Life Cycle of Videos in DTube with Distributed Storage Network (IPFS)

Final report for CSE 534 Fundamentals of Computer Networks - Fall 2021

ZhengYu Wu (111601422), YuHao Wang (114408039)

zhenwu@cs.stonybrook.edu, yuhaowang@cs.stonybrook.edu

https://github.com/w774908117/CSE-534-Project-DTube

Abstract

DTube, a video streaming platform that aims to build a distributed and decentralized platform by utilizing distributed services such as IPFS. InterPlanetary File System(IPFS), a peer-to-peer hypermedia protocol, with its decentralized design model, is expected to become the next-generation Internet standard that replaces the HTTP protocol. We collected DTube's daily hot, trending, and new videos over a month. We found out that most of DTube's videos are YouTube's redirect. We also found that most of the video is not reachable through regular IPFS clients, but we found out that DTube's IPFS still uses IPFS as storage and it has a lower connection time.

1 Introduction

With the surging fact that today's Internet is being dominated by a few giant companies. It comes as no surprise that the web is more centralized than ever because all the data is being hosted and controlled by these companies. While a centralized structure can provide easy management and maintenance, it comes with other issues. With all the data being stored in one place it can easily be the target to be attacked thus losing all the data. In addition, it can be used as a political weapon resulting from censorship.

Interplanetary File System (IPFS) is a distributed file system used to store and access files, websites, applications, and data. IPFS supports the creation of fully distributed applications, allowing file data stored on IPFS to be quickly obtained anywhere in the world, and designed to preserve and grow humanity's knowledge by making the web upgradeable, resilient, and more open[1]. DTube is a decentralized video platform, which is derived from IPFS as the core.

DTube's goal is to build a decentralized video website similar to YouTube with IPFS. 1 video is shared every 3 minutes on DTube and they already have 189,000 registered creators and curators, 2.1 Million Monthly Unique Visitors, and 1.1 Million distributed to the community[2]. They have been making continuous progress over the years, working towards their goals step by step.

2 Problem Statement

YouTube is the largest video platform in the world. A large number of videos are uploaded to YouTube every day. But at the same time, because YouTube is centralized, YouTube's administrators can easily block some content they do not want to show or decide which content will be hot or appropriate. This undermines fairness, making the administrator be the decision-maker to issue video content. DTube uses IPFS as their web's core, which is a very innovative and great decision. Because this can ensure that the video on DTube is decentralized and persistent, and will not be easily blocked.

The purpose of DTube is very ambitious, but we don't know how much benefit they decided to use IPFS brought to us. We plan to do some exploratory research on DTube to see if the various network optimizations of the DTube have met their original expectations. For example, how effective is DTube's decentralization? How DTube utilized the source nodes and how is the life cycle of videos in DTube.

3 Experiment Setup

To deeply understand how distributed and decentralized DTube is, we set up a script to record DTube's daily hot, trending, new videos over November and across the following locations: New York (US), California (US), Belgium (EU), and Tokyo (JP). The script will first retrieve all video information within each category and identify what is the video source type. Then if the video is using IPFS the script will store the CID information and try to access the video through our IPFS client (local gateway), as well as through DTube's public IPFS gateway.

For each video we are recording the following static:

- Video Duration
- Bandwidth
- Connection Time

In addition, for each video that is accessible through a regular IPFS client, we tried to measure how many IPFS nodes the video has and what is the RTT to our vantage point.

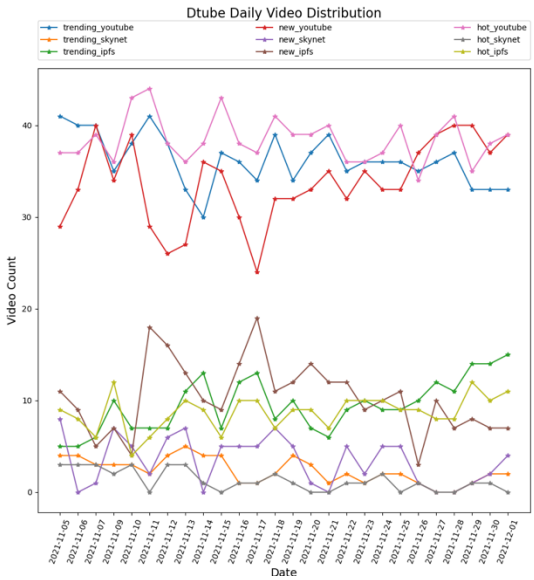
4 Evaluation

In this section, we evaluate the following questions with respect to video playing on DTube:

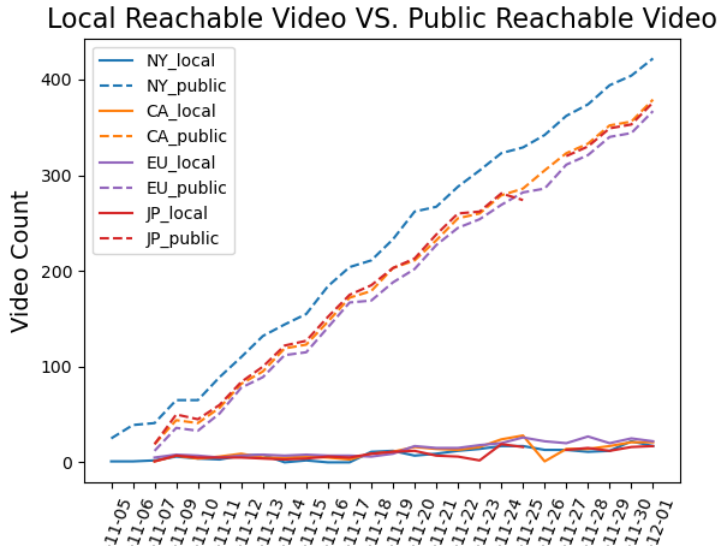
- What is DTube made of? What kind of video storage solution DTube is using?
- What is the bandwidth when playing video on DTube?
- How long does it take to start playing the video, and does it affect by geolocation?
- Will video be served by different peers and if so how many peers have the video?

4.1 Video Storage Type Distribution

The figure below shows the video storage type distribution for daily hot, new, trending. As we observed, DTube has a lot of YouTube redirects, about twice as many as other types. This raises our concern of wherever DTube is moving in the direction that they promised to build a distributed video platform. We also observed that IPFS in general has a higher number of videos than Skynet.



In addition, we compared the public gateway reachable video with the local gateway (regular IPFS client) reachable video count.



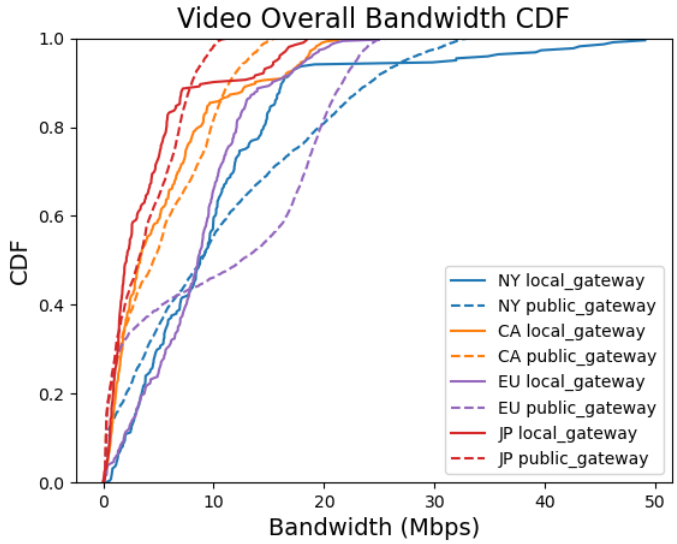
The figure above shows the video count for the local gateway and public gateway. Most of the video is not reachable through a regular IPFS client, but it has video CID embedded to DTube's site. We suspect that DTube uses a private IPFS network and clusters to store the video. Upon request to the public gateway, the gateway will find the video corresponding to the CID and serve it to the client. This explains why the video has IPFS CID but is not reachable through the regular IPFS finding process.

4.2 Video Playing Bandwidth

In this section, we analyzed the bandwidth regarding local gateway and public gateway. We also analyzed the bandwidth with respect to video reachable through regular IPFS clients and DTube's public gateway.

