Bandwidth Allocation in Peer-to-Peer File Sharing Networks

Albert Creus-Mir [a], Ramon Casadesus-Masanell [b], Andres Hervas-Drane [c,*]

[a] *Universitat Politècnica de Catalunya, EUETIB. Comte d'Urgell 187, Barcelona 08036 (Spain)*

[b] *Harvard Business School. Morgan Hall 233, Soldiers Field Rd. Boston, MA 02163 (USA)*

[c] *Harvard Business School. Baker Library B20H, Soldiers Field Rd. Boston, MA 02163 (USA)*

Abstract

We present a model of bandwidth allocation in a stylized peer-to-peer file sharing network with s peers (*sharers*) who share files and download from each other and f peers (*freeriders*) who download from sharers but do not contribute files. Assuming that upload bandwidth is scarcer than download bandwidth and efficient allocation, we compute the expected bandwidth obtained by each peer. We show that (i) while the exact formula is complex, $s/(s+f)$ is a good approximation and (ii) sharers (freeriders) obtain bandwidth larger (smaller) than $s/(s+f)$. The paper constitutes a first step towards a general analytical foundation for scarce resource allocation in peer-to-peer file sharing networks.

Key words: Peer-to-Peer, Network formation, Resource allocation, Congestion

1 Introduction

Consider a peer-to-peer (p2p) file sharing network where content offered for download is all of similar value to downloaders. Peers in the p2p network may

* Corresponding author. Phone: (617) 496 5725; Fax: (617) 495 0699.

act as *sharers* (by contributing files for others to download) or as *freeriders* (by downloading from others but not making files available for download). In a world where upload bandwidth is scarce, a natural question arises: how is total upload bandwidth allocated among sharers and freeriders?

In this paper we present a stylized model of a p2p file sharing network to address this question. Our model of p2p assumes that:

(1) Each sharer provides one unit of upload bandwidth;
(2) All peers have at least one unit of download bandwidth capacity;
(3) Every peer connects to one sharer only;
(4) A sharer may *not* connect to herself;
(5) Bandwidth obtained from a sharer is allocated equably amongst all peers connected to her.

We refer to a set of links connecting peers to sharers as an *allocation.* A *stable* allocation is one where no peer can be made strictly better off by connecting to a different sharer. We assume all stable allocations arise in the network with equal probability.

For example, consider a p2p network with three peers only: two sharers and one freerider. In this case, there are two stable allocations (see Figure 1). Notice that in this simple example, sharer 1 *must* download from sharer 2 (and vice versa). Given this, the freerider (peer 3 in the figure) is indifferent between downloading from sharer 1 or from sharer 2. In both cases, the freerider obtains download bandwidth equal to 1/2. Depending on the sharer to which the freerider connects to, sharers may end up with download bandwidth of 1 or 1/2. Under the assumption of equiprobability of stable allocations, sharers wind up with expected bandwidth equal to 3/4 while the freerider gets 1/2 only.

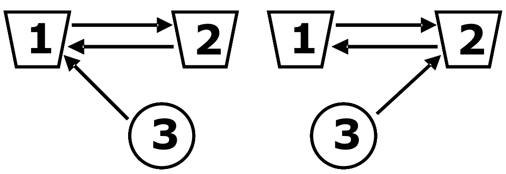


Fig. 1. Stable allocations with two sharers and one freerider. The arrow originating from a given peer indicates the sharer from whom that peer obtains files.

In this paper we consider p2p networks with an arbitrary number of peers and derive an exact formula for the expected bandwidth obtained by sharers and freeriders, where expectations are taken across *all* stable allocations. We show that, just as in the example above, the expected bandwidth obtained by sharers is always larger than that available to freeriders. Sharers can be allocated to fewer sources as they face the constraint of not connecting to themselves.

We also show that #sharers/#peers is a good approximation to the expected bandwidth obtained by both sharers and freeriders. Sharers (freeriders) always obtain expected bandwidth larger than (smaller than) #sharers/#peers. And as the size of the network grows, the difference between expected bandwidth and #sharers/#peers quickly decreases. Already in a network of size 10, the expected bandwidth obtained by sharers and freeriders differs from #sharers/#peers by, at most, 10^{-4}. And when network size is 100, the difference is always less than 10^{-6}. Because the approximation has an exceedingly simple form, it provides a foundation for applied theoretical work on p2p. The exact formula, on the other hand, is discouragingly complex and it is unusable for all practical purposes.

While the present paper is devoted to the study of bandwidth allocation in p2p file sharing networks, we believe our approach is quite general and the results can be extended to other scarce rival resources in p2p networks. The remainder of the paper is organized as follows. Section 2 relates our contribution to the literature. Section 3 states the problem formally. In Section 4 we introduce two reductions and proceed to count all stable allocations. We present the formula in Section 5 and discuss some properties. Section 6 provides graphical representations. In the final section, we present intuition and discuss limitations.

2 Literature

Several empirical studies have examined the topology of p2p file sharing networks, employing different techniques to gather information. Network crawlers are deployed in Asvanund et al. [1], Chu, Labonte and Levine [2], Ripeanu and Foster [3], and in Saroiu, Gummadi and Gribble [4]. While in Gummadi et al. [5] and Sen and Wang [6], p2p traffic is recovered from a trace performed at the physical network's backbone. Together, both approaches have helped characterize the main properties of these networks with respect to peer availability and activity patterns, data traffic paths, and bandwidth bottlenecks.

A parallel literature has focused on constructing theoretical models of p2p file sharing. Contributions have emerged from Computer Science (Buragohain, Agrawal and Suri [7], Feldman et al. [8] and Golle, Leyton-Brown and Mironov [9]) and Economics (Antoniadis, Courcoubetis and Mason [10], Cunningham, Alexander and Adilov [11] and Krishnan et al. [12]). This line of work explores the incentives of peers to contribute resources to p2p networks. Several papers present mechanisms to induce higher resource contribution levels.

