

# Predicting the Impact of Measures Against P2P Networks: Transient Behavior and Phase Transition

Eitan Altman, *Fellow, IEEE*, Philippe Nain, Adam Schwartz, *Senior Member, IEEE*, and Yuedong Xu

**Abstract**—The paper has two objectives. The first is to study rigorously the transient behavior of some peer-to-peer (P2P) networks whenever information is replicated and disseminated according to epidemic-like dynamics. The second is to use the insight gained from the previous analysis in order to predict how efficient are measures taken against P2P networks. We first introduce a stochastic model that extends a classical epidemic model and characterize the P2P swarm behavior in presence of free-riding peers. We then study a second model in which a peer initiates a contact with another peer chosen randomly. In both cases, the network is shown to exhibit phase transitions: A small change in the parameters causes a large change in the behavior of the network. We show, in particular, how phase transitions affect measures of content providers against P2P networks that distribute nonauthorized music, books, or articles and what is the efficiency of countermeasures. In addition, our analytical framework can be generalized to characterize the heterogeneity of cooperative peers.

**Index Terms**—Branching process, epidemics, mean field, peer-to-peer (P2P), phase transition.

## I. INTRODUCTION

IN RECENT years, the global Internet along with peer-to-peer (P2P) networks have created huge opportunities for free and open access to popular culture such as music, films, and e-books. The networking research community has been involved in creating new P2P protocols that give incentives to file sharing. However, these developments that brought a huge increase in Internet traffic turned out to involve more than technological aspects. This technology found itself in the heart of a harsh debate on copyright and on ethical issues. The main problem with the P2P technology has been the fact that a large part of P2P traffic consisted of copyrighted content. Two main approaches emerged in the conflict over access to copyrighted materials. The first tries to fight such access and proposes legal actions against it. We call this approach the confrontation one, as it fights demand and sharing of unauthorized copyrighted content. The second is a cooperative approach that seeks to profit

from such demand in a way that would benefit all actors, including the creators of the content. The conflict between both approaches involves three fronts: the legislation, the ethical, and the economic front.

On the legislation front, the confrontation approach has been gaining ground. In several countries, inspection mechanisms and laws including sanctions against infringements have been introduced (see [1]–[5]). Those in favor of the cooperative approach have proposed to introduce a volunteer-based tax to be paid by those internauts that wish to continue to have access to copyrighted content. This is called the “global license” approach. The revenues of the tax would be distributed among copyright holders and disseminators. This approach is already widely used for handling private copies of copyrighted material from the radio and television. It is directly included in the price of video cassettes or other storage devices. This approach, while widely supported by authors and performers (see, e.g., [3]), did not gain the support of the music and film industry. Though gaining ground, the confrontation policy faces the huge monitoring cost to provide credible evidence for unauthorized downloads [10].

The economic impact of nonauthorized downloads of copyrighted material is not clear. On one hand, some measure the impact on content owners by the loss of revenue that is computed by multiplying the number of nonauthorized downloads by the average price of the downloaded file. Others measure the economic impact of nonauthorized downloads in terms of the amount of money spent by a file sharer in consuming music. These come with contrasting conclusions. Indeed, a recent study [6] commissioned by the Dutch government argues that there is no direct relationship between downloading files protected by copyright and purchasing music in physical format. One of the findings points out that file sharers are not more or less willing to buy music than other people, and those file sharers that buy music do not buy more or less music than non-file sharers, but they acquire more value added products. This confirms the empirical finding of [7].

On the ethical side, those in favor of the confrontation approach have launched publicity campaigns trying to associate to unauthorized downloads of copyrighted content the image of stealing CDs from the shelf of a shop. A radically different approach had been expressed already two centuries ago by Victor Hugo, who wrote the following:

A book, as a book, belongs to the author, but as a thought, it belongs, without exaggerating, to the human mind. All minds are entitled. If one of the two rights, the right of the writer and the right of the human mind, were to be sacrificed, it would certainly be the right of the writer, because the public interest is our concern [8].

Manuscript received August 30, 2011; revised May 25, 2012; accepted August 11, 2012; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor S. Fahmy. Date of publication October 16, 2012; date of current version June 12, 2013. The work of A. Schwartz was supported in part by the US-Israel BSF under Grant 2008466 and the Technion’s fund for promotion of research. Part of this work appeared in the IEEE International Conference on Computer Communications (INFOCOM), Shanghai, China, April 10–15, 2011. (*Corresponding author: Y. Xu.*)

E. Altman, P. Nain, and Y. Xu are with the Maestro Project-Team, INRIA Sophia Antipolis, 06902 Sophia Antipolis, France (e-mail: eitan.altman@inria.fr; philippe.nain@inria.fr; yuedong.xu@gmail.com).

A. Schwartz is with the Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: adam@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNET.2012.2217505

Our goal in this paper is not to judge which approach is better, nor is it to examine the methods used today to detect infringement, whose reliability has been shown to be quite questionable [9]. Instead, we aim to study the efficiency of steps to reduce piracy as a function of measures related to the popularity of the contents and the cooperation of P2P users. There have been a couple of works addressing a similar issue. Authors of [1] and [4] have analyzed the impact of the effort, the authorities, or content owners, invested in: 1) reducing file uploading in P2P networks, and 2) reducing the demand for files, on the availability of files, and, more generally, on the operation of the P2P networks. The stationary analysis there is based on an M/G/ $\infty$  queuing 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 request or departure rate in a P2P network be reduced in order to have a significant change in file availability? To achieve that, we consider abstract models of a torrent in simplified P2P networks, where a large number of peers are interested in a file that is initially available at a small fraction of the population.

Our models are formulated as epidemic-type processes of file dissemination. We consider both *cooperative* peers, which are those that make a file available to other peers as soon as they obtain the file, and *free-riders*, who leave the system immediately after obtaining the file. To understand the impact of measures against the cooperative sharing behavior, we parameterize the degree of free-riding in the system as well as the degree of cooperation.

The P2P dynamics is modeled by a Markov chain (Section II) that is approximated in two specific regimes. The first (Section III) is the early stage when a large fraction of the population does not yet have the file. The system is then well approximated by a branching process. In the case that there is a positive probability of not getting extinct in the first regime, the system is shown in Section IV to move with some nonzero 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. A similar fluid limit is studied in Section VII for the case of limitation on uplink or downlink speed. We briefly state our contributions.

*Modeling and Approximating the Transient Behavior:* Our first important contribution is to show in what sense each of the above two models approximates the original Markov chain, and how to use both in order to get the whole transient behavior of the P2P network. This is in contrast with all other models of P2P networks that we know of, which either use only a branching process approach [25] or use only an epidemic mean-field approximation [22]. The latter approach (of using only the mean-field limit) is shown to provide a tight approximation when the initial number of peers with the file scales linearly with the total size  $N$  of the population of peers. With a fixed initial number of nodes that does not scale with  $N$ , there is a positive probability of early extinction (see Section VIII for detail) for any set of system parameters, and this probability cannot be predicted by the mean-field limit alone.

*Analysis and Identifying Phase Transitions:* We first study a P2P model that corresponds to the epidemic-like file dissemination (Sections II–IV). We then study a second model (Section VII) in which, at random times, each peer

contacts another peer randomly chosen within the set of existing peers. In both cases, we show the existence of phase transitions: A small change in the parameters causes a large change in the network behavior.

A phase transition occurs both in the branching model for the extinction time and in the epidemic model for the file availability. In the branching process, the existence of two phases was not known to Galton and Watson (considered as the founders of branching processes) and was only discovered and proved later in [11]. In the epidemiology community, the phase transition was already known in [12] for a model equivalent to our first model without the free-riders. We generalize our analytical results to the swarm composed of heterogeneous cooperative peers (Section VI). For the second model [13], we show the existence of two phase transitions, one for the file availability and the other one for the maximum torrent size.

*Application:* In Section V, we present a counteraction against unauthorized file sharing in the presence of illegal publishers. We evaluate the impact of measures against Internet piracy on the performance of P2P systems in Section VIII (see Fig. 13).

The accuracy of the various approximations is investigated in Section VIII, related studies are discussed in Section IX, and concluding remarks are given in Section X.

## II. MODEL

### A. Assumptions

Assume there is a population of  $N$  peers interested in a single file. Let  $Y(t)$  be 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 *noncooperative* peers. Once a cooperative peer has acquired the file, it stays in the network for a random time distributed according to an exponential random variable (rv) with parameter  $1/\mu \geq 0$  and then leaves the network. During the lingering time of a cooperative peer with the file, it participates in the file dissemination. A noncooperative peer, also called a *free-rider*, leaves the network at once when it receives the file. Note that “free-riders” in our context is an abstract description of noncooperative behaviors, which is different from that in the current BitTorrent system.

Let  $X_c(t)$  and  $X_f(t)$  denote the number of cooperative peers without the file and the number of free-riders (necessarily without the file) at time  $t$ , respectively. Define the process  $\mathbf{Y} := \{(Y(t), X_c(t), X_f(t)), t \geq 0\}$ . Let  $(Y(0), X_c(0), X_f(0))$  denote the initial state of  $\mathbf{Y}$  that has  $Y(0) + X_c(0) + X_f(0) = N$ . Let the ratio of various types of peers be  $(y_0, x_{c,0}, x_{f,0}) := \left(\frac{Y(0)}{N}, \frac{X_c(0)}{N}, \frac{X_f(0)}{N}\right)$ . For simplicity, we introduce new variables  $N_c = X_c(0)$  and  $N_f = X_f(0)$ . For clarity, the notations are listed in Table I.

