Predicting the Impact of Measures against P2P Networks on the Transient Behavior of a Torrent

Eitan Altman[1], Philippe Nain[1], Adam Shwartz[2] and Yuedong Xu[1]

[1]: INRIA, BP 93, 06902 Sophia Antipolis, France
{Eitan.Altman,Philippe.Nain,Yuedong.Xu}@sophia.inria.fr
[2]: Department of Electrical Engineering Technion, Haifa 32000, Israel
adam@ee.technion.ac.il

Abstract

We analyze in this paper the transient behavior of a simplified model for torrents in P2P networks. We consider a large number of peers interested in a file which is initially available at a small fraction of the population. Our model has as parameters the number of interested peers, the rate of arrival of download request, the rate at which peers propose files for upload, the degree of the free riding phenomenon etc. Our goal is to predict the impact of measures against P2P networks on their performance. We provide a Markov model that describes the detailed dynamics of the system. We analyze the model first under the assumption of an infinite population—an approach based on a branching process. This process is shown to provide a bound for the behavior of the system with finite population. We obtain the extinction probabilities for this process as well as approximations to the expected time till extinction. We then study the case of a finite population for which we establish the mean field limit. We show that the limit simplifies to a well studied epidemic model in case of no free riders. The fluid limit allows us to evaluate the behavior of the expected duration of a torrent, the expected number of nodes that eventually receive the file and the maximal availability along with the time at which it is achieved. We finally use these results to evaluate the efficiency of measures against unauthorized file sharing and their impact on file bundling.

Keywords:
Peer-to-peer, File availability, Markov branching process, Mean field approximation.

1. Introduction

Along with the worldwide penetration of the Internet, a huge demand has appeared to copyrighted music and movies that have been accessible for free over the Internet. While benefitting a very large internaut community as well as potentially providing higher benefits for Internet access providers, it seems unclear whether the creators and the copyright owners have gained anything from this unregulated access. Two main opposing sets of approaches have been proposed to protect the copyright owners. The first consists of fighting against the nonauthorized access whereas the second one, on the contrary, aims at finding cooperative solutions that would benefit both the Internauts as well as all the other economic actors. An example of a cooperative solution is a flat taxation that

would allow Internauts to pursue downloading freely music and films, and that would distribute the tax money between the copyright owners. This cooperative approach has several difficulties in its implementation; a major one is how to distribute the tax income fairly. A major drawback of the confrontation policy is the huge monitoring effort that it requires and that seems not to provide credible evidence for unauthorized downloads [17]. In order to assess the efficiency of non-cooperative measures against unauthorized downloads, the authors of [23, 22] have analyzed the impact of the effort, of the authorities or of content provider companies, invested in (i) reducing file uploading in P2P networks and in (ii) reducing the demand for files, on the availability of files and, more generally, on the operation of the P2P networks. The analysis there was based on a stationary model.

In this paper we are interested in predicting the impact of measures, as described in the previous paragraph, on the *transient* behavior of torrents. By how much should the upload or download rate in a P2P network be reduced in order to have a significant change in file availability? To do that, we consider in this paper a simplified model of a torrent in a simplified P2P network. We consider a large number of peers interested in a file which is initially available at a small fraction of the population. Our model accounts for the standard epidemic type process of file dissemination. In addition it accounts for the free riding phenomenon and for network characteristics.

We provide a Markov description of the system and then approximate its behavior in two specific regimes. In the first, we consider a very large population that does not yet have the file. The system is then well approximated by a branching process. We compute for this regime the probability of extinction and provide upper and lower bounds to the expected time until extinction. In the case that there is a positive probability of not getting extinct, the system is shown to move with some non-zero probability to a second regime in which, for the case of a sufficiently large population size, its dynamics is close to the solution of a differential equation that is well known in the area of propagation of epidemics [14]. The analysis allows us to compute approximations for the expected duration of a torrent, the expected number of nodes that eventually receive the file and the maximal availability along with the time at which it is achieved. Most importantly, these two regimes are complimentary so as to characterize the file sharing more comprehensively.

P2P protocols often try to provide incentives for sharing a file by making it available to other peers, once it is downloaded by a peer. This is considered by the P2P community to be a socially cooperative behavior whereas a peer leaving the system immediately after receiving the file is considered to be a free rider. From the legal point of view, making unauthorized copies of copyrighted files available to other users is a more severe copyright infringement and is more severely punished than downloading unauthorized copies. To understand the impact of measures against the cooperative sharing behavior, we quantify the degree of free-riders in the system as well as the degree of cooperation (the latter is based on the model proposed in [16]). Another interesting phenomenon in P2P swarms is that files are usually bundled together for downloading. Our model can also be extended to study the dynamic and deterministic behaviors of epidemic-like file sharing with bundling.

The structure of the paper is as follows. In section 2 we model the P2P network as a jump Markov process. We then provide two approximations of this model for the case of a large number of peers. The first, in Section 3, is based on a branching process.

It is used for the case that the fraction of nodes with the file is close to zero. The second, in Section 4, is based on an ordinary set of differential equations, used when the fraction of nodes with the file is bounded away from zero. Section 5 studies the phase transitions of P2P file sharing with a fixed request rate. We evaluate the impact of measures against Internet piracy on the performance of P2P systems and discuss the impact of coutnermeasures (bundling) in section 6. We end with a concluding section.

2. Model

2.1. Assumptions

Assume there is a population N of (possibly mobile) peers interested in a single file. We denote by $Y(t)$ the number of peers that possess the file at time t. A peer acquires the file when it encounters another peer that has the file. We will consider two types of peers: *cooperative* and *non-cooperative* peers. Once a cooperative peer has acquired the file it stays in the network for a random time distributed according to an exponential rv with parameter $1/\mu \geq 0$ and then leaves the network. During the time a cooperative stays in the network with the file it participates in the file dissemination. A non-cooperative peer, also called a *free-rider*, leaves the network at once when it receives the file.

Let $Y(t)$, $X_c(t)$ and $X_f(t)$ denote the number of cooperative peers with the file, the number of cooperative peers without the file and the number of free-riders (necessarily without the file) at time t, respectively. Let $(y_0, x_{c,0}, x_{f,0}) := (Y(0), X_c(0), X_f(0))$ denote the initial state of $\mathbf{Y}$.

We assume that $y_0 + x_{c,0} + x_{f,0} = N$ and denote by $x_{c,0} := N_c$ the number of cooperative peers without the file at time $t = 0$. We further assume that $y_0 \in \{1, \ldots, N - 1\}$ ($y_0 = 0$ implies that $Y(t) = 0$ for all t and $y_0 = N$ implies that $X_c(t) = X_f(t) = 0$ for all t, two cases which have no interest). Let r denote the ratio of cooperative peers that has $r = (y_0 + x_{c,0})/N$.

It is assumed that it takes an exponential time with rate $\lambda > 0$ for a peer without a file to encounter a cooperative peer with the file. When two such peers meet, the cooperative peer transmits the file to the other peer. The transmission of the file is always supposed to be successful. We assume that the file transmission time is neglectible compared to the time it takes for two peers to meet and therefore this time is taken to be zero.

All the rvs introduced so far are assumed to be mutually independent. As a consequence, if $Y(t) = k$ then any peer without the file will meet a cooperative peer with the file after a time that is distributed according to the minimum of k independent and exponential rvs with rate λ, that is after a time distributed according to an exponential rv with rate λk. Let $\rho := \lambda N_c / \mu$.

Measures of the authorities or of content provider companies against file sharing systems may have an impact on the decrease in the population N interested in the file and an increase in the fraction of free riders among the population interested in the file. It can however have an impact also on the behavior of cooperative peers that would leave the system sooner (i.e. μ is expected to increase). Our model combines an epidemic type propagation of the file together with a description of the free riding behavior.

We first consider (Section 2.2) the case where all peers are fully cooperative in the sense that $\mu = 0$ and $X_f(0) = 0$ (no free riders). $\mu = 0$ implies that cooperative peers do not leave the network after receiving the file. We then move to the general case where $\mu > 0$ and $X_f(0) \geq 0$ (Section 2.3).

3

2.2. Fully cooperative network

When all peers are fully cooperative (i.e. $\mu = 0$ and $X_f(0) = 0$) the population of peers remains constant and equal to N, that is, $Y(t) + X_c(t) = N$ at any time t. The network dynamics can be represented by the process $\mathbf{Y} = \{Y(t), t \geq 0\}$.

This is a finite-state continuous-time Markov process with non-zero transitions given by

$$Y(t) \to Y(t) + 1 \quad \text{with rate} \quad \lambda Y(t)(N - Y(t)). \tag{1}$$

In other words the process $\mathbf{Y}$ is a pure birth Markov process on the state-space $\{y_0, \ldots, N\}$ with N an absorbing state which is reached when all peers have the file.

