Peer-Assisted Content Distribution Aided by Video Popularity Evolution Model

Mohamad Dikshie Fauzie* Achmad Husni Thamrin* Jun Murai† *Graduate School of Media and Governance †Faculty of Environment and Information Studies
Keio University, 252-0882 Kanagawa, Japan
dikshie@sfc.wide.ad.jp husni@ai3.net jun@wide.ad.jp

***Abstract*—In this paper, we present peer-assisted CDN model, the peer-to-peer networks that can be used to help CDN to deliver streaming media. Our peer-assisted CDN model is aided by Youtube VoD views popularity model called CPPro. We have comparatively evaluated CPPro through trace-driven simulations with synthetic workloads. We use three scenarios in our simulation which are (a) the video popularity in the CPPro system follows the global popularity of the video. (b) the video popularity in the CPPro system is lagged behind the global popularity by several weeks. (c) the video popularity in the CPPro is unrelated to the global popularity. Our results show that CPPro gives lower number of replicas while maintaining same number of peers contribution compare to counterpart work. We also do the significance to the number of replicas using the Kolmogorov-Smirnov statistic on two samples and we find our results are significant.**

***Index Terms*—Internet Video on Demand (VoD), P2P, CDN, and Caching.**

I. Introduction

Streaming contents, especially video, represents a significant fraction of the traffic volume on the Internet, and it has become a standard practice to deliver this type of content using Content Delivery Networks (CDNs) such as Akamai and Limelight for better scaling and quality of experience for the end users. For example, Youtube uses Google cache and MTV uses Akamai in their operations.

With the spread of broadband Internet access at a reasonable flat monthly rate, users are connected to the Internet 24 hours a day and they can download and share multimedia content. P2P (peer to peer) applications are also widely deployed. In China, P2P is very popular; we see many P2P applications from China such as PPLive, PPStream, UUSe, Xunlei, etc. [1]. Some news broadcasters also rely on P2P technology to deliver popular live events. For example, CNN uses the Octoshape [2] solution that enables their broadcast to scale and offer good video quality as the number of users increases.

From the Internet provider point of view, the presence of so many always-on users suggests that it is possible to delegate a portion of computing, storage and networking tasks to the users, thus creating P2P networks where users can share files and multimedia content. Starting from file sharing protocols, P2P architectures have evolved toward video on demand and support for live events.

Alternatively, video contents can be efficiently distributed on services offered by managed network architectures and CDN companies. The major issues of CDN are high deployment cost and good but not unlimited scalability in the long term. Given the complementary features of P2P and CDN, in recent years some hybrid solutions have been proposed and applied to operational CDN [3]–[5] to take the best of both approaches. In Peer assisted CDN, users can download content from CDN nodes from or other users or peers. A user may cache the content after download to serve requests from other users. Due to the complexity of the behavior of peers, the processing should be done in the home gateway user where the ISP can control it.

In this work, we will revisit Guo et al.'s PROP, [6] as a basis to evaluate peer-assisted CDN and propose an improvement to the model for the PROP called CPPro (this system is called CPPro abbreviated from our technical term "CDN-P2P Project"). In PROP's utility function, the difference between very popular videos and unpopular video is very difficult to differentiate. For an unpopular video, $f(p)$ will be very close to $f(p_{min})$ thus $f(p)-f(p_{min})$ will be very close to 0 then the utility function becomes very small. For a very popular video, $f(p)$ will be very close to $f(p_{max})$, thus $f(p_{max})-f(p)$ will be very close to 0 and the utility function becomes very small. We will take Youtube as an example of an Internet VoD service model. In the Youtube service model, we can get data such as (1) the time when a video is uploaded and (2) the number of accesses or number of views. We can get such data using Youtube's API. Borghol et al., [7] used the above information to estimate when a video will become very popular. They divided a video's popularity into three phases: before-peak phase, at-peak phase, and after-peak phase. We will use an estimate of a video's popularity phases for helping PROP. We will explain video popularity in Sec.III. Our contribution is as follows: (1) We use the idea of VoD view popularity model to aid the PROP. To the best of our knowledge, the combination of the PROP and the VoD view popularity model is new. (2) From simulation-based experiments, we find that peer contributions in CPPro are almost as good as PROP while the numbers of replicas are lower, achieving the same performance with fewer resources.

Our paper presentation as follows: (1) we describe related work in sect.II; (2) we explain detail of Youtube popularity evolution model in sect.III; (3) we explain the caching strategy for CDN and peer in sect.IV; (4) we explain our simulation design, simulator, and its evaluation in sect.V. Finally, we present our conclusions in section VI.

II. Related Work

Content Distribution Networks with peer assist have been successfully deployed on the Internet, such as Akamai [8], [3] and LiveSky [9]. The authors of [8] examine the risks and benefits of peer-assisted content distribution in Akamai and measure the effectiveness of its peer-assisted approach. The authors of [3] conclude from two real world traces that hybrid CDN-P2P can significantly reduce the cost of content distribution and can scale to cope with the exponential growth of Internet video content. Yin et al. [9] described commercial operation of a peer-assisted CDN in China. LiveSky solved several challenges in the system design, such as dynamic resource scaling of P2P, low startup latency, ease of P2P integration with the existing CDN infrastructure, and network friendliness and upload fairness in the P2P operation. Xu et al. [10] used game-theory to show the right cooperative profit distribution of P2P can help the ISP to maximize the utility. Their model can easily be implemented in the context of current Internet economic settlements. Misra et al. [11] also mentioned the importance of P2P architecture to support content delivery networks. The authors use cooperative game theory to formulate simple compensation rules for users who run P2P to support content delivery networks.