We consider an abstract P2P network in which the file acquisition is via random contact between pairwise peers. When two such peers meet, the cooperative peer transmits the file to the other peer. 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. The transmission of the file is always supposed to be successful. This model describes a general P2P swarm without a tracker, and even the spreading of a file in current Internet. It is inspired by the contact process in [13] and [28]. One of the main differences lies in that a peer contacts all

TABLE I  
GLOSSARY OF MAIN NOTATION

Notation	Definitions
<b>General</b>	
$Y(t)$	Total number of peers with file at time $t$
$X_c(t)$	Number of cooperative peers without file at time $t$
$X_f(t)$	Number of free riders at time $t$
$N$	$Y(0) + X_c(0) + X_f(0)$ , total number of peers at $t = 0$
$N_c$	$X_c(0)$
$N_f$	$X_f(0)$
$y_0$	$Y(0)/N$
$x_{c,0}$	$X_c(0)/N = N_c/N$
$x_{f,0}$	$X_f(0)/N = N_f/N$
$r$	$(X_c(0) + Y(0))/N$ , initial ratio of cooperative peers
$\lambda$	Pairwise peer contact rate
$\mu$	Peer departure rate (= degree of cooperation)
$d$	$\lambda N_c + \mu$
$\rho$	$\lambda N_c / \mu$
<b>Section III – Branching model</b>	
$Y_b(t)$	Number of peers with file at time $t$
$p_0$	Prob. that an infected peer dies
$p_2$	Prob. of an infected peer replaced by two infected peers
$T_b(k)$	Extinction time with $k$ infected peers at $t = 0$
$G_k(t)$	Prob $\{T_b(k) < t\}$ (branching)
$q_k$	Extinction prob. with $k$ infected peers at $t = 0$
<b>Section IV – Mean-field model</b>	
$\beta$	$N\lambda$
$\theta$	$\beta/\mu$
<b>Section V-B</b>	
$\alpha$	Investment level of the content owner against P2P
$T_N$	Delay for a peer to get the file
$P_N(t)$	Prob $\{T_N < t\}$
$Y_N^*$	Number of permanent publishers
$y^*$	$Y_N^*/N$
$h(\alpha)$	Utility of content owner
<b>Section VI-A</b>	
$h_i(s_1, s_2)$	Generating function of offsprings of a class- $i$ infected peer
$\bar{q}_i(t)$	Survival probability of file at time $t$ when a single class- $i$ peer obtains file initially
$q_i^*$	$\bar{q}_i(t)$ as $t \rightarrow \infty$
$G_{k_1, k_2}(t)$	CDF of extinction time with two classes of peers

other connected peers in the system instead of only one random peer periodically. The file transmission time is assumed to be negligible compared to the time it takes for two peers to meet, and therefore this time is taken to be zero. The “meeting time” has various interpretations in practical systems. We hereby give two examples. First, the time for a peer to get a file from another peer is composed of two parts, the duration of contacting this peer and the duration of downloading the content. When a small file (e.g., MP3 music or e-book) is released, the downloading time through P2P networks is less than several minutes. However, the time that a peer is interested in this content can be much longer. Second, in mobile or online P2P swarms, the disseminated items are usually small messages, or links of contents cached in some temporary file servers. Each peer stays idle for an exponentially distributed time, and then contacts other peers. It can obtain the content immediately. The same assumption can also be found in [34].

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  $\lambda$ , that is after a time distributed according to an exponential rv with rate  $\lambda k$ .

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.,  $\mu$  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 II-B) 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 II-C).

### B. 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  $\{Y(t), t \geq 0\}$ .

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

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

In other words, the process  $\{Y(t), t \geq 0\}$  is a pure birth Markov process on the state space  $\{y_0, \dots, N\}$ , where state  $N$  is an absorbing state that is reached when all peers have the file.

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

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

The right-hand side of (2) cannot be expressed as a function of  $m(t)$ , thereby ruling out the possibility of finding  $m(t)$  in closed-form as the solution of an ordinary differential equation (ODE).

Let  $\lambda$  be written as  $\lambda = \beta/N$  and that  $\lim_{N \rightarrow \infty} N^{-1}Y(0) = y_0 \in (0, 1]$ . Then, for large  $N$ ,  $m(t)$  is well approximated by  $Ny(t)$ , where  $y(t)$  is obtained as the unique solution of the ODE [14, Theorem 3.1]

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

where  $f(u) := \beta u(1 - u)$  and  $y(0) = y_0 \in (0, 1]$  (conditions (3.2)–(3.4) in [14, Theorem 3.1] are clearly satisfied). It is found that

$$y(t) = \frac{y_0}{(y_0 + (1 - y_0) \exp(-\beta t))}, \quad t \geq 0. \quad (4)$$

This is a well-known instance (see, e.g., [23]) of what is known as mean-field approximation, a theory that focuses on the solution of ODEs obtained as limits of jump Markov processes [14]. The ODE (3) has been extensively used in epidemiology studies, where  $y(t)$  represents the fraction of infected patients at time  $t$  when the population is of size  $N$ .

Proposition 1, whose proof appears in the Appendix, states that the mean-field approximation is an upper bound for  $E[Y(t)]$ .

*Proposition 1:*  $E[Y(t)] \leq N y_0 / (y_0 + (1 - y_0) e^{-\beta t}) \forall t \geq 0$ .

### C. General Network

We consider the general network defined in Section II-A. Define the vector  $X(t) = \begin{pmatrix} X_c(t) \\ X_f(t) \end{pmatrix}$ , 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 II-A, it is seen that the process  $\mathbf{Y} = \{(Y(t), X(t)), t \geq 0\}$  is a finite-state Markov process whose nonzero transitions are given by

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t) + 1 \\ X(t) - e_c \end{pmatrix} \quad \text{with rate } \lambda Y(t) X_c(t) \quad (5)$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t) - 1 \\ X(t) \end{pmatrix} \quad \text{with rate } \mu Y(t) \quad (6)$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t) \\ X(t) - e_f \end{pmatrix} \quad \text{with rate } \lambda Y(t) X_f(t). \quad (7)$$

Throughout this paper, we will assume that  $\lambda > 0$  and  $\mu > 0$ .

The process  $\mathbf{Y}$  takes its values in the set

$$\mathcal{E} := \{(i, j, k), 0 \leq i \leq y_0 + N_c, 0 \leq j \leq N_c, 0 \leq j + k \leq N - y_0\}.$$

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.

An explicit characterization of the transient behavior of the absorbing Markov process  $\mathbf{Y}$  is a difficult task due both to the presence of nonlinear and nonhomogeneous transition rates in the state variables and to the dimension of  $\mathbf{Y}$ . In this paper, we will instead develop two approximations of the Markov process  $\mathbf{Y}$ . The first one, in Section III, will consist in replacing  $X_c(t)$  by  $N_c = X_c(0)$  in the transition rate (5), which will introduce a birth and death Markov branching process. As expected, this (so-called) branching approximation will lose its accuracy as the ratio  $X_c(t)/N_c$  decreases.

The second approximation, in Section IV, will use an asymptotic argument as  $N \rightarrow \infty$  based on a mean-field approximation of  $\mathbf{Y}$ . This approximation is justified if the initial state of  $\mathbf{Y}$  is of the order of  $N$ . Both the branching and the mean-field approximation approaches will allow us to approximate key characteristics of  $\mathbf{Y}$  such as the probability of disappearance of the file, the time before all files disappear, the maximum number of cooperative peers in the network, and the fraction of peers that eventually receive the file.

### III. BRANCHING APPROXIMATION

Let  $\mathbf{Y}_b := \{Y_b(t), t \geq 0\}$  be a Markov process on  $\mathbb{N} := \{0, 1, \dots\}$  (the subscript b refers to ‘‘branching’’) with nonzero transition rates given by

$$Y_b(t) \rightarrow Y_b(t) + 1 \quad \text{with rate } \lambda Y_b(t) N_c \quad (8)$$

$$Y_b(t) \rightarrow Y_b(t) - 1 \quad \text{with rate } \mu Y_b(t) \quad (9)$$

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 nonincreasing in  $t$ , a quick comparison between (5)–(7) and (8) and (9) indicates that the process  $\mathbf{Y}_b$  should dominate the process  $\mathbf{Y}$ . This bounding result is formalized and proved in Proposition 2.

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  $t > 0$ .

The Markov process  $\mathbf{Y}_b$  is an absorbing continuous-time birth and death process on  $\mathbb{N}$  with absorbing state 0. Because its transition rates are linear functions of the system state, this is also a continuous-time Markov *branching process* [18], 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  $\mathbf{Y}_b$ .

#### A. Extinction Probability

As previously observed, the process  $\mathbf{Y}_b$  is a birth and death branching process [18, Ch. V]. Each peer of this process has a probability of change in the interval  $(t, t+h)$  given by  $\mu h + o(h)$  with  $d = \lambda N_c + \mu$ ; with probability  $p_0 = \mu/d$ , an infected peer dies (a peer leaves) and with probability  $p_2 = \lambda N_c/d$ , an infected peer with the file is replaced by two infected peers (a peer receives the file). With certain abuse of notation, we let  $f(s) = p_0 + p_2 s^2$  be the probability generating function (p.g.f.) of the birth–death process.