4.2.1 Overall Bandwidth

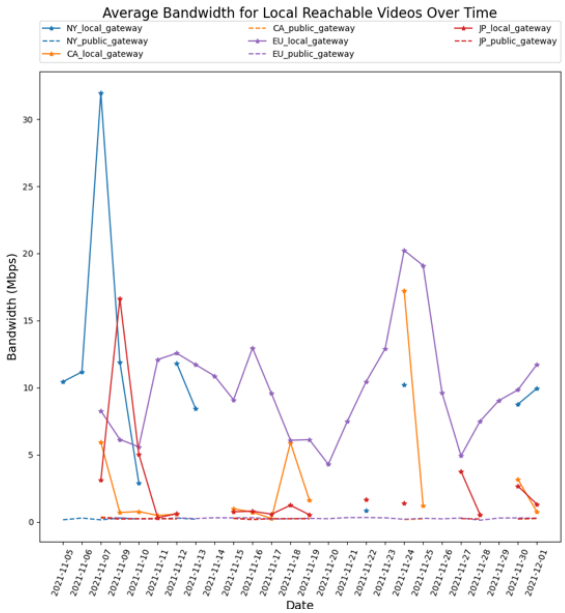
In order to study whether there is a difference in the impact of local gateways and public gateways on the overall bandwidth in different regions, we drew the CDF image with the bandwidth as the abscissa, as follows



This figure showed the overall bandwidth CDF distribution among all videos we recorded from the local gateway and public gateway. We observed that NY had the highest bandwidth and JP had the smallest bandwidth. More interestingly, almost every public gateway's bandwidth is higher than the local gateway's bandwidth. This indicates that DTube's public gateway may have cache in the gateway or DTube has less overhead when using their own IPFS network. The EU had a strange curve on the public gateway we suspect may be due to updates on the script and causing inaccurate data.

4.2.2 Bandwidth with Reachable Videos

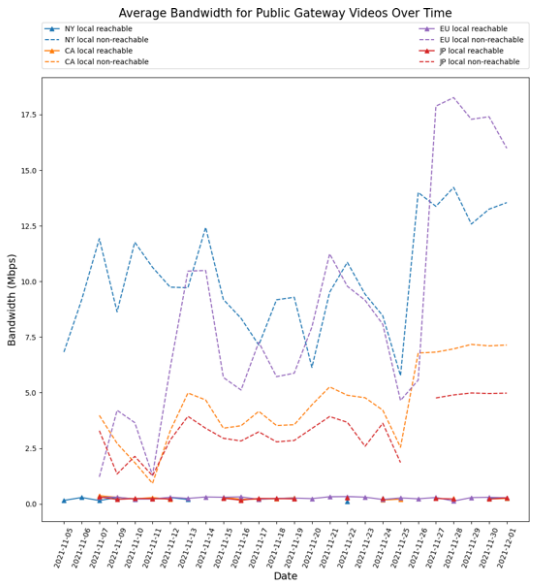
To further examine the Bandwidth regarding the video that is reachable through the IPFS network, we compared the average bandwidth for local gateway and public gateway from beginning to end.



The figure above shows how local reachable video's bandwidth compares to its public gateway bandwidth. We observed that in general local gateway has a much higher bandwidth compared to the public gateway when accessing the same video. This aligned with our over bandwidth figure and further strengthened our suspicion on DTube's public gateway has more overhead than directly using IPFS clients.

** Note: The discontinuity in the graph indicate that the local gateway was not able to reach the content*

However, the public gateway bandwidth is extremely low compared to the overall public gateway bandwidth. We wonder what is the difference between the public gateway bandwidth among local reachable video and non-reachable video.



This figure showed a comparison between average public bandwidth (local gateway reachable) and public bandwidth (local gateway not-reachable). We observed that the bandwidth is higher when the video is not reachable. We suspect that when a video is reachable through the IPFS network, the DTube gateway will follow the same way as a traditional IPFS client and go through the finding process thus causing additional overhead.

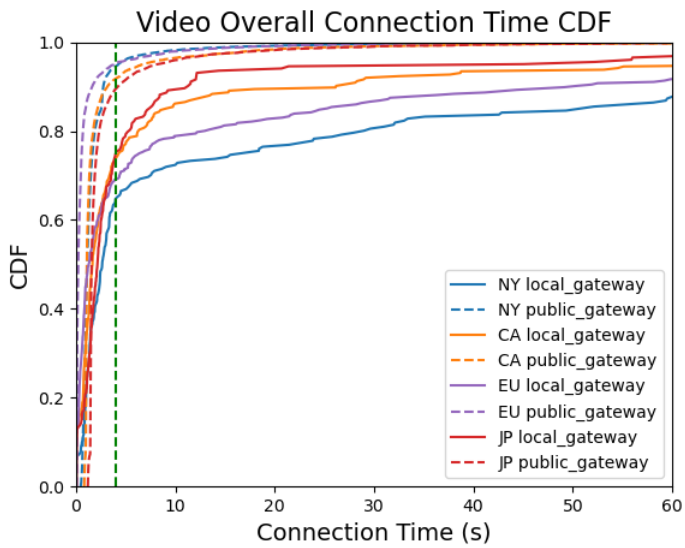
** Note: The discontinuity in the graph indicate that the local gateway was not able to reach the content*

4.3 Video Connection Time

In this section, we examined the connection time respect with to the public gateway and local gateway. In addition, we examined the connection time with videos that are reachable through the IPFS client.

