Service Routing in Multi-ISP Peer-to-Peer Content Distribution: Local or Remote?

Parimal Parag[1], Srinivas Shakkottai[1], and Ishai Menache[2]

[1] Department of Electrical and Computer Engineering
Texas A & M University, College Station, TX 77843, USA
{parimal,sshakkot}@tamu.edu,
[2] Department of Electrical and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
t-ismena@microsoft.com

Abstract. The popularity of Peer-to-Peer (P2P) file sharing has resulted in large flows between different ISPs, which imposes significant transit fees on the ISPs in whose domains the communicating peers are located. The fundamental tradeoff faced by a peer-swarm is between free, yet delayed content exchange between intra-domain peers, and inter-domain communication of content, which results in transit fees. This dilemma is complex, since peers who possess the content dynamically increase the content capacity of the ISP domain to which they belong.
In this paper, we study the decision problem faced by peer swarms as a *routing-in-time* problem with *time-varying capacity*. We begin with a system of two swarms, each belonging to a different ISP: One swarm that has excess service capacity (a *steady-state* swarm) and one that does not (a *transient* swarm). We propose an asymptotically accurate fluid-approximation for the stochastic system, and explicitly obtain the optimal policy for the transient swarm in the fluid regime.
We then consider the more complex case where multiple transient swarms compete for service from a single steady-state swarm. We utilize a proportional-fairness mechanism for allocating capacity between swarms, and study its performance as a non-cooperative game. We characterize the resulting Nash equilibrium, and study its efficiency both analytically and numerically. Our results indicate that while efficiency loss incurs due to selfish decision-making, the actual Price of Anarchy (PoA) remains bounded even for a large number of competing swarms.

Key words: peer-to-peer, overlay network, game theory, price of anarchy

1 Introduction

Recent trends suggest estimate that 35-90% of bandwidth is consumed by peer-to-peer (P2P) file-sharing applications [1–3]. While there has been some drop in the fraction of P2P traffic for file distribution [4], there has been increased use of P2P for video streaming in systems such as PPLive [5] and QQLive [6]. Thus,

P2P applications are likely to increase in number as they prove to be a relatively cheap means of media distribution from a content distributor's perspective.

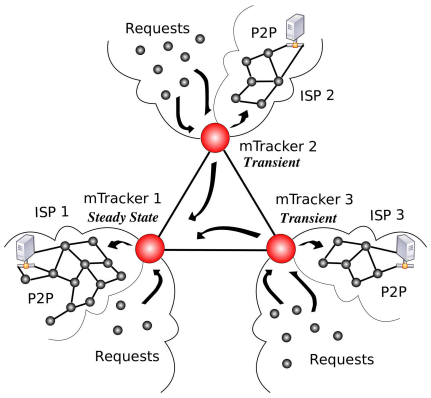


Fig. 1. The MultiTrack architecture described in [7]. The system uses multiple BitTorrent Trackers to achieve a desired delay-tariff tradeoff appropriate to the system state at that time instant. In this paper, we explicitly account for the dynamics of the P2P swarm capacities as a function of time.

P2P networks attempt to keep delays small by leveraging as much end-user bandwidth and storage as possible. However, they are often oblivious of the transit tariff that they impose on the hosting Internet Service Providers (ISPs) due to such optimizations. The pricing architecture of the Internet is tiered, wherein a lower tier ISP must pay a higher tier ISP (from which it obtains service) a tariff for all traffic entering or exiting its domain [8]. Since such lower tier ISPs are usually the ones that provide Internet connectivity to end-users, P2P communication between end-users across ISP domains causes significant tariffs for both of the terminal ISPs. Thus, there appears to be an implicit conflict between P2P applications that seek to find appropriate (lowest delay) peers regardless of the ISP domain in which such a peer might be located, and ISPs that seek to keep traffic localized within their domains. Such conflict has led to efforts by some ISPs to restrict P2P traffic [9].

The most popular P2P system nowadays is BitTorrent [10], which uses a system of *Trackers* to enable peers to find each other. When a peer without the content (called a *leech*) enters the system, it obtains a list of peers that it can communicate with from such a Tracker. The set of peers that is controlled by the Tracker in this fashion is known as the *peer swarm* (or P2P swarm) associated with that Tracker. While the original BitTorrent Tracker is ISP agnostic, Figure 1 depicts a system called MultiTrack described in [7] in which each ISP domain has a different P2P swarm for the same piece of content, with admission to each one being controlled by a separate mTracker. While a single seed is enough for whole swarm to receive the piece of content using P2P methods, it may take a long time to do so. Hence, the mTracker controlling an overloaded swarm (transient swarm) could request service from an mTracker controlling a swarm that has spare capacity (steady-state swarm). This leads to a natural tradeoff

between minimization of delay versus transit tariff. However, [7] performs an instantaneous optimization and does not model the phenomenon that when a peer obtains the content it becomes a *seed server*, effectively increasing the peer-swarm capacity with time.