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

$$T_b(k) = \min\{t > 0 : Y_b(t) = 0\}.$$

Let  $G_k(t) := P(T_b(k) < t)$  be the cumulative distribution function (CDF) of  $T_b(k)$ . Given  $Y_b(0) = k$ , the extinction probability,  $q_k$ , is given by  $q_k = G_k(\infty)$ . The CDF of  $T_b(1)$  is obtained from the differential equation  $\frac{dG_1(t)}{dt} = -\lambda(G_1(t) - f(G_1(t)))$  implied by [18, Eq. (7.3), p. 104]. Thus, we have

$$G_1(t) = \frac{1 - e^{-\mu(1-\rho)t}}{1 - \rho e^{-\mu(1-\rho)t}}, \quad t \geq 0 \quad (10)$$

where  $\rho$  defined as  $\rho = \lambda N_c / \mu$  is not 1. From (10), we find

$$q_1 = \min\{1, 1/\rho\}. \quad (11)$$

In other words, the extinction will be certain iff  $\rho < 1$ . Since all peers behave independently of each other, we have  $q_k = q_1^k = \min\{1, 1/\rho^k\}$  and

$$G_k(t) = G_1(t)^k = \left( \frac{1 - e^{-\mu(1-\rho)t}}{1 - \rho e^{-\mu(1-\rho)t}} \right)^k, \quad t \geq 0. \quad (12)$$

When  $\mathbf{Y}_b$  is a critical branching process (i.e.,  $\rho = 1$ ), the extinction probability is given by

$$G_k(t) = \left( \frac{\mu t}{1 + \mu t} \right)^k, \quad t \geq 0. \quad (13)$$

#### B. Expected Time to Extinction

Assume that  $\rho < 1$  (extinction is certain). The expected extinction time is equal to  $E[T_b(k)] = \int_0^\infty (1 - G_k(t)) dt$ . In particular

$$E[T_b(1)] = -\frac{\log(1 - \rho)}{\mu\rho}. \quad (14)$$

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

$$P(T(y_0) > t) = P(Y(t) > 0) \leq P(Y_b(t) > 0) = G_{y_0}(t).$$

In particular  $E[T(y_0)] \leq E[T_b(y_0)]$ , so that  $E[T(1)] \leq -\frac{\log(1-\rho)}{\mu\rho}$  from (14) for  $\rho < 1$ .

When there are more than one seeds at time 0, we can compute the expected extinction time in a similar way. Using some transformations in the integral [19], we obtain

$$\begin{aligned} E[T_b(k)] &= \int_0^\infty (1 - G_k(t)) dt \\ &= \int_0^1 \frac{1 - k^\omega}{\mu + (\mu + \lambda N_c)\omega + \lambda N_c \omega^2} d\omega. \end{aligned} \quad (15)$$

#### IV. MEAN-FIELD APPROXIMATION

In this section, we investigate the behavior of the process  $\mathbf{Y}$  defined in Section II-C 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 [14] for the theory and [21], [23], and [28] for recent applications in the area of file-sharing systems.

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

$$Y(0) = N y_0 \quad X_c(0) = N x_{c,0} \quad X_f(0) = N x_{f,0} \quad (16)$$

with  $y_0 + x_{c,0} + x_{f,0} = 1$ . (The analysis below holds under the weaker condition  $\lim_{N \rightarrow \infty} N^{-1}(Y(0), X_c(0), X_f(0)) = (y_0, x_{c,0}, x_{f,0})$ .)

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 nonzero transition rates of the process Markov process  $\mathbf{Y}$  out of state  $Y = (Y_1, Y_2, Y_3)$ . We have [cf. (5)–(7)]

$$\begin{aligned} g(Y, Y + v_1) &= \frac{\beta}{N} Y_1 Y_2, \quad g(Y, Y + v_2) \\ &= \mu Y_1, \quad g(Y, Y + v_3) = \frac{\beta}{N} Y_1 Y_3 \end{aligned}$$

which can be rewritten as

$$g(Y, Y + v_i) = N f\left(\frac{Y}{N}, v_i\right), \quad i = 1, 2, 3 \quad (17)$$

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  $u = (u_1, u_2, u_3)$ .

We may therefore use [14, Theorem 3.1] (it is easily seen that conditions (3.2)–(3.4) in [14] are satisfied) to obtain that the rescaled process  $N^{-1}\mathbf{Y}$  converges in probability as  $N \rightarrow \infty$ , uniformly on all *finite* intervals  $[0, T]$ , to the solution  $(y, x_c, x_f)$ ,  $0 \leq y, x_c, x_f, y + x_c + x_f \leq 1$ , of the system of ODEs

$$\frac{d}{dt} \begin{pmatrix} y \\ x_c \\ x_f \end{pmatrix} = \begin{pmatrix} y(\beta x_c - \mu) \\ -\beta y x_c \\ -\beta y x_f \end{pmatrix} \quad (18)$$

with initial condition  $(y_0, x_{c,0}, x_{f,0})$ .

In particular, for any finite  $t$ , the solution  $y, x_c, x_f$  of (18) will approximate the fraction of peers with the file, the fraction of

cooperative peers without the file and the fraction of free-riders, respectively, at time  $t$ . The accuracy of this approximation will increase with  $N$ , the total number of peers.

#### A. Peers That Never Receive the File: A Phase Transition

The fraction of cooperative peers  $x_c$  and the fraction of free-riders  $x_f$  that do not have the file monotonically decrease (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 address 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 peers that never receive the file.

Let  $\theta := \beta/\mu$ . From the first two equations in (18) we obtain  $x_c$  as

$$\frac{dy}{dx_c} = -1 + \frac{1}{\theta x_c}. \quad (19)$$

The solution of this differential equation is

$$x_c + y = \theta^{-1} \ln x_c + \phi(\theta) \quad (20)$$

where  $\phi(\theta) := x_{c,0} + y_0 - \theta^{-1} \ln x_{c,0}$ . Let  $y^{\max}$  be the maximum ratio of cooperative peers with the file. According to the first equation in (18),  $y^{\max}$  is reached when  $x_c = \theta^{-1}$  if  $\theta > 1$  and is expressed as

$$y^{\max} = -\theta^{-1}(1 + \ln \theta) + \phi(\theta). \quad (21)$$

When  $\theta \leq 1$ ,  $y^{\max}$  is reached when  $x_c = x_{c,0}$  (i.e., at time  $t = 0$ ). On the other hand, as  $t \rightarrow \infty$ ,  $y$  is approaching 0 (since we have assumed that  $\mu > 0$ ) so that, from (20),  $x_c(\infty)$  satisfies the equation

$$x_c(\infty) - \theta^{-1} \ln(x_c(\infty)) - \phi(\theta) = 0. \quad (22)$$

It is easily seen that this equation has a unique solution in  $(0, x_{c,0})$  [note that  $x_c(t) \leq x_{c,0}$  for any  $t$  since  $x_c$  is non-increasing from the second equation in (18)]. From (18), we find that  $x_f(t) = \frac{x_{f,0}}{x_{c,0}} x_c(t)$  for all  $t$ .

As recalled earlier, the mean-field approximation only holds for finite  $t$ , and there is therefore no guarantee that it will hold when  $t = \infty$ , namely, that  $N\mathbf{Y}^{-1}$  will converge in probability to  $(0, x_c(\infty), x_{f,0}x_c(\infty)/x_{c,0})$  as  $N \rightarrow \infty$ . However, due to the particular structure of the infinitesimal generator of  $\mathbf{Y}$ , this convergence takes place as shown in [24, Sec. 5.2]. We consider an alternative rescaled Markov process  $\tilde{\mathbf{Y}} := \{(\tilde{Y}(t), \tilde{X}_c(t), \tilde{X}_f(t)), t \geq 0\}$  with generator  $\tilde{g}(\cdot, \cdot) = g(\cdot, \cdot)/Y_1$  and same state space as  $\mathbf{Y}$ , so that starting from the same initial condition, the terminal values of  $X_c(t)$  and  $\tilde{X}_c(t)$  (resp.  $X_f(t)$  and  $\tilde{X}_f(t)$ ) will have the same distribution. The mean-field approximation for  $\tilde{\mathbf{Y}}$  shows that the solution of the associated ODEs is given by  $(0, x_{c,0}e^{-\beta\tau}, x_{f,0}e^{-\beta\tau})$  for any  $t \geq \tau$ , with  $\tau$  the unique solution in  $(0, \infty)$  of  $x_{c,0} + y_0 = x_{c,0}e^{-\beta\tau} + \mu\tau$ , from which the result follows.

In summary, as  $N$  is large, the fraction of cooperative (resp. free-riders) peers that will never receive the file is approximated by  $x_c^{\min} := x_c(\infty)$  (resp.  $x_f^{\min} := x_{f,0}x_c^{\min}/x_{c,0}$ ), where  $x_c(\infty)$  can be (numerically) calculated from (22).

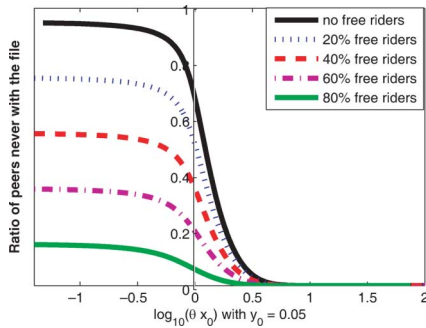


Fig. 1. Ratio of cooperative peers (as  $N$  is large) that never receive the file as a function of  $\log_{10}(\theta x_{c,0})$ .

We are interested in whether there is an abrupt change in content availability (i.e.,  $x_c(\infty)$ ) with the parameter  $\theta$ . Obviously, if  $\theta$  is 0, all the cooperative peers that do not have the file at time 0 will never receive it. To find a phase transition, we approximate  $\log(x_c(\infty))$  in (22) by using its Taylor extension at  $x_{c,0}$  and obtain

$$x_c(\infty) \approx \left( \left( \frac{1}{\theta} - x_{c,0} - y_0 \right) + \frac{1}{2\theta} \left( \frac{x_c(\infty)}{x_{c,0}} - 1 \right)^2 \right) / \left( \frac{1}{\theta x_{c,0}} - 1 \right).$$

Since the expression  $\frac{1}{\theta} \left( \frac{x_c(\infty)}{x_{c,0}} - 1 \right)^2$  is bounded, the phase transition happens at  $\theta = 1/x_{c,0}$ .

