Incentives for P2P-Assisted Content Distribution: If You Can't Beat 'Em, Join 'Em

Vinod Ramaswamy, Sachin Adlakha, Srinivas Shakkottai, Adam Wierman*

Abstract

The rapid growth of content distribution on the Internet has brought with it proportional increases in the costs of distributing content. Adding to distribution costs is the fact that digital content is easily duplicable, and hence can be shared in an illicit peer-to-peer (P2P) manner that generates no revenue for the content provider. In this paper, we study whether the content provider can recover lost revenue through a more innovative approach to distribution. In particular, we evaluate the benefits of a hybrid revenue-sharing system that combines a legitimate P2P swarm and a centralized client-server approach. We show how the revenue recovered by the content provider using a server-supported legitimate P2P swarm can exceed that of the monopolistic scheme by an order of magnitude. Our analytical results are obtained in a fluid model, and supported by stochastic simulations.

1 Introduction

The past decade has seen the rapid increase of content distribution using the Internet as the medium of delivery [1]. Users and applications expect a low cost for content, but at the same time require high levels of quality of service. However, providing content distribution at a low cost is challenging. The major costs associated with meeting demand at a good quality of service are (i) the high cost of hosting services on the managed infrastructure of CDN [2], and (ii) the additional costs associated with the fact that digital content is easily duplicable, and hence can be shared in an illicit peer-to-peer (P2P) manner that generates no revenue for the content provider. Together, these factors have led content distributors to search for methods of defraying costs.

One technique that is often suggested for defraying distribution costs is to use legal peer-to-peer (P2P) networks to supplement provider distribution [3, 4]. It is well documented that the efficient use of P2P methods can result in significant cost reductions from the perspective of ISPs [2, 5]; however there are substantial drawbacks as well. Probably the most troublesome is that providers fear losing control of content ownership, in the sense that they are no longer in control of the distribution of the content and worry about feeding illegal P2P activity.

Thus, a key question that must be answered before we can expect mainstream utilization of P2P approaches is: *How can users that have obtained content legally be encouraged*

*V. Ramaswamy and S. Shakkottai are with the Dept. of Electrical and Computer Engineering, Texas A&M University, Email: {vinod83, sshakkot}@tamu.edu. S. Adlakha and A. Wierman are with the Dept. of Computing and Mathematical Sciences, Caltech, Email: {adlakha,adamw}@caltech.edu.

to reshare it legally? Said in a different way, can mechanisms be designed that ensure legitimate P2P swarms will dominate the illicit P2P swarms?

In this paper, we investigate a "revenue sharing" approach to this issue. Motivated by recent work using lottery schemes to promote societally beneficial conduct [6], we suggest that users can be motivated to reshare the content legally by allowing them to share the revenue associated with future sales. This can be accomplished through either a lottery scheme or by simply sharing a fraction of the sale price.

Such an approach has two key benefits: First, obviously, this mechanism ensures that users are incentivized to join the legitimate P2P network since they can profit from joining. Second, less obviously, this approach actually damages the illicit P2P network. Specifically, despite the fact that content is free in the illicit P2P network, since most users expect a reasonable quality of service, if the delay in the illegitimate swarm is large they may be willing to use the legitimate P2P network instead. Thus, by encouraging users to reshare legitimately, we are averting them from joining the illicit P2P network, reducing its capacity and performance; thus making it less likely for others to use it.

The natural worry about a revenue sharing approach is that by sharing profits with users, the provider is losing revenue. However, the key insight provided by the results in this paper is that the bootstrapping provided by the second effect described above provides a magnification of the initial revenue sharing "investments", which turns out to provide exponential gains in revenue for the provider even when very little revenue sharing is used.

More specifically, the contribution of this paper is to develop and analyze a model to explore the revenue sharing approach described above. Our model (see Section 2) is a fluid model that builds on work studying the capacity of P2P content distribution systems. The key novel component of the model is the competition for users among an illicit P2P system and a legal content distribution network (CDN), which may make use of a supplementary P2P network with revenue sharing. The main results of the paper (see Section 3) are Theorems 1 and 2, which highlight the order-of-magnitude gains in revenue extracted by the provider as a result of participating in revenue sharing. Further, Corollary 3 highlights that the optimal amount of revenue sharing is quite small. In addition to the analytic results, to validate the insights provided by our asymptotic analysis of the fluid model we also perform numerical experiments of the underlying finite stochastic model. Figures 3(a), 3(b), 3(c) and 3(d) summarize these experiments, which highlight both that the results obtained in the fluid model are quite predictive for the finite setting and that there are significant beneficial effects of revenue sharing. In particular, the examples that we present indicate that revenue extraction gains between 25% to 180% are possible through appropriate revenue sharing.

There is a significant body of prior work modeling and analyzing P2P systems. Perhaps the most related work from this literature is the work that focuses on server-assisted P2P content distribution networks [7–12] in which a central server is used to "boost" P2P systems. This boost is important since pure P2P systems suffer poor performance during initial stages of content distribution. In fact, it is this initially poor performance that our revenue sharing mechanism exploits to ensure that the legitimate P2P network dominates.

Two key differentiating factors of the current work compared to this work are: (i) We model the impact of competition between legal and illegal swarms on the revenue extraction of a content provider. (ii) Unlike most previous works on P2P systems, we consider a time varying viral demand model for the evolution of demand in a piece of content based on the Bass diffusion model (see Section 2). Thus, we model the fact that interest in content grows as interested users contact others and make them interested.

With respect to (i), there has been prior work that focuses on identifying the relative value of content and resources for different users [13, 14]. For instance, [13] deals with creating a content exchange that goes beyond traditional P2P barter schemes, while [14] attempts to characterize the relative value of peers in terms of their impact on system performance as a function of time. However, to the best of our knowledge, ours is the first work that considers the question of economics and incentives in hybrid P2P content distribution networks.

With respect to (ii), there has been prior work that considers fluid models of P2P systems such as [15–17]. However, these all focus on the performance evaluation of a P2P system with constant demand rate. As mentioned above, a unique facet of our approach is that we explicitly make use the transient nature of demand in our modeling. In the sense of explicitly accounting for transient demand, the closest work to ours is [11]. However, [11] focuses only on jointly optimizing server and P2P usage in the case of transient demand in order to obtain a target delay guarantee at the lowest possible server cost.

The remainder of the paper is organized as follows. We first introduce the details of our model in Section 2. Then, Section 3 summarizes analytic and numeric results, the proofs of which are included in the appendix. Finally, Section 4 provides concluding remarks.

2 Model overview

Our goal is to model the competition between illicit peer-to-peer (P2P) distribution and a legitimate content distribution network (CDN), which may make use of its own P2P network. Our model is a fluid model, and there are four main components:

(i) The evolution of the demand for content. As mentioned in the introduction, a key feature of this paper is that we consider a realistic model for the evolution of demand, specifically, the Bass diffusion model.

(ii) The model of user behavior, which allows the user to strategically choose between attaining content legally or illegally based on the price and performance of the two options.

(iii) The model of the illicit P2P system.

(iv) The model of the legal CDN and its possibility to use "revenue sharing".

We discuss these each in turn in the following.

2.1 The evolution of demand

Models of the dynamics of demand growth for innovations date back to the work of Griliches [18] and Bass [19]. The most widely used model for dynamics of demand growth is the Bass diffusion model which describes how new products get adopted as potential users interact with users that have already adopted the product. Such word of mouth interaction between users and potential users is very common in the Internet and we use a version of Bass diffusion model that only has word of mouth spreading.

In our setting we have two key pieces of notation: N, the total size of the population, and $I(t)$, the number of users that are interested in the content at time t. We model that each interested user "attempts" to cause a randomly selected user to become interested in the content.[1] At any time t, there are $N - I(t)$ users that could potentially be interested

[1] Note that these "attempts" should not be interpreted literally, but rather as the natural diffusion of interest in the new content through the population.

in the content. Thus, the probability of finding such a users is $(N - I(t))/N$. Assuming that an interested user can interact with other users at rate 1 per unit time, we get that the rate at which interested users increase is given by the following differential equation:

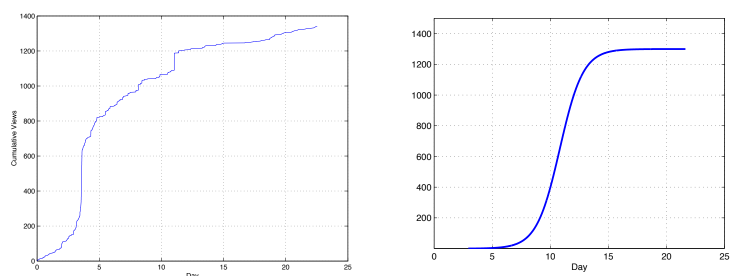
$$\frac{dI(t)}{dt} = \left(\frac{N - I(t)}{N}\right) I(t). \qquad (1)$$

The above differential equation can be easily solved and yields the so-called *logistic function* as its solution.

$$I(t) = \frac{I(0)e^t}{1 - (1 - e^t)\frac{I(0)}{N}}, \qquad (2)$$

where $I(0)$ is the number of user that are interested in the content at time $t = 0$.

