

An Adaptive Multi-channel P2P Video-on-Demand System using Plug-and-Play Helpers

*Hao Zhang
Minghua Chen
Abhay Parekh
Kannan Ramchandran*



Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-111

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-111.html>

July 28, 2010

Copyright © 2010, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

An Adaptive Multi-channel P2P Video-on-Demand System using Plug-and-Play Helpers

Hao Zhang[†], Minghua Chen[§], Abhay Parekh[†] and Kannan Ramchandran[†]

[†]Department of EECS

University of California at Berkeley, Berkeley, CA
{zhanghao, parekh, kannanr}@eecs.berkeley.edu

[§]Department of Information Engineering

The Chinese University of Hong Kong, Hong Kong
minghua@ie.cuhk.edu.hk

Abstract

We present a multi-channel P2P Video-on-Demand (VoD) system using “plug-and-play” helpers. Helpers are heterogeneous “micro-servers” with limited storage, bandwidth and number of users they can serve simultaneously. Our proposed system has the following salient features: (1) it minimizes the server load; (2) it is distributed, and requires little or no maintenance overhead and which can easily adapt to system dynamics; and (3) it is adaptable to varying supply and demand patterns across multiple video channels irrespective of video popularity. Our proposed solution jointly optimizes over helper-user topology, video storage allocation and bandwidth allocation. The combinatorial nature of the problem and the system demand for distributed algorithms makes the problem uniquely challenging. By utilizing Lagrangian decomposition and Markov chain approximation based arguments, we address this challenge by designing two distributed algorithms running in tandem: a primal-dual storage and bandwidth allocation algorithm and a “soft-worst-neighbor-choking” topology-building algorithm. Our scheme provably converges to a near-optimal solution, and is easy to implement in practice. Simulation results validate that the proposed scheme achieves minimum server load under highly heterogeneous combinations of supply and demand patterns, and is robust to system dynamics of user/helper churn, user/helper asynchrony, and random delays in the network.

I. INTRODUCTION

Our paper is motivated by the following characteristics of online video traffic:

- The amount of video traffic is growing exponentially [4], [10]: YouTube estimates that 24 hours of video are uploaded to its site every minute; thousands of films and TV shows are available for streamed viewing from sites such as Netflix, Amazon and iTunes. Cisco projects that video will account for 60% of the Internet traffic by 2013.
- The demand for video titles is time-varying and heavy-tailed [6]. The need for on-demand video delivery, where users can “channel surf” by switching across the menu of thousands of available videos, is rapidly upsurging.

A fundamental challenge in supporting such a large and diverse on-demand infrastructure is the degree to which this infrastructure can be distributed and maintained at low cost. It seems clear that architectures biased towards centralized distribution are not likely to be scalable as video adoption grows. YouTube, for example, pays millions of dollars per month only on bandwidth costs [10]. Peer-to-Peer (P2P) systems save cost and adjust load automatically, yet they do not provide acceptable quality of service (QoS) for the viewers of all but the most popular videos.

To match the heavy-tailed demand patterns at low maintenance cost, researchers and engineers have introduced the concept of helpers and explored the design of helper-assisted P2P VoD systems. Helper nodes are “micro-servers,” which, in the system scale, have only limited individual resources of storage and bandwidth for the video streaming service. In the PPSstream system for example [1], each peer dedicates about 1 GB of its local storage to cache previously watched videos and helps serve users, therefore reducing the load on the central server. The concept of helpers has also been used in other P2P streaming applications including Xunlei [3] and PPLive [2]. In this paper, we envision an ecosystem in which a variety of devices, including set top boxes, inexpensive PCs and small servers such as CDNs, are incentivized to participate in an economy of helpers.

Minimizing the server load in a helper-assisted P2P VoD system is a challenging problem under practical constraints. First, the extreme range of videos and the sheer amount of content makes it impractical to store

and serve every video on any individual helper. Given a number of distributed helper nodes with limited resources of storage and bandwidth in the system scale, it is important to answer the questions of “what fraction of what videos should be stored on each helper to optimize overall system demand patterns” and “how much bandwidth should helpers allocate to each of their requesting users?” Second, due to practical connection overheads, there are limits on how many users each helper can simultaneously connect to and vice versa. Therefore, an important question is “how should we build optimal helper-user overlay topology?” Since the number of helper-user topology configurations is exponential in the number of nodes, topology building is a challenging combinatorial problem. Third, video popularity is time-varying, and helper/user may randomly join and leave the system, making it difficult to keep track of the supply and demand patterns in real time. It is desirable that the system can adapt to these fluctuations with minimum overhead.

In this paper, we present a helper-assisted P2P VoD system to solve these challenges. Our system has the following distinguishing attributes:

- It minimizes the server load. To do this, we jointly optimize over helper’s video storage and bandwidth allocations, and helper-user topology. By utilizing Lagrangian decomposition and Markov chain approximation based arguments, we design two distributed algorithms running them in tandem: a primal-dual resource allocation algorithm and a “soft-worst-neighbor-choking” topology-building algorithm.
- The proposed algorithm is fully distributed and easy to implement. Peers can dynamically optimize system resources by taking actions based on only local information yet being able to achieve global optimality with provable convergence. Helpers are “plug-and-play”, i.e., a newly deployed helper will automatically connect to a set of interested users and load balance its storage and bandwidth according to the up-to-date supply and demand patterns with minimum maintenance requirement.
- Thanks to the simple distributed solution, our proposed system is adaptable to varying supply and demand patterns across multiple video channels irrespective of video popularity. Our system achieves minimum server load under highly heterogeneous combinations of supply and demand patterns, and is robust to system dynamics of user/helper churn, user/helper asynchrony, and random delays in the network.

Simulation results validate the feasibility and effectiveness of the proposed algorithms and offer new insights into building practical P2P VoD systems.

II. RELATED WORK

Distributed bandwidth allocation for P2P streaming was studied in several works [13], [15]. Wu and Li proposed a rate allocation scheme for a single-video P2P live streaming application without helpers and showed its convergence and optimality [15]. Wang et al. proposed a distributed solution to minimize the weighted sum of server load and non-ISP-friendly traffic [13] under a single video scenario. The problem of being able to switch streams (channels) has been studied in the live P2P streaming case by Wu et al. [16] who propose the concept of view-coupling [14], [18]. When there is storage constraint, Huang and et al. [11] suggested using the proportional replication strategy, i.e., replicating video storage in the system proportionally to their demand. However, this ignores the available system bandwidth for the videos, and can result in poor performance for videos with low demand as was observed by Wang and Lin [12]. Optimal multi-channel on-demand solutions with both bandwidth and storage constraints are challenging primarily because it is difficult to effectively aggregate instantaneous user demand and keep track of available system resources distributively.

The concept of helpers has been addressed by several authors [14], [17], [18]. Wang et al. studied a single-video helper-assisted P2P live streaming scenario and proposed that each helper downloads only one coded packet of the segment that is currently being streamed. Simulation results showed that the proposed system can achieve significantly improved streaming bitrate without incurring additional server load. Zhang et al. [17] and He et al. [18] individually proposed similar concepts of using helpers in a single-video P2P VoD application to boost system performance. He et al. also proposed a distributed algorithm to allocate bandwidth among helpers and users assuming a given helper assignment and a fixed helper-user connection topology. However, all of these works focus on a single video scenario without considering helpers’ storage constraints imposed by the sheer amount of aggregate video volume when tackling the common scenario having an arbitrarily large number of videos. Furthermore, fixed overlay topology was assumed and was not optimized for the overall system performance. BitTorrent [21] uses

a “worst-neighbor-choking” algorithm to update overlay topology, in which users periodically connects to a new and randomly chosen neighbor and chokes the worst performing neighbor. In contrast, we propose a “soft-worst-neighbor-choking” algorithm and prove its optimality.