This model may have various interpretations. Two of them are given below:

- Delay Tolerant Network: each peer corresponds to a mobile terminal. Every two nodes are within transmission range of each other at time epochs that form a Poisson process with parameter λ. It is assumed that the file is short so that the transmission of the file occurs instantaneously when two mobiles are within transmission range and only one has the file. See e.g. [7, 19, 24] for related work.

- Parallel transmission over a lossy network: each node with the file transmits it periodically to all those that do not have the file. The network is assumed to be subject to large error rates or erasures, so that when transmitting the file from node i to node j, the probability to receive it correctly is small. Erasures are assumed to be independent and the channels are assumed to be memoryless in the following sense: failure or success in a given transmission of the file from i to j at a given time is independent of the events of the history of failures and successes of files sent previous to that one.

Define $m(t) := E[Y(t)]$, the expected number of peers with the file at time t. Standard algebra shows that

$$\frac{dm(t)}{dt} = \lambda E[Y(t)(N - Y(t))], \quad t > 0. \tag{2}$$

Unfortunately, the right-hand side of (2) does not express as a function of $m(t)$, thereby ruling out the possibility of finding $m(t)$ in closed-form as the solution of an ODE.

Assume that λ is written as $\lambda = \beta/N$ and that $\lim_{N \to \infty} N^{-1}Y(0) = y^* \in (0,1)$. For instance one can take $Y(0) = Ny^*$. Then, for large N, $m(t)$ is well-approximated by $Ny(t)$ where $y(t)$ is obtained as the unique solution of the ODE [2, Thm 3.1]

$$\frac{dy(t)}{dt} = f(y(t)), \quad t > 0, \tag{3}$$

where $f(u) := \beta u(1 - u)$ and $y(0) = y^* \in (0,1]$. It is found that

$$y(t) = \frac{y^*}{(y^* + (1 - y^*)\exp(-\beta t))}, \quad t \geq 0. \tag{4}$$

This is a well-known instance (see e.g. [24]) of what is known as mean-field approximation, a theory that focuses on the solution of ODEs obtained as limits of jump Markov processes [11]. The ODE (3) has been extensively used in epimediology studies, where $y(t)$ represents the fraction of infected patients at time t when the entire population is of size N.

We show below that the mean-field approximation is an upper bound for $E[Y(t)]$. To the best of our knowledge this result is new.

Proposition 1. $E[Y(t)] \leq Ny^*/(y^* + (1-y^*)e^{-\beta t})$ *for any $t \geq 0$.*

Proof: Throughout the proof $t > 0$ is fixed. By Jensen's inequality, $E[Y(t)(N-Y(t))] \leq E[Y(t)](N - E[Y(t)])$ so that from (2)

$$\frac{dz(t)}{dt} \leq f(z(t))) \tag{5}$$

where $z(t) := E[Y(t)]/N \in (0,1]$.

From (5) there exists a function $\omega(t)$, $0 \leq \omega(t) \leq 1$, such that

$$\frac{dz(t)}{dt} = \omega(t)f(z(t)). \tag{6}$$

Rewritting (3) and (6) as

$$\frac{dy(t)}{f(y(t))} = dt \quad \text{and} \quad \frac{dz(t)}{f(z(t))} = \omega(t)dt$$

and integrating both sides of these equations yields (with $y(0) = z(0) := a \in (0,1)$)

$$\int_a^{z(t)} \frac{du}{f(u)} = \int_0^t \omega(t) \quad \text{and} \quad \int_a^{y(t)} \frac{du}{f(u)} = t. \tag{7}$$

Integrals in (7) are well defined as long as $f(y(u))$ and $f(z(u))$ do not vanish for $a \leq u \leq t$. Let show that this is true.

First, note that $f(u)$ vanishes only if $u = 0$ or if $u = 1$. We have assumed that $0 < a = y(0) < 1$ so that $f(a) > 0$. Next, we see from (4) that $y(u) < 1$ for any $u > 0$. Last, $z(u) = 1$ for some $u = u^*$ would imply that $Y(u^*) = N$ a.s. for any $u \geq u^*$ which is clearly not true. This shows that the integrals in (7) are well defined

Eqns in (7) and $0 \leq \omega(t) \leq 1$ imply

$$0 \leq \int_a^{z(t)} \frac{du}{f(u)} \leq \int_a^{y(t)} \frac{du}{f(u)}$$

which in turn implies that $z(t) \leq y(t)$ or, equivalently, $E[Y(t)] \leq Ny^*/(y^* + (1-y^*)e^{-\beta t})$ by using (4) and the definition of $z(t)$. ∎

2.3. General network

We consider the general network defined in Section 2.1.

Define the vector $X(t) = \binom{X_c(t)}{X_f(t)}$, where we recall that $X_c(t)$ is the number of cooperative nodes in the system who do not have the file at time t and $X_f(t)$ is the number of free-riders in the system at time t (by definition, none of these have the file at time t). Let $e_c = (1,0)$ and $e_f = (0,1)$. Under the statistical assumptions made in Section 2.1 it is seen that the process $\mathbf{Y} := \{(Y(t), X(t)), t \geq 0\}$ is a finite-state Markov process whose non-zero transitions are given by

$$\binom{Y(t)}{X(t)} \to \binom{Y(t)+1}{X(t)-e_c} \quad \text{with rate } \lambda Y(t)X_c(t), \tag{8}$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \to \begin{pmatrix} Y(t) - 1 \\ X(t) \end{pmatrix} \text{ with rate } \mu Y(t), \tag{9}$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \to \begin{pmatrix} Y(t) \\ X(t) - e_f \end{pmatrix} \text{ with rate } \lambda Y(t) X_f(t). \tag{10}$$

The process $\mathbf{Y}$ takes its values in the set $\mathcal{E} := \{(i, j, j), 0 \le i \le y_0 + N_c, 0 \le j \le N_c, 0 \le j + k \le N - y_0\}$. Furthermore, all states in $\mathcal{E}$ of the form $(0, j, k)$ are absorbing states since there are no more transitions when the file has disappeared.

Note that the process $\mathbf{Z} := \{(Y(t), X_c(t)), t \ge 0\}$ is also an absorbing finite-state Markov process whose non-zero transitions are given in (8)-(9).

An explicit characterization of the transient behavior of the absorbing Markov process $\mathbf{Y}$ is a very difficult task due to the presence of non-linear and non-homogeneous transition rates in the state variables and to the dimension of $\mathbf{Y}$. The same is true for $\mathbf{Z}$.

If N is not too large, a numerical approach is possible. We refer the interested reader to [8] where this approach was used in a related context.

In this paper we will instead develop two approximations/bounds of the process $\mathbf{Y}$. The first one will consist in replacing $X_c(t)$ by $N_c = X_c(0)$ in the transition rate (8), which will introduce a birth and death Markov branching process. Clearly, this (so-called) branching approximation will loose its accuracy as the ratio $X_c(t)/N_c$ changes.

The second approximation will use an asymptotic argument as $N \to \infty$ based on a mean-field approximation of the process $\mathbf{Y}$. Such an approximation is accurate only if $Y(0) = y_0$ is of the order of N.

Both approaches will allow us to approximate some key quantitive characteristics of the process $\mathbf{Y}$ such as the probability of disappearance of the file, the time before all files disappear, the maximum number of cooperative peers in the network, the percentage of peers that eventually receive the file, etc.

Remark 1. *Our general model is inspired by the one in [16]. Let us briefly point out some differences between our model and the one in [16].*

- *The model of [16] assumes that cooperative users remain only a finite time in the system as a free-riding phenomenon and thus calls μ a measure for free-riding. Our model has in addition the parameter $\eta := N_c/(N - y_0)$ that quantifies free riding. This motivated us to define η as a free riding measure and μ as a measure of cooperation.*

- *The model for interaction between nodes is different. This results in a different rate of increase of the number of nodes with the file.*

- *In [16] there are some seeders that always remain in the system. Therefore unlike our case, in [16] all peers eventually receive the file.*

3. Branching approximation

Let $\mathbf{Y}_b = \{Y_b(t), t \ge 0\}$ be a Markov process on $\mathbf{N} := \{0, 1, \ldots\}$ (the subscript b refers to "branching") with non-zero transition rates given by

$$Y_b(t) \to Y_b(t) + 1 \qquad \text{with rate } \lambda Y_b(t) N_c \tag{11}$$

$$Y_b(t) \to Y_b(t) - 1 \qquad \text{with rate } \mu Y_b(t) \tag{12}$$

where we recall that N_c is the number of cooperative peers without the file at time $t = 0$.

Since $X_c(t)$, the number of cooperative peers without the file at time t, is non-increasing in t, a quick comparison between (8)-(10) and (11)-(12) indicates that the process $\mathbf{Y}_b$ should dominate the process $\mathbf{Y}$. This bounding result is formalized and proved in the proposition below.

