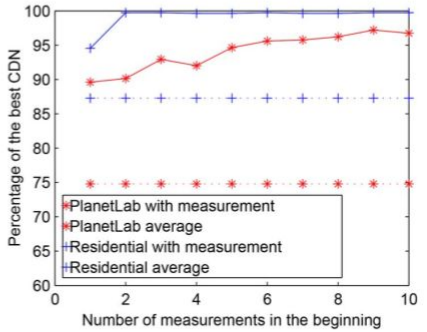
Alternate Video Delivery Strategies

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

Below graph represents the percentage of best CDN (The probability that the selected CDN is best CDN)

We can observe that the probability increases with increase in number of measurements



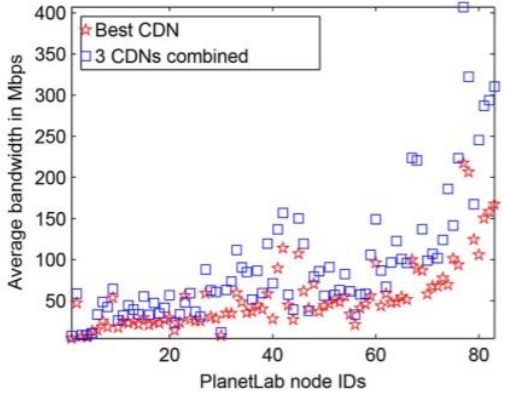
Effect of number of measurements.

Alternate Video Delivery Strategies.. contd

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.

From the below graph, we can observe that the average bandwidth in the case of 3 CDN's combined is always higher than the average bandwidth of the best CDN



Conclusion

1. As a whole we have analysed architecture of netflix and hulu and provided a detailed analysis describing how they work and performed multiple experiments which output how various CDN's work
2. We also provided many performance metrics and how they vary in vaious kind of scenarios. Under metrics we also checked how bandwidth throughput varies
3. We also proposed alternative CDN selection strategy to improve the efficiency of exising models