Though this model is simplistic, it is a useful qualitative summary of the spread of content. To highlight this, Figure 1 (taken from [11]) highlights a similar behavior in a data trace from CoralCDN [20], a CDN hosted at different university sites. The figure shows the cumulative demand for a home video of the Asian Tsunami seen over a month in December 2005. For comparision, the figure on the right shows the model in equation (2). The qualitative usefulness of the Bass model has been verified empirically in other settings as well. For example, [21] shows that a variant accurately represents the penetration of CDs using data collected internationally [22]. There are many other examples of the use of the Bass model for forecasting the dissemination of innovations [23–25], and hence it can be considered as canonical [26][2].



(a) Demand data (b) Demand in Bass model

Figure 1: (a) shows the cumulative demand for a file over one month on Coral CDN (Dec 2005–Jan 2006). (b) shows the cumulative demand seen in a Bass diffusion.

In the current model, for analytic reasons, we are not able to work with the exact Bass model. Thus, we approximate the logistic curve as follows:

$$I(t) = \begin{cases} \frac{NI(0)e^t}{N - I(0) + I(0)e^t} & 0 \le t \le T_1 \quad : \text{Phase } 1 \\ I_2 = N/\ln N & T_1 < t \le T_2 \quad : \text{Phase } 2 \\ I_3 = \frac{N}{2} & T_2 < t \le T_3 \quad : \text{Phase } 3 \\ I_4 = N & T_3 < t < T_4 \quad : \text{Phase } 4, \end{cases} \qquad (3)$$

where we have $T_1 = \ln(N/(I(0)\ln N))$, $T_2 = \ln(N/I(0))$, $T_3 = 2\ln(N/I(0))$ and $T_4 = 3\ln(N/I(0))$.[3] Notice that the first stage is the exact Bass diffusion, while the other stages

[2] Indeed, the original work by Bass is one of the ten most cited papers in Management Science, and was republished in 2004 to illustrate its impact.

[3] Note that the value of T_1 has been chosen such that $\lim_{N\to\infty} I(T_1) = N/\ln N$.

are order sense approximations of the actual expression. Though this model is approximate, it yields the same qualitative insight as the original model.

2.2 The progression of a user

In order to capture the strategic behavior of users in the face of competition between a legitimate CDN using P2P and an illicit P2P network our model is necessarily complex. Figure 2 provides a broad overview of the user behavior in the system, which we explain in detail in the following. In what follows, we develop a fluid (differential equation) model to describe our system. However, we note that it is straightforward to show that the stochastic version of the model would give the same results in an order sense, as we did in earlier work [12]. Also, all the simulations in this paper are conducted using a full stochastic model, and their accurate match with the fluid model further indicates its accuracy.

Let us explain the model through tracking the progression of a user. We term an initial user that wants, but has not yet attained, the content a *Wanter (W)*. When a Wanter arrives to the system, it has two options: get content from the illicit P2P system for free or get content from the legitimate system for a price p. We assume that the Wanter wishes to obtain content as quickly and cheaply as possible, and so she first approaches the illicit P2P swarm and then only attains the content from the legitimate system if the content is not attained a reasonable time interval (one infinitesimal clock tick in our model) from the illicit P2P. This cycle repeats, if necessary, until the content is attained. In some sense, this is the worst-case for the legitimate provider since the illicit source is tried first.

Once the Wanter has attained the content, we assume the the Wanter acts strategically when deciding its next action. If a Wanter obtains the content legally, then the Wanter has three options: (i) It might decide to use the content to assist the illicit P2P swarm, i.e., go *Rogue (R)*. We denote the probability this happens by $\rho < 1$. (ii) It might decide to assist the legitimate P2P swarm (if one exists) as a *Booster (B)*. We denote the probability of this event by $\beta < 1$. Note that $\beta = 0$ if no legal P2P is used. (iii) Or, it may simply *Quit(Q)* and leave the system. If a Wanter obtains the content illegally the options are similar: it can either aid the illicit swarm as a *Fraudster (F)*, or *Quit(Q)* and leave the system. We denote the probability that a Wanter that has obtained the content illegally becomes a Fraudster by $\kappa < 1$.

Note that the goal of revenue sharing is to incentivize Wanters to become Boosters after attaining content legally, rather than going Rogue. The hope is that the revenue invested toward reducing the number of "early adopters" that go Rogue keeps the illicit P2P swarm from growing enough to provide good enough quality of service to dominate the legitimate swarm.

To model this system more formally, we introduce the following notation. Let $N_w(t)$ be the number of Wanters at time t, i.e., the number of users who have not yet attained the content, and assume $N_w(0) = 0$. Further, let $N_l(t)$ and $N_i(t)$ be the number of users with legal and illegal copies of the content at time t. Note that the total number of interested users at any time t satisfies the following equation

$$I(t) = N_w(t) + N_l(t) + N_i(t) \tag{4}$$

We can break this down further by noting that the number of Rogues, Fraudsters, and Boosters in the system at time t (denoted by $N_r(t)$, $N_f(t)$, and $N_b(t)$ respectively) is:

$$N_r(t) = \rho N_l(t) \tag{5}$$

$$N_f(t) = \kappa N_i(t) \tag{6}$$

$$N_b(t) = \beta N_l(t), \tag{7}$$

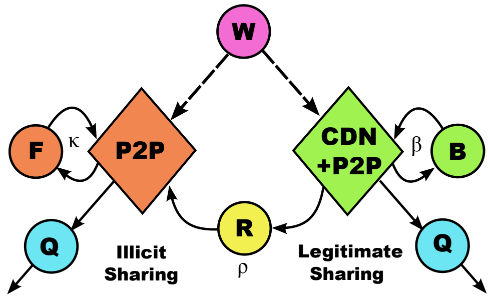


Figure 2: An overview of the progression of a user through the systems. The labels are defined as follows: W - Wanter, F - Fraudster, R - Rogue, B - Booster, and Q - Quit.

with $\rho + \beta < 1$. The rest of legal and illegal users leave the system.

The key remaining piece of the model is to formally define the transition of Wanters to holders of illegal/legal content, i.e., the evolution of $N_i(t)$ and $N_l(t)$. However, this evolution depends critically on the model of the two systems, and so we describe it in the next section.

2.3 System models

We discuss in detail the illicit and legitimate system models below. The factors in these models are key determinants of the choice of a Wanter to get the content legally or illegally. When modeling the two systems, we consider a fluid model, and so the performance is determined primarily by the capacity of each system, i.e., the combination of the initial seeds and the Fraudsters/Boosters that choose to join (and add capacity). However, other factors also play a role, as we describe below. Throughout, we model the upload capacity of a user as being 1.

2.3.1 The illicit P2P system

There are two components to the model of the illicit P2P network: (i) the efficiency of the network in terms of finding content, and (ii) the initial size of the network and its growth.

Let us start with (i). To capture the efficiency of the P2P system, we take a simple qualitative model. When attaining the content illegally, a Wanter must contact either a Rogue or a Fraudster. We let $\eta(t)$ capture the probability of a Wanter finding a Rogue or a Fraudster when looking for one instantaneous time slot. We will consider two cases: an efficient P2P and an inefficient P2P. In an ***efficient P2P***, we model $\eta(t) = 1$, with the understanding the the P2P allows easy lookup of content and all content is truthfully represented. In contrast, for an ***inefficient P2P***, we model

$$\eta(t) = (N_r(t) + N_f(t))/N,$$

where recall that N is the total population size. This corresponds to looking randomly within the user population for a Rogue or Fraudster. Neither of these models is completely realistic, but we choose them with the goal of upper and lower bounding the true efficiency of an illicit P2P system.

