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

Link of our github or code repository:

Abstract

Inter Planetary 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 lasted for about a month to obtain video network data from Dtube, Youtube and other websites, and then processed and analyzed the data in various types. We found that DTube has more Youtube redirected videos than other types. DTube does not advertise different content based on geo-location. We put other results of exploratory research in 4 Evaluation and 5 Conclusion.

1 Introduction

IPFS is a distributed file system used to store and access files, websites, applications, and data. The one-sentence description of the official website is: “a peer-to-peer hypermedia protocol”. 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.

IPFS vs HTTP

HTTP was a great invention of the Internet 20 years ago. At present, almost all Internet information transmission uses the HTTP protocol.

The progress of Internet technology has never stopped, and even has been accelerating. As the Internet becomes larger and larger, the existing HTTP protocol has exposed more and more drawbacks. For example, because under the HTTP protocol, data is stored centrally, once the central server is attacked, the data may be stolen, leaked, or even deleted; news of various servers being hacked is not uncommon, and service providers have to use various types of With expensive security solutions, the cost of preventing

attacks is getting higher and higher; in addition, once the centralized server is accessed centrally, it will cause congestion and cause the access speed to be very slow.

The birth of IPFS perfectly solves the problems of HTTP. IPFS stores data in a distributed manner, which makes hackers lose their attack targets and makes data more secure. IPFS adopts distributed multi-point transmission, which greatly improves the transmission speed of IPFS and saves about 60% of network bandwidth.

In addition to making data more secure and faster, IPFS also has a very important feature. IPFS uses hash deduplication to store data, which will greatly reduce the cost of data storage. This will help IPFS truly defeat the underlying commercial power of HTTP. Safer and faster may not be able to replace the old technology, but if it can reduce the cost at the same time, it will subvert the traditional model more quickly. This is the same as the replacement of traditional mail by e-mail. This is an unstoppable technological development process.

2 Problem Statement

With surging fact that toady's Internet is being dominated by few giant companies. Which comes to no surprises that the web is more centralized than ever because all the data is being host and controlled by these companies. While centralized structure can provide easy management and maintenance, but it comes with other issues. With all the data being stored in one place it can easily being the target to be attack thus losing all the data. In addition, it can be used as a political weapon resulting censorship. Interplanetary File System (IPFS), a distributed storage system which provides a general solution for decentralization. In this project, we are focusing life cycle of videos on DTube, a video hosting platform that utilizes IPFS. One of the issues with current popular video hosting site such as, YouTube, is that all the videos are hosted on site. Which means that the platform controls what to be shown or not. This can be used by censorship to block particular content in different region. However, with DTube and IPFS, each video is hosted by an IPFS node and it can duplicate across the IPFS network. As long as one node has the content, it can be shared throughout the network; therefore, it is nearly impossible to be blocked or deleted. Our goal in this project is to understanding how videos are being served in DTube, and what is the life cycle of each video.

3 Experiment Setup

We first need to grab the relevant data of the video from the relevant websites, and then process and analyze the data.

To record information about videos in DTube, we used script to retrieve DTube's daily popular video's CID. Then for each video we retrieved them through IPFS network via the CID. We recorded the data each day with same videos' CID from November 5 to December 1, 2021.

The features of CID we recorded are as follows：

cid: The hash value of vedio in DTube.

dur: Describe it here

ts:

upload_date:

category:

local_check_ts:

public_check_ts:

local_data:

public_data_overhead:

public_data_download_time:

public_data_file_size:

public_data_vedio_length:

public_data_stall_time:

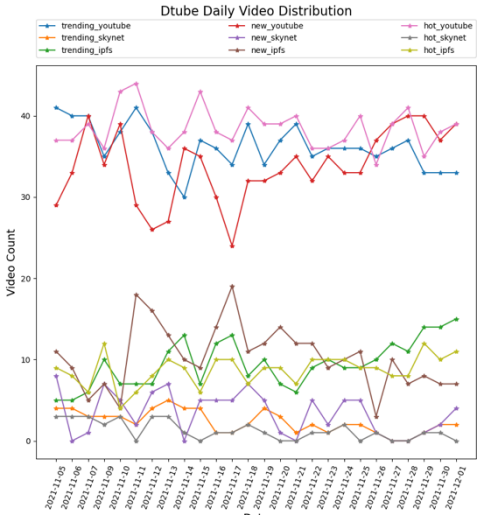
public_data_bandwidth:

We processed, analyzed, and visualized all the data from November 5 to December 1, 2021, and presented all the specific results on the next sheet.