The dynamic evolution of the peer-swarm capacity results in a problem of "routing-in-time", where it is required to route traffic in a system of *time-varying capacity* so as to minimize costs. However, unlike a general routing-in-time problem, there are predictive models that describe the evolution of the swarm capacity [11–13], which are employed in the present paper. Since each peer is hosted within a particular ISP domain, once served, that peer could become a seed server in its host domain. Hence, each mTracker must take a decision on whether to keep a request local (and potentially incur a delay cost) or to forward it to a different ISP domain (and incur a transit tariff) with the knowledge that a request forwarded to a different domain at a particular time instant could result in a new local seed once that request has been satisfied. In this paper, we will consider two important questions pertaining to routing in P2P swarms:

- What is the appropriate routing in time profile that would minimize delay plus tariff costs in a transient swarm?
- Assuming there are multiple transient swarms competing for capacity available at a steady-state swarm, how should be capacity divided, and what are the consequences of non-cooperative competition for capacity?

1.1 Related Work

While the original BitTorrent Tracker was ISP agnostic, there have been several attempts to enable Trackers to become ISP-aware. Papers such as [7, 14–16] seek for the right tradeoff between delay costs and transit tariffs by attempting to keep traffic local whenever possible. For example, [15] suggests to optimize for minimum tariffs, and then develops heuristics for allowing a certain fraction of peers to be non-local so as to ensure that delays are not excessively large. In [7] the objective is to design a distributed control scheme called MultiTrack, that would achieve a desired delay-tariff tradeoff in a distributed fashion. However, as stated above, none of these references considers the evolution of peer-swarm capacity as a function of time.

We implicitly make the assumption that the newly created seeds would be willing to serve content to the leeches in their ISP domain. Incentives for such a seeds to share content may be provided by trading a local currency in exchange for files, see, e.g., [17, 18]. The objective in this paper is to optimize over the predicted capacity of P2P swarms, assuming that a mechanism for such trade is in place.

1.2 Contribution and Content

In this paper, we study a P2P model consisting of multiple ISP domains, each associated with a Tracker and a peer swarm that has N peers. Each swarm

consists of peers that possess the content (seeds) and leeches, who may become seeds over time. We assume that one of the ISPs has a peer swarm that is in steady-state, meaning that its offered load is lower than the available content-capacity. Such peer swarm can act as a content server to boost the performance of those swarms that have not yet reached steady state. However, a transit tariff must be paid for access to remote swarms, and Trackers must take decisions on whether to request remote service or keep traffic local.

In Section 2, we develop a general stochastic model for a system with a single transient swarm. We propose a deterministic fluid model which is amenable for analysis, and has a similar behavior as the stochastic model as the swarm size N becomes large. In Section 3, we use the fluid model to show the optimality of an intuitive remote service profile, which is to request the entire capacity from the steady state swarm until a "stopping time", after which no more service should be requested. We obtain an explicit expression for this stopping time. We demonstrate through simulations (Section 4) that the optimal routing policy obtained for the fluid model is near-optimal with respect to the original stochastic model. In Section 5, we adopt a proportional-fairness like mechanism for dividing the capacity of a steady state swarm between multiple transient swarm. Such a mechanism naturally gives rise to a non-cooperative game between the swarms, each of which selfishly decides on a bid and a stopping time. We analyze the resulting game and provide bounds on the "Price of Anarchy" under symmetry assumptions. We supplement the analysis of the non-cooperative game with simulations in Section 6, which indicate that the experienced efficiency loss in terms of overall user cost is not more than 30%, and much less for a small number of swarms.

2 Single Transient Swarm

We first consider the case of a single transient swarm with N peers interested in a certain piece of content, and a single steady-state swarm that has a total upload capacity of NC distributed among N seeds. Thus, C is the maximum "per-requester capacity" available from the steady-state swarm. The transient swarm can make a time-varying request for service $C(t)$ from the steady state swarm. There are two ways in which the leeches in the transient swarm can obtain the content of interest. First, they contact other peers in the transient swarm uniformly and at random, and if the contacted peer has the file (i.e., it is a seed) the content may be downloaded. Alternatively, based on the choice of $C(t)$, seeds in the steady-state swarm are directed by the Tracker controlling their swarm to contact the Tracker controlling the transient swarm. This Tracker selects a peer that does not possess the content (i.e., a leech) from the transient swarm, and the seed from the steady-state swarm uploads the content to this peer. Below, we will develop a stochastic model corresponding to above dynamics, and then simplify it to a deterministic fluid model.

2.1 Stochastic System Model

We can think of the transient swarm as a graph $G = (V, E)$, with vertex set V corresponding to the set of peers, and edge set E corresponding to a communication link between two peers. By assumption, we have $N = |V|$. For simplicity of presentation, we consider a fully-connected graph G. However, our conclusions can easily be generalized to ensemble of randomly chosen K-regular graphs. Let $\mathcal{P}(t) \subseteq V$ be the set of seeds in the transient swarm at some time t, such that there are $P(t) = |\mathcal{P}(t)|$ seeds.