This theoretical literature, however, has so far failed to incorporate insights from the structure of the network. With just a few exceptions (see Jian and

MacKie-Mason [13], for instance), the structure of p2p networks is generally not modeled, and, as a result, the resource allocation rules considered lack a foundation. In our view, p2p research can benefit from explicitly considering the properties emerging from the underlying network structure. One main difficulty in pursuing this approach, however, is the general lack of analytical tractability of network games. To this end, the approximation result derived in our analysis can be readily applied in theoretical work.

Casadesus-Masanell and Hervas-Drane [15], for example, use the approximation to construct a model of a p2p file sharing network with endogenous sharing that competes against a for-profit firm that offers content on a client-server architecture at positive prices. Building on the foundation for bandwidth allocation presented here and incorporating the costs associated with file sharing, the authors show that such a model can explain important stylized facts identified in the empirical literature.

3 Setup

Let N, S, and F, be the set of all peers, the set of sharers (peers that contribute bandwidth), and the set of freeriders (peers that do not contribute) respectively. Let n, s, and f, be the cardinalities of these sets. Every peer is either a sharer or a freerider. That is, $F = N \backslash S$ and $f = n - s$. Recall that a *stable* allocation is one where no peer can be made strictly better off by reallocating her to another sharer. Obviously, given N and S there are multiple stable allocations. And whenever the number of peers is not divisible by the number of sharers, the bandwidth obtained from the network will differ across peers.

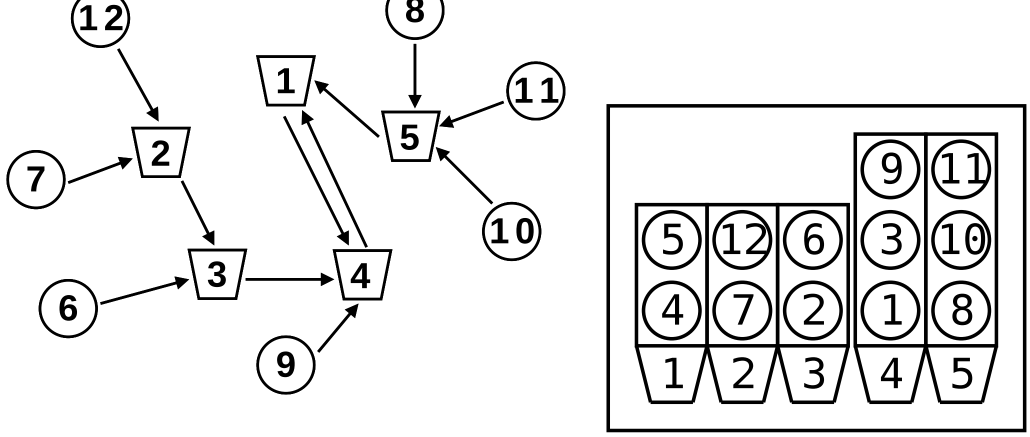


Fig. 2. Two alternative representations of an allocation.

There are two alternative, equivalent ways to represent the p2p networks that we study. First, we can use a network representation such as that of Figure 1. Nodes are peers and arrows indicate the sharer from whom a given peer obtains files. Second, we can use urns and balls. Balls are peers (sharers and freeriders) and there are as many urns as sharers. Balls in an urn are all those peers downloading from a given sharer. Figure 2 presents both representations for a given allocation in a network with $n = 12$ and $s = 5$. In what follows, we use the 'urns and balls' representation as it is more directly related to the mathematical development below.

Let $b := s-(n \bmod s)$, and $a := n \operatorname{div} s$. In every stable allocation we must have a peers allocated to b sharers and the rest $(a+1)$ allocated to the remaining $s-b$ sharers. Therefore, $n = ba+(s-b)(a+1)$. The following figure illustrates these zones.

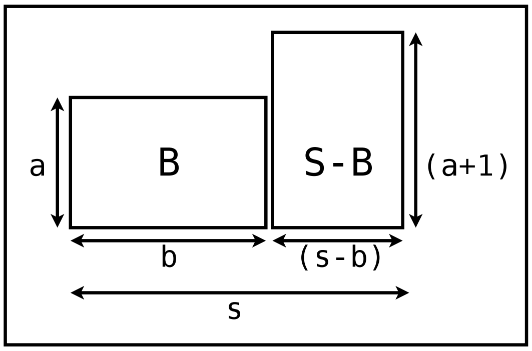


Fig. 3. Relevant zones.

To compute the *expected* bandwidth obtained by freeriders and sharers, we begin by computing the bandwidth obtained by each peer in every stable allocation. Let x be a peer, and let $G(x)$ be the bandwidth obtained by the peer in any given allocation. $G(x)$ is a random variable that can take two values only, $1/(a+1)$ and $1/a$. For any given stable allocation, we define the set B of 'fortunate' peers as $B = \{x|G(x) = 1/a\}$ and the set of 'unfortunate' peers as $S - B = \{x|G(x) = 1/(a+1)\}$. Notice that the specific peers in B and $S - B$ depend on the particular allocation under consideration.

The *expected* bandwidth obtained across all peers, both sharers and freeriders, is $E := E(G(x)) = s/n$. Obviously, $1/(a+1) < s/n \leq 1/a$. We may also consider the conditional expectations $E_S := E(G(x)|x \in S)$ and $E_F := E(G(x)|x \in F)$ (these are the expected bandwidths obtained by sharers and freeriders, respectively). If sharers were to connect to themselves the calculation would be trivial. In this case, symmetry implies that $E_S = E_F = E = s/n$.

As sharers do not connect to themselves in peer-to-peer networks, however, the conditional expectations E_S and E_F turn out to be different. Nevertheless, the total bandwidth available in the network, s, must still be equal to the number of sharers times E_S plus the number of freeriders times E_F. That is, $s/n \cdot E_S + (n-s)/n \cdot E_F = s/n$. Therefore, if we compute E_S, we can immediately obtain E_F.