Next, with respect to (ii), we model the initial condition for the illicit network with $N_i(0) = 0$, since the assumption is that the content has not yet been released, and therefore is not yet available in the illicit P2P swarm. From this initial condition, $N_i(0)$ evolves as follows:

$$\frac{dN_i(t)}{dt} = \min\left\{\eta(t)\left(N_w(t) + \frac{dI(t)}{dt}\right), N_r(t) + N_f(t)\right\}, \tag{8}$$

The interpretation of the above is that $N_r(t) + N_f(t)$ is the current capacity of the illicit P2P and $\eta(t)(N_w(t)+dI(t)/dt)$ is the fraction of the Wanters (newly arriving and remaining in the system) that find the content in the illicit P2P network. The min then ensures that no more than the capacity is used.

2.3.2 The legitimate CDN

As discussed in the introduction, our goal in this work is to contrast the revenue attained by a CDN that uses P2P and revenue sharing with one that does not use P2P. Thus, there are two key factors in modeling the legitimate CDN: (i) the amount revenue sharing used, and (ii) the initial size of the CDN and its growth, which depends on the presence/absence of the legal P2P.

Let us start with (i). Suppose that the purchase price of a copy of the content is p. Hence, a user that wishes to obtain a legal copy of the content must pay the content generator the sum p through some kind of online banking system. We consider a simple model for revenue sharing where a user receives ϵp for each piece of content it distributes when taking part in the legitimate network as a Booster. Thus, $\epsilon = 0$ corresponds to no revenue sharing. Note that this could potentially be implemented on a system such as BitTorrent by simply keeping track of amount uploaded by each peer[4]. The value ϵ can be viewed either as a share of the revenue from each download or as the expected payoff of a lottery scheme operated by the CDN.

Intuitively, κ is fixed regardless of ϵ, since once a Wanter gets the content illegally, whether it becomes a Fraudster or not is independent of revenue sharing. The key consequence of revenue sharing is on the "rogue factor" ρ. We make the assumption

$$\rho + \beta = \kappa.$$

By fixing $\rho + \beta$, we are assuming that the likelihood of a user joining a P2P swarm (either legal or not) is fixed irrespective of ϵ, and only *which* P2P swarm is joined is affected by revenue sharing. The idea is that increased revenue sharing should limit the likelihood of a Wanter going rogue after attaining the content legally.

It is difficult to predict the exact impact of this effect; however to qualitatively capture this, we model ρ as a decreasing function of ϵ. To make the analysis tractable, we use the specific form

$$\rho = \kappa N^{-\epsilon},$$

which captures the desired qualitative effect.

Next, with respect to (ii), unlike for the illicit P2P swarm, the legitimate network does not start empty. This is because it has a set of dedicated servers at the beginning which are then (possibly) supplemented using a P2P network. We denote by C_N be the capacity of the dedicated CDN servers when the total population size is N. Note that this capacity must scale with the total population size to ensure that the average wait time for the users is small. As shown in [11], a natural scaling that ensures no more that $O(\ln \ln N)$ delay is to have the capacity $C_N = \Theta(N/\ln N)$. Based on this, we adopt

$$C_N = \frac{N}{\ln N}$$

[4]BitTorrent Trackers already collect such information in order to gather performance statistics.

in this work. Additionally, we assume $N_l(0) = I(0)$ and $I(0) \in \Theta(1)$.

Given these initial conditions, $N_l(t)$ evolves as follows:

$$\frac{dN_l(t)}{dt} = \begin{cases} C_N + \beta N_l(t), & N_w(t) > 0, \\ \min\left\{C_N + \beta N_l(t), \frac{dI(t)}{dt} - \frac{dN_i(t)}{dt}\right\} & N_w(t) = 0. \end{cases} \tag{9}$$

The interpretation for the above is that if there are a positive number of Wanters remaining in the system, then the full current capacity of the CDN can be used to serve them, i.e., $C_N + \beta N_l(t)$. However, if there are no "leftover" Wanters, arriving Wanters that are not served by the illicit P2P ($\frac{dI(t)}{dt} - \frac{dN_i(t)}{dt}$) are served up to the capacity of the CDN.

3 Results

Given the model, we can now investigate the impact "revenue sharing" has on the revenue attained by the CDN. The goal of the work is to highlight that the revenue shared serves as an "investment" which pays itself back many times over as a result of the damage it causes to the capacity/performance of the illicit P2P network.

To characterize the revenue attained by the CDN, we use the *fractional revenue* attained, which is defined as follows:

Definition 1 *The* ***fractional revenue****, R, attained by the CDN is defined as*

$$R = \frac{N_l(T_4)p(1-\epsilon)}{Np} = \frac{N_l(T_4)(1-\epsilon)}{N} \tag{10}$$

Recall that T_4 is the final point of time in the evolution of demand, and so this can be interpreted as an approximation[5] of the revenue attained by the CDN divided by the maximal revenue the CDN could have achieved.

Using this metric, we look at the impact of revenue sharing in two settings: when the CDN competes against inefficient illicit P2P sharing and when it competes against efficient illicit P2P sharing. Recall, that our models for these two cases are meant to serve as upper and lower bounds on the true efficiency of an illicit P2P system.

Note that the theorems stated below characterize only the asymptotic growth of the fractional revenue. However, the proofs of these theorems, presented in Appendicies A and B, actually characterize the exact growth.

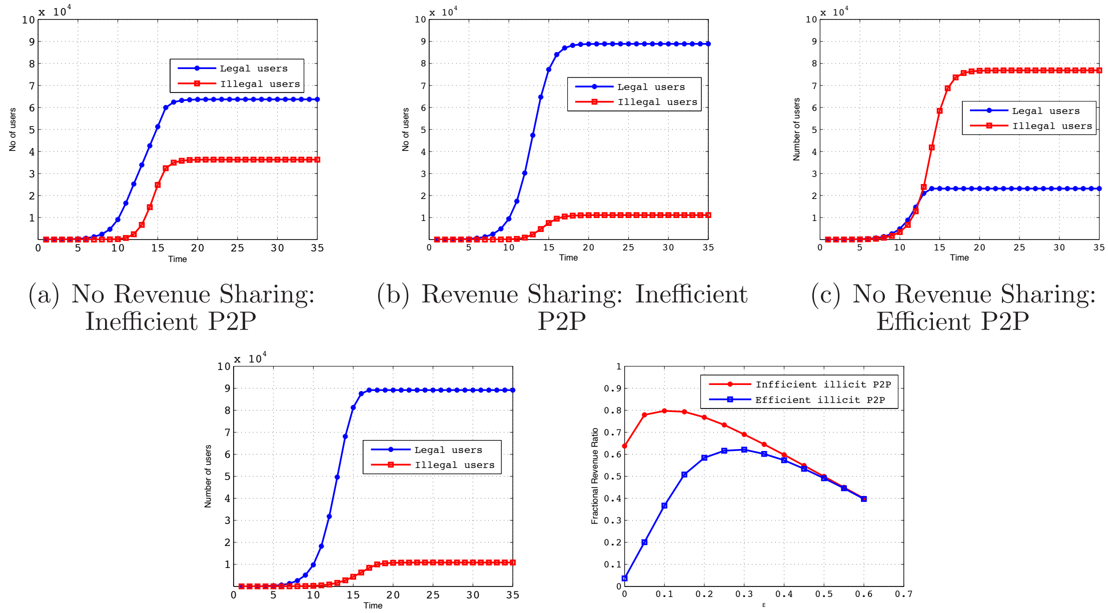
Let us first consider the case of an inefficient, illicit P2P.

Theorem 1 *Let $\rho + \beta = \kappa$ and $\rho = \kappa N^{-\epsilon}$. The fractional revenue attained by the content provider in the presence an inefficient, illicit P2P is*

$$R \in \Omega\left(\frac{1-\epsilon}{\ln N}\left(\ln\ln N + (\ln N)^{1-N^{-\epsilon}}\right)\right) \tag{11}$$

Further, when $\epsilon = 0$,

$$R \in \Theta\left(\frac{\ln\ln N}{\ln N}\right). \tag{12}$$



(a) No Revenue Sharing: Inefficient P2P (b) Revenue Sharing: Inefficient P2P (c) No Revenue Sharing: Efficient P2P

(d) Revenue Sharing: Efficient P2P (e) Optimal Revenue Sharing

Figure 3: Numerical simlation of different cases. In all cases revenue sharing increases the profit obtained by the content provider.

The interpretation of this theorem is striking. When no revenue sharing is used the fractional revenue attained by the content provider is exponentially small, $\Theta\left(\frac{\ln\ln N}{\ln N}\right)$. However, when revenue sharing is used, the fractional revenue grows by orders of magnitude. Admittedly, the exact form of the improvement is dependent on our assumption that $\rho = \kappa N^{-\epsilon}$ (which facilitates analysis); however the qualitative comparison is not dependent on this form. Further, as we show later in this section, despite the asymptotic nature of the result the qualitative insight is visible even in finite, discrete systems.