The idea of telco- or ISP-managed CDN has been proposed in recent years. The complexity of the CDN business encourages telcos and ISPs to manage their own CDN, rather than allow others to run CDNs on their networks. It has been shown that it is cost effective [12] [13]. Kamiyama et al. [14] proposed optimally ISP operated CDN. Kamiyama et al. mentioned that, in order to deliver large and rich Internet content to users, ISPs need to put their CDNs in data centers. The locations are limited while the storage is large, making this solution effective; using optimum placement algorithm based on real ISP network topologies. The authors found that inserting a CDN into an ISP's ladder-type network is effective in reducing the hop count, thus reduce total link cost. Based on the author definition: Ladder-type network is a network with a maximum degree under 10. Cisco has initiated an effort to connect telco- or ISP-managed CDNs to each other, to form a CDN federation [12] using open standards [15]. They argue that the current CDN architecture is not close enough to the users and ISPs can fill this position.

The idea of utilizing the user's computation power to support ISP operation is not new. The Figaro project [16] proposed the residential gateway as an integrator of different networks and services, becoming an Internet-wide distributed content management for a proposed future Internet architecture [16]. Cha et al., [17] performed trace analysis and found that an IPTV architecture powered by P2P can handle a much larger number of channels, with lower demand for infrastructure compared to IP multicast. Jiang et al. [18] proposed scalable and adaptive content replication and request routing for CDN servers located in users' home gateways. Maki et al., [19] propose traffic engineering for peer-assisted CDN to control the behavior of clients, and present a solution for optimizing the selection of content files. Mathieu et al., [20] are using data gathered from France telecom network to calculate reduction of network load if customers are employed as peer-assisted content delivery.

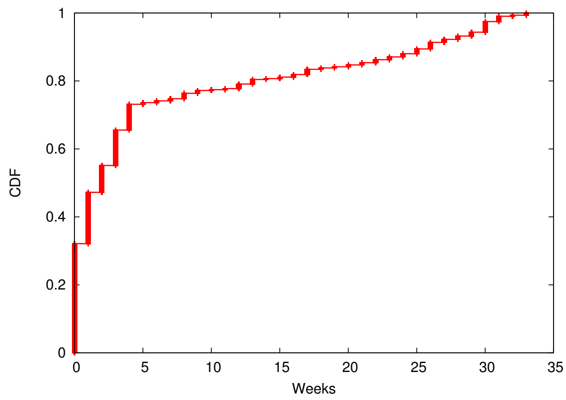


Fig. 1: Time to peak empirical distribution data from [7].

Guo et al., [6] work's PROP is closest with our work. PROP uses local system (local counter) to calculate the segment popularity in peer-assisted proxy system. PROP uses popularity for proxy cache replacement strategy. In peer side, the author use utility function for cache replacement strategy. A utility function assigns numerical value to outcomes, in such a way that outcomes with higher utility are always preferred of outcomes with lower utilities. In PROP's utility function, the difference between very popular videos and unpopular video is very difficult to differentiate. The utility function is also function from popularity. While the authors successfully show that the results are very good, the peer-assisted system behavior over time is not explain because the author focus on properties such as proxy cache size variations and peer cache size variations. The explanation of the optimal number of replicas is not also clear because unavailable information when the snapshot is taken. In our work, we complement Guo et al., [6] work with VoD viewing popularity evolution model and describe the behavior of the peer-assisted CDN over the time.

III. Determining Internet VoD Popularity Phase

The objective of determining the Internet VoD popularity phase is to determine whether a video is at before-peak, at-peak, or after-peak phase, to be used by peers in their caching strategy. For this purpose we use the Youtube content popularity dataset from Borghol et. al., [7] which contains data about 29791 videos, including the view count and upload time, during 36 weeks of measurements. Figure 1 is the cumulative distribution function (CDF) of the time-to-peak distribution from Borghol et. al., [7] which shows that around three-quarters of the videos peak within the first six weeks after upload. The time-to-peak is exponentially distributed up to the sixth week, and it is uniform beyond the sixth week. Borghol et al., [7] define time-to-peak for a video as its age (time since upload) at which its weekly viewing rate is the highest during measurement (from the first week until end of measurement). Because we know the peak time (at-peak phase) of every video, we can also find the before-peak phase

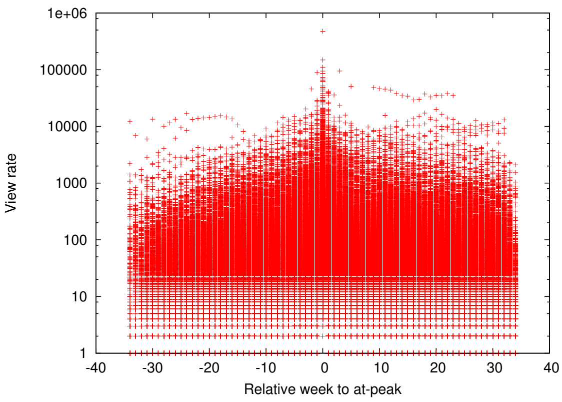


Fig. 2: View rate distribution versus week relative to at-peak phase week for every video, where y-axis in log scale. Every point lies in negative x-axis mean view rate of every video in before-peak phase. Every point lies in x-axis$=0$ mean view rate of every video at-peak phase. Every point lies in positive x-axis mean view rate of every video in after-peak phase. As we see in this graph, while fig.1 mentioned that 75% of videos reach at-peak within six weeks, we also see that some vides reach at-peak after six weeks. Data from [7].

and after-phase of every video. Borghol et al., [7] mentioned that view count evolution follows beta distribution. For detail we refer the readers to [7].