Because every stable allocation is assumed to be equiprobable, to compute E_S we need to count all of them and compute the average of $G(x)_{x \in S}$ for each. Let H be the total number of stable allocations. Let h_i be the number of these allocations with i sharers in B. Notice that for all these allocations, the average of $G(x)_{x \in S}$ is exactly the same. Therefore, we can consider h_i as the histogram of this value. Then,

$$E_S = \sum_{i=0}^{s} \left(\frac{1}{a} \frac{i}{s} + \frac{1}{a+1} \frac{s-i}{s} \right) \frac{h_i}{H}$$

or

$$E_S = \frac{a + \sum_{i=0}^{s} \frac{i}{s} \frac{h_i}{H}}{a(a+1)}.$$

4 Count

In this section we compute h_i. That is, we orderly enumerate and count all the stable allocations that have i sharers in B. Due to the magnitude of this task, we proceed by reducing and decomposing the problem.

4.1 First reduction

Given n and s, consider the set of all stable allocations. We can divide this set into subsets depending on which sharers support B (having only a peers allocated to them). These subsets are disjoint, and they add up to the entire set of stable allocations, so they are classes. Specifically, there are $\binom{s}{b}$ classes.

Each of these classes has the same number of allocations. To see this, notice that there are bijections between the classes, obtained by changing the names of the elements. Consider class r and let h_i^r be the number of allocations with i sharers in B for class r. Notice that for all r, the value of h_i^r (for every i) is the same. Therefore, by studying one single allocation we can obtain the total $h_i = \binom{s}{b} h_i^r$. Thus, to simplify the analysis, we will study the class in which the sharers supporting B are the first ones. The next figure illustrates the approach.

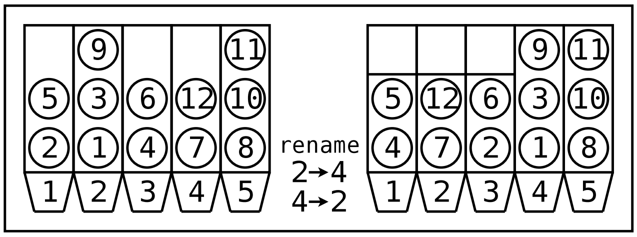


Fig. 4. The first reduction is to rename sharers to get a clearer view.

4.2 Second reduction

Given n and s, consider the set of all stable allocations where B is supported by the first sharers. We now compute how many ways we can assign the n peers between the two zones B and $S - B$.

4.2.1 Counting reduction 2

In general, there are $\binom{n}{ba}$ possibilities. There are three exceptions:

(1) When $b = s$ there is 1 possibility.
(2) When $b = 1$ there are $\binom{n-1}{ba}$ possibilities.
(3) When $b = s - 1$ there are $\binom{n-1}{ba-1}$ possibilities.

These three cases will always be exceptions to the general formula we derive. They are illustrated in the following figure. The last two cases are special because they require that a given sharer be assigned to 'the other' zone.

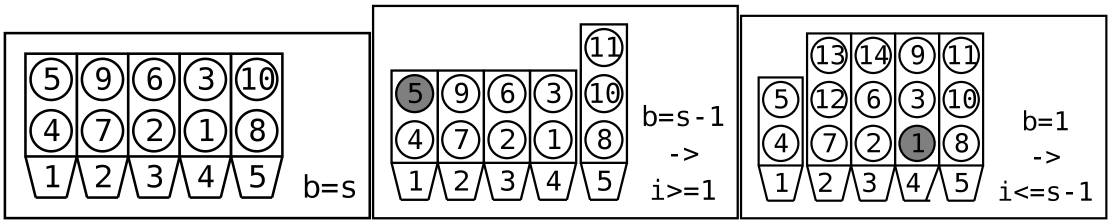


Fig. 5. The three exceptions.

In the general case, $1 < b < s - 1$, there are $\binom{n-s}{ba-i}\binom{s}{i}$ possibilities that have exactly i sharers in B. Using known formulae, adding up we obtain

$$\binom{n}{ba} = \sum_{i=\max\{0,s+ba-n\}}^{\min\{s,ba\}} \binom{n-s}{ba-i}\binom{s}{i},$$

where the limits come from observing the problem and ensuring the combinatorial numbers are well defined.

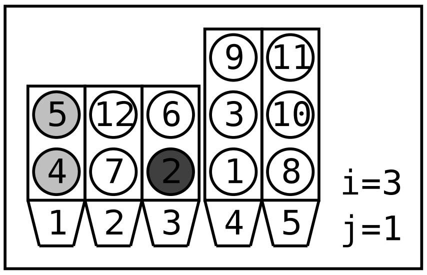


Fig. 6. Freeriders are plotted in white, easy sharers in grey and complicated sharers in black.

As illustrated in the figure above, j sharers out of i may be 'complicated.' A complicated sharer is one that supports B and happens to be in B. We have to take this into account in order to not assign a complicated sharer to herself. An allocation that has a sharer assigned to herself is called a 'coincidence.' There are $\binom{n-s}{ba-i}\binom{s-b}{i-j}\binom{b}{j}$ possibilities that contain exactly j 'complicated' sharers. Again using known formulae, we add up and obtain

$$\binom{n}{ba} = \sum_{i=\max\{0,s+ba-n\}}^{\min\{s,ba\}} \binom{n-s}{ba-i} \sum_{i=\max\{0,b+i-s\}}^{\min\{b,i\}} \binom{s-b}{i-j}\binom{b}{j}.$$

4.2.2 Relationship with the original problem

For every possibility of the reduced problem, we now count how many possibilities exist in the original problem. We proceed by assigning the peers in B to the b sharers that support them.