Assumption 1 *We assume that the capacity requested by the transient-swarm is piece-wise constant with time. Thus, time is divided into* phases $j \in \mathbb{N}_0$ *with*

$$C(t) = C_j \quad \forall t \in [T_{j-1}, T_j), \text{ where } T_{-1} = 0 \text{ and } C_j \leq C \text{ for all } j.$$

We denote the time spent in phase i as $\tau_i = T_i - T_{i-1}$. We will see later that our results hold for a general $C(t)$ as well. Leeches in the transient swarm contact each other uniformly and at random and download the file at rate η if it is available with the contacted peer. Also, in the j^{th} phase the seeds in the steady state swarm are randomly directed to serve one of the leeches in the transient swarm with an upload rate C_j each. We create a tractable stochastic model by making the following simplifying assumptions.

Assumption 2 *We assume that each leech in the transient swarm is equipped with a clock, which ticks at a random interval that is an independent and exponentially distributed random variable X_k, with mean η^{-1} for peer $k \notin \mathcal{P}(t)$ in the transient swarm. Similarly each seed l in the steady-state swarm $l \in \{1, 2, \ldots, N\}$ has a clock that ticks at times that are denoted by exponential i.i.d. random variables Y_l with mean C_j^{-1} in the j^{th} phase.*

When its clock ticks, each leech in the transient swarm contacts a neighbor uniformly at random. If the contacted peer k happens to be a seed (i.e., $k \in P(t)$), the leech downloads the content. When the clock of a seed in the steady state swarm ticks, it contacts its Tracker that directs it (via the Tracker controlling the transient swarm) to one of the leeches in the transient swarm, and uploads the content. A final assumption completes the model.

Assumption 3 *We assume that content is downloaded instantaneously.*

The assumption implies that an if a leech contacts a peer without the file of interest, it has to wait for a random interval of time before trying again. This allows for load balancing at seeds, since the number of leeches contacting each seed would be finite with high probability. Our approach provides a lower bound on system performance. The following lemma characterizes the evolution of the number of seeds as a continuous-time Markov chain (CTMC).

Proposition 1. *Let $P(t)$ be the number of seeds at any time t, in the single transient swarm model described above, that satisfies Assumptions 1, 2, and 3.*

If we assume that phase changes can occur only at the instants when $P(t)$ increases, then the number of seeds $P(t)$ evolves according to a CTMC as depicted in Figure 2, where the state-dependent jump-rate in phase j is

$$\lambda_{P(t)} = NC_j + \frac{P(t)(N - P(t))}{N}\eta, \qquad P(t) \in \{1, \ldots, N\}.$$

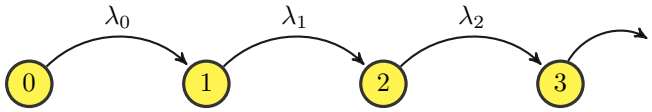


Fig. 2. Continuous time Markov chain governing the evolution of the number of seeds in the transient swarm.

We show that as the peer-population N grows large, the evolution of the number of seeds $P(t)$ in the transient swarm can be modeled deterministically by the following differential equation

$$\frac{dP(t)}{dt} = \lambda_{P(t)}. \tag{1}$$

To this end, we show that for large N, the time-interval to increase the number of seeds from m to n in the deterministic model is identical to the corresponding time-interval in the stochastic model with probability one.

Theorem 1. *For a user population N, we denote the time-interval $T_{mn}(N) = t_{n+1} - t_m$ for number of seeds to increase from m to $n+1$ under the stochastic model under consideration. As user population N grows large*

$$\lim_{N\to\infty} T_{mn}(N) = \int_{\frac{m}{N}}^{\frac{n}{N}} \frac{dy}{\eta y(1-y) + C_j} \quad \textit{with probability } 1. \tag{2}$$

Notice (2) *is the integral form of the expected differential equation in* (1)*.*

2.2 Delay Calculation Using the Fluid Model

As before, let $P(t)$ be the number of seeds in the transient swarm at t. From Theorem 1, the evolution of $P(t)$ is given by (1), restated below:

$$\frac{dP}{dt}(t) = \eta\frac{P(t)}{N}\left(N - P(t)\right) + NC(t). \tag{3}$$

We consider the evolution of $P(t)$ until $P(t) = N - 1$, at which point we say that the remaining one leech will be served in constant time. Let $y(t) = P(t)/N$ be the fraction of seeds in the system at time t. We denote the fraction of seeds at the beginning of phase i by $y_{i-1} = y(T_{i-1})$, with the convention $T_{-1} = 0$. We further define fraction $\alpha_i \triangleq (\theta_i - \eta y_{i-1})/(\eta y_{i-1} - \theta_i')$, where $\theta_i \geq \theta_i'$ are the solutions to the quadratic equation $\theta^2 - \eta\theta - \eta C_i = 0$. The explicit evolution of $y(t)$ in time is presented in the following lemma.

Proposition 2. *We can write $y(t)$ in terms of positive difference $\Delta\theta_i = \theta_i - \theta'_i$ as*

$$y(t) = \frac{\theta'_i}{\eta} + \frac{\Delta\theta_i/\eta}{1 + \alpha_i e^{-\Delta\theta_i(t-T_{i-1})}} \qquad t \in [T_{i-1}, T_i). \tag{4}$$

For a finite number of $K+1$ phases, we can compute the total delay seen by all the leeches as the difference in area between curves N and $P(t)$ for the interval $[0, T_K]$, where $P(T_K) = N - 1$[1]. That is, we have $T_0 \leq T_j \leq T_K$ such that $y(T_K) = 1 - 1/N$. The average per-requester delay D seen by leeches is the area between the curve $d(t) = 1$ (which is the demand curve, since the all leeches demand service at time zero) and $y(t)$, (the service curve) which can be expressed in terms of rates $\eta, \{(\theta_j, \theta'_j)\}$ and fractions $\{\alpha_j\}$.