A word on the notation: a real-valued rv Z_1 is stochastically smaller than another real-valued rv Z_2, denoted as $Z_1 \leq_{st} Z_2$, if $P(Z_1 > x) \leq P(Z_2 > x)$ for all x.

Proposition 2. *If $Y(0) \leq Y_b(0)$ then $Y(t) \leq_{st} Y_b(t)$ for any $t > 0$.*

Proof. The proof relies on a classical coupling argument. On a common probability space we recursively construct two Markov processes $\mathbf{U} = \{U(t) = (U_1(t), U_2(t)), t \geq 0\}$ and $\mathbf{V} = \{V(t), t \geq 0\}$, $U_1(t), U_2(t), V(t) \in \mathbf{N}$, by generating transitions as follows: assume that a transition has occured in at least one system at time t and let $U(t+)$ (resp. $V(t+)$) be the state of $\mathbf{U}$ (resp. $\mathbf{V}$) just after this transition. Let τ be an exponentially distributed rv with rate $r := \lambda V(t+)N_c + \mu V(t+)$ so that, conditioned on $V(t+)$, it is independent of the history of processes $\mathbf{U}$ and $\mathbf{V}$ up to time t. If $V(t+) = 0$ no more transition will occur in both systems after time $t+$. If $V(t+) > 0$, the next transition will occur at time $t + \tau$: with probability $\lambda U_1(t+)/r$, $U_1(t+) = U_1(t)+1$, $U_2(t+) = U_2(t)-1$ and $V(t+) = V(t)+1$, with probability $\lambda(V(t+) - U_1(t+))/r$, $U(t+) = U(t)$ and $V(t+) = V(t) + 1$, with probability $\mu U_1(t+)/r$, $U_1(t+) = U_1(t) - 1$, $U_2(t+) = U_2(t), V(t+) = V(t) - 1$, and with probability $\mu(V(t+)-U_1(t+))/r$, $U(t+) = U(t)$ and $V(t+) = V(t)-1$. This construction holds as long as $U_1(t) \leq V(t)$ for all $t \geq 0$.

Assume that $U_1(0) \leq V(0)$. We readily deduce from the above construction that $U_1(t) \leq V(t)$ for all $t > 0$, from which we conclude from the coupling theorem [9] that $U_1(t) \leq_{st} V(t)$. The proof is concluded by noting that the process $\mathbf{U}$ (resp. $\mathbf{V}$) is statistically identical to the process $\mathbf{Z}$ (resp. $\mathbf{Y}_b$). ∎

The Markov process $\mathbf{Y}_b$ is an absorbing continuous-time birth and death process on $\mathbf{N}$ with absorbing state 0. Because its transition rates are linear functions of the system state, this is also a continuous-time Markov *branching process* [6], namely, a process in which at any time t each member of $Y_b(t)$ evolves independently of each other.

The next section specializes known results of the theory of branching processes to the process $\mathbf{Y}_b$.

3.1. Extinction probability

As previously observed the process $\mathbf{Y}_b$ is a birth and death branching process [6, Chapter]. Each object (peer) of this process has a probability of change in the interval $(t, t + h)$ given by $bh + o(h)$ with $\theta = \lambda N_c + \mu$; with probability $p_0 = \mu/\theta$ an object dies (a peer leaves) and with probability $p_2 = \lambda N_c/\theta$ an object is replaced by two objects (a peer receives the file).

Given $Y_b(0) = k$ the extinction time T_k is defined by

$$T_b(k) = \min\{t > 0 \,:\, Y_b(t) = 0\} \tag{13}$$

and the extinction probability by

$$q_k = P(T_b(k) < \infty). \tag{14}$$

7

In our context $T_b(k)$ is the first time where the file disappears given that there are k copies of the file a time $t = 0$.

Let $F(s,t) := \sum_{i \geq 0} P(Y_b(t) = i)s^i$ be the generating function of the process $\mathbf{Y}_b$. Define $W(s) := \theta(f(s) - s) = \lambda N_c s^2 - (\lambda N_c + \mu)s + \mu$, $h(s) =: p_0 + p_2 s^2$.

If $Y(0) = 1$, we know from [6, Thm 10.1] that $F(s,t)$ satifies

$$\frac{\partial F(s,t)}{\partial} = W(F(s,t)) \tag{15}$$

and that the probability of extinction, q_1, is the smallest nonnegative root of the equation $s = h(s)$. We find

$$q_1 = \min\{1, 1/\rho\} \tag{16}$$

where we recall that $\rho = \lambda N_c/\mu$. More generally

$$q_k = (\min\{1, 1/\rho\})^k = \min\{1, (1/\rho)^k\}, \quad k \geq 1 \tag{17}$$

since the objects behave independently of each other.

The extinction will be certain iff $\rho \leq 1$ or equivalently iff $\lambda N_c \leq \mu$. Whenever $\rho < 1$ (resp. $\rho = 1$, $\rho > 1$) a branching process is said to be *subcritical* (resp. critical, supercritical).

3.2. Time to extinction

When the extinction is certain it is interesting to characterize some statistical properties of the extinction time $T_b(\cdot)$. Let D denote the event $Y_b(t) = 0$ for some $t > 0$ and let $q = P(D)$. We know that $q = q_k$ given in (17) if $Y_b(0) = k$.

The following is shown in [15]:

Proposition 3. *If $Y_b(0) = 1$, the conditional expectation and variance of the extinction time are expressed by*

$$E[T_b(1)|D] = \int_0^{q_1} \frac{q_1 - s}{q_1 W(s)} ds \tag{18}$$

and

$$V(T_b(1)|D) = 2 \int_0^{q_1} \frac{1}{W(s)} \int_s^{q_1} \frac{q_1 - r}{q_1 W(r)} dr ds - (E[T_b(1)|D])^2 \tag{19}$$

provided these expectations exist.

Because the file will disappear with probability one in the original process $\mathbf{Y}$, we are motived by considering the approximated process $\mathbf{Y}_b$ when its extinction is also certain ($q_1 =$) and the process is subcritical (i.e. $\rho < 1$). In this case, by Proposition 3

$$E[T_b(1)] = -\frac{1}{\lambda N_c} \log(1 - \rho). \tag{20}$$

Next, we look at the scenario where k peers have the file at time 0, namely, $Y_b(0) = k$. If we define $T_b^i(k)$ to be the extinction time of all offsprings of file $i = 1, \ldots, k$, then

$$P(T_b(k) \leq t) = P\left(\max_{1 \leq i \leq k} \{T_b^i(k)\} \leq t\right) = P(T_b(1) \leq t)^k \tag{21}$$

since in a branching process all objects behave independently of each other.

Combining the identity $F(0,t) = P(T_b(1) \leq t)$ and (21) gives

$$E[T_b(k)] = \int_0^\infty (1 - F(0,t)^k)dt = \int_0^1 \frac{1 - u^k}{W(u)}du \qquad (22)$$

in the subcritical case. The latter equality is obtained by using the change of variable $u = F(0,t)$ together with the relation (cf. (15)) $dt = dF(0,t)/W(F(0,t)) = du/W(u)$.

Let us now come back to the original process $\mathbf{Y}$.

Define $T(y_0) := \inf\{t : Y(t) = 0\}$, the first time when the file has disappeared from the network given that $Y(0) = y_0$. When $Y(0) = Y_b(0) = y_0$, Proposition 2 implies that

$$P(T(y_0) > t) = P(Y(t) > 0) \leq P(Y_b(t) > 0) = P(T_b(y_0) > t)$$

from which we deduce that

$$E[T(y_0)] \leq E[T_b(y_0)] = \int_0^1 \frac{1 - u^{y_0}}{W(u)}du$$

when $\rho < 1$. In particular, $E[T(1)] \leq -\frac{1}{\lambda N_c} \log(1 - \rho)$ from (20).

4. Mean-field approximation

In this section we investigate the behavior of the process $\mathbf{Y}$ defined in Section 2.3 as N, the number of peers, gets large. We first show that this behavior (to be made more precise) is well approximated by a deterministic limit solution of an ODE, an approach known as mean-field approximation. See [11] for the theory and [3, 12, 24] for recent applications in the area of file sharing systems.

Like in Section 2.2 we assume that the pairwise contact rate, λ, is of the form $\lambda = \beta/N$ with $\beta > 0$. We also assume that the initial state of $\mathbf{Y}$ is given by

$$Y(0) = Ny^*, \quad X_c(0) = Nx_c^*, \quad X_f(0) = Nx_f^* \qquad (23)$$

with y^*, x_c^* and x_f^* arbitrary nonnegative constants such that $y^* + x_c^* + x_f^* = 1$. [The analysis below holds under the weaker condition $\lim_{N \to \infty} N^{-1}(Y(0), X_c(0), X_f(0)) = (y^*, x_c^*, x_f^*)$.]