Despite the similarity in the definitions of  $\rho$  in Section III and of  $\theta x_{c,0}$  in the present section, the phase transition at  $\rho = 1$  is different in nature from that at  $\theta x_{c,0} = 1$ . The former indicates whether or not the file will be extinct, while the latter will drastically impact the final size of the torrent.

Fig. 1 displays the mapping  $\log_{10}(\theta x_{c,0}) \rightarrow x_c^{\min}$  for  $x_{c,0} \in \{0.01, 0.1, 0.3, 0.5, 0.9\}$  and  $y_0 = 0.05$ . The curves for  $x_c^{\min}$  are monotonically decreasing in  $x_{c,0}$  (the curve that intersects the vertical axis close to 1 is the one corresponding to  $x_{c,0} = 0.01$ , and so on.). For each curve, we note the existence of a phase transition at  $\theta x_{c,0} = 1$ , which is more pronounced as the ratio of cooperative peers increases.

### B. Combining the Branching and the Epidemic Model

The mean-field approximation is accurate for large  $N$  if the initial state scales with  $N$  linearly. In the case that  $N$  is very large but the initial condition does not scale with  $N$  (e.g.,  $Y(0) = 1$ ,  $X_c(0) + X_f(0) = N - 1$ ), we can do the following. Fix some  $N_0$  much smaller than  $N$  but larger than 1. Use the branching process approximation until the number of peers with the file is  $N_0$ . Then, switch to the epidemic model. (For the branching process, we recall that given that there is no extinction, the population size grows exponentially fast.)

## V. CONTROL ACTIONS AGAINST P2P NETWORKS

In this section, we first investigate the major findings in the analysis of content availability. A set of control actions are proposed to protect copyrighted files against P2P file sharing.

### A. Observations on File Availability and Countermeasures

Before proposing the counteractions against illegal P2P swarms, we investigate the impact of measures on file availability. The main question is how does a decrease or increase in one of the system parameters affect measures such as the following:

- 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.

According to the analysis in Sections III and IV, all above measures depend on the ratio  $\frac{\lambda}{\mu}$  (or  $\frac{\beta}{\mu}$  equivalently). A small ratio  $\frac{\lambda}{\mu}$  means a poor availability of the file.

By introducing countermeasures, the peers without the file might have difficulties (i.e., taking a longer period) in finding other peers with the file, or the peers with the file may provide downloading service for a short period (increasing  $\mu$ ).

To enforce the countermeasures, we need new policies or new strategies. For example, if the content owners use pollution attack, the peers without the file observe a large amount of useless “copies.” Hence, the rate of contacting other peers,  $\lambda$ , may decrease. Similarly, a P2P community might allow comments on a file. If the content owner places a lot of adverse comments (e.g., low quality or resolution of an MP3/Video file), the number of interested users  $N$  and the contact rate  $\lambda$  may decrease. Meanwhile, if new threatening policies are enforced either by the lawmaker or by the content owners, the peers with the file may provide downloading for a shorter period. An extreme case is that a peer leaves the system immediately after obtaining the file.

As mentioned above, countermeasures are various in forms. In what follows, we consider one particular scenario in which a small number of persistent illegal publishers resides in the swarm. They aim to spread the copyright protected file as fast as possible in the P2P swarm. To combat with undesirable file sharing, the content owner presents one simple method, namely *cooperation control*. The basic idea is to discourage the degree of cooperation of peers with the file. Here, the contact rate  $\lambda$  is deemed as an intrinsic parameter of P2P swarms that can hardly be changed technically.

### B. Control of Cooperation

We introduce the cooperation control to prevent the dissemination of copyrighted files. We aim to reduce the degree of cooperation (i.e., increasing  $\mu$ ) so that the delay of obtaining the file is increased. To achieve this goal, the content owner can invest a certain amount of money in the very beginning to discourage the cooperation of peers. The cooperation control is a case of confrontation strategy. In other words, we are focusing on this unilateral action of the content owner against unauthorized file dissemination.

We consider the same model as in Section II-C, but we now assume that all peers are cooperative and that there is a number  $Y_N^* > 0$  of permanent publishers, where the subscript  $N$  refers to the total number of peers in the system at time  $t = 0$ . The pairwise contact rate is  $\lambda = \beta/N$ . Denote by  $\alpha$  the investment level of the content owner against P2P networks. The departure rate is an increasing function of  $\alpha$ ,

denoted by  $\mu(\alpha)$ . We denote by  $Y_N(t)$  the number of nonpermanent publishers, and by  $X_N(t)$  the number of peers without the file at time  $t$ . Observe that  $Y_N(0) + X_N(0) = N - Y_N^*$ . If  $\lim_{N \rightarrow \infty} Y_N(0)/N = y_0$  and  $\lim_{N \rightarrow \infty} X_N(0)/N = x_0$ , which implies that  $\lim_{N \rightarrow \infty} Y^*/N = 1 - x_0 - y_0 := y^*$ , then, by Kurtz's result [14], the rescaled process  $N^{-1}(Y(t), X(t))$  converges in probability as  $N \rightarrow \infty$ , uniformly on all finite intervals  $[0, T]$ , to the solution of the ODEs

$$\frac{d}{dt} \begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \beta(y + y^*)x - \mu(\alpha)y \\ -\beta(y + y^*)x \end{pmatrix} \quad (23)$$

with initial state  $(y_0, x_0)$ . From now on, we will assume that  $y^* > 0$ .

Consider an arbitrary peer without the file at time  $t = 0$ , and denote by  $T_N$  the time that elapses before it receives it. Let  $P_N(t) := P(T_N < t)$ . Similarly to [23, p. 6], we find

$$\frac{dP_N(t)}{dt} = \beta(1 - P_N(t)) \frac{E[Y_N(t)] + Y_N^*}{N}. \quad (24)$$

Solving for  $P_N(t)$  gives

$$P_N(t) = 1 - e^{-\beta \int_0^t \frac{E[Y_N(s)] + Y_N^*}{N} ds}, \quad t \geq 0. \quad (25)$$

Hence

$$E[T_N] = \int_0^\infty (1 - P_N(t)) dt = \int_0^\infty e^{-\beta \int_0^t \frac{E[Y_N(s)] + Y_N^*}{N} ds} dt.$$

From the above, we know that  $(E[Y_N(t)] + Y_N^*)/N \rightarrow y + y^*$  as  $N \rightarrow \infty$  for every  $t > 0$ , so that from (25)

$$\lim_{N \rightarrow \infty} (1 - P_N(t)) = e^{-\beta \int_0^t (y(s) + y^*) ds} \quad (26)$$

for every  $t > 0$ . On the other hand,  $\lim_{N \rightarrow \infty} Y_N^*/N = y^*$  implies that for  $0 < \epsilon < y^*$  there exists  $N_\epsilon$  such that  $Y_N^*/N > y^* - \epsilon$  for all  $N > N_\epsilon$ . Therefore, from (25)

$$1 - P_N(t) \leq e^{-\beta(y^* - \epsilon)t} \quad (27)$$

for  $N > N_\epsilon, t > 0$ . Since the right-hand side (r.h.s.) of (27) is integrable in  $[0, \infty)$ , (26) and (27) allow us to apply the bounded convergence theorem to conclude that

$$\bar{T}(\alpha) := \int_0^\infty \lim_{N \rightarrow \infty} (1 - P_N(t)) dt = \int_0^\infty e^{-\beta \int_0^t (y(s) + y^*) ds} dt.$$

The objective of the content owner is to choose an investment level  $\alpha \geq 0$  which will maximize its utility

$$h(\alpha) := \bar{T}(\alpha) - \alpha. \quad (28)$$

To understand the impact of cooperation control on the delay, we present numerical studies in Section VIII.

## VI. TAKING HETEROGENEITY INTO CONSIDERATION

In the preceding analysis, we assume that peers are either free-riding or cooperative with the same effort. A natural question is whether our methodologies can handle the swarm with heterogeneous types of cooperative peers. In this section, we classify the cooperative peers into two groups, the *long-term* (or

class-1) and the *short-term* (or class-2) cooperative peers. The former group has a small departure rate  $\mu_1$ , while the latter has a large departure rate  $\mu_2$ , (i.e.,  $\mu_2 > \mu_1$ ). Here, we only consider the random contact model without free-riders. Later on, we will show that our analytical framework can be easily generalized to incorporate the case with more classes and with the random contact model in Section VII.

Let  $N_1$  and  $N_2$  be the numbers of class-1 and class-2 peers without the file in the beginning (the term ‘‘cooperative’’ is removed since all the peers are cooperative). We denote by  $X_i(t)$  (resp.  $Y_i(t)$ ) the number of class- $i$  peers without (resp. with) the file, for  $i \in \{1, 2\}$ . Define the stochastic process  $\mathbf{H} := \{X_1(t), X_2(t), Y_1(t), Y_2(t), t > 0\}$ . Let  $\{X_1(0), X_2(0), Y_1(0), Y_2(0)\}$  be the initial state of the process  $\mathbf{H}$ . Let  $N$  be the total number of peers in the swarm that has  $N = X_1(0) + X_2(0) + Y_1(0) + Y_2(0)$ . Given the assumptions in Section II, a state transition of  $\mathbf{H}$  is independent of all the past states, which is Markovian. With certain abuse of notations, we define two vectors of states as  $X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}$  and  $Y(t) = \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}$ . Define a set of constant vectors  $e_1 = \{1, 0, -1, 0\}^T$ ,  $e_2 = \{0, 1, 0, -1\}^T$ ,  $f_1 = \{-1, 0, 0, 0\}^T$  and  $f_2 = \{0, -1, 0, 0\}^T$ . The state transition of the process  $\mathbf{H}$  is described as follows:

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} + e_i \quad \text{with rate } \lambda(Y_1(t) + Y_2(t))X_i(t) \quad (29)$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} + f_i \quad \text{with rate } \mu_i Y_i(t). \quad (30)$$

To investigate the transient behaviors of the process  $\mathbf{H}$ , we still resort to the branching process and the mean-field approximations. However, the single-type Markov branching process is no longer applicable. This is because a class- $i$  peer with the file may give birth to an offspring of either class- $i$  or class- $j$  in each contact. We propose to study the early extinction of the file using multitype Markov branching process.