Our algorithm for determining a requested video is quite simple. As we know the age of a requested video in week number and its view rate, the algorithms looks for Borghol dataset with the same week, one week before, and one week after for the same view rate. We are averaging the relative week to at-peak values in the same week, one week before, and one week after.

Suppose a video v in the Borghol dataset has a viewing rate $r_v(t)$, $0 <= t < t_f$, and $r_v(t)$ peaks at t_{vp}. The data is transformed by including the relative time-to-peak, such that each data point is a 2-tuple: video rate and relative time to peak:

$$r_p^v(t) = \langle r_v(t), t_p(t) \rangle \quad (1)$$

$$t_p(t) = t - t_{vp}. \quad (2)$$

Figure 2 shows the Borghol's dataset with time axis for each video is translated by t_{vp} to the left, such that each video peaks at time 0.

In determining the phase of a requested video e with known age t_e and view rate r_e at t_e, we find the three r_v data points whose rates are closest to r_e at t_e, $(t_e - 1)$, and $(t_e + 1)$ for all videos from Borghol dataset, and then average the t_p of the three data points. The phase of the requested video e is estimated to be before-peak, at-peak, or after-peak based on whether the average is negative, 0, or positive. The view rate r_e of a video is calculated by subtracting the view counts at the time of the current and the previous video requests. The pseudo code for averaging the t_p is shown in algorithm 1.

But when a video is being requested for the first time, the phase can only be estimated using the age of the video. In

Algorithm 1 Averaging relative weeks from the nearest neighbor points

Require: dataset that consist of weeknumber, viewrate, and relative week to at peak.

$t, r_v, t_p \leftarrow$ read Borghol's dataset {where t is week number (1st column), r_v is view rate (2nd column), and t_p is relative week to at-peak values (3rd column).}
$len \leftarrow 29791 * 36$ {we have 29791 videos * 36 weeks thus we have 1072476 rows and 3 columns of Borghol's dataset.}
$t_e \leftarrow time_{cur} - time_{upload}$ {calculate week number of a requested video which is current time a requested video substract by upload time of a requested video.}
$r_e \leftarrow viewcount_{cur} - viewcount_{lastweek}$ {calculate view rate of a requested video which is current view count of a requested video substract by last week view count of a requested video.}
$t_e^{before} \leftarrow (t_e - 1)$ {at one week before.}
for $i = 0$ to len **do**
  if $t == t_e$ and $r_v == r_e$ **then**
    $t_p^{before} \leftarrow t_p$ {put t_p value into t_p^{before} list.}
  end if
end for
$t_e^{at} \leftarrow (t_e)$ {at same week.}
for $i = 0$ to len **do**
  if $t == t_e$ and $r_v == r_e$ **then**
    $t_p^{at} \leftarrow t_p$ {put t_p value into t_p^{at} list.}
  end if
end for
$t_e^{after} \leftarrow (t_e + 1)$ {at one week after.}
for $i = 0$ to len **do**
  if $t == t_e$ and $r_v == r_e$ **then**
    $t_p^{after} \leftarrow t_p$ {put t_p value into t_p^{after} list.}
  end if
end for
$t_p^{final} \leftarrow average(t_p^{before}, t_p^{at}, t_p^{after})$
if $t_p^{final} < 0$ **then** {if relative week to at-peak value is less than 0, estimate a requested video as before-peak.}
  $phase \leftarrow before$
else if $t_p^{final} == 0$ **then** {if relative week to at-peak value is equal to 0, estimate a requested video as at-peak.}
  $phase \leftarrow at$ {if relative week to at-peak value is more than 0, estimate a requested video as after-peak.}
else
  $phase \leftarrow after$
end if

Algorithm 2 Determine phase for the first access a requested video

Require: t_e and time-to-peak distribution
$len \leftarrow 35$
for $i = 0$ to len **do**
  draw integer random number between 0 and 35 respect to time-to-peak distribution: $d \leftarrow draw\_integer\_random\_number()$
end for
$total \leftarrow 0$
for $i = 0$ to t_e **do**
  $total \leftarrow total + count(d, i)$ {counting how many each integer random number and sum those}
end for
$estphase \leftarrow total/36$
if $estphase > 0.75$ **then**
  phase ← after-peak
else if $estphase \leq 0.75$ and $estphase > 0.5$ **then**
  phase ← at-peak
else
  phase ← before-peak
end if

this case, we draw 36 weighted random integer numbers s_i, $0 <= i <= 35$ where the weighted factor is from time-to-peak phase CDF as shown in fig. 1 then calculate the count of each integers between 0 and t_e from the drawn numbers then divide the result by 36, i.e., $estphase = \sum_0^{t_e} \frac{count(i,s)}{36}$. The number 36 come from the duration of measurement and each week has its own probability as we shown in fig. 1. This result represents the estimated phase. From time to peak distribution, 50% of video reach peak within four weeks. At that level, we expect that half of videos may reach at-peak and half of videos are not yet reach at-peak. Therefore we put 0.5 as low threshold. Still from the same time to peak distribution 75% of video reach peak within six weeks, and beyond six week the distribution is considered, it means there are not much additional view count. In other words, beyond six weeks we consider videos reach after-peak phase. Therefore we put 0.75 as high threshold. The pseudo code for this purpose is shown in algorithm 2.