Let $v_1 = (1, -1, 0)$, $v_2 = (-1, 0, 0)$ and $v_3 = (0, 0, 1)$. Denote by $g(Y, Y + v_i)$, $i = 1, 2, 3$, the non-zero transition rates of the process Markov process $\mathbf{Y}$ out of state $Y = (Y_1, Y_2, Y_3)$. We have (cf. (8)-(10))

$$g(Y, Y + v_1) = \frac{\beta}{N}Y_1Y_2, \quad g(Y, Y + v_2) = \mu Y_1, \quad g(Y, Y + v_3) = \frac{\beta}{N}Y_1Y_3$$

which can be rewritten as

$$g(Y, Y + v_i) = Nf\left(\frac{Y}{N}, v_i\right), \quad i = 1, 2, 3 \qquad (24)$$

where $f(u, v_1) = \beta u_1 u_2$, $f(u, v_2) = \mu u_1$ and $f(u, v_3) = \beta u_1 u_3$ for $u = (u_1, u_2, u_3)$.

9

We may therefore use Theorem 3.1 in [11] to obtain that the rescaled process $N^{-1}\mathbf{Y}$ converges in probability as $N \to \infty$, uniformly on all finite intervals $[0, T]$, to the solution of the system of ODEs

$$\frac{d}{dt} \begin{pmatrix} y \\ x_c \\ x_f \end{pmatrix} = \begin{pmatrix} y(\beta x_c - \mu) \\ -\beta y x_c \\ -\beta y x_f \end{pmatrix} \tag{25}$$

with the initial conditions $y(0) = y^*$, $x_c(0) = x_c^*$, $x_f(0) = x_f^*$.

In other words, the solution $y(t)$, $x_c(t)$, $x_f(t)$ of (25) will give an approximation of the fraction of peers with the file at time t, of the number of cooperative peers without the file at time t and of the number of free-riders at time t, respectively. The accuracy of this approximation will increase with N, the total number of peers.

4.1. A change of time

The system of ODEs in (25) has a unique solution [2]. Unfortunately, this unique solution cannot be obtained in closed-form due to the presence of non-linear terms in the r.h.s. of (25).

By using an implicit change of time similar to that in [10], we transform (25) into a linear system of ODEs, from which we will be able to derive some interesting properties related to the behavior of the P2P network. As a starting point, observe that since $y(0) = y_0/N > 0$, $y(t)$ cannot be equal to zero for some finite t (Hint: $y(t) \geq y(0)e^{-\mu t}$ by integrating the first equation in (25)). For any finite t, we may therefore define

$$s(t) = \int_0^t \frac{1}{y(s(r))} \, dr \ . \tag{26}$$

This is a differential equation for s (note that it appears in the argument of y), which has a unique solution since y is Lipschitz continuous and $y(t) > 0$ [2]. The function s is increasing, and we shall use it to define a time change. Define $\bar{v}(t) = v(s(t))$ with a similar convention for the components of v. Then $\bar{v}(0) = v(0)$ and by the chain rule,

$$\frac{d\bar{v}(s(t))}{dt} = \frac{d\bar{v}(s(t))}{ds} \cdot \frac{ds(t)}{dt} = \frac{d\bar{v}(s(t))}{ds} \cdot \frac{1}{y(s(t))} \ . \tag{27}$$

But dv/ds is exactly the expression on the right of (25) except that the time is now $s(t)$. Thus the equation for s in (26), the chain rule in (27) and (25) give

$$\begin{pmatrix} \dot{\bar{y}}(t) \\ \dot{\bar{x}}_c(t) \\ \dot{\bar{x}}_f(t) \end{pmatrix} = \begin{pmatrix} \beta \bar{x}_c(t) - \mu \\ -\beta \bar{x}_c \\ -\beta \bar{x}_f \end{pmatrix} \tag{28}$$

with $(\bar{y}(0), \bar{x}_c(0), \bar{x}_f(0)) = (y(0), x_c(0), x_f(0))$.

This is a linear equation which can be solved explicitly.

4.2. Peers that never receive the file: a phase transition

The fraction of cooperative peers x_c and free-riders x_f that do not have the file monotonely decreases (this is true also for the original system) to some limit values.

They can continue decreasing until there are no copies of the file in the system, namely until $y = 0$.

The first question we wish to ask is whether these limits are close to 0 or are large. In other words, we wish to know whether all (or almost all) peers interested in the file are able to obtain it or not. If the answer is no, then we shall be interested in computing the fraction of those that will never receive the file.

From the second equation in the system (28) we obtain $\bar{x}_c$ as

$$\bar{x}_c(t) = x_c^* \, e^{-\beta t} \tag{29}$$

so that from the first equation we find

$$\bar{y}(t) = y^* + x_c^* \left(1 - e^{-\beta t}\right) - \mu t \ . \tag{30}$$

Define $\theta = \beta/\mu$ and $H = y^* + x_c^*$.

It is easily seen that the r.h.s. of (30) has a unique zero in $(0, \infty)$. This zero, denoted by t_0, is given by

$$t_0 = \frac{1}{\beta} \left[\theta H + \mathrm{LambertW} \left(-\beta x_c^* \, e^{-\beta H}\right) \right] , \tag{31}$$

where the Lambert W-function is the inverse function of $f(w) = w e^w$ (see e.g. [4]).

We note that our time change transforms the infinite time interval (y does not reach 0 in finite time) into the finite interval $[0, t_0)$.

Since $s(t)$ maps the interval $[0, t_0)$ onto the interval $[0, \infty)$ we conclude from (29) that $x_c^{\min}$, defined as the smallest value of $x_c(t)$, is obtained at $t = t_0$, and is given by

$$x_c^{\min} = x_c^* \exp(-\beta t_0). \tag{32}$$

Similarly, $x_f^{\min}$, the smallest value of $x_f(t)$, is obtained at $t = 0$, and given by (Hint: solve the third equation in (28))

$$x_f^{\min} = x_f^* \exp(-\beta t_0). \tag{33}$$

It is worth noting (see (31)) that both $x_c^{\min}$ and $x_f^{\min}$ are only function of the parameter θ.

By using the Taylor series expansion $\mathrm{LambertW}(x) = x - x^2 + o(x^2)$ [4] we get from (31)

$$t_0 \approx \frac{\theta y(0)}{\beta} \tag{34}$$

as θ is small, so that from (32) and (33)

$$x_c^{\min} \approx x_c^* \, e^{-\theta y^*} \quad \text{and} \quad x_f^{\min} \approx x_f^* \, e^{-\theta y^*} \tag{35}$$

as θ is small.

The constants $x_c^{\min}$ and $x_f^{\min}$ are nothing but the limits of $x_c(t)$ and $x_f(t)$ as $t \to \infty$, respectively. This raises the issue of the validity of the mean-field approximation at $t = \infty$ since, as recalled earlier, Theorem 3.1 in [11] only holds for finite-time intervals. The result in [26, Sec. 5.2] shows that $N^{-1}\mathbf{Y}_N(\infty)$ indeed converges in probability to

11

$(0, x_c(\infty), x_f(\infty))$ as $N \to \infty$, which will thereby justify to approximate the fraction of cooperative peers (resp. free-riders) that will never receive the file by $x_f^{\min}$) (resp. $x_f^{\min}$).

Figure 1 (left) displays the mapping $\theta \to x_c^{\min}$ for $x_c^* \in \{0.01, 0.1, 0.3, 0.5, 0.9\}$. The curves for $x_c^{\min}$ are monotone decreasing in x_c^* (the curve that intersects the vertical axis close to 1 is the one corresponding to $x_c(0) = 0.01$, and so on.). The right part of figure 1 is the same as the left one except that it uses a logarithmic scale for the horizontal axis: the horizontal axis is given by $\log_{10} \theta$ instead of θ.
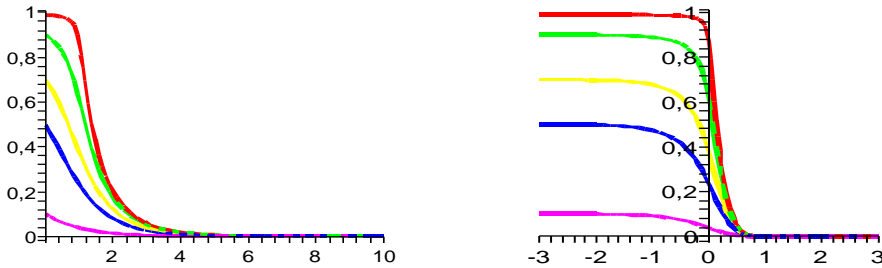
Figure 1: The ratio of cooperative peers that never receive a copy the file (vertical axis) as a function of $\theta = \beta/\mu$ (left figure) and of $\log_{10} \theta$ (right)

We see from Fig. 1 (right) a clear phase transition that occurs at $\theta = 1$ (i.e. when $\beta = \mu$). $x_c^{\min}$ converges to x_c^* (see (35)) that depends on the initial condition) as θ decreases to zero, and approaches zero as θ increases beyond 1 (or equivalently as λ increases beyond μ).