4 Evaluation

4.1 Video Distribution Summary

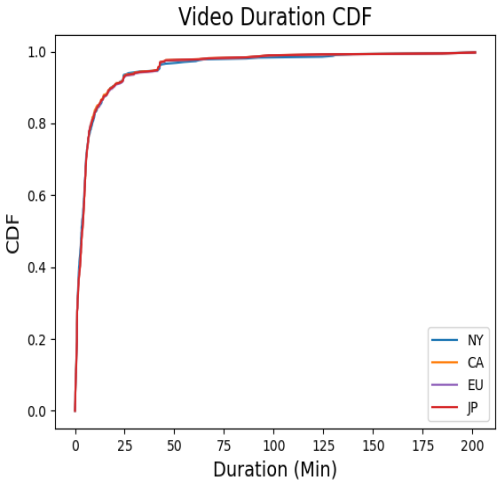
We first counted the daily changes in the count of trending, hot and new videos on Youtube, skynet and DTube, and visualized the results, as shown in the figure below.



This figure summarized the distribution of videos, which showed the daily hot, new, trending videos distribution.
As we observed that DTube has a lot of YouTube redirection about twice as more than other type.

4.2 Video Duration CDF

In the second step, we studied the video duration in different regions and plotted the result as a CDF image, as shown in the figure below

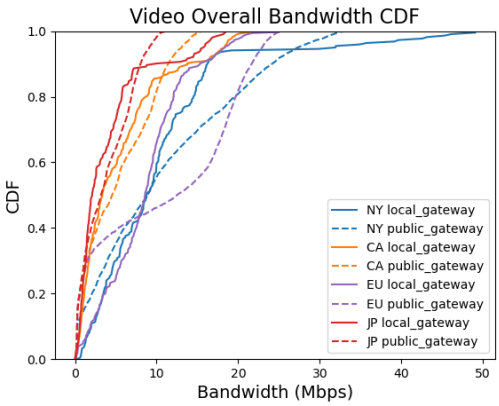


This figure showed the video duration CDF distribution, we can observe that about 90% of the video is below 25 min.

All the location aligned almost perfectly indicate that DTube does not advertise different content based on geo-location.

4.3 Video Bandwidth CDF

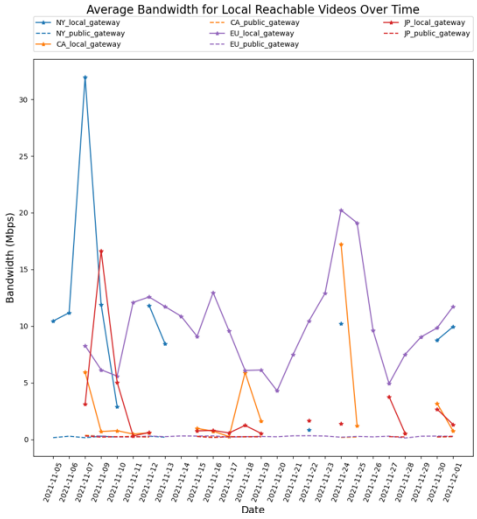
Next, in order to study whether there is a difference in the impact of local gateways and public gateways on the overall bandwidth in different regions, we drew the CDF image with the bandwidth as the abscissa, as follows



This figure showed the overall bandwidth CDF distribution among all video we recorded from local gateway and public gateway.
We observed that NY had the highest bandwidth and JP had the smallest bandwidth.
The EU had a strange curve on public gateway we suspect maybe due to updates on the script and causing inaccurate data.