Next, let us consider the case of an efficient, illicit P2P system.

Theorem 2 *Let $\rho + \beta = \kappa$, with $\rho = \kappa N^{-\epsilon}$ and $\kappa \in (C_N/N, 1 - C_N/N)$. The fractional revenue attained by the content provider in the presence an efficient, illicit P2P is*

$$R \in \Omega\left(\frac{1-\epsilon}{\ln N}\left(\frac{(\ln N)^{1-N^{-\epsilon}}-1}{1-N^{-\epsilon}}\right)\right). \tag{13}$$

Further, when $\epsilon = 0$,

$$R \in \Theta\left(\frac{\ln\ln N}{\ln N}\right). \tag{14}$$

[5]This is an approximation (actually a lower bound) since it assumes that only $(1-\epsilon)p$ is attained from *all* sales, rather than having the full p attained from distribution from the dedicated servers. This difference does not affect the asymptotic order of the fractional revenue, which is what we are interested in characterizing.

Again, the the impact of revenue sharing in Theorem 2 is striking. The fraction of revenue obtained by the content provider rises by an order of magnitude when revenue sharing is used. Interestingly, the efficiency of the illicit P2P does not impact the asymptotic order of the fractional revenue when revenue sharing is not used, since in both the efficient and inefficient case it is $\Theta\left(\frac{\ln\ln N}{\ln N}\right)$. However, the efficiency of the illicit P2P does affect the fractional revenue attained by revenue sharing. In particular, it causes a $1 - N^{-\epsilon}$ factor change in the fractional revenue attained; however this has almost no effect on the asymptotic growth. So, the benefits of revenue sharing are robust to the efficiency of the P2P system. Note that for technical reasons, Theorem 2 requires an extra assumption on κ.

Given Theorems 1 and 2, it is natural to ask about the optimal amount of revenue sharing a CDN should use. One might worry that sharing a significant amount of revenue is necessary in order to properly incentivize users. Using the scalings in Theorems 1 and 2 it is possible to characterize the asymptotically optimal revenue sharing $\epsilon = \epsilon_N$. Interestingly, it turns out that the order of the optimal revenue sharing is the same in both the case of an inefficient and an efficient illicit P2P, and that a very small revenue share is optimal. Formally, we have the following corollary. The proof is omitted due to space constraints.

Corollary 3 *Let $\rho + \beta = \kappa$ with $\rho = \kappa N^{-\epsilon_N}$. In the presence of either an inefficient P2P or an efficient P2P with $\kappa \in (C_N/N, 1 - C_N/N)$, the choice of ϵ_N that achieves optimal fractional revenue is*

$$\epsilon_N \in \Theta\left(\frac{\ln\ln\ln N}{\ln N}\right). \tag{15}$$

Since Theorems 1 and 2 and Corollary 3 rely on a fluid model, and characterize only the asymptotic growth rate of the fractional revenue and the optimal revenue sharing, we present numerical simulations to verify the qualitative insights in discrete systems with finite N.

To simulate the underlying discrete stochastic system, we assume time is discrete and that there are $N = 100,000$ users in the system. At each time slot, each user picks a Poisson distributed number (with mean 1) of other users to spread interest to. The server has a FIFO policy with service rate $C = 8000 \approx N/\ln N$.

Figures 3(a) and 3(b) illustrate the case of an inefficient illicit P2P system with and without revenue sharing. We use $\kappa = 0.75$, $\epsilon = 0$ in the no revenue sharing setting (Figure 3(a)) and $\epsilon = 0.1$ in the case of revenue sharing (Figure 3(b)). In the case of no revenue sharing, the number of legal users eventually present in the system is $63,000$, while the value anticipated from our theoretical analysis, specifically by Corollory 7, is $60,100$. In the revenue sharing setting, the final number of legal users in the system is $88,888$, while the value predicted by Lemma 6 is $79,100$. The simulation results validate the insights obtained from our theoretical analysis. The key point in this figure is that even when the illicit P2P is extremely inefficient, there is significant revenue that can be gained from revenue sharing. In fact, the fractional revenue increases by more than 25%.

Next, we move to the case of an efficient illicit P2P. Figures 3(c) and 3(d) illustrate the case of an efficient illicit P2P system with and without revenue sharing. We use $\kappa = 0.4$, $\epsilon = 0$ in the no revenue sharing setting (Figure3(c)) and $\epsilon = 0.25$ in the case of revenue sharing (Figure 3(d)). In the case of no revenue sharing, the number of legal users eventually present in the system is $23,153$, while the value anticipated from our theoretical analysis, specifically by Corollory 10, is $45,920$. In the revenue sharing setting, the final number of legal users in the system is $89,151$, while the value predicted by Lemma 9 is $96,380$.

The key point here is that when the illicit P2P is efficient, the gain from revenue sharing increases. In particular, the gain perceived by the content provider in terms of fractional revenue is over 180%. Note that this detailed contrast was not evident in the asymptotic results in Theorems 1 and 2.

Finally, Figure 3(e) illustrates the impact of the amount of revenue sharing on the fractional revenue ratio of the CDN in the cases of inefficient and efficient illicit P2Ps. We use $\kappa = 0.75$ in the simulation. The key point to observe in the figure is that there is a clear optimal amount of revenue sharing for the provider. In both cases, this amount is fairly small, however, it is clearly desirable to share more revenue in the presence of an efficient illicit P2P than in the presence of an inefficient illicit P2P. In fact, sharing nearly zero percent of the revenue still provides fairly close to the optimal fractional revenue in the inefficient case, while one must share more than 10% of the revenue to be near-optimal in the case of an efficient, illicit P2P.

4 Conclusion

Our goal in this paper is to quantify the ramifications of coopting legal P2P content sharing, not only as a means of reducing costs of content distribution, but, more importantly, as a way of hurting the performance of illegal P2P file sharing. The model that we propose internalizes the idea that demand for any content is transient, and that all content will eventually be available for free through illegal file sharing. The objective then is not to cling to ownership rights, but to extract as much revenue from legal copies as possible within the available time. We develop a revenue sharing scheme that recognizes the importance of early adopters in extending the duration of time that revenue may be extracted. In particular, keeping users from "going rogue" (becoming seeds in illegal networks) by allowing them to extract some revenue for themselves (and so defray part of their expense in purchasing the content in the first place), provides *order sense improvements* in the extractable revenue. We realize that our paradigm is contrary to the "conventional wisdom" of charging *more* rather than *less* to early adopters, and also to discourage file sharing using legal threats. However, as many recent studies have demonstrated, incentives work better than threats in human society, and adoption of our revenue sharing approach might result in a cooperative equilibrium between content owners, distributors and end-users. Future work includes a characterization of the exact value of users based on their times of joining the system, as well as considering content streaming, which requires strict quality of service guarantees.

References

[1] C. Labovitz, D. McPherson, and S. Iekel-Johnson, "2009 Internet Observatory report," in *NANOG-47*, October 2009.

[2] W. B. Norton, "Internet Video: The Next Wave of Massive Disruption to the U.S. Peering Ecosystem," 2007, Available at http://www.equinix.com.

[3] "Pando Networks, Inc." http://www.pando.com/, 2011.

[4] P. Rodriguez, S. Tan, and C. Gkantsidis, "On the feasibility of commercial, legal P2P content distribution," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, pp. 75–78, 2006.

[5] T. Karagiannis, P. Rodriguez, and K. Papagiannaki, "Should Internet service providers fear peer-assisted content distribution?" in *Proc. of the 5th ACM SIGCOMM conference on Internet Measurement*, 2005.

[6] D. Merugu, B. Prabhakar, and N. Rama, "An incentive mechanism for decongesting the roads: A pilot program in Bangalore," in *Proc. NetEcon, ACM Workshop on the Economics of Networked Systems*, July 2009.

[7] J. Wang, C. Yeo, V. Prabhakaran, and K. Ramchandran, "On the role of helpers in peer-to-peer file download systems: Design, analysis and simulation," in *Proc. IPTPS*, Feb. 2007.

[8] E. Setton and J. Apostolopoulos, "Towards quality of service for peer-to-peer video multicast," in *Proc. ICIP*, Sept. 2007.

[9] S. Liu, R. Zhang-Shen, W. Jiang, J. Rexford, and M. Chiang, "Performance bounds for peer-assisted live streaming," in *Proc. ACM SIGMETRICS*, June 2008.