If there were no 'complicated' sharers, this would be a simple multinomial problem with formula $\frac{(ab)!}{(a!)^b}$. However, because there are j complicated sharers, we perform the same count but need to subtract all those possibilities with coincidences. For example, if $j = 1$,

$$\begin{aligned}\frac{(ab)!}{(a!)^b} - \frac{(ab-1)!}{(a-1)!(a!)^{b-1}} &= \frac{(ab)!}{(a!)^b} - \frac{a}{ab}\frac{(ab)!}{(a!)^b} \\ &= \frac{(ab)!}{(a!)^b}\left(1 - \frac{a}{ab}\right).\end{aligned}$$

If $j = 2$, we subtract twice the cases with one coincidence, but add once the cases with two coincidences,

$$\begin{aligned}
&\frac{(ab)!}{(a!)^b} - 2\frac{(ab-1)!}{(a-1)!^1!(a!)^{b-1}} + 1\frac{(ab-2)!}{(a-1)!^2(a!)^{b-2}} = \\
&= \frac{(ab)!}{(a!)^b}(1 - 2\frac{a^1}{(ab)} + 1\frac{a^2}{(ab)(ab-1)}) = \\
&= \frac{(ab)!}{(a!)^b}\binom{2}{0}\frac{a^0(ab-0)!}{(ab)!} - \frac{(ab)!}{(a!)^b}\binom{2}{1}\frac{a^1(ab-1)!}{(ab)!} + \frac{(ab)!}{(a!)^b}\binom{2}{2}\frac{a^2(ab-2)!}{(ab)!}.
\end{aligned}$$

In this process, we are applying the principle of inclusion-exclusion. The problem is similar to a *derangement*. By analogy, we define the *Multinomial Derangement number*,

$$MD(b,a) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{b}\binom{b}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!},$$

and the *Generalized Multinomial Derangement number*,

$$GMD(b,a,j) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{j}\binom{j}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!}.$$

Notice that $MD(b,a) = GMD(b,a,b)$, and $MD(b,1) = Derangements(b) = !b$.

As we have to take into account both zones, B and $S-B$, for each possibility of the reduced problem there are $GMD(b,a,j)GMD(s-b,a+1,(s-b)-(i-j))$ possibilities of the original problem.

5 Formula

Consider the original problem described in Section 3. Let n be the number of peers, s the number of sharers, $a = n \operatorname{div} s$ and $b = s - (n \bmod s)$.

0) For $b = s$, we have that $E_S = 1/a = s/n$. Otherwise, if $H := \sum_i h_i$,

$$E_S = \frac{a + \sum_{i=0}^{s}\frac{i}{s}\frac{h_i}{H}}{a(a+1)}.$$

1) For $b = 1$, $i \in [\max\{0, s+ba-n\}, \min\{s-1, ba\}]$.
2) For $b = s-1$, $i \in [\max\{1, s+ba-n\}, \min\{s, ba\}]$.

3) Otherwise, $1 < b < s-1$, $i \in [\max\{0, s+ba-n\}, \min\{s, ba\}]$.

Where

$$h_i = \binom{n-s}{ba-i} \sum_{j=\max\{0,b+i-s\}}^{\min\{b,i\}} \binom{s-b}{i-j}\binom{b}{j} GMD(b,a,j) \cdot \\ \cdot GMD(s-b, a+1, (s-b)-(i-j)),$$

and

$$GMD(b,a,j) = \frac{(ab)!}{(a!)^b} \sum_{k=0}^{j} \binom{j}{k} (-1)^k a^k \frac{(ab-k)!}{(ab)!}.$$

5.1 Properties

If sharers can connect to themselves we obtain $E_S = s/n$. To see this, notice that in this case the $GMD()$ factors disappear and

$$\sum_i \frac{i}{s}\frac{h_i}{H} = \frac{\sum_i \frac{i}{s}\binom{n-s}{ba-i}\binom{s}{i}}{\binom{n}{ba}} = \frac{\sum_i \binom{(n-1)-(s-1)}{(ba-1)-(i-1)}\binom{(s-1)}{(i-1)}}{\binom{n}{ba}} = \frac{\binom{n-1}{ba-1}}{\binom{n}{ba}} = \frac{ab}{n}.$$

Therefore, the difference is given by the $GMD()$ factors.

Also note that $(ab)!/(a!)^b$ is not relevant within $GMD()$. Because this expression does not depend on i or j, it also appears in $H := \sum_i h_i$. Therefore it cancels out.

Two efficient ways to compute $GMD()$ recursively are as follows. Let

$$
\begin{aligned}
d_k :&= \binom{j}{k}(-1)^k a^k \frac{(ab-k)!}{(ab)!} \\
&= \{\frac{(j-0)...(j-k+1)}{12...k}\}\{(-1)...(-1)\}\{aa...a\}\{\frac{11...1}{(ba-0)...(ba-k+1)}\}, \\
f_k :&= \frac{(-1)(j-k+1)(a)}{(k)(ba-k+1)},\ f_0 := 1, \\
d_k :&= f_k d_{k-1},\ d_0 := 1, \\
g_k :&= f_{j-k}(1+g_{k-1}),\ g_0 := f_j.
\end{aligned}
$$

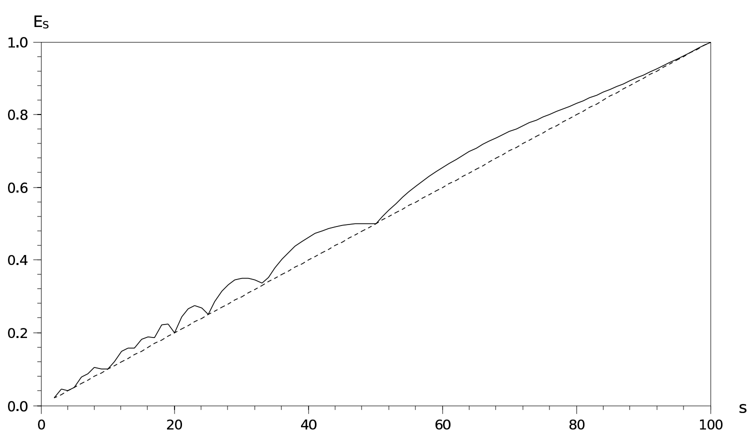
Then,

$$GMD(b,a,j) = \frac{(ab)!}{(a!)^b}\sum_{k=0}^{j} d_k = \frac{(ab)!}{(a!)^b} g_j.$$