There are a number of works on practical P2P VoD system design. Annapureddy and et al. proposed a P2P VoD system called Redcarpet [19]. The authors proposed an efficient video block dissemination algorithm in a mesh-based P2P system, and showed that pre-fetching and network coding techniques can greatly improve system performance. Simulation results showed their system can achieve small start-up time and smooth video playback. Huang et al. studied [11] the challenges and the architectural design issues of a large-scale P2P-VoD system based on the experiences of a real system deployed by PPLive. Such challenges include coordinating content storage distribution, content discovery, and peer scheduling. There are also a number measurement studies of practical P2P VoD systems [9], [10].

Our work contributes to the VoD literature in several respects: First, we take practical considerations of bounded user/helper connections and propose a distributed algorithm that optimizes the overlay topology building. Second, we target the problem of multi-channel helper-assisted P2P VoD system with both storage and bandwidth constraints and jointly optimize their allocation. Third, our distributed solution is provided with provable analytics that allows for videos to be served efficiently and with effective response time irrespective of video popularity. Our system is “plug-and-play”, and can easily handle user/helper dynamics and video demand pattern changes and it does so with minimal central co-ordination.

III. PROBLEM SETUP

We first formulate the problem of minimizing the server load in a static helper-assisted VoD system, and then design distributed solutions that allows the system to be adaptive to varying demand patterns in dynamic situations with minimum maintenance overhead. Table I lists the relevant notations.

A. Problem Overview and Assumptions

Consider a VoD system where M videos are served by a dedicated central server and a group of helpers. The dedicated central server fills in any required system deficit to guarantee the real-time streaming requirements of all the users. Each video m has constant streaming rate r_m , duration l_m and size $V_m = r_m l_m$, $m = 1, 2, \dots, M$. There are I_m users in each video session m , and every user i_m watches only one video at a time. Consider J helpers in the system, and each helper j has storage capacity S_j and upload bandwidth capacity B_j , $j = 1, 2, \dots, J$.

We assume that the download bandwidths do not represent a system bottleneck, as is true in typical Internet access scenarios today. We also assume decoupled roles of users and helpers, i.e., users can only request service and helpers can only provide service. In the case that users can also help redistribute the video content, one can still conceptually separate their roles as service-request and service-offering entities, as proposed by Zhang et al. [18] and Wu et al. [16]. We will refer to both users and helpers as “peers”.

Due to practical limits in connection overhead, we consider that each user i_m (helper j) can *simultaneously* connect to at most $N_{i_m}^{\max}$ (N_j^{\max}) neighbor nodes from its candidate neighbor set \mathbb{N}_{i_m} (\mathbb{N}_j). Denote by set \mathbb{C} the entire possible helper-user topology configurations, where each configuration $c \in \mathbb{C}$ is a set of links which connect all the users and helpers and which satisfy the bounded-neighbors constraints. Denote by $\mathbb{N}_{i_m}^c$ ($\mathbb{N}_{j,m}^c$) the set of active helpers (users) that user i_m (helper j) connects to under configuration c . Helper j 's upload rate to user $i_m \in \mathbb{N}_{j,m}^c$ under c is denoted by $x_{j i_m}^c$. For simplicity, we will drop the superscript c in $x_{j i_m}^c$ and use $x_{j i_m}$ instead, but it should be easy to clarify based on the context.

We break video m into segments, each having k packets of equal size, e.g., 1KB. A helper increases (decreases) its stored portion of video m , by downloading (offloading) at the unit of one packet per segment for all video m 's segments. Each packet that the helper stores for each segment is coded using a random linear combination of the k original packets of that segment. In this way, any coded packets from the helpers are equally useful to users in need of the corresponding segments. Consequently, a helper with f_{jm} fraction of the video m can supply at the rate of $f_{jm} r_m$ to each user in video session m , regardless of users' playback times and what they receive from other helpers. The above coding and storage arrangement simplifies the system design as well as the problem formulation.

Table I
KEY NOTATIONS

Notation	Definition
M	total number of videos
I_m	total number of users watching video m
J	total number of helpers
r_m, l_m, V_m	video m 's streaming rate, duration and size
\mathbb{C}	set of feasible overlay configurations
\mathbb{N}_{i_m}	user i_m 's helper neighborhood
\mathbb{N}_j	helper j 's user neighborhood
$\mathbb{N}_{i_m}^c$	set of helpers connected to user i_m under c
$\mathbb{N}_{j,m}^c$	set of users connected to helper j under c
B_j, S_j	upload, storage capacity of helper j
x_{ji_m}	upload rate from helper j to user i_m
f_{jm}	fraction of video m stored by helper j , in $[0, 1]$
k_{ji_m}	helper j 's availability price to user i_m
λ_j, μ_j	bandwidth, storage prices of helper j

Note: we use bold-type to denote vectors.

B. Problem Formulation

Minimizing the server load is equivalent to maximizing the sum of helpers' effective contribution to all the users: $\max_{c, \mathbf{f}, \mathbf{x}} \sum_{m=1}^M \sum_{i_m=1}^{I_m} \min(\sum_{j \in \mathbb{N}_{i_m}^c} x_{ji_m}, r_m)$. Here we implicitly assume that helper's download is a transient cost and is negligible compared to their contribution. Incorporating the constraints, we arrive at the following optimization problem:

$$\max_{c, \mathbf{f}, \mathbf{x}} \sum_{m=1}^M \sum_{i_m=1}^{I_m} \min(\sum_{j \in \mathbb{N}_{i_m}^c} x_{ji_m}, r_m) \quad (1)$$

$$\text{s.t. } x_{ji_m} \leq f_{jm} r_m, \forall j, m, i_m \in \mathbb{N}_{j,m}^c \quad (2)$$

$$\sum_{m=1}^M \sum_{i_m \in \mathbb{N}_{j,m}^c} x_{ji_m} \leq B_j, \forall j \quad (3)$$

$$\sum_{m=1}^M f_{jm} V_m \leq S_j, \forall j \quad (4)$$

$$0 \leq f_{jm} \leq 1, \forall j, m \quad (5)$$

$$c \in \mathbb{C} \quad (6)$$

Constraints (2) are such that each helper's total upload to neighboring user i_m who is viewing video m cannot exceed its available service rate for video m . Constraints (3), (4) and (5) are feasibility constraints on bandwidth and storage. (6) are combinatorial, representing the bounded-neighbor helper-user topology constraints. The above problem is a joint storage, bandwidth, and helper-user topology optimization problem, and is challenging to solve even in a centralized manner due to its combinatorial nature. In the next section, we design distributed algorithms and prove its convergence to a near-optimal solution.

IV. DISTRIBUTED SOLUTION

Constraints (2)~(5) are independent of constraint (6), which allows us to decompose it into a resource allocation problem and a topology building problem and solve them in tandem. We present our solutions to each problem in the following subsections.

A. Storage and Bandwidth Allocation

Fixing topology c , and assigning Lagrangian variables k_{ji_m} to constraints (2), λ_j to constraints (3), and μ_j to constraints (4), we obtain the following partial Lagrangian:

$$\begin{aligned} \min_{\mathbf{k}, \lambda, \mu} \max_{\mathbf{0} \leq \mathbf{f} \leq \mathbf{1}, \mathbf{x}} \sum_{m, i_m} \min_{j \in \mathbb{N}_{i_m}^c} (\sum_{j \in \mathbb{N}_{i_m}^c} x_{ji_m}, r_m) + \sum_{j, m, i_m} \lambda_j (B_j - x_{ji_m}) \\ + \sum_{j, m, i_m} k_{ji_m} (f_{jm} r_m - x_{ji_m}) + \sum_j \mu_j (S_j - \sum_m f_{jm} V_m) \end{aligned} \quad (7)$$

whose optimal solution is denoted by $U(c)$. For simplicity, we abbreviated $\sum_{m=1}^M \sum_{i_m=1}^{I_m}$ as \sum_{m, i_m} , and $\sum_{j=1}^J \sum_{m=1}^M \sum_{i_m \in \mathbb{N}_{j,m}^c}$ as \sum_{j, m, i_m} . By rearranging the terms, the above problem can be solved successively in the primal and dual variables. Specifically, given \mathbf{k}, λ and μ , we have the subproblem in \mathbf{x} and \mathbf{f} :

$$\begin{aligned} \max_{\mathbf{x}} \sum_{m, i_m} (\min_{j \in \mathbb{N}_{i_m}^c} (\sum_{j \in \mathbb{N}_{i_m}^c} x_{ji_m}, r_m) - \sum_{j \in \mathbb{N}_{i_m}^c} (\lambda_j + k_{ji_m}) x_{ji_m}) \\ + \max_{\mathbf{0} \leq \mathbf{f} \leq \mathbf{1}} \sum_{j, m} (r_m \sum_{i_m \in \mathbb{N}_{j,m}^c} k_{ji_m} - \mu_j V_m) f_{jm} \end{aligned} \quad (8)$$

The unique structure of the partial Lagrangian allows us to use a simple primal-dual algorithm [8] as its solution, which we state in the following theorem.