4.4 Bandwidth Local VS Public

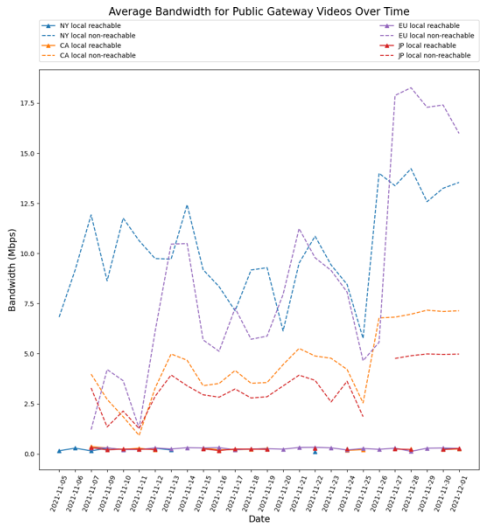
We are also very interested in the difference between the bandwidth of the local gateway and the public gateway, so we access the same video through the local gateway and the public gateway, record their bandwidth, and draw the following figure



This is Bandwidth Local VS Public figure showed an more deep analysis on how local reachable video's bandwidth compare to its public gateway bandwidth.
We observed that in general local gateway has much higher bandwidth compare to public gateway when access the same video.
* Note: The discontinuity in the graph indicate that the local gateway was not able to reach the content

4.5 Bandwidth Public VS Public

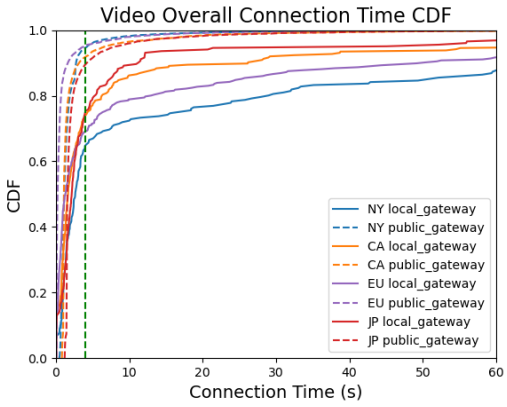
Next, we will study the difference in average bandwidth when local gateway is reachable and local gateway non-reachable, as shown in the figure below



This figure showed a comparison between average public bandwidth (local gateway is reachable) and public bandwidth (local gateway non-reachable).
We observed that the bandwidth is higher when the video is not reachable.
We suspect that when DTube gateway does not have the video it goes through the same IPFS routing thus causing additional hops
* Note: The discontinuity in the graph indicate that the local gateway was not able to reach the content

4.6 Video Connection Time CDF

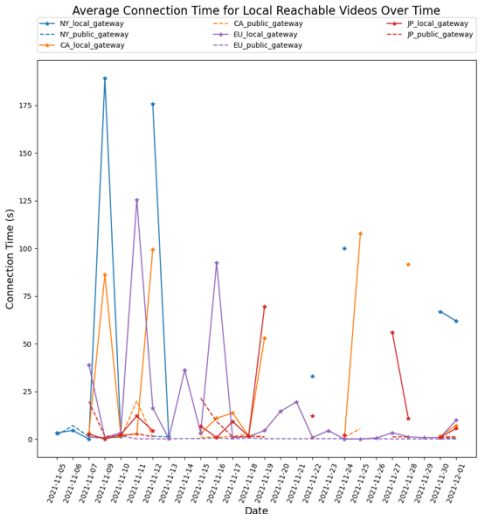
Next, we study the difference in connection time between the public gateway and the local gateway,



This figure showed overall connection time between local gateway and public gateway.
We observed that the overall public gateway has less connection time and this is expected as it is being served from single host.

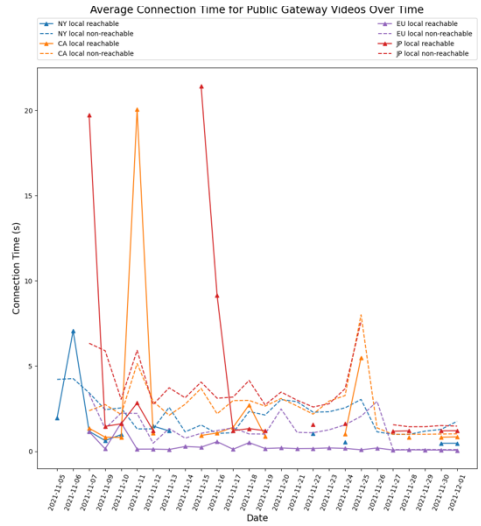
4.7 Connection Time Local VS Public

Next, we will study the difference between public gateway connection time and local gateway connection time by studying how different the local gateway reachable video's connection time between local gateway and public gateway.