### A. Two-Type Markov Branching Approximation

We approximate the process  $\mathbf{H}$  by the corresponding 2-type Markov branching process  $\mathbf{H}_b$ . Assuming that the number of peers without the file does not change over time (i.e.,  $X_1(t) = N_1$  and  $X_2(t) = N_2$ ), we obtain the process  $\mathbf{H}_b$  with linear state transition rates

$$Y_i(t) \rightarrow Y_i(t) + 1 \quad \text{with rate } \lambda(Y_1(t) + Y_2(t))X_i(0) \quad (31)$$

$$Y_i(t) \rightarrow Y_i(t) - 1 \quad \text{with rate } \mu_i Y_i(t). \quad (32)$$

Different from Section III, the offspring of a peer in the 2-type branching process is expressed as a vector. For example, a class- $i$  peer with the file produces either one class- $j$  (i.e., itself) and one class- $i$  infected peer, or two class- $i$  infected peers (including itself) in a birth event. When this peer leaves the swarm, a death event happens in a particular type of peer. Obviously, a class- $i$  infected peer cannot give birth to two class- $j$  offsprings. Here, we define four constant vectors to describe the offspring vectors of the process  $\mathbf{H}_b$ :  $O_1 = \{0, 0\}^T$ ,

$O_2 = \{1, 1\}^T$ ,  $O_3 = \{2, 0\}^T$ , and  $O_4 = \{0, 2\}^T$ . The  $j$ th element of vector  $O_l$  represents the number of class- $j$  offspring,  $\forall j \in \{1, 2\}$  and  $l \in \{1, 2, 3, 4\}$ . The offspring of a class-1 infected peer is in the set  $\{O_1, O_2, O_3\}$ , and that of a class-2 infected peer is in the set  $\{O_1, O_2, O_4\}$ . Denote by  $p_i^l$  the probability of a class- $i$  infected peer to have an offspring  $O_l$ ,  $\forall l \in \{1, 2, 3, 4\}$ , when an event happens on the class-1 peers. Define  $N_{1\cup 2} := N_1 + N_2$ . We obtain

$$\begin{aligned} p_1^1 &= \frac{\mu_1}{\lambda N_{1\cup 2} + \mu_1} & p_1^2 &= \frac{\lambda N_2}{\lambda N_{1\cup 2} + \mu_1} \\ p_1^3 &= \frac{\lambda N_1}{\lambda N_{1\cup 2} + \mu_1} & p_1^4 &= 0. \end{aligned}$$

Similarly

$$\begin{aligned} p_2^1 &= \frac{\mu_2}{\lambda N_{1\cup 2} + \mu_2} & p_2^2 &= \frac{\lambda N_1}{\lambda N_{1\cup 2} + \mu_2} \\ p_2^4 &= \frac{\lambda N_2}{\lambda N_{1\cup 2} + \mu_2} & p_2^3 &= 0. \end{aligned}$$

The generating functions of an infected peer in class 1 and 2 can be written as

$$h_1(s_1, s_2) = p_1^1 + p_1^2 s_1 s_2 + p_1^3 s_1^2 \quad (33)$$

$$h_2(s_1, s_2) = p_2^1 + p_2^2 s_1 s_2 + p_2^4 s_2^2 \quad (34)$$

where  $s = \{s_1, s_2\}$  is a set of variables. Denote by  $\bar{q}_i$  (complementary to  $q_i$ ) the survival probability of the file at time  $t$  when a single class- $i$  peer obtains the file initially. Denote by  $d_i$  the rate of an event happening in class  $i$  ( $i \in \{1, 2\}$ ). According to the method in [20], the survival probabilities are the solutions of the following ODEs:

$$\begin{aligned} \frac{1}{d_i} \frac{d\bar{q}_i(t)}{dt} &= -\bar{q}_i + \sum_{l=1}^4 p_i^l (1 - (1 - \bar{q}_i)^{l_i} (1 - \bar{q}_j)^{l_j}) \\ &= -\bar{q}_i + \frac{\lambda N_i}{d_i} (2\bar{q}_i - \bar{q}_i^2) + \frac{\lambda N_j}{d_i} (\bar{q}_i + \bar{q}_j - \bar{q}_i \bar{q}_j) \end{aligned} \quad (35)$$

for  $i, j \in \{1, 2\}$ . The constant  $d_i = \lambda N_i + \mu_i$  is the Poisson rate of an event happening among the peers of class- $i$ . The (35) can be rewritten as

$$\begin{aligned} \frac{d}{dt} \begin{pmatrix} \bar{q}_1 \\ \bar{q}_2 \end{pmatrix} &= \begin{bmatrix} \lambda N_1 - \mu_1 & \lambda N_2 \\ \lambda N_1 & \lambda N_2 - \mu_2 \end{bmatrix} \cdot \begin{pmatrix} \bar{q}_1 \\ \bar{q}_2 \end{pmatrix} \\ &\quad - \begin{pmatrix} \lambda(N_1 \bar{q}_1^2 + N_2 \bar{q}_1 \bar{q}_2) \\ \lambda(N_2 \bar{q}_2^2 + N_1 \bar{q}_1 \bar{q}_2) \end{pmatrix} \\ &\leq \begin{bmatrix} \lambda N_1 - \mu_1 & \lambda N_2 \\ \lambda N_1 & \lambda N_2 - \mu_2 \end{bmatrix} \cdot \begin{pmatrix} \bar{q}_1 \\ \bar{q}_2 \end{pmatrix}. \end{aligned} \quad (36)$$

The inequality holds componentwise since the variables  $\bar{q}_1(t)$  and  $\bar{q}_2(t)$  are nonnegative. Define  $R$  to be the above  $2 \times 2$  constant matrix. One can see that  $R$  is irreducible. Let  $z$  be an eigenvalue of the matrix  $R$ . The extinction of the file in the 2-type branching process depends on the sign of the eigenvalues (see [20, Theorem 2.1]). The determinant of  $R$  is expressed as

$$\Delta = z^2 - (\lambda N_{1\cup 2} - \mu_1 - \mu_2)z + (\mu_1 \mu_2 - \lambda N_1 \mu_2 - \lambda N_2 \mu_1).$$

When both eigenvalues are negative, the file will disappear in the approximated branching process definitely. Given the

determinant  $\Delta$ , the process  $\mathbf{H}_b$  is subcritical (extinct for sure) if the parameters satisfy

$$\begin{cases} \mu_1 + \mu_2 - \lambda N_{1\cup 2} > 0 \\ \mu_1 \mu_2 - \lambda N_1 \mu_2 - \lambda N_2 \mu_1 > 0. \end{cases} \quad (37)$$

We next compute the asymptotic survival probability of a super-critical P2P swarm. Denote by  $\bar{q}_i^*$  the survival probability of file extinction as  $t \rightarrow \infty$  if the single seed belongs to class- $i$  initially. When the process  $\mathbf{H}_b$  is super-critical,  $\bar{q}_i^*$  can be solved numerically by letting the left side of (35) be 0, that is

$$(\lambda N_{i,c} - \mu_i) \bar{q}_i^* + \lambda N_j \bar{q}_j^* - \lambda N_i (\bar{q}_i^*)^2 - \lambda N_j \bar{q}_i^* \bar{q}_j^* = 0 \quad (38)$$

for  $\bar{q}_i^*, \bar{q}_j^* \in [0, 1], \forall i, j \in \{1, 2\}, i \neq j$ . Define  $q_i(t)$  to be the corresponding extinction probability at time  $t$ , and  $\bar{q}_i^*$  to be that as  $t \rightarrow \infty$ . We can easily obtain  $q_i(t) = 1 - \bar{q}_i(t)$  and  $q_i^* = 1 - \bar{q}_i^*, \forall i \in \{1, 2\}$ .

Next, we compute the extinction probability of the process  $\mathbf{H}_b$  with multiple seeds initially. Suppose that there are  $k_i$  peers with the file at time 0 in class- $i$ . Define  $G_{k_1, k_2}(t)$  to be the CDF of file extinction at time  $t$ . Because all the ancestors (infected peers at time 0) behave independently, the file extinction probability is given by

$$G_{(k_1, k_2)}(t) = (q_1(t))^{k_1} (q_2(t))^{k_2}. \quad (39)$$

## B. Mean-Field Analysis

When the number of peers  $N$  gets large, the Markov process  $\mathbf{H}$  can be approximated by the solutions of mean-field ODEs. Let  $x_i(t)$  (resp.  $y_i(t)$ ) be the ratio of class- $i$  peers without (resp. with) the file at time  $t$ . The initial state of the scaled process  $N^{-1}\mathbf{H}$  is given by  $\{y_{1,0}, y_{2,0}, x_{1,0}, x_{2,0}\} = \left\{ \frac{Y_1(0)}{N}, \frac{Y_2(0)}{N}, \frac{X_1(0)}{N}, \frac{X_2(0)}{N} \right\}$ . We apply Kurtz theorem [14] and obtain the rescaled process  $N^{-1}\mathbf{H}$  that converges to the solution of the following ODEs on all finite time intervals as  $N \rightarrow \infty$ :

$$\frac{d}{dt} \begin{pmatrix} y_1 \\ y_2 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \beta(y_1 + y_2)x_1 - \mu_1 y_1 \\ \beta(y_1 + y_2)x_2 - \mu_2 y_2 \\ -\beta(y_1 + y_2)x_1 \\ -\beta(y_1 + y_2)x_2 \end{pmatrix}. \quad (40)$$

The last two equations in (40) give rise to  $x_2(t) = \frac{x_{2,0}}{x_{1,0}} x_1(t)$ . In the P2P swarm with two types of cooperative peers, the rescaling method in [24, Sec. 5.2] does not work. Hence, there is no guarantee that this mean-field approximation holds as  $t$  approaches infinity.