4.3. Maximum file availability

From (30) we see that $\bar{y}$, and therefore y, reaches its maximum at $t = t_m$ with

$$t_m = (1/\beta) \log(\theta x_c^*) \tag{36}$$

provided that $\theta^{-1} < x_c^*$ (i.e. $\bar{y}$ is increasing at time $t = 0$).

Let us determine the $y^{\max} := y(t_m)$. The first two equations in (25) give

$$\frac{dy}{dx_c} = \frac{dy/dt}{dx_c/dt} = \frac{1}{\theta x_c} - 1$$

so that

$$y(x_c) = \theta^{-1} \log(x_c/x_c^*) - x_c + x_c^* + y^* \tag{37}$$

with the initial condition $y(x_c^*) = y^*$.

On the other hand, we see from the first equation in (25) that if $x_c(0) > \theta^{-1}$ then y reaches its maximum at $x_c = \theta^{-1}$. Introducing this value in (37) shows that $y^{\max}$, the maximum value of y, is given by

$$y^{\max} = -\theta^{-1} \left(1 + \log(\theta x_c^*)\right) + x_c^* + y^* \tag{38}$$

when $\theta^{-1} < x_c^*$.

12

The second and third equation in (25) give

$$\frac{dx_f}{dx_c} = \frac{x_f}{x_c}$$

which implies that

$$x_f(x_c) = x_c + x_f^* - x_c^*$$

with the initial condition $x_f(x_c^*) = x_f^*$. In particular, when $y(x_c)$ reaches its maximum (at point $x_c = \theta^{-1}$ as shown above) then $x_f = \theta^{-1} + x_f^* - x_c^*$.

In conclusion, for large N, the time before the file disappears from the network can be approximated by t_0 in (31), the number of cooperative peers (resp. free-riders) that will never receive the file can be approximated by $Nx_c^* \exp(-\beta t_0)$ (resp. $Nx_f^* \exp(-\beta t_0)$) or equivalently (see (23)) by $Y(0) \exp(-\beta t_0)$ (resp. $X_c(0) \exp(-\beta t_0)$), the maximum number of files in the network (file availability) can be approximated by $y^{\max}$ in (38), the time at which this maximum is reached can be approximated by t_m in (36), and the number of cooperative peers (resp. free-riders) in the network when the file availability is the largest can be approximated by θ^{-1} (resp. $\theta^{-1} + x_f^* - x_c^*$).

5. P2P with a fixed request or transmission rate per node

5.1. Model

We consider here the P2P model described by the following non-zero transitions

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \to \begin{pmatrix} Y(t) + 1 \\ X(t) - 1 \end{pmatrix} \text{ with rate } \tfrac{\lambda Y(t) X_c(t)}{Y(t) + X_c(t)}, \quad \begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \to \begin{pmatrix} Y(t) - 1 \\ X(t) \end{pmatrix} \text{ with rate } \mu Y(t). \tag{39}$$

It is further assumed that there is a permanent seed node ("publisher") that always has a copy of the file. (While the impact of a permanent seed on the mean field limit while vanish, its presence implies that there is no extinction at the early phase of the epidemics, in contrast with the situation in Section 3.)

This model was recently introduced and studied in [16] (free riders are not considered here so we shall omit the index c or f from X). This differs from our previous model by **normalizing the rate of increase** in the number of peers with a copy of the file by the sum $Y(t) + X(t)$. The interpretation in [16] is that of a P2P network in which a peer without a file sends a request at a fixed rate λ to a randomly selected peer. The probability that the peer who receives the request has the file is $X(t)/[Y(t) + X(t)]$. Thus the total rate at which requests for a file are received is $\lambda Y(t) X(t)/[Y(t) + X(t)]$. When a request is received, the file is sent to the source of the request instantaneously. More generally, the normalization of the rates can be used to model slow uplink or slow downlink speeds of peers.

The corresponding ODE is given by

$$\frac{dy(t)}{dt} = -\mu y(t) + \frac{\lambda x(t) y(t)}{x(t) + y(t)}, \quad \frac{dx(t)}{dt} = -\frac{\lambda x(t) y(t)}{x(t) + y(t)}. \tag{40}$$

The justification for this model via Kurtz theorem [11] continues to hold. Taking the sum $\dot{x} + \dot{y}$ and dividing by $\dot{x}$, we obtain [16, Equation (10)]

$$y(x) = -x + c^* x^{\mu/\lambda}, \tag{41}$$

13

where the constant c^* is taken to be 1. Let the initial condition of the ODEs be $y(0) = y^*$ and $x(0) = x^*$ with $0 \le x^*, y^* \le 1$ and $x^* + y^* = 1$. To make the solution of the ODEs an approximation of the Markov process, the initial state of $(X(0), Y(0))$ is supposed to be (Nx^*, Ny^*).

5.2. Phase transitions

The authors of [16] consider $y^* = 0, x^* = 1$ as initial condition. Substituting in (41) this gives the solution $c^* = 1$. We shall consider in contrast an arbitrary initial state.

We note that $y^* = 0$ is an equilibrium point of (40) for any $x^* > 0$. We shall show however that this is not the only equilibrium of (40). We shall derive a second equilibrium and will identify to which of the equilibria will the system converge, as a function of μ and λ. Let $x^* \in (0, 1)$ and consequently by (41) $c^* = (x^*)^{-\mu/\lambda} \in (1, \infty)$. But then

$$y(x) = -x + \left(\frac{x}{x^*}\right)^{\mu/\lambda} \tag{42}$$

and if we look for $y^* = 0$ we obtain

$$x = (x^*)^{\mu/(\mu-\lambda)} \tag{43}$$

We thus obtained a second equilibrium. We can observe two phase transitions:

- Phase transition around $\lambda = \mu$: The second equilibrium describes the behavior of the system (in case of nonzero initial y) only if $\mu > \lambda$. Indeed, if $\mu < \lambda$, (43) gives $x > 1$ and it is obtained at negative time. It is thus not related to our problem (in which $x \le 1$ and only $t \ge 0$ is considered). We show that there is in fact a sharp phase transition that occurs between the two regions $\mu > \lambda$ and $\mu < \lambda$.

- We note from (40) that x is strictly decreasing (as a function of time) as long as it as well as y are positive, and that there is not limit (at positive time) for the equations unless y converges to 0. Now

$$\frac{dy(t)}{dt} = \frac{dy}{dx}\frac{dx}{dt}. \tag{44}$$

  So y is unimodular if and only if $dy/dx < 0$ at $x = 0$. Now

$$\frac{\partial y}{\partial x} = -1 + \frac{\mu}{x^*\lambda}\left(\frac{x}{x^*}\right)^{\frac{\mu}{\lambda}-1} \tag{45}$$

  is negative at $t = 0$ if and only if $\frac{\mu}{x^*\lambda} < 1$ (since $x(0) = x^*$). This gives one a phase transition (from monotonicity to unimodularity). As t increases x decreases and so for $t > 0$ large enough y will start decreasing. On the other hand, y is always monotone (is never unimodular) if $\mu > \lambda$.

Computing the maximum torrent size.

We saw that y^* is already the maximum if and only if $\mu/(\lambda x^*) \ge 1$. So we consider the unimodular case $\mu < \lambda x^*$ so that the maximum of y is obtained for $t > 0$, at a point (x, y) where $dy/dx = 0$. This gives

$$\left(\frac{x}{x^*}\right)^{\mu/\lambda-1} = \frac{x^*\lambda}{\mu} \quad \text{and thus} \quad x = x^*\left(\frac{x^*\lambda}{\mu}\right)^{\lambda/(\mu-\lambda)}$$

14

$$y_{\max} = -x^* \left(\frac{x^*\lambda}{\mu}\right)^{\frac{\lambda}{\mu-\lambda}} + \left[\left(\frac{x^*\lambda}{\mu}\right)^{\frac{\lambda}{\mu-\lambda}}\right]^{\frac{\mu}{\lambda}} = (x^*)^{\frac{\mu}{\mu-\lambda}} \left(\frac{\lambda}{\mu}\right)^{\frac{\lambda}{\mu-\lambda}} \left(\frac{\lambda}{\mu} - 1\right) \qquad (46)$$

(which is positive as $\lambda > \mu$.)

Size of the Epidemy: do all receive the file?

If $\mu > \lambda$ then (43) shows that y approaches 0 (and as we know—monotonically) but x approaches a finite limit $0 < x_\infty < 1$, which depends on μ, λ and on the initial conditions as can be seen from the expression for $\dot{y}$ in (40). The final value of the number of peers who have not received the file satisfies, in this case, $x \downarrow 0$ as $\mu \downarrow \lambda$.