Theorem 1. *The following resource allocation algorithm converges to the optimal solution to problem (7):*

$$\begin{cases} \dot{x}_{ji_m} = \alpha (g_{x_{ji_m}} - (\lambda_j + k_{ji_m}))_{x_{ji_m}}^{[0, +\infty)}, \quad \forall j, m, i_m \in \mathbb{N}_{j,m}^c \\ \dot{f}_{jm} = \beta (\sum_{i_m \in \mathbb{N}_{j,m}^c} k_{ji_m} - l_m \mu_j)_{f_{jm}}^{[0, 1]}, \quad \forall j, m \\ \dot{\lambda}_j = \gamma (\sum_{m=1}^M \sum_{i_m \in \mathbb{N}_{j,m}^c} x_{ji_m} - B_j)_{\lambda_j}^{[0, +\infty)}, \quad \forall j \\ \dot{\mu}_j = \delta (\sum_{m=1}^M f_{jm} V_m - S_j)_{\mu_j}^{[0, +\infty)}, \quad \forall j \\ \dot{k}_{ji_m} = \varepsilon (x_{ji_m} - f_{jm} r_m)_{k_{ji_m}}^{[0, +\infty)}, \quad \forall j, m, i_m \in \mathbb{N}_{j,m}^c. \end{cases} \quad (9)$$

where $h_y^{[a,b]} = \begin{cases} \min(0, h), & y \geq b; \\ h, & a < y < b; \\ \max(0, h), & y \leq a, \end{cases}$ $g_{x_{ji_m}}$ is the partial derivative of the function g with respect to x_{ji_m} and $g = \min(\sum_{j \in \mathbb{N}_{i_m}^c} x_{ji_m}, r_m)$, and $\alpha, \beta, \gamma, \delta, \varepsilon$ are small step sizes.

Proof: At optimal, the following KKT conditions of the Lagrangian in (7) should also hold:

$$\begin{cases} (g_{x_{ji_m}^*} - (\lambda_j^* + k_{ji_m}^*))_{x_{ji_m}^*}^{[0, +\infty)} = 0 \\ (\sum_{i_m \in \mathbb{N}_{j,m}^c} k_{ji_m}^* - l_m \mu_j^*)_{f_{jm}^*}^{[0, 1]} = 0 \\ \lambda_j^* (\sum_{m=1}^M \sum_{i_m \in \mathbb{N}_{j,m}^c} x_{ji_m}^* - B_j) = 0 \\ \mu_j^* (\sum_{m=1}^M f_{jm}^* V_m - S_j) = 0 \\ k_{ji_m}^* (x_{ji_m}^* - f_{jm}^* r_m) = 0 \end{cases}$$

where \mathbf{x}^* and \mathbf{f}^* are the primal optimal and λ^*, μ^* and \mathbf{k}^* are the dual optimal. Strictly speaking, g is not differentiable at $\sum_{j \in \mathbb{N}_{i_m}^c} x_{ji_m} = r_m$, where we simply let $g_{x_{ji_m}} = 0$ in practice without affecting the performance. Denote by $\mathbf{y} = (\mathbf{x}, \mathbf{f}, \lambda, \mu, \mathbf{k})$ and by $\mathbf{y}^* = (\mathbf{x}^*, \mathbf{f}^*, \lambda^*, \mu^*, \mathbf{k}^*)$. To prove that $\mathbf{y} \rightarrow \mathbf{y}^*$, we propose the following

generalized energy function:

$$V(\mathbf{y}) = \frac{1}{2\alpha}\|\mathbf{x} - \mathbf{x}^*\|^2 + \frac{1}{2\beta}\|\mathbf{f} - \mathbf{f}^*\|^2 + \frac{1}{2\gamma}\|\lambda - \lambda^*\|^2 \\ + \frac{1}{2\varepsilon}\|\mathbf{k} - \mathbf{k}^*\|^2 + \frac{1}{2\delta}\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|^2$$

and show that (a) $V(\mathbf{y}) > 0 \forall \mathbf{y} \neq \mathbf{y}^*$ and $V(\mathbf{y}^*) = 0$; (b) $\dot{V}(\mathbf{y}) \leq 0 \forall \mathbf{y}$ and $\dot{V}(\mathbf{y}^*) = 0$.

(a) is obvious since $V(\mathbf{y})$ is summation of quadratic terms. To show (b), we derive $\dot{V}(\mathbf{y})$:

$$\dot{V}(\mathbf{y}) = \sum (x_{ji_m} - x_{ji_m}^*)(g_{x_{ji_m}} - (\lambda_j + k_{ji_m}))_{x_{ji_m}}^{[0,+\infty)} \\ + \sum (f_{jm} - f_{jm}^*)(\sum k_{ji_m} - l_m \mu_j)_{f_{jm}}^{[0,1]} \\ + \sum (\lambda_j - \lambda_j^*)(\sum x_{ji_m} - B_j)_{\lambda_j}^{[0,+\infty)} \\ + \sum (k_{ji_m} - k_{ji_m}^*)(x_{ji_m} - f_{jm} r_m)_{k_{ji_m}}^{[0,+\infty)} \\ + \sum (\mu_j - \mu_j^*)(\sum f_{ji_m} V_m - S_j)_{\mu_j}^{[0,+\infty)}$$

by applying partial derivatives and plugging in the dynamic system equations. For simplicity, we have omitted the sets over which the terms are summed up. It is easy to see that $\dot{V}(\mathbf{y}^*) = 0$. Now we can upper-bound $\dot{V}(\mathbf{y})$ as follows:

$$\dot{V}(\mathbf{y}) \leq \sum (x_{ji_m} - x_{ji_m}^*)(g_{x_{ji_m}} - (\lambda_j + k_{ji_m})) \\ + \sum (f_{jm} - f_{jm}^*)(\sum k_{ji_m} - l_m \mu_j) \\ + \sum (\lambda_j - \lambda_j^*)(\sum x_{ji_m} - B_j) \\ + \sum (k_{ji_m} - k_{ji_m}^*)(x_{ji_m} - f_{jm} r_m) \\ + \sum (\mu_j - \mu_j^*)(\sum f_{ji_m} V_m - S_j) \\ = \sum (x_{ji_m} - x_{ji_m}^*)(g_{x_{ji_m}} - g_{x_{ji_m}^*}) \\ + \sum (x_{ji_m} - x_{ji_m}^*)(g_{x_{ji_m}^*} - (\lambda_j^* + k_{ji_m}^*)) \\ + \sum (f_{jm} - f_{jm}^*)(\sum k_{ji_m}^* - l_m \mu_j^*) \\ + \sum (\lambda_j - \lambda_j^*)(\sum x_{ji_m}^* - B_j) \\ + \sum (k_{ji_m} - k_{ji_m}^*)(x_{ji_m}^* - f_{jm}^* r_m) \\ + \sum (\mu_j - \mu_j^*)(\sum f_{ji_m}^* V_m - S_j) \\ \leq \sum (x_{ji_m} - x_{ji_m}^*)(g_{x_{ji_m}} - g_{x_{ji_m}^*}) \\ + 0 + 0 + 0 + 0 + 0 \leq 0$$

where the first inequality is obtained by dropping the $_{y}^{[a,b]}$ terms, and the second inequality is obtained by applying the set of KKT conditions. The last set inequality holds due to the concavity of the function $\min(\sum x_{ji_m}, r_m)$ over x_{ji_m} .