[10] M. Chen, M. Ponec, S. Sengupta, J. Li, and P. A. Chou, "Utility maximization in peer-to-peer Systems," in *Proc. ACM SIGMETRICS*, June 2008.

[11] S. Shakkottai and R. Johari, "Demand aware content distribution on the Internet," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, April 2010.

[12] P. Parag, S. Shakkottai, and I. Menache, "Service routing in multi-ISP peer-to-peer content distribution: Local or remote?" in *Proc. of GameNets*, 2011.

[13] C. Aperjis and R. Johari, "A peer-to-peer system as an exchange economy," in *Proc. GameNets*, Oct. 2006.

[14] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié, "Incentivizing peer-assisted services: A fluid Shapley value approach," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 1, 2010, pp. 215–226.

[15] D. Qiu and R. Srikant, "Modeling and performance analysis of BitTorrent-like peer-to-peer networks," in *Proc. ACM SIGCOMM*, Aug. 2004.

[16] L. Massoulie and M. Vojnovic, "Coupon replication systems," in *Proc. ACM SIGMETRICS*, June 2005.

[17] 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.

[18] A. Griliches, "Hybrid Corn and the Economics of Innovation," *Science*, vol. 132, pp. 275–280, 1960.

[19] F. M. Bass, "A new product growth model for consumer durables," *Management Science*, vol. 15, pp. 215–227, 1969.

[20] M. J. Freedman, E. Freudenthal, and D. Mazières, "Democratizing content publication with Coral," in *Proc. NSDI*, Mar. 2004.

[21] H. Boswijk and P. Franses, "On the econometrics of the Bass diffusion model," *Journal of Business and Economic Statistics*, vol. 23, no. 3, pp. 255–268, 2005.

[22] R. Bewley and W. Griffiths, "The penetration of CDs in the sound recording market: issues in specification, model selection and forecasting," *International Journal of Forecasting*, vol. 19, no. 1, pp. 111–121, 2003.

[23] S. Sundqvist, L. Frank, and K. Puumalainen, “The effects of country characteristics, cultural similarity and adoption timing on the diffusion of wireless communications,” *Journal of Business Research*, vol. 58, no. 1, pp. 107–110, 2005.

[24] C. Chu and J. Pan, “The forecasting of the mobile Internet in Taiwan by diffusion model,” *Technological Forecasting and Social Change*, vol. 75, no. 7, pp. 1054–1067, 2008.

[25] H. Hu and X. Wang, “Evolution of a large online social network,” *Physics Letters A*, vol. 373, no. 12-13, pp. 1105–1110, 2009.

[26] G. Moore, *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers, Rev edition.* New York, NY: HarperBusiness, 1999.

A Proof of Theorem 1

To prove Theorem 1, we go through a sequence of intermediate results characterizing the number of legal/illegal users at the transition points of the approximate Bass model.

We start by characterizing the number of legal and illegal users at the end of Phase 1.

Lemma 4 *In the presence of an inefficient, illicit P2P, the number of illegal and legal users at the end of Phase* 1 *of the approximate Bass model are given by*

$$N_i(T_1) = \left(\frac{\rho I(0)}{\kappa-\rho} + \frac{N\rho}{(\kappa-\rho)^2}\right)\exp\left(B_N\right) - \frac{I(T_1)\rho}{\kappa-\rho} - \frac{N\rho}{(\kappa-\rho)^2} \tag{16}$$

$$N_l(T_1) = I(T_1) - N_i(T_1), \tag{17}$$

$$\text{where} \quad I(T_1) = \frac{N}{\ln N}\frac{N}{N - I(0) + (N/\ln N)} \quad \text{and} \quad B_N = \left(\frac{(\kappa-\rho)}{N}(I(T_1) - I(0))\right).$$

Note that in the above, we have allowed κ, ρ, and β to be arbitrary. In fact, in this case, β is inconsequential since the full amount of interested users can be served by the dedicated capacity of the CDN. Note that in the case when $\rho = \kappa$, things simplify considerably.

Corollary 5 *Let $\rho = \kappa$. In the presence of an inefficient, illicit P2P, the number of illegal and legal users at the end of Phase* 1 *of the approximate Bass model are given by*

$$N_i(T_1) = \frac{\kappa(I^2(T_1) - I^2(0))}{2N} \quad \text{and} \quad N_l(T_1) = I(T_1) - N_i(T_1),$$

where $I(T_1) = \frac{N}{\ln N}\frac{N}{N-I(0)+(N/\ln N)}$.

We now prove the lemma.

Proof Lemma 4. From (3), the population of interested users in phase I is given by

$$I(t) = \frac{NI(0)e^t}{N - I(0) + I(0)e^t}. \tag{18}$$

From the above equation, it is easy to verify that the rate of growth of interested users is less than the server capacity C_N, i.e., $dI(t)/dt \leq C_N$. Thus, any interested user is served instantaneously either by a legal or illegal mechanism. Hence, the number of Wanters in the system is zero, i.e, $N_w(t) = 0$. Therefore, it follows from (4) that $N_l(t) + N_i(t) = I(t)$.

Next, from equation (8), we get that

$$\frac{dN_i(t)}{dt} = \min\left\{\eta(t)\frac{dI(t)}{dt}, N_r(t) + N_f(t)\right\} \overset{(a)}{=} \eta(t)\frac{dI(t)}{dt}, \tag{19}$$

where the equality (a) follows from the definition of $\eta(t)$ and the fact that $dI(t)/dt \leq C_N < N$. Because we are considering an inefficient P2P, we have

$$\eta(t) \;=\; \tfrac{N_r(t)+N_f(t)}{N} \overset{(b)}{=} \tfrac{\rho N_l(t)+\kappa N_i(t)}{N} \overset{(c)}{=} \tfrac{\rho(I(t)-N_i(t))}{N} + \tfrac{\kappa N_i(t)}{N} = \tfrac{\rho I(t)}{N} + \tfrac{(\kappa-\rho)N_i(t)}{N}. \tag{20}$$

where equality (b) follows from (5), (6) and the equality (c) follows from the fact that $N_l(t) = I(t) - N_i(t)$. Substituting the above result in equation (19), we get

$$\frac{dN_i(t)}{dt} = \frac{dI(t)}{dt}\frac{\rho I(t)}{N} + \frac{dI(t)}{dt}\frac{(\kappa-\rho)N_i(t)}{N}.$$

The solution of the above differential equation with the initial condition, $N_i(0) = 0$, is given by

$$N_i(t) = \left(\frac{\rho I(0)}{\kappa-\rho} + \frac{N\rho}{(\kappa-\rho)^2}\right)\exp\left(\frac{(\kappa-\rho)}{N}(I(t)-I(0))\right) - \frac{\rho I(t)}{\kappa-\rho} - \frac{N\rho}{(\kappa-\rho)^2}.$$

The number of illegal users at the end of Phase 1 can be obtained by evaluating the above expression at $t = T_1$. The remaining population get the content legally, i.e, $N_l(T_1) = I(T_1) - N_i(T_1)$. ■

Now that we have characterized the number of legal and illegal users at the end of Phase 1, we can move to Phases 2-4. Unfortunately, the resulting number of legal and illegal users at the end of these phases is much more complicated. However, much of this complicated form is only necessary to specify the exact analytic values. Once we focus on the asymptotic form (as in Theorem 1), it simplifies considerably.

Before stating the result, we need to introduce a considerable amount of notation. This notation stems from the fact that we do not analyze the exact process of $N_l(t)$ and $N_i(t)$. Instead, we define a processes $\bar{N}_l(t)$ and $\bar{N}_i(t)$ which bounds $N_l(t)$ and $N_i(t)$ and analyze these processes. Importantly, the bounding processes are equivalent to the original processes when $\beta = 0$, i.e., the case of no revenue sharing.