IV. System Description

A. System Overview

The main components of the system are: (1) a CDN and (2) a set of peers which are self organized into a P2P overlay network. Each peer in the system has two functions. First, a peer is a client that requests a video. Second, a peer is a contributor that shares the cached video with other peers in the system. Peers control the number and utilization of their connection based on current resources availability. In fig.3, we describe the process of a peer that requests a video which derived from PROP. When a video is requested for the first time, the CDN is responsible for delivering the requested video. When a CDN receives a query for the same video, a CDN will find suitable peers that currently have a copy of the requested video. The CDN then returns information about these peers to the querying peer.

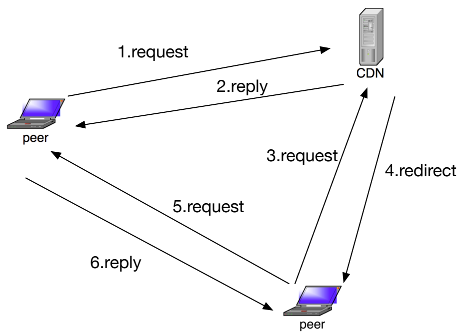


Fig. 3: Peer assisted CDN works as follows: when a peer requests a video, it always goes to a CDN server (step 1). The CDN provides the videos to the peer (step 2). If there is another peer request same video, that request will go to CDN (step 3). A CDN will check its record to see if there is some peers cache that requested video. If there is some peers cache that requested video, a CDN will reply with redirect message that asking a peer to download requested video from other peer (step 4). If there s no peers have requested video, a CDN will serve the video. A peer then can request the video to other peer and get the video (step 5 and step 6).

B. Peer caching strategy

A peer stores the requested videos into its cache based on the value of a utility function, i.e., in a peer with a full cache a newly requested video will replace one or more videos in the cache whose utility values are smaller than that of the new video. The utility function reflects the popularity of a video in the system that considering number of copy of its video or replica. PROP's u value itself lies in interval $[0, 2]$ Guo et al., [6]. Our utility function u is:

$$u = \frac{(f(p) - f(p_{min}))(f(p_{max}) - f(p))}{r^{\alpha+\beta}} + z(t) \qquad (3)$$

where the first term is the utility function of PROP, and the second term, $z(t)$, is the phase factor that we introduced in CPPro, where $z(t)$ is large if the video phase is at-peak. p, p_{min}, and p_{max} are the popularity of the video, the minimum popularity, and the maximum popularity in the system. We follow [6] for $f(p) = log(p)$. The popularity of a video is the determined by the view rate of the video. r is the number of replicas. α and β are adjustment factor [6].

Following [6], we can calculate p as follows:

$$p = min\left(\frac{n_i^r}{t_i^r - t_a^i}, \frac{1}{t_{cur} - t_i^r}\right) \qquad (4)$$

Where n_i^r is number of requested video, t_i^r is last time the video is requested, t_a^i is the time when the video is accessed for the first time, and t_{cur} is the current time. To able to track the simulation, we use default value from PROP thus we refer the readers to [6] for the details.

The $z(t)$ values are set as follows:

$$z(t) = \begin{cases} 0.15 & \text{if phase estimation is before-peak} \\ 0.47 & \text{if phase estimation is at-peak} \\ 0.38 & \text{if phase estimation is after-peak} \end{cases} \quad (5)$$

where the numbers are based on the proportion of video counts in the Youtube dataset before the peaks, around the peaks, and after the peaks. $z(t)$ for at-peak is determined by the total proportion of video counts from 1-week-before until 1-week-after the peak, which accounts to almost half of the total. The value of $z(t)$ is assigned after we finish to determine a video popularity phase. For example: if we determine a video popularity phase is at-peak, then we assign $z(t) = 0.47$.

V. Evaluation

We performed simulations to compare the performance of CPPro against PROP in two main metrics: (1) peer contributions in video delivery, and (2) number of video replicas in peers. The peer contribution is the number of videos served by a peer instead of by the CDN server. The number of video replicas shows how much storage is used in peers to contribute in video delivery. The peer contribution metric is related to the byte-hit-ratio. The byte-hit-ratio is defined as the total bytes of content served by peers normalized by the total bytes of video all peers and the CDN consume. With more peer contributions, we will have higher byte-hit-ratio because peer can get content from other peers. However, because we only interested in peer performance, we compare peer contribution between PROP and CPPro. Contribution ratio of peer to total contribution (comparing to CDN contribution) becomes irrelevant in this case.

A. Simulation Design

This peer-assisted CDN is simulated using an event driven simulation implemented in Python. Peers request videos from a video catalog where the peer request as well as the videos in the catalog are generated using certain distributions.