6 Plots

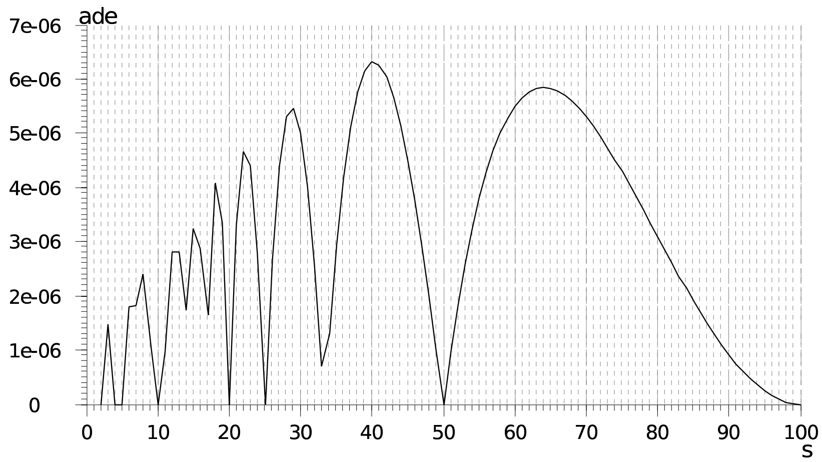
In this section we present a few plots to illustrate the properties of E_S, the expected bandwidth obtained by sharers. We initially fix the number of peers to $n = 100$, increase the number of sharers from $s = 2$ to 100, and plot E_S. The plot reveals that $E_S \geq s/n$ (the diagonal line). E_S is a curve with several peaks, always above s/n. Of course, $E_S > s/n$ implies that $E_F < s/n$. This follows from the fact that $s/n \cdot E_S + (n-s)/n \cdot E_F = s/n$. The difference between E_S and s/n is small. In the following plot we have augmented 10,000 times the difference to make it visible.

.



Plot 1. $E_S \geq s/n$

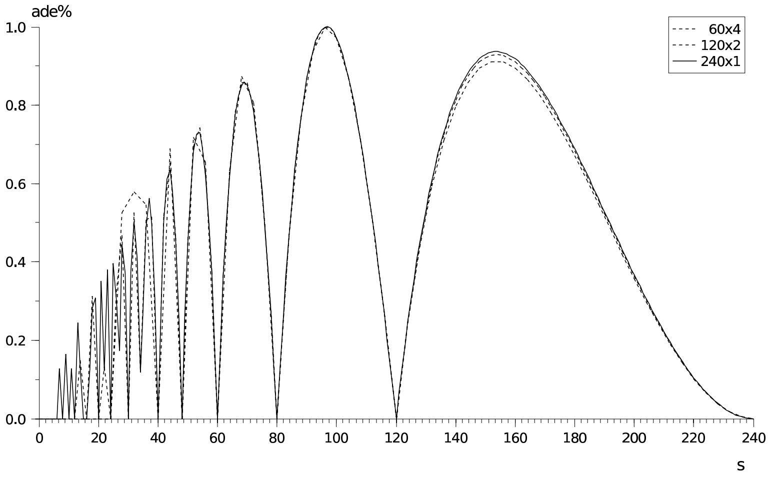
We now plot the absolute difference, $ade = E_S - s/n$. We should note that, in order to generate a statistical plot of ade by the Montecarlo method, one needs to be especially careful not to end up with a biased generator. A full 3D plot is included in the Appendix. Although a plot of the relative difference in percentage terms between E_S and s/n could be more informative, such a curve has peaks that vary wildly in size and cannot be drawn well.



Plot 2. A reference plot of ade.

As we increase n, the differences between E_S and s/n maintains a similar pattern, only the magnitudes differ. The next plot shows ade for different values of n ($n = 60$, $n = 120$, and $n = 240$), normalized in order for the largest peak to reach 1. We have also stretched the horizontal axis. For example, the plot for $n = 60$ has its horizontal axis stretched 4 times. And the plot for $n = 120$ has its horizontal axis stretched twice. The similarity between the three plots is remarkable.

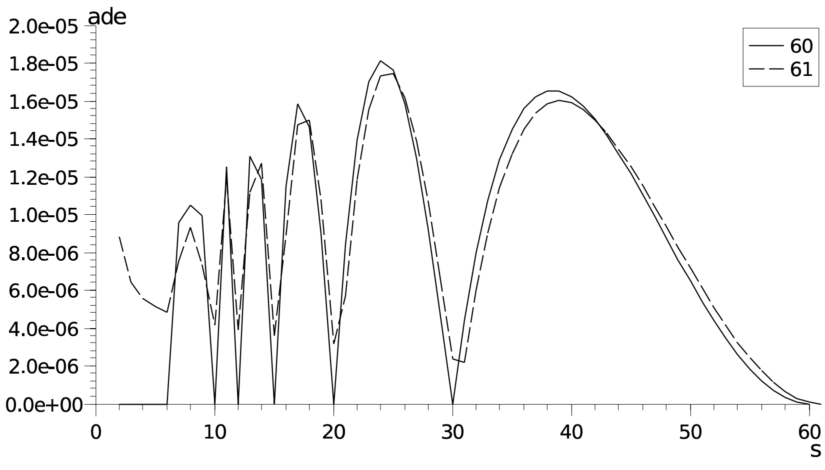
.



Plot 3. The *ade* curve has almost the same shape everywhere.

As n changes, the most pronounced differences in the shape of *ade* occur at the values of s where *ade* approaches zero ($n/2$, $n/3$, $n/4$,...). If n is divisible by s, then *ade* is equal to zero. Otherwise, $ade > 0$. The following plot shows *ade* for $n = 60$ and $n = 61$. 60 is divisible by 2, 3, 4, 5, 6, 10, 12, 15, 20, and 30. At all these points *ade* is zero. 61, however, is prime. In this case *ade* never reaches zero.