## VII. P2P WITH A FIXED REQUEST RATE PER NODE

### A. Model

In this section, we will consider a slight variation of the model in [13]: There are  $N$  peers at time  $t = 0$ , at least one of them having a file. Each peer without the file sends a request for the file to another peer selected at random. These requests are initiated at Poisson rate  $\lambda > 0$ . It is assumed that a peer with the file leaves the system after an exponentially distributed random duration with rate  $\mu > 0$ . All these rvs are mutually independent. Let  $Y(t)$  (resp.  $X(t)$ ) be the number of peers with the file (resp. without the file) at time  $t$ . We have  $Y(0) + X(0) = N$  with  $Y(0) \geq 1$ . Under the above



assumptions,  $\mathbf{Z} := \{(Y(t), X(t)), t \geq 0\}$  is a Markov process on the set  $\mathcal{E} := \{(y, x) \in \{0, 1, \dots, N\}^2 : 0 \leq y + x \leq N\}$ . Let  $q(z, z')$ ,  $z = (y, x)$ ,  $z' = (y', x') \in \mathcal{E}$ , denote its generator. Nonzero transition rates are given by

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t)+1 \\ X(t)-1 \end{pmatrix} \quad \text{with rate } \frac{\lambda Y(t)X(t)}{Y(t)+X(t)} \quad (41)$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \rightarrow \begin{pmatrix} Y(t)-1 \\ X(t) \end{pmatrix} \quad \text{with rate } \mu Y(t). \quad (42)$$

This model differs from our previous model in that the rate of increase is normalized by the total number of peers in the system. More precisely, the rate in (41) follows from the fact that with probability  $Y(t)/(Y(t)+X(t))$ , a peer without the file will contact a peer with a file at time  $t$  (the latter implicitly assumes that a peer may contact itself, as otherwise this probability would be  $Y(t)/(Y(t)+X(t)-1)$ ); the reason for doing this will next become apparent. Note that this assumption will have no effect when  $N$  gets large, so that the total rate of increase of the number of peers with the file is  $\lambda Y(t)X(t)/(Y(t)+X(t))$  at time  $t$ .

The same model is considered in [13] with the difference that in [13] there is one permanent publisher, thereby implying that all peers will receive the file. These authors show that the mean broadcast time is  $O(N)$  if  $\lambda < \mu$ , and is  $O(\log(N))$  if  $\lambda > \mu$ . Thus, there is a phase transition at  $\lambda = \mu$ .

In this section, we will instead focus on: 1) the file extinction probability and expected extinction time; 2) the fraction of peers that will receive the file (in the absence of a permanent publisher, this fraction is not always equal to 1); and 3) the maximum torrent size (maximum number of copies of the file at one time) as  $N$  is large. In all cases, we will show the existence of phase transitions.

### B. Branching Approximation

With certain abuse of notations, we denote by  $\mathbf{Y}_{\tilde{b}}$  a Markov process on  $N = \{0, 1, \dots, N\}$  that has

$$Y_{\tilde{b}}(t) \rightarrow Y_{\tilde{b}}(t) + 1 \quad \text{with rate } \lambda Y_{\tilde{b}} \quad (43)$$

$$Y_{\tilde{b}}(t) \rightarrow Y_{\tilde{b}}(t) - 1 \quad \text{with rate } \mu Y_{\tilde{b}}(t). \quad (44)$$

Here, the subscript  $\tilde{b}$  denotes the branching process with a fixed request rate per node. Compared to  $\mathbf{Z}$ , the Markov process  $\mathbf{Y}_{\tilde{b}}$  assumes  $X(t) \gg Y(t)$ . Because the continuous birth–death process  $\mathbf{Y}_{\tilde{b}}$  has linear rates of state transition, it is also a continuous-time *Markov branching process*.

To find the extinction probability and the expected time until extinction, one can resort to the analysis in Section III. Define  $\xi := \lambda/\mu$ . All the analytic results in Section IV can be tailored for the process  $\mathbf{Y}_{\tilde{b}}$  if  $\rho$  is replaced by  $\xi$ .

### C. Mean-Field Approximation

Our analysis will use Kurtz's theorem [14, Theorem 3.1] like in Section IV. Note, however, that both metrics 1) and 2) above require to use the mean-field limit as  $t \rightarrow \infty$ , something that Kurtz's result does not cover.

To overcome this difficulty, we will use the same argument as in [24] (see also Section IV, where this argument was already

used), taking advantage of the particular structure of the infinitesimal generator of the process  $\mathbf{Z}$ . More specifically, it is seen that the generator of  $\mathbf{Z}$  writes in the form  $q(z, z') = y\tilde{q}(z, z')$  for  $z = (y, x)$ ,  $z' = (y', x') \in \mathcal{E}$ , where nonzero transition rates are given in (41) and (42).

Let  $\tilde{\mathbf{Z}} = \{(\tilde{Y}(t), \tilde{X}(t)), t \geq 0\}$  be a Markov process with generator  $\tilde{q}(z, z')$  and state space  $\mathcal{E}$  (same state space as  $\mathbf{Z}$ ).

Since  $\tilde{\mathbf{Z}}$  has been obtained by changing the timescale of  $\mathbf{Z}$ , the final values of  $\tilde{X}(t)$  and of  $X(t)$  will have the same distribution (note that the final state of  $\tilde{Y}(t)$  and  $Y(t)$  is always zero since states  $(0, \cdot)$  are all absorbing states), and so will have the maximum torrent size.

Since the generator  $\tilde{q}(z, z')$  can be written as  $\tilde{q}(z, z') = Nf(z/N, z'/N)$  (this is where the assumption that a peer may contact itself is useful) and since conditions (3.2)–(3.4) in [14, Theorem 3.1] are clearly satisfied, we may apply [14, Theorem 3.1] to obtain that, at any finite time  $t$ ,  $N^{-1}(\tilde{Y}(t), \tilde{X}(t))$  converges in probability as  $N \rightarrow \infty$  to the solution  $(\tilde{y}, \tilde{x})$ ,  $0 \leq \tilde{y}, \tilde{x} \leq 1$ ,  $\tilde{y} + \tilde{x} \leq 1$ , of the ODEs

$$\dot{\tilde{y}} = -\mu + \lambda\tilde{x}/(\tilde{x} + \tilde{y}), \quad \dot{\tilde{x}} = -\lambda\tilde{x}/(\tilde{x} + \tilde{y}) \quad (45)$$

given that  $\lim_{N \rightarrow \infty} N^{-1}(\tilde{Y}(0), \tilde{X}(0)) = (\tilde{y}(0), \tilde{x}(0))$ . Let  $(y_0, x_0) := (\tilde{y}(0), \tilde{x}(0))$ . We will assume that  $0 < y_0 < 1$  and  $y_0 + x_0 = 1$  [the case  $y_0 = 0$  (resp.  $y_0 = 1$ ) has no interest since it corresponds to a P2P network with no file at any time (resp. where all peers have the file at time  $t = 0$ )].

### D. Phase Transitions

Adding both ODEs in (45) yields  $\dot{\tilde{y}}(t) + \dot{\tilde{x}}(t) = -\mu t + 1$ . Plugging this value back into (45) gives  $\dot{\tilde{x}}(t) = x_0(1 - \mu t)^\xi$  for  $0 \leq t < 1/\mu$  and, by continuity,  $\tilde{x}(t) = x_0(1 - \mu t)^\xi$  for  $0 \leq t \leq 1/\mu$  with  $\tilde{x}(1/\mu) = 0$ .

In order to approximate the fraction of peers that will never receive the file as  $N$  is large, one needs to find the first time  $\tau > 0$  where either  $\tilde{x}(\tau) = 0$  or  $\tilde{y}(\tau) = 0$ . This time  $\tau$  is easy to find, as shown below.

We already know that  $\tilde{x}(t) > 0$  for  $0 \leq t < 1/\mu$  and  $\tilde{x}(1/\mu) = 0$ , so that we only need to focus on the zeros of  $\tilde{y}(t)$  in  $[0, 1/\mu]$ . By writing  $\tilde{y}$  as  $\tilde{y}(t) = (1 - \mu t)(1 - x_0(1 - \mu t)^{\xi-1})$ , we conclude that the smallest zero of  $\tilde{y}$  in  $[0, 1/\mu]$  is  $(1 - x_0^{1/(1-\xi)})/\mu$  if  $\xi < 1$ , and is  $1/\mu$  if  $\xi \geq 1$ . Therefore,  $\tau = (1 - x_0^{1/(1-\xi)})/\mu > 0$  if  $\xi < 1$  and  $\tau = 1/\mu$  if  $\xi \geq 1$ . Introducing this value of  $\tau$  in  $\tilde{x}(t)$  yields  $\tilde{x}(\tau) = x_0^{1/(1-\xi)}$  if  $\xi < 1$  and  $\tilde{x}(\tau) = 0$  if  $\xi \geq 1$ . In other words, as  $N$  is large, all peers will get the file if  $\xi \geq 1$ , and a fraction  $x_0^{1/(1-\xi)}$  of them will not if  $\xi < 1$ . In other words, we observe a phase transition at  $\xi = 1$ : All peers will get the file if  $\xi \geq 1$ , and a fraction  $\xi \rightarrow x_0^{1/(1-\xi)}$  will not if  $\xi < 1$ .