1) Video Catalog: Each video in the catalog has the following properties: video-id,size,upload time, final view count, view count function parameters. View count parameters are the distribution parameters. The final view count is the total number of views of a video at the end of simulation and it is generated using uniform distribution sampling from Borghol dataset ranging from 27 until 1069385. The upload time interval is a Poisson process with $\lambda = 1$. The video size is generated using a uniform distribution. Because of the very weak relationship between video size and popularity [21] and because our work focuses on the impact of the popularity aspect on the utility function rather than storage optimization we believe that the choice to assign a random uniform video size from the Youtube dataset does not have an effect to our results. The view rate progression from the upload time until the end of simulation time is modeled using a Beta distribution [7]. As Borghol et al., [7] showed that view rate of a video can be modeled using beta distribution we can calculate α and β parameters. We can use Beta distribution mode formula $m = \frac{\alpha-1}{\alpha+\beta-2}$ to calculate α or β. In this case, we choose α random uniform between 1 and 3 and m is view rate at-peak, thus we can get $\beta = \frac{\alpha-1}{m} - \alpha - 2$.

TABLE I: Parameters variations

	A	B	C	D
number of requests	5001101	5001101	15000617	15000617
number of videos	5000	15000	5000	15000

TABLE II: The difference of request satisfied by peer between CPPro and PROP.

Scenario	Parameter	Difference
A	Parameter A	1.0%
A	Parameter B	2.0%
A	Parameter C	1.9%
A	Parameter D	1.2%
B	Parameter A	1.2%
B	Parameter B	2.3%
B	Parameter C	0.6%
B	Parameter D	1.4%
C	Parameter A	1.7%
C	Parameter B	3.5%
C	Parameter C	1.0%
C	Parameter D	2.1%

2) Peer Request Generator: Peers request videos from the catalog using a Poisson process with $\lambda = 1$ [22] for the inter-arrival time. For the requested videos there are three scenarios A,B,and C. In Scenario A, the video popularity in the peer-assisted CDN system follows the global popularity of the video. In Scenario B, the video popularity in the peer-assisted CDN system lags four weeks behind the global popularity of the video. We choose four weeks lags based on the probability from time-to-peak distribution that half of videos are already reach peak within four weeks. In Scenario C, the video popularity in the peer-assisted CDN system does not follow the global popularity of the video. We use a Zipf distribution with rate$= 0.9$ for this purpose [23].

3) Simulation Parameters: We have four variations of parameter (from A to D) used in simulation as shown in table I.

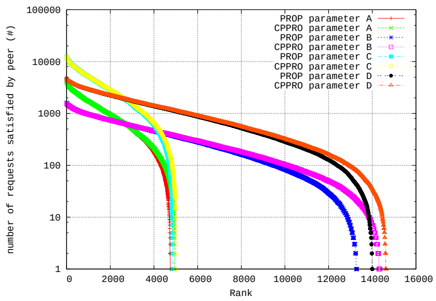
We also have several fixed value for other parameters as follows:

- Length: 360 days.
- Video size: uniform random between 1MB and 200MB.
- Peer storage capacity: 500MB.
- CDN storage capacity: 10000MB.
- Number of peers: 10000.
- Peer's caching strategy: CPPro, PROP.

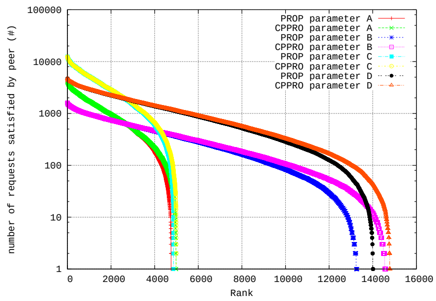
Finally, we compare our results to PROP [6] implementation.

B. Result and Discussion

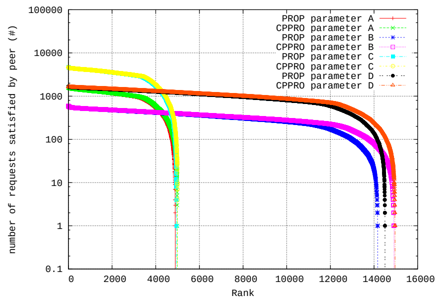
Figure 4 shows the peer contributions in each scenario ranked by the number of videos served by each peer for parameter A, B, C, and D. They exhibit a similar pattern and only differ in the tails for all scenarios, where CPPro gives higher, albeit insignificant, results. The difference between CPPro and PROP for parameter A,B,C,and D are shown in table **??**.



(a) Scenario A.

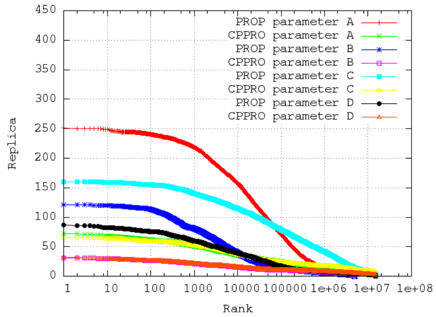


(b) Scenario B.

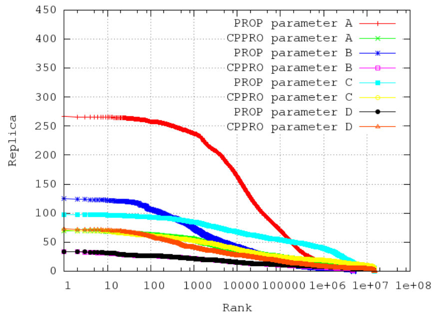


(c) Scenario C.