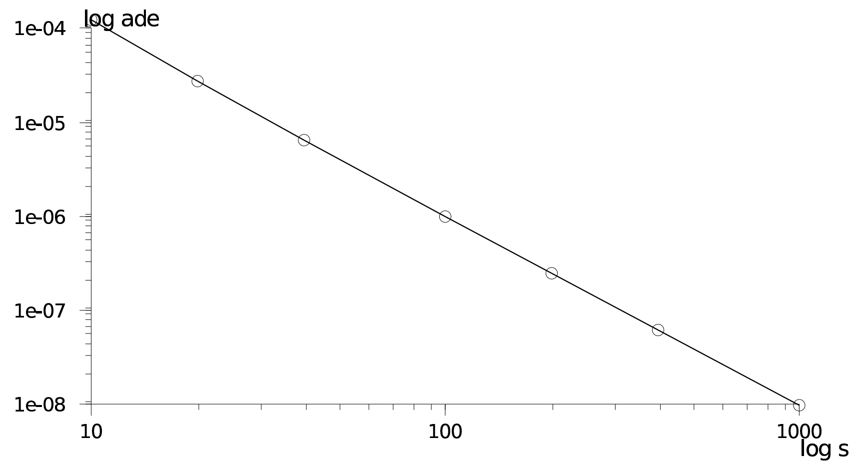
.



Plot 4. The biggest differences are on the zeros.

The fact that *ade* is similar for all n allows us to tabulate *ade* for a given n and then extrapolate its value for other n. In this way, we can obtain approximate values of *ade* without costly calculations.

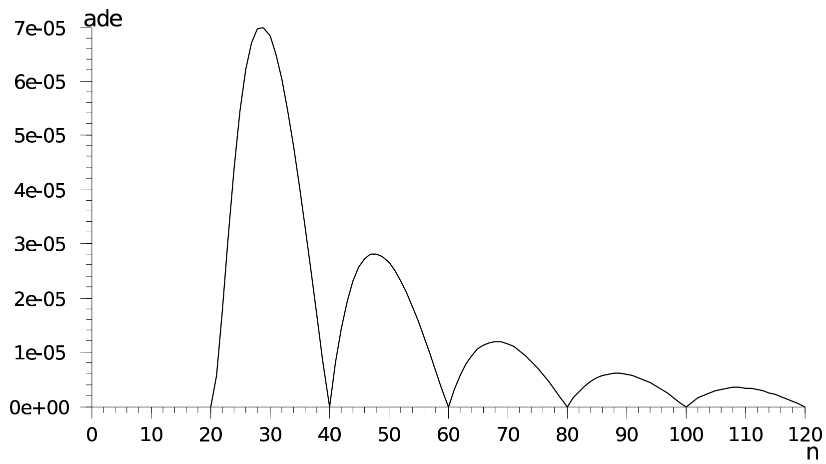
It is of interest to evaluate how *ade* varies as we increase n; or how the bandwidth difference between sharers and freeriders evolves as the size of the network increases. We next plot the values of *ade* at the last peak as we increase both n and s, thus maintaining the proportion of sharers.



Plot 5. The *ade* curve's last peak decreases at an exponential rate of -2 as new peers arrive.

The plot shows that *ade* decreases at a quadratic rate. This suggests a quick method to approximate the value of *ade* given s and n: look up $s_{100} = s/n\,100$ in the second plot, and apply the formula $ade = ade_{100}(100/n)^2$.

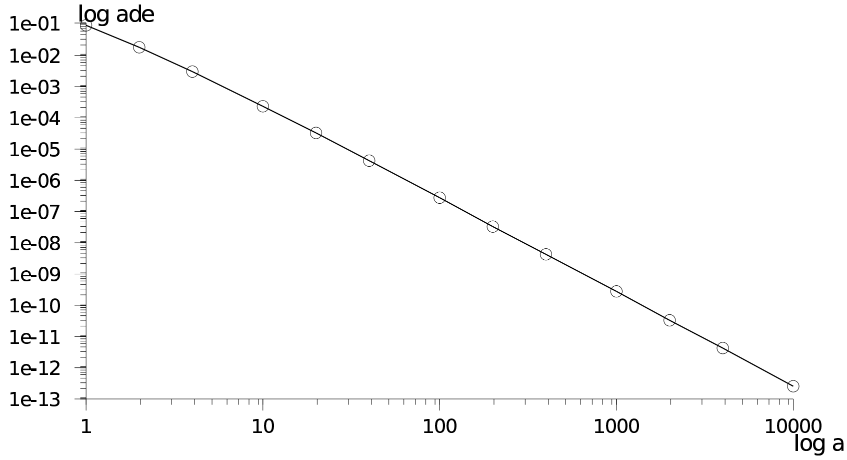
It is also interesting to see how *ade* evolves for a given s as n increases. In the following plot, we set $s = 20$ and let n vary from 20 to 120.



Plot 6. The *ade* curve when only freeriders arrive.

The next plot shows how the values at the peaks of the *ade* curve decrease for this case. The decrease rate is also exponential, but even larger with a -3 exponent.

.



Plot 7. The peaks of the *ade* curve decrease at an exponential rate of -3 as new freeriders arrive.

These results suggest that *ade* quickly converges to 0 as n increases.

7 Discussion

7.1 Connectivity constraints

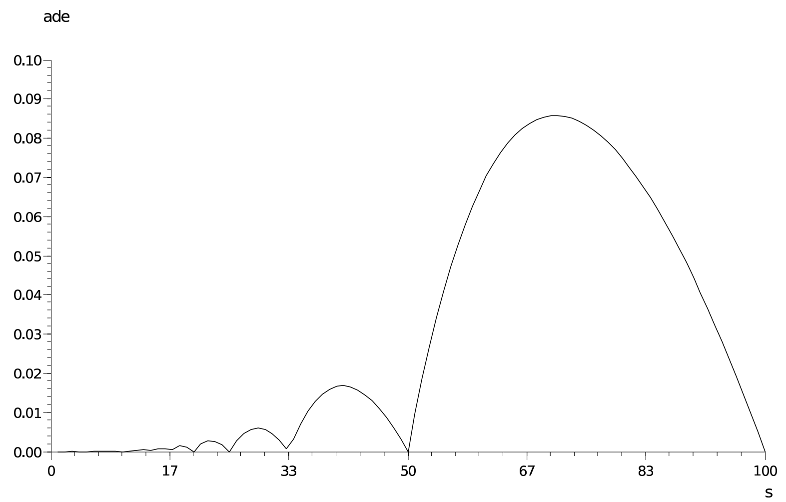
As we have just shown, the difference in expectations E_S and E_F is due to the constraint that a sharer cannot connect to herself. Furthermore, not only $E_S \neq E_F$, but $E_S \geq E_F$. The constraint 'helps' sharers; they end up better off (in expected terms). Counterintuitively, peers with more options available (freeriders) are worse off than those with less options (sharers).