Proposition 3. *The average per-requester delay D_i in phase i can be expressed as*

$$D_i = \frac{\theta'_i \tau_i}{\eta} - \frac{1}{\eta} \ln\left(\frac{\eta y_{i-1} - \theta'_i}{\eta y_i - \theta'_i}\right). \tag{5}$$

The aggregate average per-requester delay in $K+1$ finite phases is $D = \sum_{i=0}^{K} D_i$.

Now that we have expressions for the delay experienced in each phase, we optimize over delay and transit tariff in the following section.

3 Single Swarm Optimization

Let the transit tariff per unit traffic be denoted by p. Since this value is fixed, we can equivalently assume that the Tracker controlling the transient-swarm asks for a per-requester capacity $C(t)$ from the steady-state swarm, at a rate of p per unit capacity. The value of the per-requester capacity must be chosen such that a linear combination f of average per-requester delay D and transit tariff per unit traffic is minimized. That is, we wish to minimize

$$f = D + p \int_{t=0}^{\infty} C(t) dt. \tag{6}$$

Consider the case of piece-wise constant $C(t)$ with two non-zero phases. The problem reduces to the following optimization problem

$$\begin{aligned} &\text{minimize } f(\tau_0, \tau_1) \triangleq D + p\,(C_0\tau_0 + C_1\tau_1) \\ &\text{such that } C_0, C_1 \leq C. \end{aligned} \tag{7}$$

We now present two results that lead us to the intuitively appealing conclusion that remote capacity usage is necessary only in the first phase, and the amount of capacity used during that one phase to should be the maximum possible. We present these results without the proof due to space limitations.

[1] Such a computation of total delay is valid for any work conserving policy.

Lemma 1. *The cost function f associated with the ISP is minimized for problem* (7) *when phase 1 is stopped at time T_1 such that $y_1 = y^*$, where*

$$y^* = \min\left\{\frac{1}{\eta p}, \frac{1}{\eta p'}\right\} \text{ where } p'^{-1} = \eta\left(1 - \frac{1}{N}\right). \tag{8}$$

Lemma 2. *The average per-requester delay D is minimized for problem* (7) *when the phase-interval with the smaller remote service-rate is zero.*

We next characterize the stopping time after which remote capacity usage is detrimental. Proof is omitted due to space constraints.

Theorem 2. *The cost function f associated with supporting a P2P swarm is minimized for the problem* (7)*, if the maximum remote service capacity C available is utilized till an optimum stopping time τ^*, such that $y(\tau^*) = y^*$ defined in* (8)*. In other words, setting remote service-rates as $C_0 = C, C_1 = 0$, and phase-change times as $\tau_0 = \tau^*$ and $\tau_1 = 0$ minimizes f. The optimal stopping-time τ^* in terms of ϕ, ϕ', such that $\phi + \phi' = \eta$ and $\phi\phi' = -\eta C$, is*

$$\tau^* = \frac{1}{\Delta\phi} \ln\left(\alpha \frac{(\eta y^* - \phi')}{(\phi - \eta y^*)}\right), \text{ where } \Delta\phi = \phi - \phi'. \tag{9}$$

We can also characterize the optimal per-user delay

$$D = \frac{\phi'\tau^*}{\eta} + \frac{1}{\eta} \ln\left(1 - \frac{\eta y^*}{\phi'}\right) - \frac{1}{\eta} \ln\left(\eta y^* p'\right), \tag{10}$$

and the associated cost function

$$f = D + pC\tau^*. \tag{11}$$

Finally, we have the following corollary that shows that our restriction to piece-wise constant functions is actually not binding.

Corollary 1. *Let T be such that $y(T) = 1 - 1/N$. Then, the minimizer function in $\mathcal{C} = \{C(t), t \in [0,T] : C(t)$ simple $, 0 \le C(t) \le C, \forall t \in [0,T]\}$ for the following optimization problem*

$$\text{minimize } f = D + p\int_0^T C(t)dt \text{ such that } C(t) \in \mathcal{C},$$

is the following function $C(t) = C\chi_{[0,\tau^*]}(t)$ where τ^* is defined in (9)*.*

Proof. It follows from induction using Lemma 1, Lemma 2, and Theorem 2.

Given a price p for service from remote swarm, one can find the optimal stopping time τ^*. It can quickly be seen that this stopping time is a non-increasing function of price p. In fact, it stays constant for $p \le p'$ and starts decreasing when $p > p'$. We are also interested in finding how the total cost per-user for the

transient swarm increases with the price p for service from the remote swarm. It follows from equation (8), (10), and (11), that

$$\frac{df}{dp} = \begin{cases} C\tau^*(p') & p \le p' \\ \frac{Cy^*}{\eta y^*(1-y^*)+C} + C\tau^*(p) & p > p' \end{cases}. \tag{12}$$

Therefore, it is clear that the total cost per-user in the transient swarm is concave increasing in the remote service usage price p.