Before defining $\bar{N}_l$ and $\bar{N}_i$, we need some notation. Let

$$\Delta\bar{\tau}_2 = \frac{1}{\kappa \ln N Z_1}\ln\left(\frac{Z_1 + 1 - \frac{2I(T_1)}{(N/\ln N)}}{Z_1 - 1 + \frac{2I(T_1)}{(N/\ln N)}}\right) + \frac{1}{\kappa \ln N Z_1}\ln\left(\frac{Z_1+1}{Z_1-1}\right), \tag{21}$$

$$\Delta\bar{\tau}_3 = \frac{2}{\kappa Z_2}\ln\left(\frac{Z_2 + 1 - \frac{4}{\ln N}}{Z_2 - 1 + \frac{4}{\ln N}}\right) + \frac{2}{\kappa Z_2}\ln\left(\frac{Z_2+1}{Z_2-1}\right), \quad , \Delta\bar{\tau}_4 = \frac{1}{\kappa Z_3}\ln\left(\frac{Z_3+1}{Z_3-1}\right), \tag{22}$$

where $Z_1 = \sqrt{1 + \frac{4 \ln N}{\kappa}}$, $Z_2 = \sqrt{1 + \frac{16}{\kappa \ln N}}$, $Z_3 = \sqrt{1 + \frac{4}{\kappa \ln N}}$ and $I(T_1) = \frac{N}{\ln N} \frac{N}{N - I(0) + (N/\ln N)}$. In addition, let

$$\theta_1^j = \kappa \frac{I_j}{2N} + \frac{1}{2}\sqrt{\left(\frac{\kappa I_j}{N}\right)^2 + \frac{4\kappa}{\ln N}}, \quad \theta_2^j = \kappa \frac{I_j}{2N} - \frac{1}{2}\sqrt{\left(\frac{\kappa I_j}{N}\right)^2 + \frac{4\kappa}{\ln N}},$$

$$b_j = \frac{N\theta_1^j - \kappa I(T_{j-1})}{\kappa I(T_{j-1}) - N\theta_2^j} \tag{23}$$

and $\Delta\theta_j = \theta_1^j - \theta_2^j$. Note that, in the above definition, in fact $I(T_{j-1}) = I_{j-1}$ for $j = 3$ and 4. Furthermore, for $j = 2, 3$ and 4, let $d_j = (b_j + \exp(\Delta\theta_j \Delta\bar{\tau}_j))$, $q_1^j = \left(\frac{\beta\theta_2^j}{\kappa} - \frac{\beta I_j}{N}\right)$ and $q_2^j = \frac{\beta\theta_1^j}{\kappa} - \frac{\beta I_j}{N}$

Finally, we are ready to define the bounding processes used in the proof, $\bar{N}_l(t)$ and $\bar{N}_i(t)$. Let $\bar{N}_i(T_1) = N_i(T_1)$. Furthermore, during Phase j, let

$$\frac{d\bar{N}_i(t)}{dt} = \frac{\rho\bar{N}_l(t) + \kappa\bar{N}_i(t)}{N}(I_j - (\bar{N}_l(t) + \bar{N}_i(t))). \tag{24}$$

Similarly, let $\bar{N}_l(T_1) = N_l(T_1)$ and, during Phase j,

$$\frac{d\bar{N}_l(t)}{dt} = \begin{cases} C_N + \beta\bar{N}_l(t)\frac{I_j - (\bar{N}_l(t) + \bar{N}_i(t))}{N}, & \bar{N}_w(t) > 0, \\ 0, & \bar{N}_w(t) = 0. \end{cases} \tag{25}$$

where $\bar{N}_w(t) = I_j - (\bar{N}_i(t) + \bar{N}_l(t))$. Finally, let $\bar{U}(t) = \bar{N}_l(t) + \bar{N}_i(t)$.

To state the result, we use a bit more notation about these processes. Let $\bar{N}_l^1 = N_l(T_1)$ and for $j = 2, 3$, and 4 define $\bar{N}_l(T_j)$ recursively as follows:

$$\begin{aligned} \bar{N}_l^j \quad &= \bar{N}_l^{j-1}\left(\frac{1+b_j}{d_j}\right)^{\frac{\beta}{\kappa}} e^{(-q_1^j\Delta\bar{\tau}_j)} + C_N\left(\frac{b_j}{d_j}\right)^{\frac{\beta}{\kappa}} e^{\left(-q_1^j\Delta\bar{\tau}_j\right)}\left(\frac{e^{\left(q_1^j\frac{\ln b_j}{\Delta\theta_j}\right)}}{q_1^j} - \frac{1}{q_1^j}\right)\mathbf{1}_{b\geq 1} \\ &+ C_N\left(\frac{1}{d_j}\right)^{\frac{\beta}{\kappa}} e^{\left(-q_1^j\Delta\bar{\tau}_j\right)}\left(\frac{e^{\left(q_2^j\Delta\bar{\tau}_j\right)}}{q_2^j} - \frac{e^{\left(\frac{q_2^j\ln b_j}{\Delta\theta_j}\right)}\mathbf{1}_{b\geq 1}}{q_2^j}\right) - C_N\left(\frac{1}{d_j}\right)^{\frac{\beta}{\kappa}} e^{\left(-q_1^j\Delta\bar{\tau}_j\right)}\frac{1}{q_2^j}(1 - \mathbf{1}_{b\geq 1}), \end{aligned} \tag{26}$$

where $\mathbf{1}_{b\geq 1} = 1$ when $b \geq 1$ and zero otherwise.

We can now state our result characterizing the number of legal and illegal users at the end of Phases 2-4.

Lemma 6 *Let $\rho + \beta = \kappa$. In the presence of an inefficient, illicit P2P, the number of illegal and legal users at the end of Phase j, $j \in \{2, 3, 4\}$ of the approximate Bass model are given by*

$$N_l(T_j) \geq \bar{N}_l^j,$$

where equality holds when $\beta = 0$.

Proof. From the approximate Bass model (3), the evolution of demand in Phase j is, $I(t) = I_j, \quad \text{where} \quad t \in (T_{j-1}, T_j]$,and the number of Wanters in Phase j is $N_w(t) = I_j - (N_l(t) + N_i(t))$. Recall that the efficiency factor of an inefficient illicit P2P, $\eta(t)$, is given by

$$\eta(t) = \frac{N_r(t) + N_f(t)}{N} = \frac{\rho N_l(t) + \kappa N_i(t)}{N}. \tag{27}$$

The second equality follows from (5) and (6). From (8), the illegal growth rate in Phase j is

$$\frac{dN_i(t)}{dt} \overset{(a)}{=} \min\left\{\eta(t)N_w(t), N_r(t) + N_f(t)\right\} \overset{(b)}{=} \eta(t)N_w(t) \overset{(c)}{=} \frac{(\rho N_l(t)+\kappa N_i(t))(I_j-(N_l(t)+N_i(t)))}{N}. \tag{28}$$

Here (a) follows from the fact that $I(t)$ is constant in the last three phases. (b) follows from the definition of $\eta(t)$ and the fact that $N_w(t) \leq N$. (c) follows from (27).

From equation (9), the growth rate of legal users in Phase j is given by

$$\frac{dN_l(t)}{dt} = \begin{cases} C_N + \beta N_l(t), & N_w(t) > 0, \\ 0, & N_w(t) = 0. \end{cases} \tag{29}$$

The second equality follows from the fact that $\frac{dN_i}{dt} = 0$ when there are no Wanters in the system (from (28)) and $I(t)$ is constant. Let $U(t)$ be the total copies of the content in the system. Then, $U(t) = N_l(t) + N_i(t)$.

Note that the growth rate $N_l(t)$ is at least equal to C_N when $N_w(t) > 0$. In that case, it can be shown that $C_N \times (T_j - T_{j-1}) > (I(T_j) - I(T_{j-1}))$, if $I(0) < \sqrt{N}$, which is in fact true by assumption. This means that every interested user generated in any one of the last three phases can be served within that phase itself. Furthermore, Lemma 4 shows that no Wanters are left unserved after Phase 1. Therefore, we can conclude that

$$N_l(T_j) + N_i(T_j) = U(T_j) = I(T_j) = I_j. \tag{30}$$

The same arguments hold true forf $\bar{N}_l(t)$, i.e,

$$\bar{N}_l(T_j) + \bar{N}_i(T_j) = \bar{U}(T_j) = I(T_j) = I_j. \tag{31}$$

Now, we claim that,

$$N_l(T_j) \geq \bar{N}_l(T_j), \tag{32}$$

and the equality holds when $\beta = 0$.

The proof is as follows: We can derive $\frac{dN_i}{dU}$ and $\frac{d\bar{N}_i}{d\bar{U}}$ from the pair of equations (28), (29) and (24), (25) respectively. Then, it can be shown that

$$\frac{dN_i}{dU}|_{N_i=x,U=y} \leq \frac{d\bar{N}_i}{d\bar{U}}|_{\bar{N}_i=x,\bar{U}=y}, \tag{33}$$

and the equality holds when $\beta = 0$. Note that the range space of functions $U(t)$ and $\bar{U}(t)$ are identical; in fact they are equal to $[I(T_{j-1}), I(T_j)]$ in Phase j which follows from (30) and (31). Furthermore, recall that the initial values of $N_i(T_1)$ and $\bar{N}_i(T_1)$ are equal by definition. Hence, the conclusion is, $N_i(T_j) \leq \bar{N}_i(T_j)$. Then, the claim in (32) is true from the facts that $N_l(T_j) = I(T_j) - N_i(T_j)$ and $\bar{N}_l(T_j) = I(T_j) - \bar{N}_i(T_j)$.