Using (a) and (b), it follows from Krasovskii-LaSalle principle [20] that \mathbf{y} converges to the set $\mathbb{S} = \{\mathbf{y} | \dot{V}(\mathbf{y}) = 0\}$. It remains to show that the \mathbb{S} contains no trajectories other than $\{\mathbf{y} = \mathbf{y}^*\}$. Due to the space constraint, we omit that part of the proof in this paper. Interested readers can refer to [8] for details. ■

We make the following remarks:

- The resource allocation algorithm has intuitive economic explanations. Specifically, one can view k_{ji_m} and λ_j as the *video availability prices* and *bandwidth prices* which are induced by helper j 's storage and bandwidth constraints and which helper j charges user i_m . One can also view μ_j as *storage prices* that helper j has

to *pay*. Indeed, the larger the video availability and bandwidth prices are, the smaller \hat{x}_{ji_m} (that user i_m requests from helper j) is. The larger/smaller the video availability price/storage price is, the larger f_{jm} (that helper j increases for video m) is. Similarly, the values of the prices are also driven by the relative difference between the given demand and the available resources. For example, the increase in video availability price k_{ji_m} is proportional to the difference between the demand x_{ji_m} and the available rate $f_{jm}r_m$. This economic framework can be potentially extended to building incentive mechanisms into the system.

- It is also not hard to see that the primal variables \mathbf{x} and \mathbf{f} will converge to the following intuitive solutions. In problem (8), every user i_m will choose to request x_{ji_m} with the smallest combined prices ($\lambda_j + k_{ji_m}$) until it reaches the maximum possible value. If the summation of received rates has not reached r_m , it will choose to request x_{ji_m} with the second smallest combined prices. It will continue to do so until the summation of the received rates reaches r_m . Similarly, the solution for $f_{jm}, m = 1, 2, \dots, M$ can be obtained by water-filling helper j 's storage S_j in descending order of the combined prices ($r_m \sum_{i_m \in \mathbb{N}_{j,m}^c} k_{ji_m} - \mu_j V_m$), which matches with helper j 's goal to maximize its "profit".

B. Markov Approximation of Overlay Optimization

Recall that $U(c)$ is the optimal solution to problem (7). It is then left to solve:

$$\max_c U(c) \quad \text{s.t.} \quad c \in \mathbb{C} \quad (10)$$

However, the set of possible overlay configurations given peers neighborhood constraints is exponential in the number of nodes, which make the problem NP hard even in a centralized manner. To overcome this difficulty, we re-write it as follows:

$$\begin{aligned} \max_p \quad & \sum_{c \in \mathbb{C}} p_c U(c) \\ \text{s.t.} \quad & \sum_{c \in \mathbb{C}} p_c = 1 \quad \text{and} \quad 0 \leq p_c \leq 1, \forall c \in \mathbb{C} \end{aligned} \quad (11)$$

One can see that the problems (11) and (10) are equivalent: the optimal solution to problem (11) is obtained by setting $p_{c^*} = 1$ for $c^* = \arg \max_{c \in \mathbb{C}} U(c)$ and $p_c = 0$ for all other $c \in \mathbb{C}$. Relaxing the objective $\sum_{c \in \mathbb{C}} p_c U(c)$ by adding a weighted entropy term $\frac{1}{\kappa} H(p)$, where $\kappa > 0$ and $H(p) = -\sum_{c \in \mathbb{C}} p_c \log p_c$, we have the following theorem shown by Chen et al. [7]:

Theorem 2. *The optimal solution to:*

$$\begin{aligned} \max_{\mathbf{p}} \quad & \sum_{c \in \mathbb{C}} p_c U(c) - \frac{1}{\kappa} \sum_{c \in \mathbb{C}} p_c \log p_c \\ \text{s.t.} \quad & \sum_{c \in \mathbb{C}} p_c = 1 \quad \text{and} \quad 0 \leq p_c \leq 1, \forall c \in \mathbb{C} \end{aligned} \quad (12)$$

is given by:

$$p_c^* = \frac{\exp(\kappa U(c))}{\sum_{c \in \mathbb{C}} \exp(\kappa U(c))}, \quad \forall c \in \mathbb{C} \quad (13)$$

Proof: The Lagrangian of the problem (12) is given by:

$$\begin{aligned} L(p_c, \nu_c, \mu) = & \sum_{c \in \mathbb{C}} p_c U(c) - \frac{1}{\kappa} \sum_{c \in \mathbb{C}} p_c \log p_c \\ & + \sum_{c \in \mathbb{C}} \nu_c p_c + \mu(1 - \sum_{c \in \mathbb{C}} p_c), \end{aligned}$$

where ν_c and μ are the Lagrangian variables. At optimal, the following KKT conditions [?] should hold:

$$U(c) - \frac{1}{\kappa} (\log p_c^* + 1) + \nu_c^* - \mu^* = 0, \quad \forall c \in \mathbb{C},$$

where p_c^* is the primal optimal, and ν_c^* and μ^* are the dual optimal. Writing p^* as a function of ν_c^* and μ^* and applying the constraint $\sum_{c \in \mathbb{C}} p_c^* = 1$, we obtain:

$$\mu^* = \frac{1}{\kappa} \log \left(\sum_{c \in \mathbb{C}} \exp(\kappa(U(c) + \nu_c^*) - 1) \right).$$

Plugging μ^* back into p_c^* , we get:

$$\begin{aligned} p_c^* &= \frac{\exp(\kappa(U(c) + \nu_c^*) - 1)}{\exp(\kappa\mu^*)} \\ &= \frac{\exp(\kappa U(c))}{\sum_{c \in \mathbb{C}} \exp(\kappa U(c))}, \quad \forall c \in \mathbb{C}. \end{aligned}$$

■

Note that the optimal solution p_c^* is in a product-form, and thus is the stationary distribution of some time-reversible Markov Chain (MC), hence the term MC approximation. Note that as $\kappa \rightarrow +\infty$, $p_{c^*}^* \rightarrow 1$ and therefore the optimal solution of the relaxed problem (12) approaches to that of the original problem (10). It is also easy to see that for a fixed κ , the error term $\frac{1}{\kappa} H(p)$ is bounded by $\frac{1}{\kappa} \log |\mathbb{C}|$.

Our motivation behind this approximation is that it can potentially lead to distributed solutions. In this case, one can construct a MC with the overlay topology configurations as its states, and carefully design transition rates $q_{c,c'}$ such that the overall system will probabilistically jump between topology configurations while staying in the best configuration, i.e., c^* for most of the time, and that the system performance will approach to the optimal. One straightforward design of $q_{c,c'}$ is given by:

$$q_{c,c'} = \begin{cases} \frac{\tau}{\exp(\kappa(U(c)))} & c, c' \text{ satisfy } S; \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

where $\tau > 0$ is a constant, $U(c)$ is the overall system utility under state c , and S is the following set of conditions:

- $\exists \tilde{c}$ s.t. $\tilde{c} \subseteq c, \tilde{c} \subseteq c', |c \setminus \tilde{c}| = |c' \setminus \tilde{c}| = 1$;
- Link $c \setminus \tilde{c}$ and link $c' \setminus \tilde{c}$ originates from the *same* peer.

In other words, only the following state transitions $c \rightarrow c'$ are valid: a *single peer* first drops a *single* connection to one of his neighbors (from $c \rightarrow \tilde{c}$) and then randomly adds a new *single* connection from his neighborhood (from $\tilde{c} \rightarrow c'$). \tilde{c} is an auxiliary state and can be viewed as the intermediate state in which a single link from a single peer is dropped from c , where $c \setminus \tilde{c}$ represents the dropped link. It is not hard to see that $q_{c,c'}$ satisfies the detailed balanced equations $q_{c,c'} p_c^* = q_{c',c} p_{c'}^*$, thus the stationary distribution p_c^* in equation (13) can be achieved. We refer to this as the “*uniform-neighbor-choking*” algorithm because peers uniformly randomly choke neighbors in periods that depend on $U(c)$.