In conclusion, we have shown in this section that in order to minimize the total cost, a transient swarm should utilize both local P2P dissemination as well as all the capacity available from a remote steady state swarm up to a stopping time, after which the transient swarm has enough seeds that the correct decision is to not utilize the remote capacity. We found an explicit characterization of this stopping time, whose value has the intuitive property of being non increasing in transit tariff.

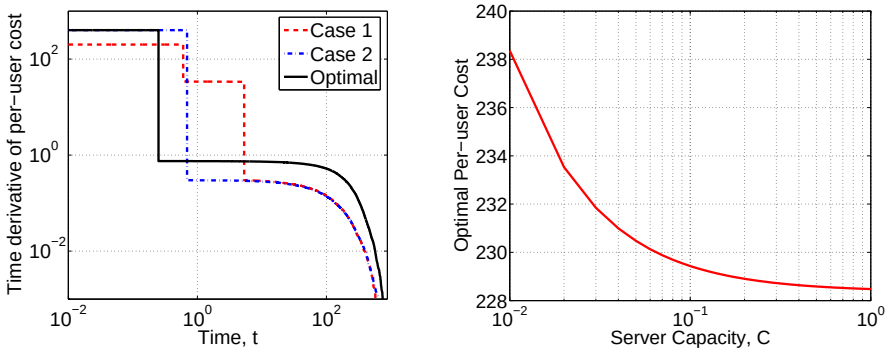
4 Stochastic Simulations of the Single Swarm

We perform simulations of the Markov chain described in Section 2.1 to illustrate the nature of the analytical results derived above. For the optimal case, we calculate the optimal stopping time based on the derivations relating to the fluid model of the previous section, but apply the resulting policy to the stochastic system described in Section 2.1. In our simulations, we took the population of peers in ISP 1 to be $N = 10^4$. The number is realistic for many P2P scenarios. Similar results hold for smaller values of N. However, the differential equations approximate the stochastic system model better with increasing number of peers. The capacity per-seed at the steady-state swarm is assumed to be $C = 1$, and the upload rate for peers in the transient swarm is $\eta = 0.01$ units. We consider the case where the transit tariff per-unit traffic follows $p = 4p'$, where $p' = \frac{N}{\eta(N-1)}$ as defined in Lemma 1 and its numerical value is 100.01 in current setting.

Our first objective is to illustrate that our optimal stopping time policy yields a lower cost than other policies. Hence, we create two arbitrary policies that are described below. We run the simulation until the transient swarm has $N - 1$ seeds. We can compute the time derivative $\frac{df}{dt}(t)$ of the per-user cost function f from the definition of per-user delay and cost function (11). Therefore, we have

$$\frac{df}{dt}(t) = 1 - y(t) + pC(t). \tag{13}$$

We plot this time derivative $\frac{df}{dt}(t)$ in Figure 3(a) for three different cases. In all cases, the area enclosed between the curves and the axes is the total cost of the system. The dashed curve denotes the first case, where requests are served in two phases by the steady-state swarm. In the zeroth phase, the remote service capacity is $C_0 = C/2$ and the phase ends when a three-tenth of the population has the desired content, i.e. $y_0 = 0.3$. In the first phase, server-rate is $C_1 = C/12$



(a) Cost evolution for different policies. (b) Per-user optimal costs.

Fig. 3. Left figure shows cost evolution for optimal vs. arbitrary stopping time. The area between curves is the total system cost. Right figure shows the optimal cost for different server capacities.

and the phase ends when seven-tenth of the user population has the desired content, i.e. $y_1 = 0.7$. The dotted curve denotes the case when there is a single phase where maximum available server-capacity C is used by the requesting swarm. However, in this case, the stopping time is not optimal, and the phase ends when seven-tenth of the user population has the desired content. The solid curve denotes the time-derivative of cost for the optimal case. Note that since the scale is log-log to illustrate the differences in the curves, the large difference between the optimal area and suboptimal area is not prominently visible.

In Figure 3(b), we plot the minimum per-user cost f for ISP 1 as a function of available per-requester server-capacity C, for the same parameters for user-population, server-capacity, upload rate, and tariff per-unit traffic. Essentially, we plot the equivalent of the fluid equation (1) in the stochastic setting when y^* has been chosen optimally as a function of constant C. The available per-requester server-capacity C takes values from 0 to 1. Clearly, the minimum per-user cost decreases and the rate of decay decreases with C as expected from the proof of Theorem 2.

We conclude that the optimal policy in the fluid system is indeed optimal in the original stochastic system, and that the parameters τ^*, y^* that can be calculated from the optimal deterministic policy are essentially optimal in the stochastic case.