This figure showed how local gateway reachable video's connection time between local gateway and public gateway.
We observed that local gateway's connection time is much higher than public gateway.
* Note: The discontinuity in the graph indicate that the local gateway was not able to reach the content

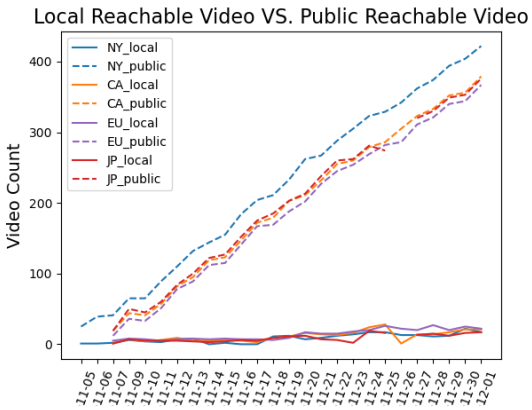
4.8 Connection Time Public VS Public



This figure show connection time between public gateway (local gateway reachable video) and public gateway(non-reachable).
We observed that in general the public gateway (non-reachable) is higher than public gateway (reachable)

4.9 Video Ratio

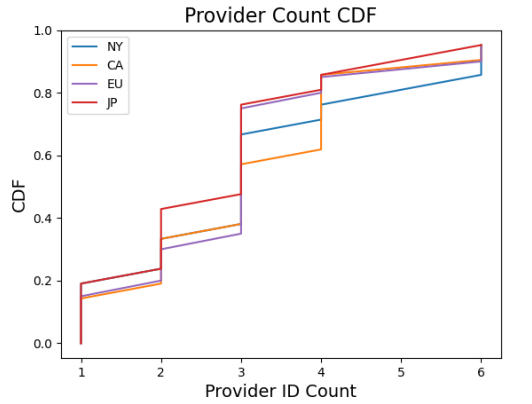
In this section, we study the number of videos between public gateway reachable video and local gateway reachable video.



This figure showed the video count between public gateway reachable video and local gateway reachable video.
It clearly showed that most of the video is not reachable through regular IPFS routing, we suspect that DTube had it private IPFS nodes and cluster, and it is only accessible through its own gateway.

4.10 Provider Count CDF

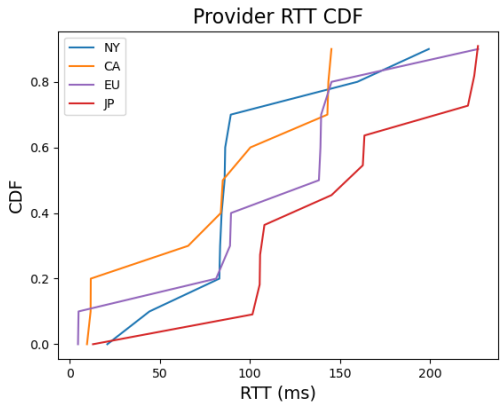
Next, we will count the number of video providers from different regions.



This figure showed the provider count CDF amount all the video that is local gateway reachable.
We observed that 50 % if the video only has 3 providers

4.11 Provider RTT CDF

This section, we will show how RTT different from different regions.



This figure showed the provider RTT CDF.

We observed that 50% of the provider has RTT of close to 100 ms.
In addition, NY and CA seems to have lower RTT we suspect most of the provider may be in the US.

In addition, we have drawn other statistical data into the following table:

Location	Total Vid	Reachable	Distribution
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 further proved our speculation of most provider is from the US and CA and NY has lower RTT

5 Conclusion

In this project, we have successfully explored some characteristics of Dtube based on IPFS.
For example:

DTube has a lot of YouTube redirection about twice as more than other type.

DTube does not advertise different content based on geo-location.

NY had the highest bandwidth and JP had the smallest bandwidth.

In general, local gateway has much higher bandwidth compares to public gateway when access the same video.

Bandwidth is higher when the video is not reachable.

When DTube gateway does not have the video, it may go through the same IPFS routing thus causing additional hops.

DTube maybe had it private IPFS nodes and cluster, and it is only accessible through its own gateway.

3 is an average count of IPFS providers for video from all regions.

Most of the provider of IPFS may be in the US.

References