Finally, the expression for $\dot{x}$ in (40) shows that x is strictly monotone decreasing, and by [16, Eq. (11)] y is also monotone. The analysis holds for the starting position $x = 1$ only, and then it is monotone increasing. For other starting points y may be unimodular decreasing and then increasing. When $\mu > \lambda$ it is actually monotone decreasing.

6. The impact of measures against P2P networks

In this section, we investigate the impact of measures against non-authorized uploading or downloading on file availability in P2P swarming systems. Our investigation concerns the model of Section 2, although we present some general insight that holds also for the model of Section 5.

We assume that such measures would aim at increasing the noncooperative degree μ or the contact rate λ. A larger μ would imply a smaller lingering time for uploading. A smaller λ can indicate a smaller interest in downloading.

Remark 2. *Another possibility is that the measures affect the number N of those interested in the torrent. The impact of a change in N on the dynamics of the mean field limit is the same as that obtained by varying λ. If the pairwise contact rate λ and the ratio of cooperative peers r are fixed, the increase of N results in the increase of β and ρ. Hence a multiplicative increase of N leads to a larger rate of growth in the number of infected peer and the increase of file availability.*

The main question is how does a decrease in one of these parameters affect measures such as

- The size of the torrent: the fraction of those who are interested in the file and are able to get a copy of it. This can be seen as a global availability measure.

- The extinction probability or the expected extinction time,

- The maximum availability: the maximum number of copies that can be found simultaneously in the system. This can be viewed as an instantaneous availability measure.

We identify cases in which a small decrease in the upload or download rate can have a significant impact on the above measures, as well as cases in which even large decrease in upload or download do not have an impact on these measures.

6.1. Some insight

We have observed in previous sections that there are various "phase transitions" in which the macroscopic behavior of the systems changes abruptly when moving from one range of parameters to another one.With $\rho = \lambda N_c/\mu$, we recall in particular, that:

- As $\rho \to 0$, the fraction of peers that does not manage to get the file as the time horizon goes to infinity, goes to a strictly positive constant, and as $\rho \to \infty$, it converges to zero. It was around $\rho = 1$ that the (quite abrupt) change occurred.

- We recall that if the system starts with a "small" number of copies of the file then it behaves for some time like a continuous time branching process. As such it can get extinct at a very early stage; the extinction probabilities of a branching process is 1 for $\rho < 1$ and are smaller than one, otherwise.

We can thus expect that if ρ is close to one, then a small effort in measures to decrease the upload rate (by increasing μ) or the download rate (by decreasing λ) can result in a large decrease in the torrent size, and in the availability. In contrast, we may expect there to be little effect if the file is very popular so that ρ is much larger than 1. We examine these points as well as other ones in the next subsections.

6.2. Counter measures: bundling

Assume that a set S_i of peers is interested in the file i, $i = 1, ..., I$. Assume that all I files are bundled together into one file which now interests the set $S_b = \cup_{i=1}^{I} S_i$. Let N_b be the number of peers in this set. A file j that was of interest for N_j peers would thus see its effective β_j (see Remark 2) increase by a factor of N_b/N_j. In the supercritical case with homogeneous arrival and departure rates, the extinction probability of file j in the bundle is reduced by N_b/N_j times at the early stage.

We conclude that the same type of radical changes in performance measures as those obtained by measures to reduce access to the file (mentioned in the previous part of this Section) can be obtained by bundling. As we saw, this type of impact would be the strongest if it is applied for files that have their ratio $\lambda N_b/\mu = \rho$ close to 1.

Bundling has already been proposed as a tools to increase availability of files in [25].

6.3. Impact of measures: a numerical analysis

In this subsection, we focus on the impact of the noncooperative degree μ and the contact rate λ. Experimental results include the time to extinction and the number of peers without a file. To facilitate the comparison, we fix one of the parameters when tuning the others. There are 400 peers in this file sharing system. We fix the noncooperative degree μ to be 1 in order to study the impact of contact rate λ on the P2P performance. Here, three sets of experiments are conducted, each of which is duplicated 200∼1000 times. We also take the ratio of cooperative peers and the initial state into consideration.

We want to know the distribution of time that all the peers with the file disappear. Note that the disappearance of the file has two modes, the "*early extinction*" and the "*late removal*". Intuitively, *early extinction* happens for small ρ while the "*late removal*" is mainly caused by departures in the closed P2P system. Figure 2 and 3 show the cumulative probability distribution of extinction time with $\lambda = 0.002$. Since $\rho = 0.8$ is

less than 1, the peers with the file die out with probability 1 in the Markov branching process. The experimental results of extinction time distribution are very close to the branching process model, which means that the file disappears at the early stage of file sharing. As illustrated in Figure 2 and 3, the presence of free riders reduces the extinction time. A cross comparison of above figures shows that a larger $Y(0)$ leads to an increase of extinction time. Figure 4 and 5 illustrate the cumulative distribution of extinction time for the case $\lambda = 0.006$. We have the following observations: i) the extinction time in this P2P network is upper bounded by the Markov branching process; ii) the CDF curve has a "step" shape; iii) the presence of free riders prolongs the time of file existence in the swarm. In this set of experiments, the ratio ρ is around 2.4 so that the file survives with probability $(1 - 1/\rho)$ if there are no free riders. Let us focus on the extinction time in Figure 4. The cumulative probability at point A is very close to that at B, while T_B is much larger than T_A. This manifests that the extinction happens before T_A is mainly caused by the early extinction in the Markov branching process. There is almost no extinction in the process of file sharing between T_A and T_B. After the number of active peers in the system achieves it peak value, the arrival rate of new requests is less than the death rate μ. The file begins to die and the extinction probability is 1 as t is large enough. One can clearly see that the cumulative probability of extinction grows rapidly from T_B to T_C. The CDF curve of $r = 1$ meets that of $r = 0.6$ at T_D, and the former is greater than the latter since then. Thus, the file lives longer in case $r = 0.6$ than in case $r = 1.0$ if it avoids the early extinction. Similar conclusions can be drawn from the experimental results when $Y(0) = 3$. The difference lies in that the increase of probability of "late removal" is steep with three initial seeds. We further evaluate our observations by setting λ to be large enough. In figure 6 and 7, the slopes of CDF increment are very steep with $\rho = 8$. Especially when there are three initial seeds, very few early extinctions occur and all the peers with that file leave the system within 5 unit time in the late removal phase.
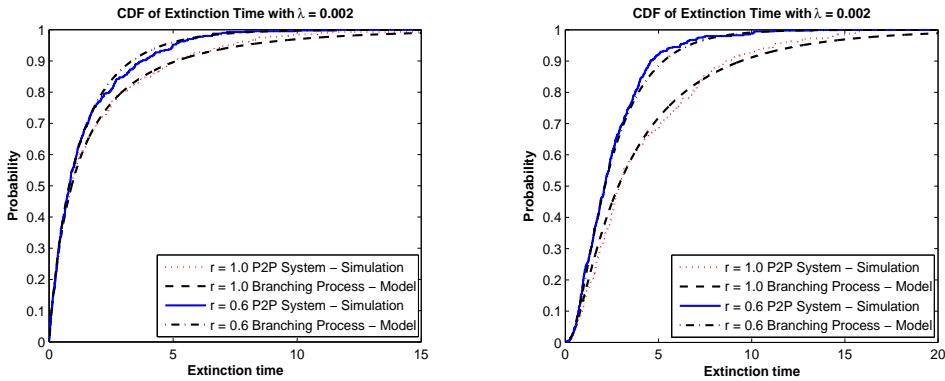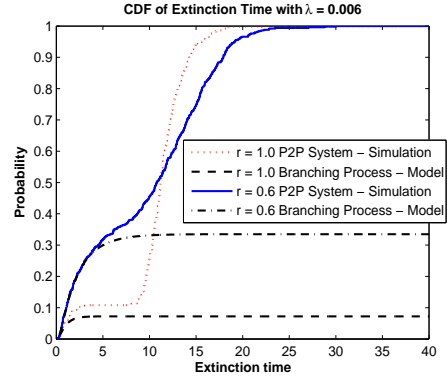

Figure 2: Single Source : $Y(0) = 1$ and $\lambda = 0.002$ Figure 3: Three Sources : $Y(0) = 3$ and $\lambda = 0.002$