4.3.1 Overall Connection Time

To understand if DTube's public gateway makes any sense for serving content to clients. We analyze the connection time between the public gateway and local gateway

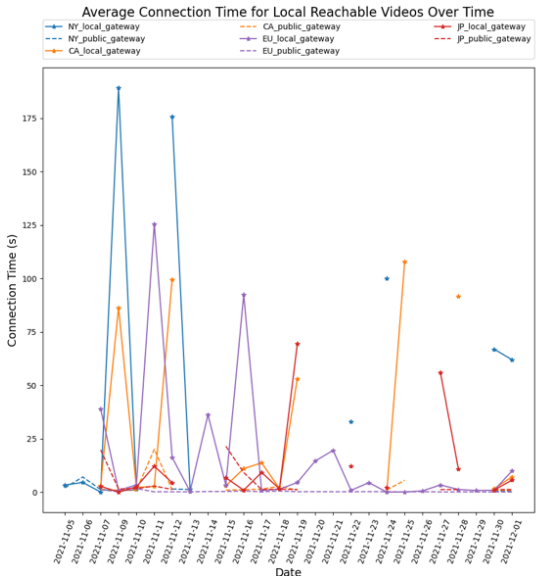


The figure above showed the overall connection time between the local gateway and public gateway. We observed that the overall public gateway has less connection time and this is surprising as every vantage point has much less connection time compared to the local gateway. However, this makes sense as the local gateway has to go through the IPFS finding process but the public gateway may have cache already and can serve back out of cache.

Connection Time with Reachable Video

4.3.2 Connection Time with Reachable Video

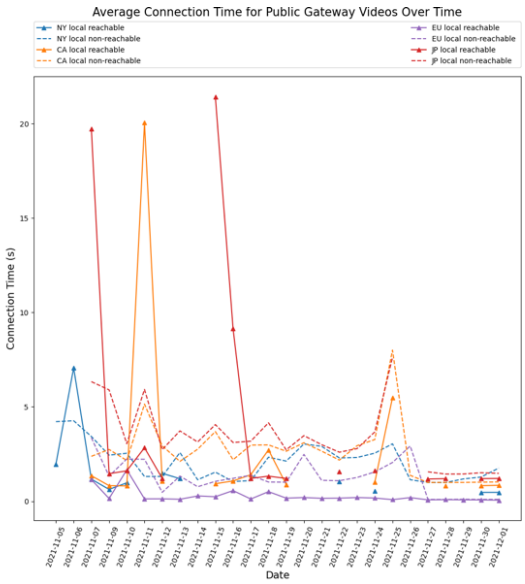
To further examine the connection time regarding the video that is reachable through the IPFS network, we compared the average connection time for local gateway and public gateway from beginning to end.



The figure above showed the local gateway reachable video's connection time between the local gateway and public gateway. We observed that the local gateway's connection time is much higher than the public gateway, which is aligned with the overall connection time CDF. We believe that this is due to the additional overhead due to the IPFS finding process.

** Note: The discontinuity in the graph indicate that the local gateway was not able to reach the content*

In addition, we examined the public gateway connection with respect to local reachable video and non-local reachable video.