Our objective is to derive an expression of $\bar{N}_l(t)$. Then, evaluate the expression at $t = T_j$ in order to obtain a lower bound on the number of legal users at the end of each Phase j. Let $\bar{\tau}_j$ be the time such that $\bar{U}(\bar{\tau}_j) = I_j$. This event happens within Phase j itself (from (31)). i.e, $\bar{\tau}_j \in (T_{j-1}, T_j]$. In addition, $\bar{N}_w(t) = 0$ when $t \in (\bar{\tau}_j, T_j]$. Adding (25) and (24), for $t \in (T_{j-1}, \bar{\tau}_j]$, we get,

$$\frac{d\bar{U}}{dt} = \left((\beta+\rho)\bar{N}_l(t) + \kappa\bar{N}_i(t)\right)\frac{(I_j - (\bar{N}_l(t)+\bar{N}_i(t)))}{N} \overset{(f)}{=} \kappa\bar{U}(t)\frac{I_j - \bar{U}(t)}{N}.$$

(f) follows from the fact that $\rho + \beta = \kappa$ and the definition of $\bar{U}(t)$ in Phase j. The above differential equation is in the form of a standard Riccatti equation, and it's solution can be written as

$$\bar{U}(t) = \frac{N\theta_{2,j}}{\kappa} + \frac{N\Delta\theta_j/\kappa}{1 + b_j e^{-\Delta\theta_j(t-T_{j-1})}}, \tag{34}$$

where $\Delta\theta_j = \theta_{1,j} - \theta_{2,j}$. $\theta_{1,j}, \theta_{2,j}$ and b_j are given by equations (23) and respectively.

Let $\Delta\bar{\tau}_j = \bar{\tau}_j - T_{j-1}$. Recall that $\bar{\tau}_j$ is the solution of the equation $\bar{U}(\bar{\tau}_j) = I_j$. Hence, from the above result, we get,

$$\bar{\tau}_j - T_{j-1} = \frac{1}{\Delta\theta_j}\ln\left(\frac{\sqrt{1+\frac{4C_N N}{\kappa I(T_j)^2}} + 1 - \frac{2I(T_{j-1})}{I(T_j)}}{\sqrt{1+\frac{4C_N N}{\kappa I(T_j)^2}} - 1 + \frac{2I(T_{j-1})}{I(T_j)}}\right) + \frac{1}{\Delta\theta_j}\ln\left(\frac{\sqrt{1+\frac{4C_N N}{\kappa I(T_j)^2}} + 1}{\sqrt{1+\frac{4C_N N}{\kappa I(T_j)^2}} - 1}\right). \tag{35}$$

The above expression yields the set of equations given by (22), by substituting $I(T_j)$ from the bass model. Now, applying the above expression in (25), for $t \in (T_{j-1}, \bar{\tau}_j]$, we get

$$\frac{d\bar{N}_l(t)}{dt} = C_N + \beta\bar{N}_l(t)\frac{I_j - (\bar{N}_l(t)+\bar{N}_i(t))}{N}.$$

A lower bound on the solution of the above differential equation is provided by Lemma 11 in Appendix C. It can be shown that $b\exp(-\Delta\theta_j\Delta\bar{\tau}_j) << 1$. Then $\bar{\tau}_j$ satisfies the condition stipulated by that lemma and a lower bound on the number of legal at the end of Phase j can be obtained by evaluating (49) at $t = \bar{\tau}_j$, which yields $\bar{N}_l^j$ in (26). When $\beta = 0$, (49) is an exact solution of the above differential equation, which completes the proof. ∎

As mentioned in the statement of Lemma 6, the inequality is exact in the case of $\beta = 0$. Additionally, in this case, the form of $N_l(T_4)$ simplifies.

Corollary 7 *Let $\beta = 0$. In the presence of an inefficient, illicit P2P, the number of illegal and legal users at the end of Phase 4 of the approximate Bass model is given by*

$$N_l(T_4) = N_l(T_1) + C_N\sum_{j=2}^{4}\Delta\bar{\tau}_j \tag{36}$$

where $N_l(T_1)$ is given by Corollary 5.

Now that we have characterized the number of legal and illegal users at the end of Phase 4 precisely, attaining the statement in Theorem 1 is accomplished by taking studying the asymptotics of the results in Lemma 6 and Corollary 7.

To begin, recall from (10) that the revenue ratio is

$$R = \frac{N_l(T_4)}{N}(1-\epsilon) \geq \frac{\bar{N}_l^4}{N}(1-\epsilon), \tag{37}$$

where $\bar{N}_l^4$ is recursively defined by (26) in terms of $\bar{N}_l^1, \bar{N}_l^2$ and $\bar{N}_l^3$. From (37) and (26), after following a few straightforward but lengthy algebraic steps, we can show that

$$R \in \Omega\left(\frac{1-\epsilon}{\ln N}\left(\ln\ln N + (\ln N)^{1-N^{-\epsilon}}\right)\right), \quad \text{if} \quad \epsilon > 0 \tag{38}$$

$$R \in \Theta\left(\frac{\ln\ln N}{\ln N}\right), \quad \text{if} \quad \epsilon = 0, \tag{39}$$

when $\beta = \kappa(1 - N^{-\epsilon})$.

B Proof of Theorem 2

The proof of Theorem 2 parallels to that of Theorem 1; therefore, due to space constraints we provide only a sketch of the proof in this section.

As in the case of Theorem 1, to prove Theorem 2, we go through a sequence of intermediate results characterizing the number of legal/illegal users at the transition points of the approximate Bass model.

To start, we characterize the number of legal and illegal users at the end of Phase 1.

Lemma 8 *Let $\kappa \in (\frac{C_N}{N}, 1 - \frac{C_N}{N})$. Then, in the presence of an efficient, illicit P2P, the number of legal users at the end of Phase 1 satisfies*

$$N_l(T_1) \geq \begin{cases} I(T_1) - \rho N \ln\left(\frac{N - I(0) + C_N}{N}\right) \; \textit{if} \quad \rho = \kappa \quad \textit{else} \\ \max\left\{0, I(T_1) - \frac{\rho N^2\left(1 - \left(\frac{N}{\ln N I(0)}\right)^{(\kappa-\rho-1)}\right)}{\ln N(N-I(0))(1-\beta)}\right\}, \end{cases} \tag{40}$$

where the equality holds when $\rho = \kappa$.

The proof of this lemma parallels that of Lemma 4 used in the proof of Theorem 1; so, due to space constraints, we omit it.

Given the characterization of the number of legal and illegal users at the end of Phase 1, we now move to Phases 2-4. Similarly to the case of inefficient illicit P2P, the differential equations characterizing the growth of legals and illegals are hard to analyze. Hence, we mimick the approach of the proof of Theorem 2 and define two processes $\bar{N}_l(t)$ and $\bar{N}_i(t)$ that bound $N_l(t)$ and $N_i(t)$ and analyze these processes. Importantly, the bounding processes are equivalent to the original processes when $\beta = 0$, i.e., the case of no revenue sharing.