5 Multiple Swarms: Collaboration or Competition?

We now consider the case of Q transient P2P swarms, each controlled by a distinct Tracker $i \in \{1, 2, \ldots, Q\}$. There is also a single steady state swarm indexed by 0 from which all the transient swarms attempt to obtain service. Thus, we have swarm 0 in steady state with total available capacity NC, while every other swarm $i \in \{1, 2, \ldots, Q\}$ starts with 0 initial service capacity and

number of leeches N. Suppose that the transit tariff to reach the steady state swarm is p for all the transient swarms (this value could also be 0). In addition to this tariff, each transient swarm i bids a value p_i indicating its willingness to pay the steady state swarm for service. The steady state swarm must use some mechanism to decide how much of its capacity to allocate to each transient swarm.

5.1 Proportional-Fairness Mechanism

We propose to use the proportional-fairness mechanism for the steady state swarm to allocate capacity amongst the transient swarms. Under this mechanism the allocation to each transient swarm i is given by

$$\textbf{Mechanism:} \quad C_i = \frac{p_i C}{\sum_{j=1}^{k} p_j}. \tag{14}$$

The mechanism is very simple to implement, and has been successfully used in communication networks for apportioning bandwidth to competing flows [19]. Further, it has been shown to have a bounded inefficiency even with strategic users that optimize against the mechanism [20]. It is therefore, a good candidate for our system of competing transient swarms.

The Tracker associated with a transient swarm i can utilize the capacity allocated to it for any duration that it chooses, and pays $p_i + p$ per unit traffic during that time. We assume that once a certain amount of capacity has been allocated to a transient swarm i, it cannot be withdrawn and reallocated to some other swarm. Such a scheme is consistent with the idea developed in the previous section that the capacity from the transient swarm is most useful during the initial stage, and also simplifies our analysis.

5.2 Definition of the Game

We utilize the same fluid approximation developed in Section 2.1 to describe the dynamics of each swarm. The cost function f_i associated with swarm i for the multiple ISP scenario can be expressed in terms of the fixed transit tariff p, the bid p_i, the per-requester delay D_i, and allocated capacity C_i as

$$f_i(p_i, p_{-i}) = D_i + (p + p_i) C_i \tau_i. \tag{15}$$

Note that in general f_i is also a function of the upload capacity η_i of peers in swarm i. We can then define a strategic game $\mathcal{G} =< \mathcal{Q}, \mathcal{P}, \mathcal{F} >$, where $\mathcal{Q}$ is the set of Trackers (players), $\mathcal{P}$ is the set of bid profiles (action sets) and $\mathcal{F} = \{f_1, f_2, \cdots, f_Q\}$.

Our first objective is to find the socially optimal way of bidding when all Trackers collaborate. The objective here is to minimize the sum of the costs incurred by all swarms. Secondly, we also wish to compute what the cost is if Trackers are selfish and act individually and rationally to arrive at a bid decision.

In the following, we follow the notation $p_{-i} = \sum_{j \neq i} p_j$. We will only analyze the symmetric case where $\eta_i = \eta$ for all swarms $i \in \{1, 2, \ldots, Q\}$, and consider the scenario where the transit tariff p is lower bounded by $p \geq p' = N/(\eta(N-1))$. Therefore, by Theorem 2, the total available capacity is used only till each ISP reaches the fraction $y_i = \frac{1}{\eta(p+p_i)}$.

5.3 Collaborative Scenario

We consider cooperating Trackers, who wish to jointly minimize the aggregate cost. Consider a set of bids $P = \{p_i : i = 1, 2, \ldots, Q\}$. Then the problem that the Trackers wish to solve is

$$\textbf{Opt:} \quad \min_{p_i \geq 0} \sum_{i=1}^{Q} f_i(P). \tag{16}$$

Theorem 3. *For collaborative scenario, optimal set of bids is* $P^* = \{0, 0, \ldots, 0\}$.

5.4 Multiplayer Game

We now consider the non-cooperative situation, where every Tracker acts according to its own self interest. We assume each Tracker makes a rational decision, assuming every other Tracker does the same. We also assume that each bid is made without knowledge of any other Tracker's bid. In this setting, we wish to find the Nash equilibrium (if it exists) of the bid p_i made by each Tracker i. Hence, each Tracker wishes to solve the following problem.

$$\textbf{Game:} \quad \min_{p_i \geq 0} f_i(p) \qquad \forall i \in \{1, 2, \ldots, Q\}. \tag{17}$$

The following theorem provides the necessary conditions for the existence of a symmetric Nash equilibrium of bids for this non-cooperative strategic game. We omit the proof in the interest of space.

Theorem 4. *If the number of competing swarms* Q *is such that* $\eta < 2C/Q$ *then the strategic game* $\mathcal{G}$ *has a pure strategy Nash Equilibrium of set of bids* $P = \{\beta, \beta, \ldots, \beta\}$.

Since the NE exists, we identify a pure strategy NE in terms of a common bid β made by all the Trackers. Note, that $h(\beta) = 0$ and τ_i is a logarithmic function of p_i (see (9)). We identify upper and lower bounds on the pure, symmetric bid in the following theorem stated without the proof.

Theorem 5. *The optimal bid* β *for the strategic game* $\mathcal{G}$ *is bounded above and below by the following values*

$$\frac{Q-1}{\eta Q + 4C} \leq \beta \leq \frac{Q-1}{\eta Q + 4C} \left(\frac{1 + \frac{p\eta(\eta+2C_i)}{4C_i}}{1 - \frac{(Q-1)\eta(\eta+2C_i)}{4C(\eta+4C_i)}} \right), \tag{18}$$

under the condition $\eta < 2C/Q$.