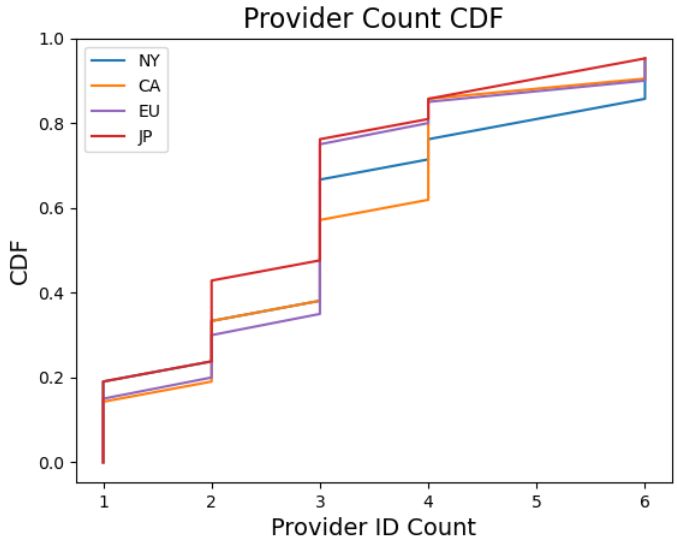
The figure above shows the connection time between the public gateway (local gateway reachable video) and the public gateway(non-reachable). We observed that in general, the public gateway (non-reachable) is higher than the public gateway (reachable) this is confusing as we expected it to be the opposite. We observed that upon initial request it has a huge connection time and connection time drops the next day. We suspect that the video is being served out of cache right after, while the rest of the non-reachable video needs to be looked up from DTube's own IPFS cluster.

4.4 Video Peers

In this section, we analyzed the video peers with each IPFS client reachable video. Then we analyzed the RTT with respect to each vantage point and Geolocation of these peers.

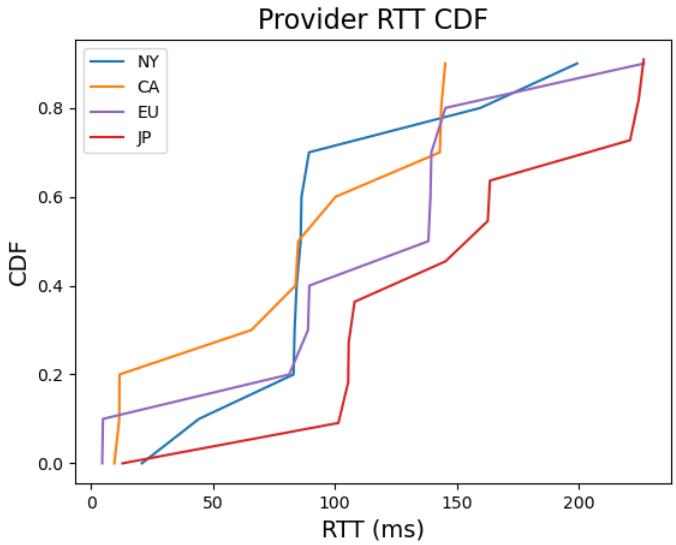
4.4.1 Multi-Peer

We analyzed how many peers are providing the video throughout the IPFS network.



The figure above showed the provider count CDF among all the videos that are reachable through the local gateway. We observed that 50 % of the video has at least 3 providers.

4.4.2 Provider's RTT and Geo Location



The figure above showed the provider's RTT CDF. We observed that 50% of the provider has an RTT of close to 100ms. In addition, NY and CA have lower RTT. We suspect most of the providers may be in the US.

In addition, we looked at the Geolocation for each peer as shown in the table below

Location	Total Vid	Reachable Vid	Distribution: (Country : Provider Count)
NY	422	21	United States: 11 Japan: 1 Netherlands 3
CA	379	21	United States: 11 Japan: 1 Netherlands 4
EU	367	20	United States: 11 Japan: 1 Netherlands 4
JP	376	21	United States: 11 Japan: 1 Netherlands 4

This data proved our speculation that most providers are from the US and CA and NY has lower RTT.

5 Conclusion

DTube aims to build a true decentralized video streaming service by utilizing distributed storage services such as IPFS. However, throughout our observation over the month, we notice that the videos are not really from distributed service providers such as IPFS or Skynet. Instead, the videos are mostly YouTube's redirects. This raises the question: is DTube moving in the direction as promised? In addition, we observed that most of the IPFS videos are not publicly reachable through regular IPFS clients, rather only accessible through DTube's public getaway. We suspect that DTube uses its own private IPFS network and IPFS cluster to manage a large number of videos being uploaded. We further examined the benefits and drawbacks of this scheme. We found that the video playing bandwidth is lower when the video is served from the local IPFS gateway, but DTube's public gateway has a lower connection time overall. Lastly, we examined the peers whose videos are reachable through the public IPFS network. We found that most of the peers are in the US and at least one peer is in Asia and Europe.

References

[1] https://ipfs.io/

[2] https://token.d.tube/