To further illustrate the effect consider the following modified model where each sharer can *only* be connected to herself. Notice that the constraint is now stronger than before; sharers have only one feasible link and connections to other sharers are no longer available. The analysis is immediate,

$$E_S = \frac{1}{s}(b\frac{1}{a} + (s-b)\frac{1}{a+1}) = \frac{1}{s}\frac{sa+b}{a(a+1)}.$$

The following plot illustrates the behavior of *ade* in the modified model.

.



Plot 8. The *ade* curve when sharers can only connect to themselves.

As shown by the graph, $ade > 0$. And in fact, the difference between E_S and s/n is now even larger. This example shows that the more constrained peers are, the better off they end up. As a result, if peers could commit not to be allocated to a given sharer, they would do so.

7.2 Number of links

Our model assumes that peers can connect to one sharer only. While in real networks peers hold multiple links, the number of simultaneous connections is generally limited. In this sense, our single-link assumption is not that unrealistic. A more general model of p2p file sharing would allow for multiple-link formation. We have considered generalizations of the model in this direction, but the complexity of the analysis raises substantially and we have been unsuccessful in generating the more general formula.

To provide some insight on the effect of multiple links we analyze the following simple case. Consider a p2p network with $s = n - 1$ sharers (there is only one freerider) and where peers hold $k = s - 1$ simultaneous links. By construction, every sharer is connected to all other sharers in all stable allocations. Hence the upload bandwidth contributed by any given sharer is always accessed by the remaining k sharers. This implies that the freerider obtains the same bandwidth across all her links in all stable allocations,

$$E_F = k\frac{1}{k+1}.$$

We can calculate the bandwidth obtained by sharers by taking into account that $s/n \cdot E_S + (n-s)/n \cdot E_F = s/n$ and $n = s+1$:

$$E_S = \frac{1+k+k^2}{(1+k)^2},$$

which implies that in this particular example $E_S > E_F$. This suggests that the asymmetry identified above is likely to hold more generally to the case of multiple links.

7.3 Equiprobability

We have assumed throughout that stable allocations are equiprobable. We next motivate this assumption in the context of a model where peers decide with whom to connect to. The nature of p2p applications suggests one-sided link formation, where peers can decide which sharer to connect to without the consent of the sharer. [1] In this setting, a simultaneous one-shot game yields a set of equilibria that coincides by definition with our set of stable allocations. Clearly, this is the set of allocations that are of interest for the analysis. [2]

A one-shot game, however, provides no insight on the relative probability of each outcome. To construct a probability distribution over this set we need to consider a sequential game, where peers decide orderly with whom to connect to. To model such a game consider a randomized connecting order with myopic peers. That is, peers take the current allocation as given when choosing their connection; no forward induction takes place (which is unfeasible given the size and complexity of p2p networks). In the model described, however, the probability distribution over the set of stable allocations depends on the fine details of the connection process.

Consider the following example. There are six peers, three sharers (1, 2 and 3) and three freeriders (4, 5 and 6). The random ordering of peers is 3, 2, 4, 5, 6, and 1. In the connecting sequence 3 connects to 2; 2 connects to 3; 4 connects to 1; 5 connects to 2; and 6 connects to 3. At this point, 1 can only connect to 2 or 3. But neither constitutes a stable allocation.

[1] See Jackson [16] for a survey of models of networks in Economics and a discussion of different models of link formation.

[2] We should note that in our model no tradeoff exists between stability and efficiency. An allocation is Pareto efficient if and only if the upload bandwidth provided by *all* sharers is being utilized in the network. That is, if the whole bandwidth provision in the network is enjoyed by peers. Clearly, if an allocation does not satisfy this condition, a peer reassigned to a free sharer can be made better off without worsening the remaining peers. Hence stable allocations are a subset of Pareto allocations.

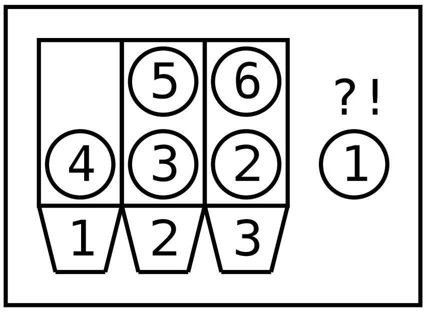


Fig. 7. Sequential connecting orderings with myopic peers may yield unstable allocations.

It turns out that constraining sharers not to connect to themselves while demanding equiprobability is somewhat equivalent to allocating them *first*, or at least, not last. This provides further intuition as to why the more constrained they are, the better off they end up.

And as shown by the example, further assumptions are required concerning the mechanism by which peers may update their links if unstable allocations arise. Due to the dependence of the solution on the fine details of the modeling choice, we have assumed equiprobability. Different mechanisms may favor either sharers or freeriders compared to the equiprobability benchmark.