It is clear that it is difficult to analytically compute Nash equilibrium β. Therefore, we make an approximation to get some insight into how the optimal bid changes with the number of transient swarms competing for the available capacity at the steady-state swarm. In the regime $\eta \ll C_i$, we have $y_i \approx C_i\tau_i$. Under this approximation, from Theorem 5 we obtain

$$\beta \approx \frac{Q-1}{\eta Q + 4C} \qquad \forall i \in \{1, \ldots, Q\}. \tag{19}$$

Notice that as the number of competing Trackers go up, so does the bid at Nash equilibrium. Such an increased bid is the price paid for uncoordinated behavior by the different Trackers. Our approximation is a lower bound on the Nash equilibrium bid. Also, it is clear from (18) that the approximation error is small when $\eta \ll C_i$ and $p(\eta + 2C_i) \ll 1$.

5.5 The Price of Anarchy

In most work on selfish decision making, it is found that individual optimization has a negative impact on the total value of a system. We observed that the lack of coordination results in a bid that is linearly increasing in number of ISPs Q. How different would the system cost be in such a scenario? Note that since are dealing with costs, a larger PoA is worse. Due to symmetry in the problem, each Tracker i bids the same value and receives a total service-rate NC/Q from the steady-state swarm. For the optimal case, when the players collaborate, this bid is 0 and the per-user cost for each transient swarm is f_i^{opt}. When the players are selfish, they bid value β additional to the base price p as tariff per-unit traffic.

Following terminology from [21], we define the "price of anarchy" as

$$\text{PoA} \triangleq \frac{\sum_{j=1}^{Q} f_i^{\text{game}}}{\sum_{j=1}^{Q} f_i^{\text{opt}}}. \tag{20}$$

However, unlike most work on the price of anarchy, we are less interested in the regime where the number of players is large. In other words, $Q \to \infty$ is less interesting to us since the number of peer swarms simultaneously competing for capacity is likely to be fairly small, although each swarm might have many thousand peers. Thus, our primary focus will be on obtaining good bounds on the PoA for relatively small values of Q, and in this regime we have the following theorem. Proof is omitted due to space constraints.

Theorem 6. *The price of anarchy (PoA) for strategic game $\mathcal{G}$ is bounded above by the following*

$$1 + \frac{2\ln\left(1+\frac{\beta}{p}\right)}{\eta f_i(p') + \frac{8C_i}{(4C_i+\eta)}\ln\left(\frac{p}{p'}\right)}$$

where β is the Nash Equilibrium bid for the game.

Therefore, if number of Trackers Q is small enough, such that $\eta \ll C_i$, then we can approximate PoA to be

$$\text{PoA} \approx 1 + \frac{2\ln\left(1+\frac{\beta}{p}\right)}{\eta f_i(p') + 2\ln\left(\frac{p}{p'}\right)}. \tag{21}$$

Notice that if we use the approximate value of β given in (19), the upper bound above is a function of known parameters of the system. Such a form is appealing since it a simple upper bound on the PoA. In the next section, we will numerically solve the game, and compare the actual price of anarchy to the bound derived above.

6 Numerical Studies of the Game

For numerical studies, we considered the symmetric case where each transient swarm has the same user population $N = 10^4$, and upload rate of seeds $\eta = .01$. The available per-user capacity at the steady-state swarm was taken to be 1. We considered the base price $p = p'$ where $p' = 100.01$ as in Section 4. We varied the number of transient swarms from 2 to 50.We plot the bid at Nash equilibrium along with the upper and lower bounds in the Figure 4. Since $\eta < 2C_i$ for the chosen values of Q, the bounds are reasonable. It must be noted that the bounds get worse with increase in number of swarms and they do not hold when number of transient swarms become large. Figure 5(a) shows the actual

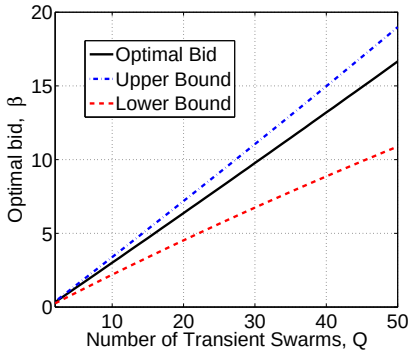
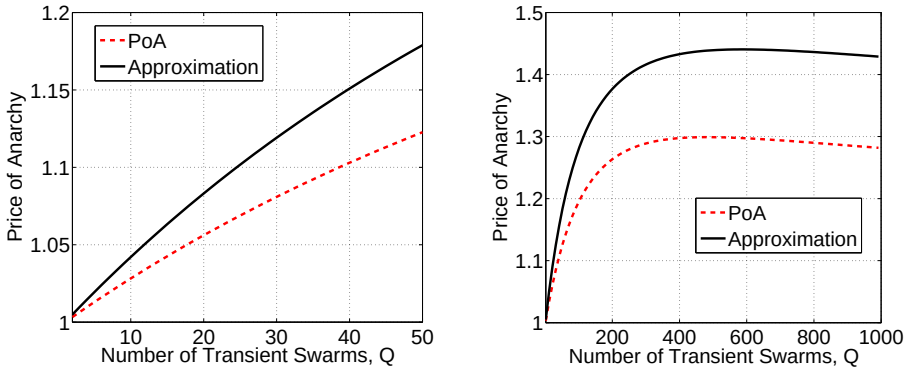


Fig. 4. Optimal bid in the symmetric case, and its upper and lower bounds.