The noncooperative degree μ decides the average lingering time of a cooperative peer after finishing the downloading. To enable the comparisons, the contact rate λ is fixed to be 0.0025. We mainly focus on the situations that ρ is less than 1 (i.e. $\mu < 1$). Figure
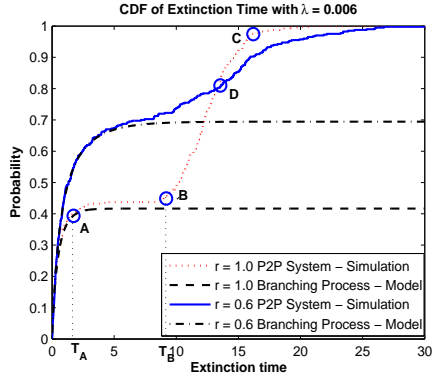
Figure 4: Single Source : $Y(0) = 1$ and $\lambda = 0.006$ Figure 5: Three Sources : $Y(0) = 3$ and $\lambda = 0.006$
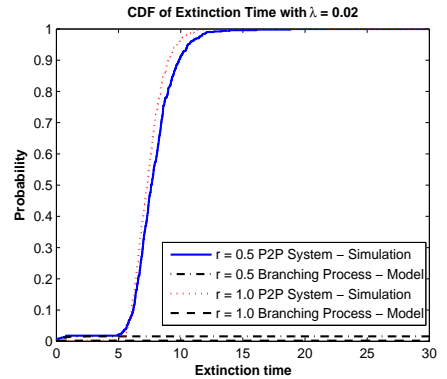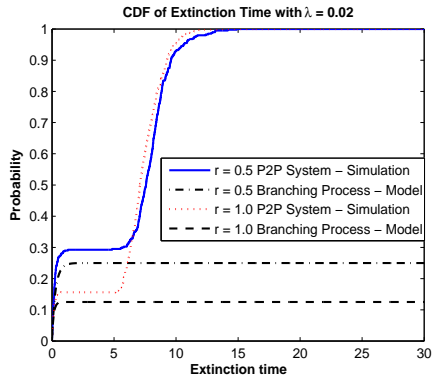

Figure 6: Single Source : $Y(0) = 1$ and $\lambda = 0.02$ Figure 7: Three Sources : $Y(0) = 3$ and $\lambda = 0.02$

8 and 9 depict the CDF curves with different parameter settings. As one can imagine, the cumulative distributions of extinction time also exhibit the "step" behavior. When μ is small, the file disappears either in the beginning or after many peers obtained it.

6.4. The accuracy of models

The goal of this subsection is to demonstrate how accurate our models match the P2P file sharing system. We perform three set of experiments, the first is to evaluate the time till extinction in the early stage, the second is to study how many peers obtain that file and the last is to verify the deterministic model. To reduce the simulation burden, we assume that there are 300 peers in the system.

The "early extinction" refers to the disappearance of the file before the number of active peers with that file reaches a peak value. According to the mean field equation in Eq.(25), an early extinction happens before $x_c(t)$ is less than $\min\{1, \mu/\lambda\}$. This principle is adopted in our experiments to measure the extinction time in the early stage. Figure 10 shows the extinction time for λ between 0.0001 and 0.003 with the initial
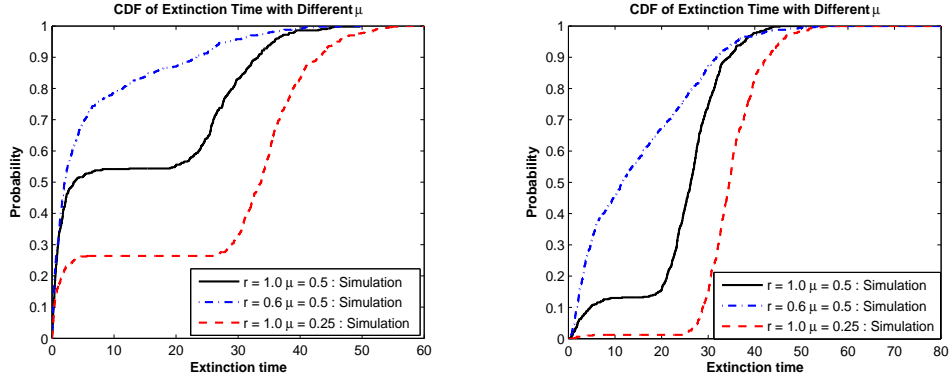
18

Figure 8: Three Sources : $Y(0) = 1$ for Different Figure 9: Three Sources : $Y(0) = 3$ for Different
μ μ

state $Y(0) = 1$. As λ grows, the contact rate becomes large so that the extinction time increases correspondingly. One can see that branching process approximates the swarming system well for small λ. This is because N_c does not change too much in the branching approximation. For some large λ, it presents almost an upper bound of extinction time because $X_c(t)$ is quite different from N_c. In figure 11, we illustrate the extinction time with three seeds at time 0. Compared with the case $Y(0) = 1$, the extinction times of $Y(0) = 3$ have been greatly extended. Next, we set λ to be 0.03 and increase μ from 0 to 15. Figure 12 shows the extinction time of the file with a single seed at time 0. The simulation results coincide with the branching process for large departure rate μ and small fraction of cooperative peers r. We measure the extinction time for the case with three seeds in figure 13. More seeds can help to extend the lifetime of the P2P swarming system for $\rho < 1$.
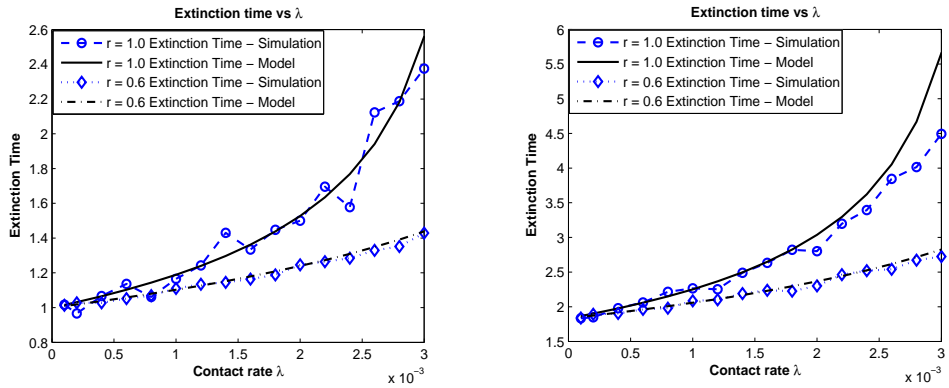

Figure 10: Early extinction time versus λ with Figure 11: Early extinction time versus λ with
$Y(0) = 1$ $Y(0) = 3$

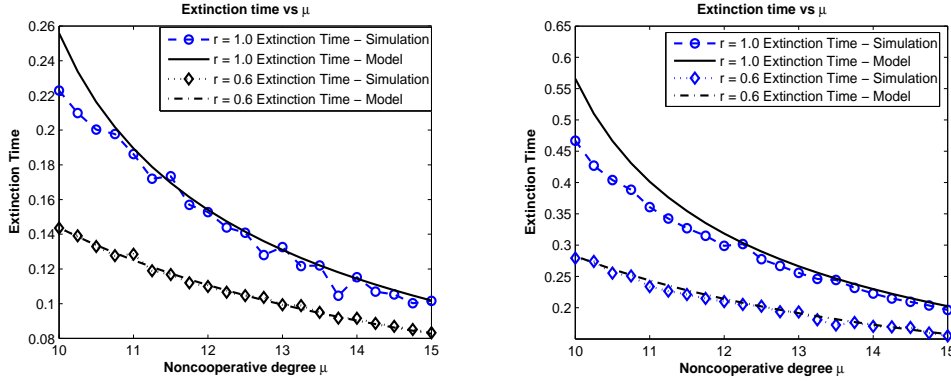In the second set of experiments, we look into the impact of free riders on file avail-

19

Figure 12: Early extinction time versus μ with $Y(0) = 1$

Figure 13: Early extinction time versus μ with $Y(0) = 3$

ability based on the mean field approximation. The performance metric is the percentage of peers that do not obtain the file. This metric of cooperative peers is the same as that of free riders simply because their contact rates are homogeneous. We first evaluate the accuracy of extinction time by varying contact rate λ. Figure 14 shows the relationship between λ and the percentage of peers without the file. Note that ten cooperative peers have the file at time 0 such that we can omit the extinction of very early stage. In the first experiment, all the peers are assumed to be cooperative, while in the second experiment half of the peers are free riders. As λ increases, more and more peers will download the file. The numbers of peer without the file of the experiment $r = 0.5$ are much larger than those in $r = 1.0$ for $\lambda \geq 0.004$. Furthermore, one can observe that the analytic model coincides with the experimental results. For the cooperative case, the gap between the model and the simulation is less than 1%. When half of the peers are free riders, the gap, though larger, is still less than 7%. Next, we investigate the impact of μ over the number of peers that never receive the file. As shown in figure 15, the ratio of peers without the file decreases if the noncooperative degree μ increases. When μ is less than 1 or greater than 11, the difference in the percentage of peers without the file is small. However, the case $r = 1$ has a much better file availability than the case $r = 0.5$ for μ in the range [1, 11]. Figure 15 also shows that the proposed model matches the simulation well when λ is 0.03 and μ increases from 0 to 15. The average errors of the percentages are less than 2% and the largest error is less than 6%.