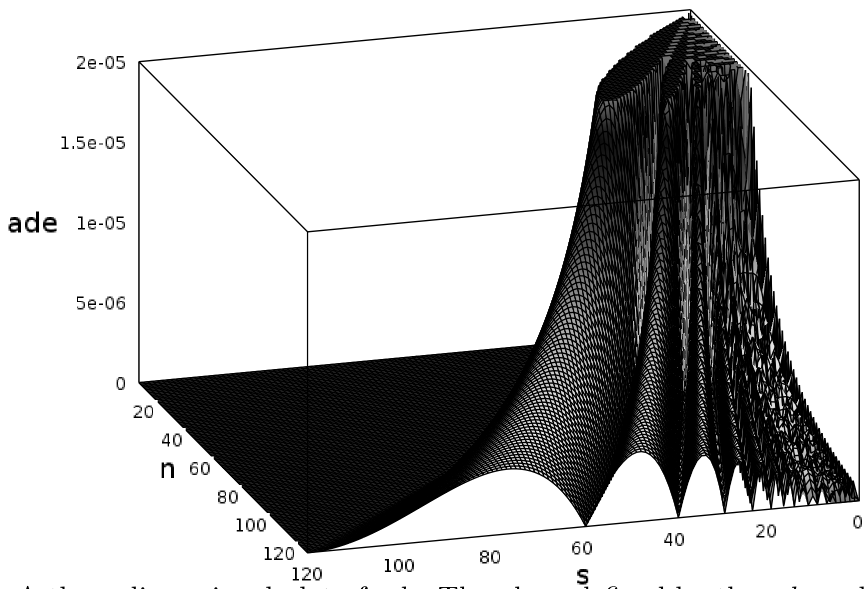
7.4 Applicability and limitations

We end with a few observations regarding the limitations of the model. An important characteristic of real networks we have not considered in the analysis is peer heterogeneity. In real networks, upload and download bandwidth capacities differ between peers. Heterogeneity in upload capacities, for example, would imply that certain sharers are able to provide higher utility to the peers that connect to them. Our model is also static, it does not consider how stable allocations are reached nor the evolution of the network over time. In real networks, peers enter and leave.

Even with these limitations, we hope to have provided a benchmark on which to construct these (and other) extensions. We believe the qualitative results obtained from our model are robust and should hold under more general specifications. A clear understanding of bandwidth allocation in p2p file sharing networks is a necessary first step towards providing a theory of incentives to contribute resources to p2p. [3]

[3] For a recent application of the s/n approximation see Casadesus-Masanell and Hervas-Drane [15].

Appendix



A three dimensional plot of *ade*. The plane defined by the *ade* and *s* axis corresponds to the *ade* curve depicted in plot 2.

Acknowledgements

We thank three anonymous referees and the Editor for comments that helped improve the manuscript. We are grateful for financial support from the HBS Division of Research, the NET Institute, Fundación BBVA, and the FPU program of the Spanish Ministry of Education and Science.

References

[1] Atip Asvanund, Karen Clay, Ramayya Krishnan, and Michael D. Smith, "An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks," *Information Systems Research* 15(2004) 155-174.

[2] Jacky Chu, Kevin Labonte, and Brian N. Levine, "Availability and Locality Measurements of Peer-to-Peer File Systems," *Proceedings of ITCom: Scalability and Traffic Control in IP Networks* (2002).

[3] Matei Ripeanu and Ian Foster, "Mapping the Gnutella Network: Macroscopic Properties of Large-Scale Peer-to-Peer Systems," *Proceedings of the 1st International Workshop on Peer-to-Peer Systems* (2002).

[4] Stefan Saroiu, Krishna Gummadi, and Steven D. Gribble, "A Measurement Study of Peer-to-Peer File Sharing Systems," *Proceedings of Multimedia Computing and Networking* (2002).

[5] Krishna P. Gummadi, Richard J. Dunn, Stefan Saroiu, Steven D. Gribble, Henry M. Levy, and John Zahorjan, "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload," *Proceedings of the 19th ACM Symposium on Operating Systems Principles* (2003).

[6] Subhabrata Sen and Jia Wang, "Analyzing Peer-to-Peer Traffic Across Large Networks," *ACM/IEEE Transactions on Networking* 12(2004) 219-232.

[7] Chiranjeeb Buragohain, Divyakant Agrawal and Subhash Suri, "A Game Theoretic Framework for Incentives in P2P Systems," *Proceedings of the 3rd International Conference on Peer-to-Peer Computing* (2003).

[8] Michael Feldman, Christos Papadimitriou, John Chuang, and Ion Stoica, "Free-riding and Whitewashing in Peer-to-Peer Systems," *Proceedings of the ACM SIGCOMM 2004 workshop on Practice and Theory of Incentives and Game Theory in Networked Systems* (2004).

[9] Philippe Golle, Kevin Leyton-Brown, and Ilya Mironov, "Incentives for Sharing in Peer-to-Peer Networks," *Lecture Notes in Computer Science* 2232(2001) 75-87.

[10] Panayotis Antoniadis, Costas Courcoubetis, and Robin Mason, "Comparing Economic Incentives in Peer-to-Peer Networks," *Computer Networks* 46(2004) 133-146.

[11] Brendan M. Cunningham, Peter J. Alexander, and Nodir Adilov, "Peer-to-Peer File Sharing Communities," *Information Economics and Policy* 16(2004) 197-221.

[12] Ramayya Krishnan, Michael D. Smith, Zhulei Tang, and Rahul Telang, "The Virtual Commons: Understanding Content Provision in Peer-to-Peer File Sharing Networks," *Carnegie Mellon Heinz School working paper* 2004-19.

[13] Lian Jian and Jeffrey MacKie-Mason, "Why Share in Peer-to-Peer Networks?," *Proceedings of the ACM EC 2006 workshop on the Economics of Networked Systems* (2006).

[14] Michael Feldman, Kevin Lai, John Chuang, and Ion Stoica, "Quantifying Disincentives in Peer-to-Peer Networks," *Proceedings of the 1st Workshop on Economics of Peer-to-Peer Systems* (2003).

[15] Ramon Casadesus-Masanell and Andres Hervas-Drane, "Peer-to-Peer File Sharing and the Market for Digital Information Goods," *Mimeo, Harvard Business School* (2006).

[16] Matthew O. Jackson (2003), "A Survey of Models of Network Formation: Stability and Efficiency," *Mimeo, California Institute of Technology* (2003).