However, one caveat of the above design is that a peer still needs to know the global information $U(c)$ that needs to be broadcast to all the peers from time to time. This burdens the system with overhead. It is desirable to have a distributed algorithm in which each peer needs only *local* information to perform such update and still achieve *global* optimality.

C. Soft-Worst-Neighbor-Choking Algorithm

Motivated by the above discussions, we propose the “*soft-worst-neighbor-choking*” algorithm by designing:

$$\bar{q}_{c,c'} = \frac{\tau}{\exp(\kappa x^{c \setminus \tilde{c}})} \quad (15)$$

where $x^{c \setminus \tilde{c}}$ is the rate on the dropped link $c \setminus \tilde{c}$, and c, \tilde{c}, c' should satisfy S . Here, the transition rates depend on only local information of link rates of peers’ active neighbors.

We now give the overall distributed algorithm. For simplicity, the algorithm is stated under the perspective of user i_m , and those at other users/helpers are similar and self-explanatory.

Topology building - “*soft-worst-neighbor-choking*”

- Initialization: User i_m randomly chooses and connects to $N_{i_m}^{\max}$ neighboring helpers from his neighborhood \mathbb{N}_{i_m} and does the following steps.
- Step 1: User i_m independently draws an exponentially distributed random variable with mean $\frac{1}{\tau(|\mathbb{N}_{i_m}| - N_{i_m}^{\max}) \sum_{j \in \mathbb{N}_{i_m}^c} \exp(-\kappa x_{ji_m})}$ and counts down to zero.
- Step 2: After the count-down expires, user i_m drops neighbor j with probability $\frac{\exp(-\kappa x_{ji_m})}{\sum_{j' \in \mathbb{N}_{i_m}^c} \exp(-\kappa x_{j'i_m})}$ and randomly chooses and connects to a new neighbor from the set $\mathbb{N}_{i_m} \setminus \mathbb{N}_{i_m}^c$. It then repeats Step 1.

We make the following remarks:

- It is easy to see that the algorithm gives $q_{c,c'}$ as in (15).
- The algorithm is fully distributed, i.e., each peer runs the algorithm independently. Compared to the “uniform-neighbor-choking” algorithm, peers only need to know local information of the link rates of their one-hop neighbors. The algorithm is intuitive: the larger the link rate, the less likely it is dropped and vice versa.
- It is worth noting that BitTorrent [21] uses a “worst-neighbor-choking” algorithm, where each peer periodically chokes the link with the worst rate. In our case, link rates are weighted exponentially. The worst link is choked with the highest probability (which approaches 1 as $\kappa \rightarrow +\infty$) while other links can also be choked occasionally, hence the term “soft-worst-neighbor-choking” algorithm. This algorithm is also generalizable to other P2P systems.

D. Performance Analysis of Soft-Worst-Neighbor-Choking

We now state the mathematical underpinnings behind such design of the algorithm and analyze its performance. It is interesting to see that the rate $x^{c,\tilde{c}}$ on link $c \setminus \tilde{c}$ can be viewed as an approximation to $U(c) - U(\tilde{c})$, which is the overall system performance difference before and after link $c \setminus \tilde{c}$ is dropped. If $x^{c,\tilde{c}} = U(c) - U(\tilde{c})$, i.e., the helper cannot re-utilize the upload rate on $c \setminus \tilde{c}$ after it is dropped, then we have:

$$\bar{q}_{c,c'} = \frac{\tau}{\exp(\kappa(U(c) - U(\tilde{c})))}$$

where c, \tilde{c}, c' satisfy S . In this case, $\bar{q}_{c,c'}$ still satisfies $\bar{q}_{c,c'} p_c^* = \bar{q}_{c',c} p_{c'}^*$ and the stationary distribution is no different from that of the uniform-neighbor-choking algorithm in (13). However, $U(c) - U(\tilde{c}) \leq x^{c,\tilde{c}}$ in general, because the rate $x^{c,\tilde{c}}$ on the dropped link maybe fully or partially re-utilized by the helper for his other neighbors. In the following, we show that under some minor assumptions, one can still achieve a stationary distribution in product form similar to that in (13).

Denote by $\omega_c = x^{c,\tilde{c}} - (U(c) - U(\tilde{c}))$ the error term resulted by approximating $(U(c) - U(\tilde{c}))$ with $x^{c,\tilde{c}}$. Depending on overlay c and the actual converged values of the storage and rate allocation algorithm, ω_c may take values anywhere in between 0 and B_{\max} , where B_{\max} is the maximum over all helpers' upload capacity. We quantize such error ω_c into $n_c + 1$ values $[0, \frac{B_{\max}}{n_c}, \frac{2B_{\max}}{n_c}, \dots, B_{\max}]$, and assume that $\omega_c = \frac{kB_{\max}}{n_c}$ with probability $\rho_{c_k}, k = 0, 1, \dots, n_c$ and $\sum_{k=0}^{n_c} \rho_{c_k} = 1$. Under these assumptions, we show the following theorem:

Theorem 3. *The stationary distribution p_c of MC with transition rates (15) is given by:*

$$p_c = \sum_{k=0}^{n_c} p_{c_k} = \frac{\sigma_c \exp(\kappa U(c))}{\sum_{c' \in \mathcal{C}} \sigma_{c'} \exp(\kappa U(c'))} \quad (16)$$

where $\sigma_c = \sum_{k=0}^{n_c} \rho_{c_k} \exp\left(\kappa \frac{kB_{\max}}{n_c}\right)$.

Proof: Consider a modified MC as follows: expand each state c of the original MC to $n_c + 1$ states $c_k, k = 0, 1, \dots, n_c$ with the following transition rates:

$$\bar{q}_{c_k, c'_k} = \frac{\tau \rho_{c'_k}}{\exp\left(\kappa \left(U(c) - U(\tilde{c}) + \frac{kB_{\max}}{n_c}\right)\right)} \quad (17)$$

where $\rho_{c'_k}, k = 0, 1, \dots, n_{c'}$, is the probability measure on expanded states and $\sum_{k'=0}^{n_{c'}} \rho_{c'_k} = 1$. Note that c_0 refers to state c with zero error. Using equation (17) and detailed balance equations $p_{c_k} \bar{q}_{c_k, c'_k} = p_{c'_k} \bar{q}_{c'_k, c_k}$, we have

$\forall c_0, c'_{k'}$:

$$\frac{p_{c_0}}{\rho_{c_0} \exp(\kappa U(c))} = \frac{p_{c'_{k'}}}{\rho_{c'_{k'}} \exp\left(\kappa \left(U(c') + \frac{k' B_{\max}}{n_{c'}}\right)\right)} = \text{const}$$

Using $\sum_{c \in \mathcal{C}} \sum_{k=0}^{n_c} p_{c_k} = 1$, we obtain:

$$p_{c_k} = \frac{\rho_{c_k} \exp\left(\kappa \left(U(c) + \frac{k B_{\max}}{n_c}\right)\right)}{\sum_{c' \in \mathcal{C}} \sum_{k'=0}^{n_{c'}} \rho_{c'_{k'}} \exp\left(\kappa \left(U(c') + \frac{k' B_{\max}}{n_{c'}}\right)\right)}$$

Denote by $\sigma_c = \sum_{k=0}^{n_c} \rho_{c_k} \exp\left(\kappa \frac{k B_{\max}}{n_c}\right)$ and we have:

$$p_c = \sum_{k=0}^{n_c} p_{c_k} = \frac{\sigma_c \exp(\kappa U(c))}{\sum_{c' \in \mathcal{C}} \sigma_{c'} \exp(\kappa U(c'))} \quad (18)$$

■

We make the following remarks:

- If $n_c = \text{const} \forall c$ and the distribution $\rho_{c_k}, k = 0, 1, \dots, n_c$ is the same $\forall c, \sigma_c = \text{const}$. In this case, $p_c = p_c^*$.
- The total variational distance $d_{TV}(\mathbf{p}^*, \mathbf{p})$ can be upper bounded by $(1 - \exp(-\kappa B_{\max}))$. This is because $p_c^* - p_c = p_c^* \left(1 - \frac{\sigma_c \exp(\kappa U(c))}{\sum_{c' \in \mathcal{C}} \sigma_{c'} \exp(\kappa U(c'))}\right)$. Since $\sigma_c \in [1, \exp(\kappa B_{\max})]$, the fractional can be lower bounded by $\exp(-\kappa B_{\max})$ and hence the result.
- In general, it is difficult to give a tight analytical lower bound on the performance $\sum_{c \in \mathcal{C}} U(c) p_c$ because σ_c is unknown. However, note that when $\kappa \rightarrow +\infty, p_{c^*} \rightarrow 1$ where $c^* = \arg \max_{c' \in \mathcal{C}} U(c')$, and the system approaches the optimal $U(c^*)$. We will show in our simulations that the algorithm performs well and is close to the optimal.

E. Discussions

Our overall scheme is fully distributed, i.e., each peer runs the algorithm independently and makes changes based on only one-hop local information. The user passes the derivative of their utility function to helpers to perform distributed resource allocation; users and helpers periodically choke their neighbors based the relative performance of their one-hop links.

The deployment of such a practical system is easy: newly deployed helper nodes can automatically connect to a set of interested users and load balance their storage and bandwidth resources using the distributed algorithms. This simple solution helps achieve minimum maintenance overhead and can easily adapt to system dynamics. When system's supply and demand pattern changes due to helper/user joining/leaving the system and video popularity shifting, the helper nodes will automatically update their content caching, allocate their bandwidth resource and dynamically change their neighborhood selections in a distributed and local manner, which will best match to the global system demand and available resources and optimize overall system utility. We summarize the simple algorithms at both classes of peer nodes in the next section.

V. SYSTEM IMPLEMENTATION

A. Back-up Server and Tracker

A centralized server with all the video content is present that acts as the "life-line" to supplement the deficit (if any) in the system. A tracker is used to keep track of all the participating peers and to assist building an overlay network. When a user/helper joins the system, it obtains from the tracker the IP addresses of a list of helpers/users in their neighborhood. It then connects to its maximum allowed number of neighbors randomly chosen from their neighborhood.

Algorithm 1 User Protocol

- 1: **Initialization:** Set $t_1, t_2 = 0$, draw $t_3 \sim \text{Exp}\left(\tau(|\mathbb{N}_{i_m}| - N_{i_m}^{\max}) \sum_{j \in \mathbb{N}_{i_m}^c} \exp(-\kappa x_{ji_m})\right)$, and iterate:
 - 2: **if** $\text{mod}(t_1, T_{i_m}) = 0$ **then**
 - 3: Count the number of received packets from each neighboring helper over the period $[t_1 - T_{i_m}, t_1 - 1]$, and update the corresponding average rate x_{ji_m} . Derive the derivative $g_{x_{ji_m}}$ of its utility function and sends them to all $j \in \mathbb{N}_{i_m}^c$.
 - 4: **end if**
 - 5: **if** $\text{mod}(t_2, \text{BUFFER_TIME}) = 0$ **then**
 - 6: Download from the server all the missing packets in the next BUFFER_TIME worth of segments.
 - 7: **end if**
 - 8: **if** $t_3 = 0$ **then**
 - 9: Drop neighbor helper j with probability $\frac{\exp(-\kappa x_{ji_m})}{\sum_{j' \in \mathbb{N}_{i_m}^c} \exp(-\kappa x_{j'i_m})}$. Randomly choose and connect to a new neighbor from the remaining neighborhood to replace j , and set $x_{ji_m} = 0$. Draw $t_3 \sim \text{Exp}\left(\tau(|\mathbb{N}_{i_m}| - N_{i_m}^{\max}) \sum_{j \in \mathbb{N}_{i_m}^c} \exp(-\kappa x_{ji_m})\right)$.
 - 10: **end if**
 - 11: $t_1 \leftarrow t_1 + 1, t_2 \leftarrow t_2 + 1, t_3 \leftarrow t_3 - 1$.
-

B. Packet Exchange Protocol

The video packets each helper stores are coded using a rateless code and downloaded from the server. To validate the proposed algorithms under asynchronous scenarios, we let each helper j maintain its own clock. Helper j also updates its bandwidth/storage allocation algorithm only in periods of T_j seconds and keeps an outgoing buffer worth of T_j seconds of its upload bandwidth capacity. Users maintain a buffer length denoted by BUFFER_TIME that covers an integer number of video segments. For a particular user, as it decodes the packets and plays the video in the segment right ahead of its playback time, it also receives packets of the next unfulfilled segment from the helpers. Upon finishing BUFFER_TIME worth of segments, the user will immediately fetch the packets from the server to fill any missing packets in the next BUFFER_TIME range. Each user also has its own clock and an bandwidth request update period T_{i_m} . The detailed packet exchange protocols for both users and helpers are described in Algorithm 1 and 2 based on the theoretical analysis given in section (IV).

VI. SIMULATION RESULTS

It is worth noting that our proposed analysis relies on a few assumptions made to make mathematical arguments simple. First, we have assumed in the algorithms that peers have *synchronized* clocks. This can be difficult to maintain in practice. Second, when solving problem (10), we have assumed that the underlying resource allocation algorithm has fully converged and ignored the different time scales. Although there exist a number of techniques that can address these issues [5], [8], they are not the focus of our paper and we omit the heavy discussions involved in the analysis. Instead, we validate the feasibility and effectiveness of the proposed algorithms by designing our simulations that capture real-world scenarios including the effects of asynchronized clocks among peers and random network delays. We also show in the simulations that the distributed scheme adapts well to system fluctuations including change in video demand patterns and peer dynamics.

A. Experimental Setup

We set total number of videos $M = 4$, helpers $J = 70$ and users $\sum_{m=1}^M I_m = 100$. Table II shows each video's streaming rate and the fraction of users watching it. Helpers have upload and storage capacities with different distributions shown in Tables III and IV. Each peer can potentially connect to every other peer in the system, but has a maximum allowed number of neighbors uniformly randomly chosen from $[3, 10]$. This setup is based on practical data in commercialized P2P systems [1]–[3], which makes it easy to test the robustness of the proposed algorithms in highly heterogenous scenarios. We also set the step sizes $\alpha = 1, \beta = 0.01, \gamma = \delta = 0.5, \varepsilon = 0.05$ for