Last, we exhibit the accuracy of the mean field approximation. Note that there are five cooperative peers with the file initially. The contact rate λ is set to 0.01 and the noncooperative degree μ varies. Figure 16 shows the dynamic growth of the average number of peers having the file along with time. The proposed model is shown to be a tight upper bound on $Y(t)$ obtained from the experiments. Figure 17 illustrates the active peers residing in the system. After the number of active copies reaches the peak value, they leave the swarming system eventually.
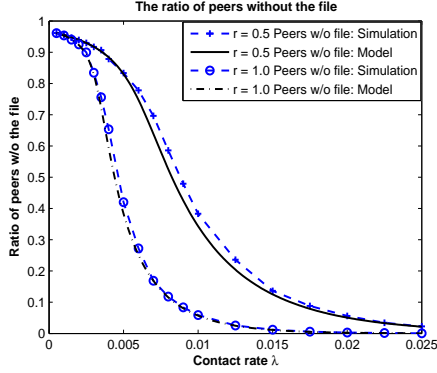
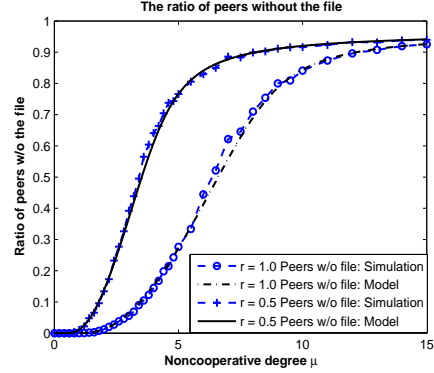Figure 14: The percentage of remaining peers w/o the file versus λ

Figure 15: The percentage of peers w/o the file versus μ
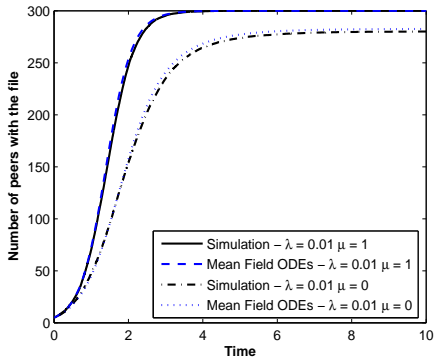

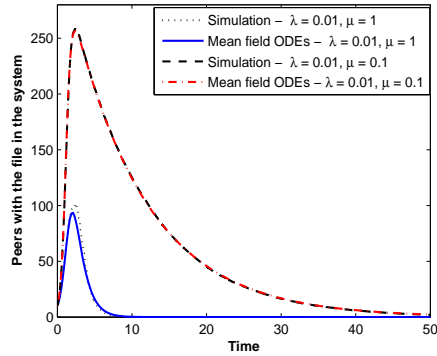Figure 16: The dynamic behavior of the number of peers that receive that file

Figure 17: The number of active peers in the system that have the file

7. Conclusion

In this paper we have proposed the use of the theory of continuous branching process as well as of the dynamics of epidemics in order to study the transient behavior of torrents that occur in P2P systems. The use of these tools allowed us to compute the probability of early extinction of the torrent as well as the expected time until that extinction, the availability of a file in the system, the maximum availability and when it occurs, and the size of the torrent (the fraction of nodes that are eventually infected). This was used for analyzing the impact of measures to decrease non-authorized Internet access to copyrighted files. We identified regimes in which the performance measures are quite sensitive to such measures and others in which the measures have very limited impact. In particular, we presented a numerical study of performance measures as a function of the system parameters. Besides, our methodology can be extended to various P2P applications, such as the content distribution networks with publishers and the P2P file bundling.

References

[1] E. Altman, S. Wong and J. Rojas-Mora, "P2P Business and Legal Models for Increasing Accessibility to Popular Culture," *Proc. of International Conference on Digital Business*, London, UK, June 17 - 19, June 2009.

[2] J. K. Hal, *Ordinary Differential Equations*, Pure and Applied Mathematics XXI, Robert E. Krieger Pub. Company, Huntington, New York, 1980.

[3] A. Chaintreau, J.-Y. Le Boudec and N. Ristanovic, "The Age of Gossip: Spatial Mean Field Regime," *Proc. ACM Sigmetrics Conf.*, Seattle, WA, USA, June 15-19, 2009.

[4] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth, "On the Lambert W Function", *Advances in Computational Mathematics*, Vol. 5, pp. 329–359, 1996.

[5] D.J. Daley and J. Gani. *Epidemic Modelling*. Cambridge Univ. Press, Cambridge, 1996.

[6] T. E. Harris, *The theory of Branching Processes*, Dover Publications, 1989 Printing, 1963 Edition.

[7] R. Groenevelt, P. Nain and G. Koole, "The Message Delay in Mobile Ad Hoc Networks," *Performance Evaluation*, Vol. 62, No. 1-4, pp. 210-228, October 2005.

[8] M. Ibrahim, A. Al Hanbali and P. Nain, "Delay and Resource Analysis in MANETS in Presence of Throwboxes," *Performance Evaluation*, Vol. 24, Issues 9-12, pp. 933-945, October 2007.

[9] T. Kamae, U. Krengel and G. L. O'Brien, "Stochastic Inequalities on Partially Ordered Spaces," *Annals of Probability*, Vol. 5, No. 6, pp. 899-912, 1977.

[10] D. G. Kendall, "Deterministic and Stochastic Epidemics in Closed Populations," *Proc. Third Berkeley Symp. on Math. Statist. and Prob.*, Vol. 4, Univ. of Calif. Press, 1956.

[11] T. G. Kurtz, "Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes," *Journal of Applied Probability*, Vol. 7, No. 1 (April 1970), pp. 49-58.

[12] L. Massoulié and M. Vojnović, "Coupon Replication Systems,", *Proc. ACM Sigmetrics Conf.*, Banff, Alberta, Canada June 6-10, 2005.

[13] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

[14] D. Mollison, *Epidemic Models: their Structure and Relation to Data*. Ed. D. Mollison, Cambridge University Press, Cambridge, 1995.

[15] P. Narayan, "On the Extinction Time of a Continuous Time Markov Branching Process," *Australian Journal of Statistics*, Vol. 24, Issue 2, pp. 160-164, 1982.

[16] R. Nunez-Queija and B. J. Prabhu. "Scaling Laws for File Dissemination in P2P Networks with Random Contacts," *Proc. 16th International Workshop on Quality of Service* (IWQoS 2008), June 2-4, 2008, University of Twente, Enschede, The Netherlands.

[17] M. Piatek, T. Kohno and A. Krishnamurthy, "Challenges and Directions for Monitoring P2P File Sharing Networks - or - Why My Printer Received a DMCA Takedown Notice," University of Washington, Technical Report,UW-CSE-08-06-01.

[18] I. Sazonov, M. Kelbertb and M. B. Gravenor, "The Speed of Epidemic Waves in a One-Dimensional Lattice of SIR Models", *Mathematical Modelling Natural Phenomena*, Vol. 3, No. 4, pp. 28-47, 2008.

[19] T. Small and Z. J. Haas, "The Shared Wireless Infostation Model: a New Ad Hoc Networking Paradigm (or where there is a whale, there is a way)," *Proc. 4th ACM Int. Symposium on Mobile Ad Hoc Networking and Computing* (MobiHoc'03), Annapolis, MD, USA, June 1-3, 2003.

[20] F. Wang *Application of the Lambert W Function to the SIR Epidemic Model*, Maplesoft Application Center.

[21] A. Weiss and A. Shwartz, *Large Deviations for Performance Analysis*. Chapman and Hall, 1995.

[22] S. Wong and E. Altman, "Restricting Internet Access: Ideology and Technology," *Proc. 2nd Int. Conf. on Communication Systems and Networks* (COMSNETS), January 5-9, 2010, Bangalore, India.

[23] S. Wong, E. Altman and M. Ibrahim, "P2P Networks: The Interplay between Legislation and Information Technology,"*Proc. Networking and Electronic Commerce Research Conf.* (NAEC) Lake Garda, Italy, October 8-11, 2009. See also INRIA Research Report RR-6889, version 2, June 2009.

[24] X. Zhang, G. Neglia, J. Kurose, D. Towsley, "Performance Modeling of Epidemic Routing,"*Computer Networks*, Vol. 51, No. 10, pp. 2859-2891, 2007.

[25] D.S. Menasche, G. Neglia, D. Towsley, and S. Zilberstein. "Strategic Reasoning About Bundling in Swarming Systems". Proceedings of the International Conference of Game Theory for Networks (GameNets) 2009.

[26] R. Darling, J. Norris. "Differential equation approximations for Markov chains,"*Probability Surveys*, Vol.5, pp. 37-79, 2008.