Let $\bar{U}(t) = \bar{N}_l(t) + \bar{N}_i(t)$. Further, define

$$\frac{d\bar{N}_l(t)}{dt} = = \begin{cases} C_N + \beta\bar{N}_l(t) & \bar{N}_w(t) > 0, \\ 0 & \bar{N}_w(t) = 0. \end{cases} \tag{41}$$

where $\bar{N}_w(t) = I(t) - \bar{U}(t)$. Also, let $\bar{N}_l(T_1)$ be equal to the lower bound on the number illegals at the end of Phase 1, derived by Lemma 8. Furthermore, in each phase, we choose the growth rate of $\bar{N}_i(t)$ as follows:
Phase 2 :

$$\frac{d\bar{N}_i(t)}{dt} = I_2 - \bar{U}(t) = I_2 - \bar{N}_l(t) - \bar{N}_i(t). \tag{42}$$

Phase 3, 4 : For $j = 3$ and 4, let

$$\frac{d\bar{N}_i(t)}{dt} = \begin{cases} \rho\bar{N}_l(t) + \kappa\bar{N}_i(t) & I(T_{j-1}) \leq \bar{U}(t) \leq \frac{I(T_j)}{1+\rho}, \\ I(T_j) - \bar{N}_l(t) - \bar{N}_i(t) & \frac{I(T_j)}{1+\rho} \leq \bar{U}(t) \leq I(T_j). \end{cases} \tag{43}$$

Finally, let $\bar{N}_i(T_1) = I(T_1) - \bar{N}_l(T_1)$. To state the results, we may need a bit more notation. Let $\bar{N}_1^1 = \bar{N}_l(T_1)$ and for $j = 2, 3$ and 4, let

$$\bar{N}_l^j = \left(\bar{N}_1^1 + \frac{N}{\ln N\beta}\right) e^{\beta\Delta\bar{\tau}_j} - \frac{N}{\ln N\beta}. \tag{44}$$

Furthermore, $\Delta\bar{\tau}_2 = \frac{1}{1+\beta}\ln\left(1 + \frac{\frac{N}{\ln N} - I(T_1)}{G_1}\right)$, $\Delta\bar{\tau}_3 = \frac{1}{1+\beta}\ln\left(1 + \frac{N\rho}{2(1+\rho)}\frac{H_3^{\frac{-\beta}{\kappa}}}{G_2}\right) + \frac{1}{\kappa}\ln(H_3)$, and $\Delta\bar{\tau}_4 = \frac{1}{1+\beta}\ln\left(1 + \frac{N\rho}{(1+\rho)}\frac{H_4^{\frac{-\beta}{\kappa}}}{G_3}\right) + \frac{1}{\kappa}\ln(H_4)$,where $G_j = \frac{\beta}{1+\beta}\left(\bar{N}_l(T_j) + \frac{N}{\ln N\beta}\right)$ for $j = 1, 2, 3$ and $H_i = \left(\frac{\frac{I_i}{1+\rho} + \frac{N}{\ln N\kappa}}{I_{i-1} + \frac{N}{\ln N\kappa}}\right)$. for $i = 3, 4$. Now, we characterize the number of legal users and illegal users at the end of Phase $2 - 4$ in the following lemma.

Lemma 9 *Let $\beta + \rho = \kappa$ and $\kappa \in (\frac{C_N}{N}, 1 - \frac{C_N}{N})$. Then, in the presence of an efficient, illicit P2P, the number of illegal users at the end of Phase j satisfies*

$$N_l(T_j) \geq \bar{N}_l^j, \tag{45}$$

and the equality holds when $\kappa = \rho$.

The proof of this lemma is too long to fit in the space provided, however its structure parallels that of the proof of Lemma 6 in the proof of Theorem 1.

As mentioned in the statement of Lemma 9, the inequality is exact in the case of $\beta = 0$. Additionally, in this case, the form of $N_l(T_4)$ simplifies.

Corollary 10 *Let $\rho = \kappa$. Then, the number of legal users at the end of Phase 4 is given by $N_l(T_4) = \bar{N}_l(T_1) + C_N \sum_{j=2}^{4} \Delta\bar{\tau}_j$, where $\bar{N}_l(T_1) = I(T_1) - \rho N \ln\left(\frac{N - I(0) + C_N}{N}\right)$.*

Now that we have characterized the number of legal and illegal users at the end of Phase 4 precisely, attaining the statement in Theorem 2 is accomplished by taking studying the asymptotics of the results in Lemma 9 and Corollary 10.

From (10), the fractional revenue ratio is

$$R = \frac{N_l(T_4)}{N}(1 - \epsilon) \geq \frac{\bar{N}_l^4}{N}(1 - \epsilon). \tag{46}$$

From Lemma 8, Lemma 9, Corollary 10 and equation (44) we can show that

$$R \in \Omega\left(\frac{(1-\epsilon)}{\ln N}\frac{(\log N)^{1-N^{-\epsilon}}-1}{1-N^{-\epsilon}}\right), \quad \text{if} \quad \epsilon > 0 \tag{47}$$

$$R \in \Theta\left(\frac{\ln\ln N}{\ln N}\right) \quad \text{if} \quad \epsilon = 0. \tag{48}$$

when $\rho = \kappa N^{-\epsilon}$ and $\beta = \kappa(1-N^{-\epsilon})$, which completes the proof.

C Technical lemmas

Lemma 11 *Consider a differential equation given by* $\frac{dy}{dt} = C_N + \frac{\beta y}{N}(I - U(t))$ *where* $U(t) = \frac{N\theta_2}{\kappa} + \frac{N\Delta\theta/\kappa}{1+be^{-\Delta\theta(t-T)}}$*.Also, assume that* $y(T)$ *is given. Then for all* $t - T > \frac{\ln b}{\Delta\theta}$*, the solution to the above differential equation satisfies the inequality*

$$\begin{aligned} y(t) \quad & \geq y(T)\left(\tfrac{1+b}{d}\right)^{\frac{\beta}{\kappa}} e^{(-q_1(t-T))} + C_N\left(\tfrac{b}{d}\right)^{\frac{\beta}{\kappa}} e^{(-q_1(t-T))}\left(\frac{e^{\left(q_1\frac{\ln b}{\Delta\theta}\right)}}{q_1} - \frac{1}{q_1}\right)\mathbf{1}_{b\geq 1} \\ & + C_N\left(\tfrac{1}{d}\right)^{\frac{\beta}{\kappa}} e^{(-q_1(t-T))}\left(\frac{e^{\left(q_2\Delta\tau_j\right)}}{q_2} - \frac{e^{\left(q_2\frac{\ln b}{\Delta\theta}\right)}}{q_2}\mathbf{1}_{b\geq 1}\right) - C_N\left(\tfrac{1}{d}\right)^{\frac{\beta}{\kappa}} e^{(-q_1(t-T))}\tfrac{1}{q_2}(1-\mathbf{1}_{b\geq 1}), \end{aligned} \tag{49}$$

where $d = (b + \exp(\Delta\theta(t-T)))$*,* $q_1 = \left(\frac{\beta\theta_2}{\kappa} - \frac{\beta I}{N}\right)$ *and* $q_2 = \frac{\beta\theta_1}{\kappa} - \frac{\beta I}{N}$ *and* $\mathbf{1}_{b\geq 1} = 1$ *if* $b \geq 1$ *and zero otherwise. Furthermore, for* $\beta = 0$*, equality holds.*

Proof. A general solution to the above differential equation is

$$y(t) = \frac{\int C_N \exp(\int P dt) + M}{\int P dt} \tag{50}$$

where $P(t) = -\frac{\beta}{N}(I - U(t))$. Taking the integral of $P(t)$, we can show that $C_N e^{\int P dt} = C_N B(t)\exp\left(\frac{\beta\theta_2}{\kappa} - \frac{\beta I t}{N}\right)t$, where $B(t) = (1 + (1/b)\exp(\Delta\theta(t-T)))^{\frac{\beta}{\kappa}}$. For $b \geq 1$, we can lower bound $B(t)$ as

$$B(t) \geq \begin{cases} 1 & t \leq \frac{\ln b}{\Delta\theta} + T \\ \left(\frac{1}{b}\right)^{\frac{\beta}{\kappa}}\exp\left(\frac{\beta}{\kappa}\Delta\theta(t-T)\right) & t > \frac{\ln b}{\Delta\theta} + T. \end{cases} \tag{51}$$

On the other hand, if $b < 1$,

$$B(t) \geq \left(\frac{1}{b}\right)^{\frac{\beta}{\kappa}}\exp\left(\frac{\beta}{\kappa}\Delta\theta(t-T)\right), \quad \forall t. \tag{52}$$

Let us define $A(t)$ as $A(t) = \int C_N e^{\int P dt} dt$. From the bound on $B(t)$, we can show that, for $t > \frac{\ln b}{\Delta\theta} + T$,

$$A(t) \quad \geq C_N e^{(q_1 T)}\frac{\exp\left(q_1\frac{\ln b}{\Delta\theta}\right)}{q_1}\mathbf{1}_{b\geq 1} + C_N e^{(q_1 t)}\left(\tfrac{1}{b}\right)^{\frac{\beta}{\kappa}}\frac{\exp\left(\frac{\beta\Delta\theta}{\kappa}(t-T)\right)}{q_2} - C_N e^{(q_1 T)}\left(\tfrac{1}{b}\right)^{\frac{\beta}{\kappa}}\frac{\exp\left(q_2\frac{\ln b}{\Delta\theta}\right)}{q_2}\mathbf{1}_{b\geq 1}.$$

where $\mathbf{1}_{b\geq 1}$ is the indicator function of the event $b \geq 1$. We obtain M from the initial conditions, then apply the above result in (50) to get (49). For $\beta = 0$, (51) and (52) hold equality, which completes the proof. ■