Algorithm 2 Helper Protocol

- 1: **Initialization:** Set $x_{ji_m} = 0$, $f_{jm} = 0$, $\lambda_j = 0$, $\mu_j = 0$ and $k_{ji_m} = 0$. Set $t_1 = 0$, draw $t_2 \sim \text{Exp}\left(\tau(|\mathbb{N}_j| - N_j^{\max}) \sum_{i_m \in \mathbb{N}_{j,m}^c} \exp(-\kappa x_{ji_m})\right)$, and iterate:
 - 2: **if** $\text{mod}(t_1, T_j) = 0$ **then**
 - 3: $x_{ji_m} \leftarrow x_{ji_m} + \alpha(g_{x_{ji_m}} - (\lambda_j + k_{ji_m}))_{x_{ji_m}}^{[0,+\infty)}$
 - 4: $f_{jm} \leftarrow f_{jm} + \beta(\sum_{i_m \in \mathbb{N}_j^m} k_{ji_m} - l_m \mu_j)_{f_{jm}}^{[0,1]}$
 - 5: $\lambda_j \leftarrow \lambda_j + \gamma(\sum_{m=1}^M \sum_{i_m \in \mathbb{N}_{j,m}^c} x_{ji_m} - B_j)_{\lambda_j}^{[0,+\infty)}$
 - 6: $\mu_j \leftarrow \mu_j + \delta(\sum_{m=1}^M f_{ji_m} V_m - S_j)_{\mu_j}^{[0,+\infty)}$
 - 7: $k_{ji_m} \leftarrow k_{ji_m} + \varepsilon(x_{ji_m} - f_{jm} T_m)_{k_{ji_m}}^{[0,+\infty)}$.
 - 8: Allocate number of packets equivalent to $x_{ji_m} T_j$ to user i_m for all i_m and put them in the outgoing buffer.
 - 9: Re-allocate its storage of videos according to f_{ji_m} .
 - 10: **end if**
 - 11: Send remaining packets in the buffer to neighbor users.
 - 12: **if** $t_2 = 0$ **then**
 - 13: Drop neighbor user i_m with probability $\frac{\exp(-\kappa x_{ji_m})}{\sum_{i'_m \in \mathbb{N}_{j,m}^c} \exp(-\kappa x_{ji'_m})}$. Randomly choose and connect to a new neighbor from the remaining neighborhood to replace i_m , and set $x_{ji_m} = 0$. Draw $t_2 \sim \text{Exp}\left(\tau(|\mathbb{N}_j| - N_j^{\max}) \sum_{i_m \in \mathbb{N}_{j,m}^c} \exp(-\kappa x_{ji_m})\right)$.
 - 14: **end if**
 - 15: $t_1 \leftarrow t_1 + 1, t_2 \leftarrow t_2 - 1$.
-

the bandwidth and storage allocation algorithm, and set $\kappa = 10, \tau = 0.01$ in the topology update algorithm. These parameters are chosen to guarantee smooth algorithm updates and small MC approximation errors.

B. Convergence in the Static Case

We first test the convergence of the storage and bandwidth allocation algorithm in the static case, where all peers stay in the system during the entire simulation time and perform no topology update. We first focus on the synchronous case, where peers share a synchronous clock and have an update period of 1 second. Figure 1(a) shows the instantaneous server load versus simulation time. Also shown as for comparison is the system's intrinsic deficit, i.e., total users' streaming rate demand minus total helpers' upload bandwidth. The initial server load is high, but it quickly drops to a stable point. The sub-figures (b) and (c) in Figure 1 show the convergence of a particular helper's (ID = 1) upload rate and storage allocation. The convergence results for the shadow prices are similar, which we omit here due to limit in space.

Table II
VIDEO STREAMING RATE DISTRIBUTION

Streaming rate (kbps)	768	896	896	1152
Fraction (%)	10	20	50	20

Table III
HELPER UPLOAD CAPACITY DISTRIBUTION

Upload (kbps)	256	384	512	640	768	896	1024
Fraction (%)	5	10	15	40	15	10	5

Table IV
HELPER STORAGE CAPACITY DISTRIBUTION

Storage (MB)	768	960	1152	1344	1536	1728	1920
Fraction (%)	5	5	10	10	20	40	10

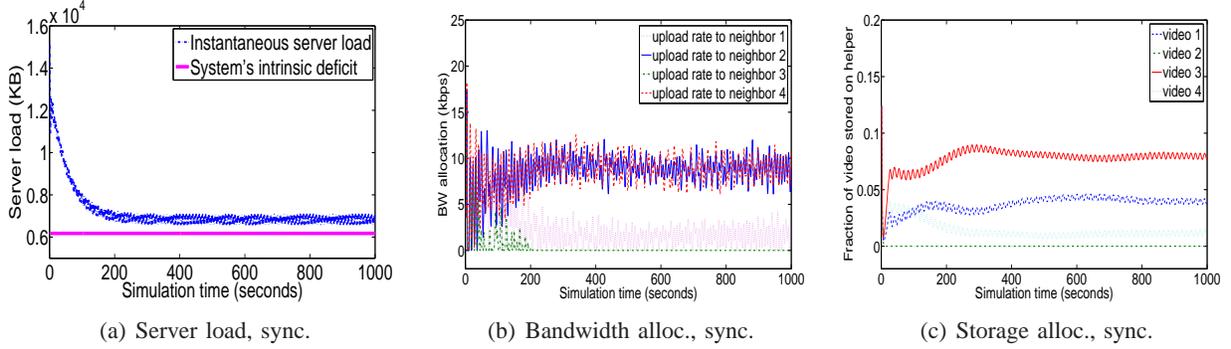


Figure 1. Convergence results of storage and bandwidth allocation algorithm in a static and synchronous setting, where no overlay topology update nor peer dynamics is present. Peers have synchronized update periods of 1 seconds. (a), (b) and (c) show the required server load, bandwidth and storage allocation for the helper (ID = 1).

C. Asynchrony and Random Network Delay

To test the robustness of the bandwidth and storage allocation algorithm in real networks, we add asynchrony and random network delay in the system. Specifically, peers have asynchronous clocks and choose update periods uniformly randomly from the set of $\{1, 3, 5, 7, 9\}$ seconds. In addition, every peer has a communication delay to every one of its neighbors randomly chosen from 1 to 5 seconds. These numbers are chosen to stress test the system. Figure 2 shows the server load, bandwidth and storage allocation for the same helper (ID = 1). Compared to Figure 1, the bandwidth allocations experiences more fluctuations, but they still center around comparable average values. The server load and helpers' storage load are quite stable, which demonstrates the robustness of the algorithm. In the following sections, our simulation experiments will apply the same asynchrony and random network delays unless mentioned otherwise.

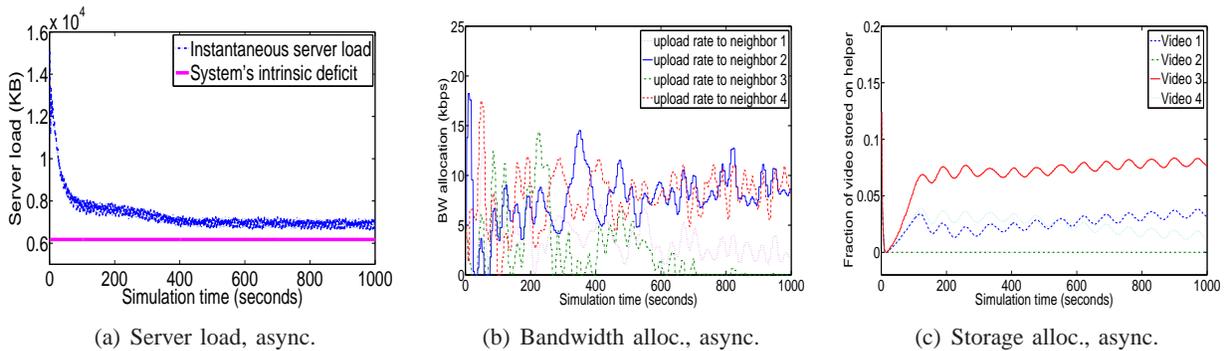


Figure 2. Convergence of storage and bandwidth allocation algorithm in the case of asynchrony and random network delays, with no overlay topology update nor peer dynamics. Peers have asynchronous clocks and random update periods uniformly drawn from $\{1, 3, 5, 7, 9\}$ seconds. Each peer also has a communication delay to his neighbors randomly chosen from $[1, 5]$. (a), (b) and (c) show the required server load, bandwidth and storage allocation for the helper (ID=1).

D. Effectiveness of the Overlay Topology Update

As is evident from Figure 1, the server load cannot reach to the minimum value of the intrinsic system deficit without topology update. This is because some helpers have poor performance to their connected neighbors and have not fully utilized their upload bandwidth. We run the distributed overlay topology update algorithm on top of the resource allocation algorithm with other parameters and configurations unchanged. Figure 3(a)(b) shows server load versus simulation time without and with overlay topology update respectively, where Figure 3(a) is simply Figure 2(a) shown again for comparison. It can be seen that overlay topology update buys approximately 14% reduction in server load and eventually achieves the intrinsic system deficit which is the theoretical lower bound of server load.