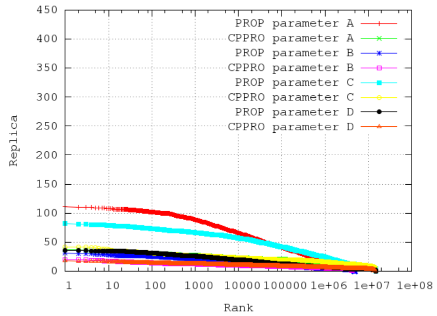
Fig. 4: Videos rank distribution delivered by peer compared between CPPro and PROP (y-axis in log-scale and y-axis unit is times).



(a) Scenario A.



(b) Scenario B.



(c) Scenario C.

Fig. 5: Comparison of available replicas between CPPro and PROP when a peer requests a video (y-axis in log-scale).

TABLE III: The difference of replicas between CPPro and PROP.

Scenario	Parameter	Difference
A	Parameter A	13%
A	Parameter B	9.4%
A	Parameter C	16.0%
A	Parameter D	0.7%
B	Parameter A	6.5%
B	Parameter B	1.4%
B	Parameter C	11.5%
B	Parameter D	1.8%
C	Parameter A	13%
C	Parameter B	4.4%
C	Parameter C	4.5%
C	Parameter D	1.6%

The difference between CPPro and PROP for parameter A in scenario A is 1.0%, 1.2% in scenario B, and 1.7% for scenario C. The difference between CPPro and PROP for parameter B in scenario A is 2.0%, 2.3% in scenario B, and 3.5% for scenario C. The difference between CPPro and PROP in scenario A is 1.9%, 0.6% in scenario B, and 1.0% for scenario C.

F

When a peer requests a video from another peer, a peer has to decide whether a peer wants to cache the video or not to cache the video based on the calculated utility function value. If a peer decided to cache a video, we called it cached event. If a peer did not decide to cache a video, we called it not-cached event. We breakdown how many cache events occur in our

TABLE IV: Percentage of Cached events and Not-Cached events in CPPro and PROP for Scenario A, B, and C.

Scenario	Parameter	Type	Cached (times)	Not-Cached (times)
A	Parameter A	CPPro	48%	52%
		PROP	68%	32%
	Parameter B	CPPro	20%	80%
		PROP	72%	28%
	Parameter C	CPPro	26%	74%
		PROP	69%	31%
	Parameter D	CPPro	27%	73%
		PROP	73%	27%
B	Parameter A	CPPro	15%	85%
		PROP	68%	32%
	Parameter B	CPPro	10%	90%
		PROP	72%	28%
	Parameter C	CPPro	11%	89%
		PROP	69%	31%
	Parameter D	CPPro	15%	85%
		PROP	73%	27%
C	Parameter A	CPPro	30%	70%
		PROP	66%	34%
	Parameter B	CPPro	26%	74%
		PROP	71%	29%
	Parameter C	CPPro	40%	60%
		PROP	67%	33%
	Parameter D	CPPro	30%	70%
		PROP	73%	27%

TABLE V: Percentage cached and not-cached events for each video popularity phase in CPPro for Scenario A, B, and C.

Scenario	Parameter	Type/Events	Before-Peak	At-Peak	After-Peak
A	Parameter A	CPPro Cached	2.2%	0.6%	45.8%
		CPPro Not-Cached	2.6%	0.7%	48.1%
A	Parameter B	CPPro Cached	0.9%	0.2%	19.4%
		CPPro Not-Cached	1.9%	0.6%	77%
A	Parameter C	CPPro Cached	0.1%	0.06%	25.21%
		CPPro Not-Cached	0.8%	0.03%	73.8%
A	Parameter D	CPPro Cached	1.6%	0.3%	24%
		CPPro Not-Cached	2.5%	0.6%	71%
B	Parameter A	CPPro Cached	0.3%	0.1%	13%
		CPPro Not-Cached	1.4%	0.6%	84.6%
B	Parameter B	CPPro Cached	0.07%	0.04%	9.7%
		CPPro Not-Cached	1.3%	0.5%	88.39%
B	Parameter C	CPPro Cached	0.02%	0.08%	4.2%
		CPPro Not-Cached	0.07%	0.02%	95.61%
B	Parameter D	CPPro Cached	0.3%	0.1%	15.3%
		CPPro Not-Cached	1.7%	0.6%	82%
C	Parameter A	CPPro Cached	6.5%	0.7%	23.4%
		CPPro Not-Cached	17.7%	2.2%	49.5%
C	Parameter B	CPPro Cached	6.9%	1.0%	17.8%
		CPPro Not-Cached	6.1%	1.1%	67.1%
C	Parameter C	CPPro Cached	0.4%	0.2%	39%
		CPPro Not-Cached	0.3%	0.2%	59.9%
C	Parameter D	CPPro Cached	9.3%	1.0%	19.0%
		CPPro Not-Cached	7.2%	1.0%	62.5%

simulation as shown in table.IV. Furthermore, we breakdown again by video popularity phase for CPROP's cached and not-cached events as shown in table. V. From table.IV, we can see that CPPro cached less videos compared to PROP for all scenarios. CPPro cached events are fewer than PROP because in CPPro the probability of cached events are fewer than not-cached events. We want to know what makes cached events in CPPro are fewer than PROP. Assume a peer requests a video with utility function u_r and the minimum utility function inside peer's cache is u_c-min. Intuitively, when a requested video in after-peak phase and videos inside peer's storage are in at-peak phase then in CPPro case high probability that a requested video will not be cached. In PROP case, since PROP does not has video popularity phase the probability of a requested video will not be cached is lower than CPPro. We can take same intuition for before-peak and at-peak phase.