Let us now turn to the maximum torrent size. As  $N$  is large, it will be approximated by the maximum of  $\tilde{y}$  over the interval  $[0, \tau]$ . A straightforward analysis of the mapping  $t \rightarrow \tilde{y}(t)$  in  $[0, \tau]$  shows the following.

- It is decreasing if  $\xi \leq 1$  or if  $\xi > 1$  and  $\xi x_0 \leq 1$ —these conditions can be merged into the single condition  $\xi \leq 1/x_0$ —so that its maximum,  $y^{\max}$ , is given by  $y^{\max} = y_0 = 1 - x_0$ ,

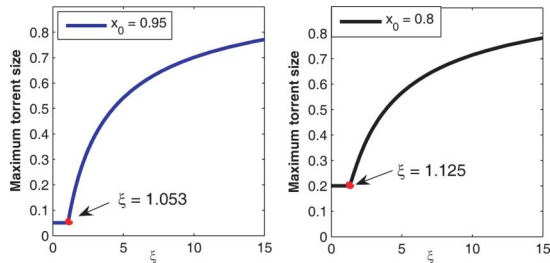


Fig. 2. Maximum torrent size over  $N$  (as  $N$  is large) as a function of  $\xi$  for  $x_0 = 0.95$  (left figure) and  $x_0 = 0.8$  (right figure).

- It is unimodal (first increasing then decreasing) if  $\xi > 1/x_0$ , with its maximum reached at  $t_1 := (1 - (\xi x_0)^{1-\xi})/\mu$  and given by

$$y^{\max} = (x_0)^{1/(1-\xi)} \xi^{\xi/(1-\xi)} (\xi - 1) > 0. \quad (46)$$

In summary, as  $N$  is large, the maximum torrent size is approximated by  $Ny^{\max}$  with  $y^{\max}$  given in (46) if  $\xi > 1/x_0$ , and  $y^{\max} = 1 - x_0$  if  $\xi \leq 1/x_0$ . This shows another phase transition (see Fig. 2) at  $\xi = 1/x_0$  (i.e., at  $\lambda x_0 = \mu$ ) in the sense that the torrent is maximum at  $t = 0$  if  $\xi \leq 1/x_0$  and is maximum at a later time if  $\xi > 1/x_0$ .

## VIII. NUMERICAL RESULTS

This section has two goals: to investigate the accuracy of the approximations developed in the previous sections (to be made more precise) and to study the impact of measures against nonauthorized uploading or downloading on the file availability in P2P swarming systems. Due to lack of space, we will not report any numerical result for the P2P model considered in Section VII; we will instead focus on the P2P model introduced in Section II-C and on its branching and mean-field approximations developed in Sections III and IV, respectively (Figs. 3–12), as well as on the optimization problem set in Section V (Fig. 13).

For each set of parameters, between 200 and 1000 discrete-event simulations of the Markov model in Section II-C have been run. In each figure (except in Figs. 7 and 8, where only simulation results are displayed, and in Fig. 13, where only mean-field results are shown), both simulation and approximation results are reported for the sake of comparison. Let  $r := (Y(0) + X_c(0))/N$  be the ratio of cooperative peers at time  $t = 0$ , and recall that  $N_c = X_c(0)$  (see Section II). The total number of peers,  $N$ , at time  $t = 0$  is equal to 400 in Figs. 3–8, 300 in Figs. 9 and 10, and 500 in Fig. 13.

This paper introduces two epidemic models that differ in their random contact behaviors. If a peer makes contacts with all other peers randomly instead of one other peer, it is called *epidemic model I* in brief. Otherwise, it is called *epidemic model II*. For the sake of clarity, we organize this section as follows. Sections VIII-A–VIII-C illustrate the extinction time and the file availability of *epidemic model I*. In Section VIII-D, we demonstrate the results of *epidemic model I* with two types of cooperative peers. Lastly, Section VIII-E shows the numerical results of *epidemic model II*.

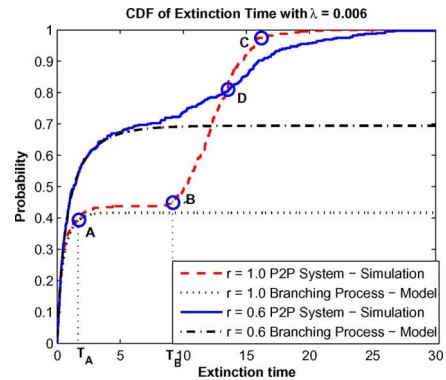


Fig. 3. CDF of extinction time for  $Y(0) = 1$ ,  $\lambda = 0.006$ ,  $\mu = 1$ .

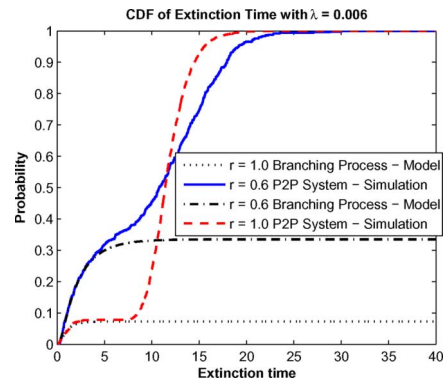
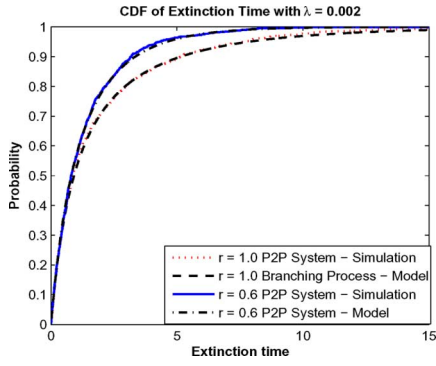
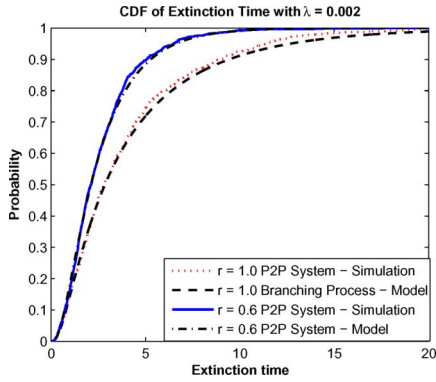
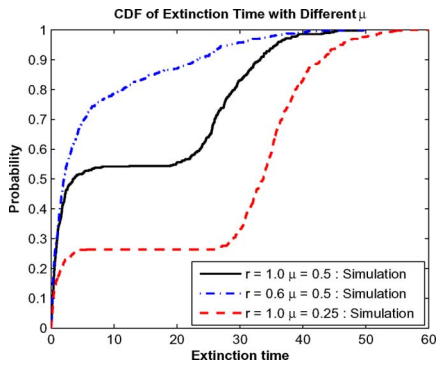
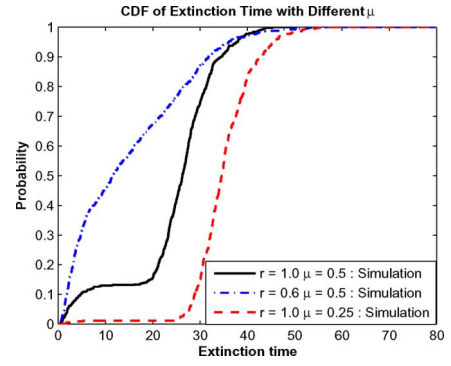
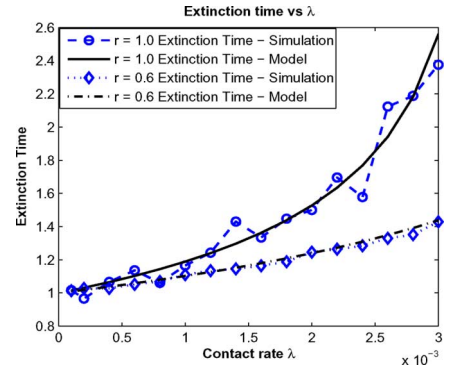
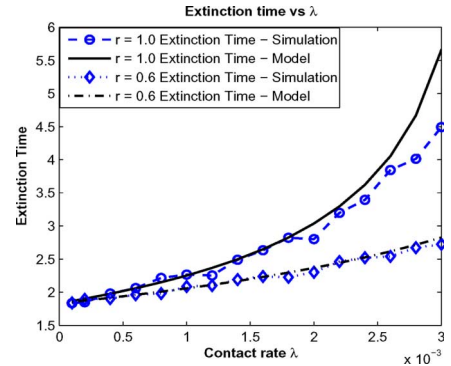


Fig. 4. CDF of extinction time for  $Y(0) = 3$ ,  $\lambda = 0.006$ ,  $\mu = 1$ .