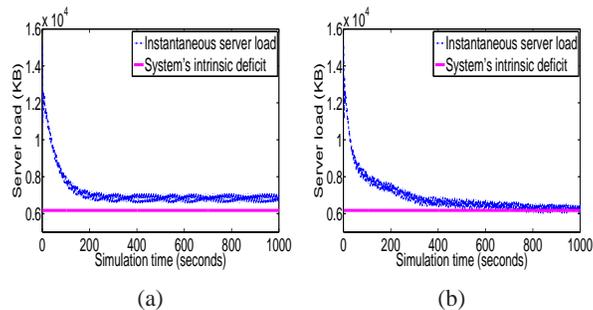


Figure 3. Server load versus simulation time. Peers have asynchronous update periods and random network delays. (a) Server load without overlay topology update; (b) server load with overlay topology update.

E. Effects of Dynamics

We show in this section that our system is “plug-and-play”, i.e., which requires minimum maintenance and which is automatically adaptive to system dynamics. Peers will only need to run their distributed algorithms regardless of system fluctuations and be able to keep updated to the supply and demand patterns across multiple channels.

We first examine the effects of peer dynamics. To do this, we add new users and new helpers that join the system following a Poisson process with mean 20. The newly joined peers will follow the demand and resource distributions listed in Tables II, III and IV. In addition, every peer will stay in the system for an exponential random amount of time with average of 200 seconds. To examine how fast the system responds to dynamics, we simulate till 1000 seconds but stop the dynamic process at the 600th second. Figure 4 (a)(b) show how the server load varies with time, without and with overlay topology updates respectively. The available system resources also change due to dynamics, as is evident from the varying intrinsic system deficit shown in the figures. It is demonstrated that the algorithm can keep updated to the dynamics. When overlay topology update is present, the system can also approach the minimum server load. Note that the instantaneous intrinsic system deficit stops at a different value in two cases, only due to the difference in the pseudo-randomness generated by the computer with and without the topology update. The results have demonstrated the robustness of the resource allocation and topology update algorithms to system dynamics.

We also use a simple example to illustrate how the system responds to changes in video demand patterns. In particular, we pick a helper (ID = 4) who has 10 neighbor users with 7 users watching video 3 and 3 users watching video 4. At $t = 300$, we let all the users in video 3 “switch channels” with half of them switching to video 4 and the other half switching to video 2. Figure 5 shows how the helper (ID = 4) responds to such change by re-allocating its storage resources. Both Figure 5 and Figure 1(c) demonstrate the helper’s “plug-and-play” feature, i.e., helpers can automatically load balance its resources given system demand patterns.

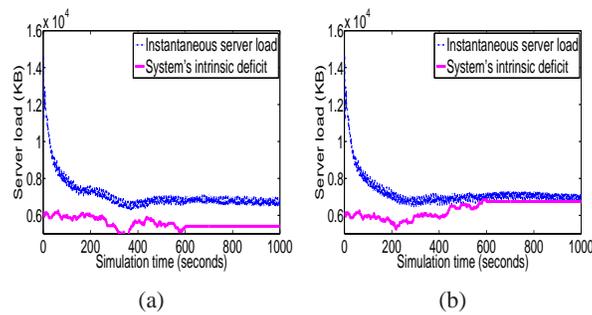


Figure 4. Effects of system dynamics on server load. A new user and a new helper will join the system every 20 seconds on average. Each peer stays for an average of 200 seconds in the system. (a) Server load without overlay topology update; (b) server load with overlay topology update.

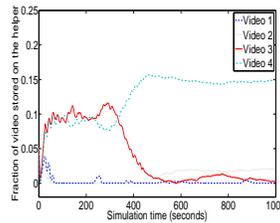


Figure 5. Effects of changes in video demand patterns. Users in video session 3 “channel switch” to video 2 and 4 with equal numbers.

VII. CONCLUSIONS

In this paper, we propose to minimize the server load in a helper-assisted multi-channel P2P VoD system. Helpers who help provide the VoD service are limited in bandwidth and storage, and each helper and user has a constraint on the maximum number of neighbors that they can connect to. This problem is critical for exploring the maximum potential of practical distributed P2P VoD systems. The mix-convex-combinatorial nature of the problem under practical constraints makes it challenging to solve even in a centralized manner. We tackle this challenge by designing two distributed algorithms running in tandem: a primal-dual resource allocation algorithm and a “soft-worst-neighbor-choking” topology building algorithm. The overall scheme is simple to implement and provably converges to a near-optimal solution. Simulation results show that our proposed algorithm minimizes the server load, and is robustness to asynchronous clock times, random network delay, video popularity changes and peers dynamics. Our proposed system design and algorithm provide useful insight to practical video content distribution applications. Possible future work includes: (1) design incentive mechanisms into the system; and (2) build a practical system prototype.

REFERENCES

- [1] <http://www.pstream.com/>.
- [2] <http://www.pplive.com/>.
- [3] <http://www.xunlei.com/>.
- [4] Cisco Visual Networking Index: Forecast and Methodology, 2008-2013. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.
- [5] A. Benveniste, P. Priouret, and M. Métivier. *Adaptive algorithms and stochastic approximations*. 1990.
- [6] Y. Boufkhad, F. Mathieu, and et al. Achievable catalog size in peer-to-peer video-on-demand systems. In *IPTPS*, 2008.
- [7] M. Chen, S. Liew, Z. Shao, and C. Kai. Markov approximation for combinatorial network optimization. In *Proc. of IEEE INFOCOM*, 2010.
- [8] M. Chen, M. Ponc, S. Sengupta, J. Li, and P. Chou. Utility maximization in peer-to-peer systems. In *Proc. ACM SIGMETRICS*, 2008.
- [9] X. Hei, C. Liang, and et al. A measurement study of a large-scale P2P IPTV system. *IEEE Transactions on Multimedia*, 9(8), 2007.
- [10] C. Huang, J. Li, and K. Ross. Can Internet video-on-Demand be profitable? In *Proc. of ACM SIGCOMM*, 2007.
- [11] Y. Huang, T. Fu, D. Chiu, J. Lui, and C. Huang. Challenges, design and analysis of a large-scale p2p-vod system. 2008.
- [12] K. Wang and C. Lin. Insight into the P2P-VoD system: performance modeling and analysis. In *Proc. of ICCCN*, 2009.

- [13] J. Wang, C. Huang, and J. Li. On ISP-friendly rate allocation for peer-assisted VoD. In *Proc. of ACM Multimedia*, pages 279–288, 2008.
- [14] J. Wang and K. Ramchandran. Enhancing peer-to-peer live multicast quality using helpers. In *Proc. of ICIP*, pages 2300–2303, 2008.
- [15] C. Wu and B. Li. On meeting P2P streaming bandwidth demand with limited supplies. In *Proc. of SPIE/ACM International Conference on Multimedia Computing and Networking*, 2008.
- [16] D. Wu, Y. Liu, and K. Ross. Queuing network models for multi-channel p2p live streaming systems. In *Proceedings of IEEE Infocom*, 2009.
- [17] Y. He, and L. Guan. Improving the streaming capacity in p2p vod systems with helpers. In *Proceedings of the 2009 IEEE international conference on Multimedia and Expo*. Institute of Electrical and Electronics Engineers Inc., 2009.
- [18] H. Zhang, J. Wang, M. Chen, and K. Ramchandran. Scaling peer-to-peer video-on-demand systems using helpers. In *ICIP*, 2009.
- [19] S. Annapureddy, C. Gkantsidis, P. Rodriguez, and L. Massoulie. Providing video-on-demand using peer-to-peer networks. In *Proc. Internet Protocol TeleVision (IPTV) workshop*, 2006.
- [20] J. LaSalle. Some extensions of Liapunov's second method. *Circuit Theory, IRE Transactions on*, 7(4):520–527, 1960.
- [21] B. Cohen. Incentives build robustness in BitTorrent. Workshop on Economics of Peer-to-Peer systems, Volum 6, Citeseer, 2003.