In CPPro, The cached events happen more frequently in scenario A comparing to scenario B, because in scenario B when we shift requests four weeks later, estimated of requested videos will probably fall at-peak phase or after-peak phase and the videos in peer's cache are also probably estimated at-peak phase or after-peak thus probability of not-cached events are higher in scenario B than scenario A. In table.V and **??**, the at-peak phase has low proportion comparing to before-peak and after-peak phase because when we estimate popularity phase of a requested video, the estimate of video to be at-peak phase occurs when average of relative week is exactly 0. In table.V and **??** the after-peak phase has high proportion comparing to before-peak and after peak phase in scenario A and B because video popularity evolution model that we use to estimate a new requested video assume that after four weeks most of video will enter after peak phase. In scenario B, its higher than scenario A since we shifted four weeks. In scenario C, while after-peak phase has higher proportion than before-peak and at-peak phase, its proportion is lower than scenario A an B because different popularity video choice. In table.V and **??**, the before-peak phase in scenario A has higher proportion comparing to scenario B because in scenario B we shift the video request time four weeks thus the probability of a requested video to be before-peak is lower than scenario A. In scenario C, the proportion before-peak phase is higher than scenario A and B. This probably due to different popularity of video choice that a requested video is estimated to be before-peak phase but the video inside peer's cache are having lower utility function.

In fig. **??** shows number of replicas available for each scenario when a peer request a video in parameter A. Maximum replica for PROP is around three times from CPPro in scenario A and B, while in scenario C the maximum value for PROP is around two times from CPPro. Number of replicas for each scenario are different between CPPro and PROP in the head of distribution and similar in the body and the tail of distribution. The difference of replica between PROP and CPPRO in scenario A is 13%, 6.5% in scenario B, and 13% in scenario C.

In fig. **??** shows number of replicas available for each scenario when a peer request a video in parameter A. The difference of replica between PROP and CPPRO in scenario A is 9%, 1.4% in scenario B, and 4.4% in scenario C.

Figure **??** shows the peer contributions in each scenario ranked by the number of videos served by each peer for parameter C. They exhibit a similar pattern and only differ in the tails for all scenarios, where CPPro gives higher, albeit insignificant, results. The difference between CPPro and PROP in scenario A is 1.9%, 0.6% in scenario B, and 1.0% for scenario C. The advantage of CPPro to PROP is evident in fig. **??** which shows the number of replicas of the requested videos at each request event.

In fig. **??** shows number of replicas available for each scenario when a peer request a video in parameter A. The difference of replica between PROP and CPPRO in scenario A is 16%, 11.5% in scenario B, and 4% in scenario C.

Figure **??** shows the peer contributions in each scenario ranked by the number of videos served by each peer for parameter D. They exhibit a similar pattern and only differ in the tails for all scenarios, where CPPro gives higher, albeit insignificant, results. The difference between CPPro and PROP in scenario A is 1.2%, 1.4% in scenario B, and 2.1% for scenario C. The advantage of CPPro to PROP is evident in fig. **??** which shows the number of replicas of the requested videos at each request event.

In fig. **??** shows number of replicas available for each scenario when a peer request a video in parameter A. The difference of replica between PROP and CPPRO in scenario A is 0.7%, 1.8% in scenario B, and 1.6% in scenario C.

VI. Conclusion and Future Work

This paper presents a scheme for peer-to-peer network can help CDN to deliver the content over the Internet. We show that by introducing z factor to utility function CPPro can maintain same peer contribution while reducing number of replicas. We found that there are no much different all scenario in term of peer contribution to deliver a video. However, we found that in the all scenarios, CPPro gives lower replicas than PROP because in CPPro, we found that probability not-cached events occur are higher than cached events, furthermore compared to PROP, probability a peer is not cache a requested video is higher in CPPro compared to PROP. Therefore, in CPPro the numbers of available replicas are lower than PROP. We also did the significance test to the number of replicas using the Kolmogorov-Smirnov statistic on two samples and we find that for all scenarios the p-values are less than 1% thus the results are significant.

Some areas of improvement that we have identified for future are: (1) since we know the CPPro can delivery a video similar to PROP with less replica, we are interested to know the energy trade-off of this peer-assisted CDN architecture in order to know how much energy saving by ISP and how much increase of energy at users home gateway side in this architecture since we have higher peer contribution. (2) Involving the different popularity model from different geographically VoD service. This is very important to see the system behavior when receiving request from different location with different popularity while still in the same global CDN service. Furthermore, we can exploit this behavior for commercial ads in VoD service.

Acknowledgment

The authors would like to thank Internet research laboratory member at Keio University and anonymous reviewers.