PoA computed numerically, and its approximation computed analytically. As expected, the PoA grows with number of transient swarms. However, when Q becomes large, we are no longer in the regime $\eta < 2C_i$. We see in Figure 5(b) that PoA remains bounded even when number of swarms become large. Further, our approximation of the PoA remains bounded as well. While we do not expect the number of competing swarms to be this large in reality, it is interesting the PoA is at most about 30% even for large Q.



(a) PoA and approximation. (b) PoA with increasing swarm size.

Fig. 5. Illustration of the price of anarchy for different swarm sizes. As the swarm size increases, it remains bounded.

7 Conclusion

We studied in this paper the basic dilemma faced by any content distributor that wishes to utilize the inherent capacity scaling effects of P2P networks, but also does not want to impose excessive transit tariffs on the ISPs hosting the peers. We showed that since a P2P network has a capacity that scales as the number of users served, the greatest gain for usage of the steady-state swarm is in the initial phase, with the duration of usage that depends on the transit tariff. We also considered the case of multiple ISPs competing for capacity, and showed that while the resulting equilibrium is suboptimal, performance is adequate. We believe that besides the specific results, the model proposed in this paper can be used for more complicated P2P interactions that we will explore in the future..

Acknowledgments

Research was funded in part by NSF grants CNS 0904520 and CNS 0963818, the Google Research Awards program, and Qatar Telecom, Doha, Qatar.

References

1. C. Fraleigh, S. Moon, B. Lyle, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level traffic measurements from the Sprint IP backbone," *IEEE Network Magazine*, vol. 17, no. 6, pp. 6–16, 2003.
2. K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," in *Proc. SOSP*, Ocober 2003.
3. E. Bangeman, "P2P responsible for as much as 90 percent of all 'Net traffic," *ArsTechnica*, Sept. 3, 2007.

4. C. Labovitz, D. McPherson, and S. Iekel-Johnson, "2009 Internet Observatory report," in *NANOG-47*, October 2009.
5. "PPLive," http://www.pplive.com/, 2009.
6. "QQLive," http://www.qqlive.com/, 2009.
7. V. Reddy, Y. Kim, S. Shakkottai, and A. L. N. Reddy, "MultiTrack: A Delay and Cost Aware P2P Overlay Architecture," in *ACM SIGMETRICS as a poster*, Seattle, WA, Jun. 2009.
8. G. Huston, *ISP Survival Guide: Stratagies for Running a Competitive ISP*. John Wiley and Sons, New York, 1999.
9. A. Broache, "FCC chief grills Comcast on BitTorrent blocking," *C|Net News.com*, Feb. 25, 2008.
10. "BitTorrent," http://www.bittorrent.com/, 2005.
11. X. Yang and G. de Veciana, "Performance of Peer-to-Peer Networks: Service Capacity and Role of Resource Sharing Policies," *Performance Evaluation: Special Issue on Performance Modeling and Evaluation of P2P Computing Systems*, vol. 63, 2006.
12. D. Qiu and R. Srikant, "Modeling and performance analysis of BitTorrent-like peer-to-peer networks," in *Proc. ACM SIGCOMM*, Portland, Oregon, USA, August 2004.
13. S. Shakkottai and R. Johari, "Demand Aware Content Distribution on the Internet," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 476–489, April 2010.
14. V. Aggarwal, A. Feldmann, and C. Scheideler, "Can ISPs and P2P users cooperate for improved performance?" *ACM Computer Communication Review*, vol. 37, no. 3, Jul. 2007.
15. H. Xie, Y. R. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz, "P4P: Portal for P2P applications," in *Proc. ACM SIGCOMM*, Aug. 2008.
16. D. R. Choffnes and F. Bustamante, "Taming the torrent: A practical approach to reducing cross-ISP traffic in p2p systems," in *Proc. ACM SIGCOMM*, Seattle, WA, Aug. 2008.
17. C. Aperjis, M. J. Freedman, and R. Johari, "Peer-assisted content distribution with prices," in *Proc. 4th ACM SIGCOMM Conference on emerging Networking EXperiments and Technologies (CoNext 08)*, Dec. 2008.
18. D. S. Menasché, L. Massoulié, and D. F. Towsley, "Reciprocity and barter in peer-to-peer systems." in *INFOCOM'10*, San Diego, CA, 2010.
19. F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.
20. R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Mathematics of Operations Research*, vol. 29, pp. 407–435, March 2004.
21. T. Roughgarden and E. Tardos, "How bad is selfish routing?" in *IEEE Symposium on Foundations of Computer Science*, 2000, pp. 93–102.