References

[1] L. Vu, I. Gupta, K. Nahrstedt, and J. Liang, "Understanding overlay characteristics of a large-scale peer-to-peer iptv system," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 6, no. 4, pp. 31:1–31:24, Nov. 2010. [Online]. Available: http://doi.acm.org/10.1145/1865106.1865115

[2] Octoshape, "Octoshape," http://www.octoshape.com/cnn-com-using-octoshapes-p2p-for-live-feed/.

[3] C. Huang, A. Wang, J. Li, and K. W. Ross, "Understanding hybrid cdn-p2p: why limelight needs its own red swoosh," in *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, ser. NOSSDAV '08. New York, NY, USA: ACM, 2008, pp. 75–80. [Online]. Available: http://doi.acm.org/10.1145/1496046.1496064

[4] H. Jiang, J. Li, Z. Li, and J. Liu, "Efficient hierarchical content distribution using p2p technology," in *Networks, 2008. ICON 2008. 16th IEEE International Conference on*, dec. 2008, pp. 1 –6.

[5] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, and B. Li, "Design and deployment of a hybrid cdn-p2p system for live video streaming: experiences with livesky," in *Proceedings of the 17th ACM international conference on Multimedia*, ser. MM '09. New York, NY, USA: ACM, 2009, pp. 25–34. [Online]. Available: http://doi.acm.org/10.1145/1631272.1631279

[6] L. Guo, S. Chen, and X. Zhang, "Design and evaluation of a scalable and reliable p2p assisted proxy for on-demand streaming media delivery," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 18, no. 5, pp. 669–682, May 2006.

[7] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing and modelling popularity of user-generated videos," *Perform. Eval.*, vol. 68, no. 11, pp. 1037–1055, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.peva.2011.07.008

[8] M. Zhao, P. Aditya, A. Chen, Y. Lin, A. Haeberlen, P. Druschel, B. Maggs, B. Wishon, and M. Ponec, "Peer-assisted content distribution in akamai netsession," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New York, NY, USA: ACM, 2013, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/2504730.2504752

[9] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, and B. Li, "Livesky: Enhancing cdn with p2p," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 6, no. 3, pp. 16:1–16:19, Aug. 2010. [Online]. Available: http://doi.acm.org/10.1145/1823746.1823750

[10] K. Xu, Y. Zhong, and H. He, "Can p2p technology benefit eyeball isps? a cooperative profit distribution answer," *CoRR*, vol. abs/1212.4915, 2012.

[11] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié, "Incentivizing peer-assisted services: a fluid shapley value approach," *SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 1, pp. 215–226, Jun. 2010. [Online]. Available: http://doi.acm.org/10.1145/1811099.1811064

[12] Cisco, "Cdn federation," 10 2012, available on http://www.cisco.com/web/about/ac79/docs/sp/CDN-Federation_Phase-2-Pilot.pdf.

[13] W. B. Norton, *The Internet Peering Playbook: Connecting to the Core of the Internet*, 2nd ed. DrPeering Press, December 2012.

[14] N. KAMIYAMA, T. MORI, R. KAWAHARA, and H. HASEGAWA, "Optimally designing isp-operated cdn," *IEICE Transactions on Communications*, vol. E96.B, no. 3, pp. 790–801, March 2013.

[15] IETF, "Cdn interconnect ietf working group," 2013, available on https://datatracker.ietf.org/wg/cdni/. [Online]. Available: https://datatracker.ietf.org/wg/cdni/

[16] "Figaro project," http://www.ict-figaro.eu/, 2012.

[17] M. Cha, P. Rodriguez, S. Moon, and J. Crowcroft, "On next-generation telco-managed p2p tv architectures," in *Proceedings of the 7th international conference on Peer-to-peer systems*, ser. IPTPS'08. Berkeley, CA, USA: USENIX Association, 2008, pp. 5–5. [Online]. Available: http://dl.acm.org/citation.cfm?id=1855641.1855646

[18] W. Jiang, S. Ioannidis, L. Massoulié, and F. Picconi, "Orchestrating massively distributed cdns," in *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, ser. CoNEXT '12. New York, NY, USA: ACM, 2012, pp. 133–144. [Online]. Available: http://doi.acm.org/10.1145/2413176.2413193

[19] N. MAKI, T. NISHIO, R. SHINKUMA, T. MORI, N. KAMIYAMA, R. KAWAHARA, and T. TAKAHASHI, "Traffic engineering of peer-assisted content delivery network with content-oriented incentive mechanism," *IEICE Transactions on Information and Systems*, vol. E95.D, no. 12, pp. 2860–2869, December 2012.

[20] B. Mathieu and Y. Levene, "Impact of ftth deployment on live streaming delivery systems," in *Computers and Communications (ISCC), 2012 IEEE Symposium on*, july 2012, pp. 000 259 –000 264.

[21] A. Abhari and M. Soraya, "Workload generation for youtube," *Multimedia Tools and Applications*, vol. 46, no. 1, pp. 91–118, 2010.

[22] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of youtube network traffic at a campus network - measurements, models, and implications," *Comput. Netw.*, vol. 53, no. 4, pp. 501–514, Mar. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2008.09.022

[23] F. Guillemin, T. Houdoin, and S. Moteau, "Volatility of youtube content in orange networks and consequences," in *Communications (ICC), 2013 IEEE International Conference on*, June 2013, pp. 2381–2385.