### A. Epidemic Model I: File Extinction Time and the Branching Approximation

In this section, we focus on Figs. 3–10. Fig. 3 (resp. Fig. 4) compares the CDF of the extinction time obtained by simulation and by the branching approximation in (12) when  $Y(0) = 1$  (resp.  $Y(0) = 3$ ),  $\lambda = 6 \cdot 10^{-3}$ ,  $\mu = 1$ , and for two values of  $r$  ( $r = 0.6$  implies that  $N_c = 239$  and  $X_f(0) = 160$ ,  $r = 1$  implies that there are no free-riders ( $X_f(0) = 0$ ) and  $N_c = 399$ ). Note that  $\rho = \lambda N_c / \mu$  is close to 2.4 when  $r = 1$  and is close to 1.43 when  $r = 0.6$ . In all cases, the simulation and the branching approximation are in close agreement up to a certain time (time  $T_b$  in Fig. 3), which, interestingly, corresponds to the extinction time in the branching model. After this time, the extinction of the file in the Markov model increases sharply (the larger  $r$ , the larger the increase). In other words, the extinction of the file in the original Markov model has two modes, an *early* extinction mode and a *late* extinction mode. The former occurs when the file disappears before the dissemination has reached its peak value (i.e., most peers do not get the file), and the latter when most peers leave the network with the file. One may also check that the branching approximation provides an upper bound for the CDF of the extinction time, as predicted by Proposition 2. Lastly, we note that when there are less cooperative peers ( $r = 0.6$ ), the file lifetime is prolonged (see, e.g., point D in Fig. 3, where simulation curves for  $r = 1$  and  $r = 0.6$  cross each other); this can be explained by the fact that there are less contact opportunities between cooperative peers when  $r = 0.6$ . The main difference between Figs. 3 and 4 lies in the increase of the probability of the late extinction that is


 Fig. 5. CDF of extinction time for  $Y(0) = 1$ ,  $\lambda = 0.002$ ,  $\mu = 1$ .

 Fig. 6. CDF of extinction time for  $Y(0) = 3$ ,  $\lambda = 0.002$ ,  $\mu = 1$ .

 Fig. 7. CDF of extinction time for  $Y(0) = 1$  and different  $\mu$ .

 Fig. 8. CDF of extinction time for  $Y(0) = 3$  and different  $\mu$ .

 Fig. 9. Early extinction time as a function of  $\lambda$  with  $Y(0) = 1$ .

 Fig. 10. Early extinction time as a function of  $\lambda$  with  $Y(0) = 3$ .

steeper with three initial seeds ( $Y(0) = 3$ ) than with one initial seed ( $Y(0) = 1$ ).

Figs. 5 and 6 show the CDF of extinction time when  $\lambda = 0.002$ . In this experiment,  $\rho = 0.8$  if  $r = 1$ , and  $\rho = 0.48$  if  $r = 0.6$ . Thus, the approximated branching process is subcritical, resulting in a quick extinction of the file. The CDF of extinction time obtained from numerical examples is very close to that obtained from the branching process model. This is to say, when  $\rho < 1$ , all the peers with the file die out at the early stage, and hence the branching process provides a very accurate approximation of extinction probability.

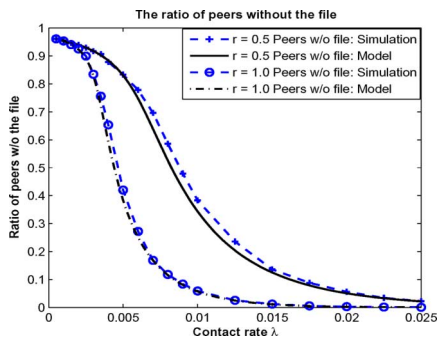
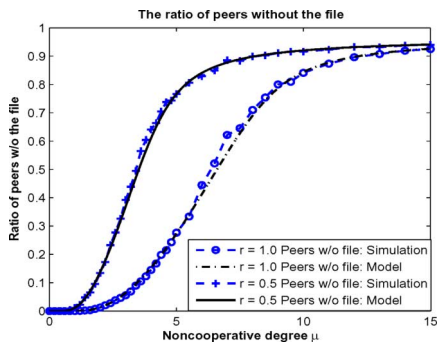
Simulation results in Figs. 7 and 8 exhibit the same *early-late* extinction pattern as in Figs. 3 and 4; they have been obtained for  $\lambda = 25 \cdot 10^{-4}$  and for two different values of  $\mu$ ,  $r$  and  $Y(0)$ .

Figs. 9 and 10 show the expected time to extinction as a function of the pairwise contact rate  $\lambda$ , in the case of an early

extinction (i.e., for small values of  $\lambda$ ), for  $\mu = 1$  and for two values of  $r$ . The curves "Model" display the mapping  $\lambda \rightarrow E[T_b(k)]$ , with  $E[T_b(k)]$  the expected extinction time in the branching process given  $Y(0) = k$  (see Section III). We observe an excellent match between the simulation and the branching approximation, thereby showing that the latter works well for early file extinction. Also note that having three seeds instead of one greatly extends the expected extinction time.

### B. Epidemic Model I: File Availability and the Mean-Field Approximation

We now look at the fraction of peers that will not acquire the file. We assume that  $Y(0) = 10$ , and we recall that  $N = 300$ . Fig. 11 (resp. Fig. 12) displays this fraction as a function of  $\lambda$  (resp.  $\mu$ ) for two different values of  $r$  ( $r = 1$  corresponding to  $X_c(0) = 290$  and  $r = 0.5$  corresponding to

Fig. 11. Fraction of peers without the file as a function of  $\lambda$ .Fig. 12. Fraction of peers without the file as a function of  $\mu$ .

$X_c(0) = 140$  and  $X_f(0) = 150$ ). In each figure, both simulation and mean-field approximation results are reported. The fraction of peers without the file is a decreasing function of the pairwise contact rate  $\lambda$  and an increasing function of the cooperation degree  $\mu$ . The mean-field approximation is obtained as the unique solution  $x_c(\infty)$  in  $(0, x_c(0))$  of (22) where the initial condition of the ODEs (18) is given by  $(y_0, x_c(0), x_f(0)) = (Y(0)/N, X_c(0)/N, X_f(0)/N)$ . In both figures, we observe a remarkable agreement between the simulation and the mean-field results [relative errors never exceed 2% when all peers are cooperative ( $r = 1$ ) and never exceed 7% when half of the peers are free-riders ( $r = 0.5$ )]. We also note that the fraction of peers without the file considered as a function of  $\lambda$  (resp.  $\mu$ ) is larger (resp. smaller) when  $r = 0.5$  than when  $r = 1$ ; this is of course not surprising since, unlike cooperative peers, free-riders do not contribute to the file dissemination.

### C. Epidemic Model I: Action Against Unauthorized File Downloading

We now evaluate the impact of actions against unauthorized file downloading. For that, we use the framework developed in Section V. Since the above simulations show that, for large  $N$ ,  $\bar{T}(\alpha)$  in Section V is a good approximation of the expected time,  $T_N$ , needed for an arbitrary peer to get the file, we only consider the utility function  $h(\alpha)$  [see (28)]. We assume that the cooperation degree  $\mu(\alpha)$  is given by  $\mu(\alpha) = \mu\alpha$ . There are 500 peers ( $N = 500$ ) at time  $t = 0$  including two persistent publishers ( $Y^* = 2$ ). We assume that  $Y(0) = 0$  so that  $X(0) = N - Y^* = 498$ . The initial condition of the ODE (23) is  $(y_0, x_0) = (0, X_N(0)/N)$  with  $y^* = 2/N$ . Fig. 13 displays the mapping  $\alpha \rightarrow h(\alpha)$  for two different values of  $\beta$  and  $\mu$ . We observe that a small investment cannot obviously postpone the expected delivery delay of the file, resulting in a decreased utility.

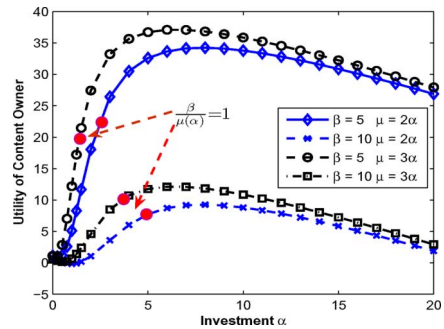


Fig. 13. Investment versus utility of content owner.

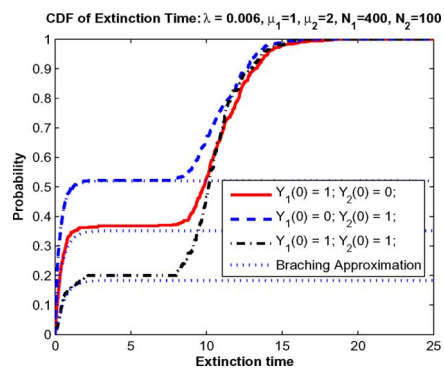


Fig. 14. Early extinction time with 80% class-1 and 20% class-2 peers.

As the investment grows, the utility of the content owner increases significantly. The curves in Fig. 13 also show how large an investment has to be to counteract P2P illegal downloading. Note the content owner can still have an increased utility when the ratio  $\beta/\mu(\alpha)$  is greater than one, as the utility is maximized across all curves when the ratio  $\beta/\mu(\alpha)$  lies between two and three.

### D. Epidemic Model I: Numerical Analysis of Two Types of Peers

In this section, we evaluate the transient behaviors of a P2P swarm where the peers have different degrees of cooperation. Fig. 14 compares the CDF of extinction obtained through multitype branching process model and the simulation. In this set of experiments, we consider a swarm of 500 peers. They are classified into two groups, class-1 with 400 peers and class-2 with 100 peers. The pairwise contact rate of the peers without the file is set to  $\lambda = 6 \cdot 10^{-3}$ . The class-1 and the class-2 peers differ in the departure rate  $\mu_1 = 1$  and  $\mu_2 = 2$  after obtaining the file. We perform three simulations: 1) one *infected* peer of class-1 at  $t = 0$ ; 2) one *infected* peer of class-2 at  $t = 0$ ; and 3) one *infected* peer of class-1 and one *infected* peer of class-2 at  $t = 0$ . In the branching approximation, the extinction probabilities of above cases are 0.35, 0.52, and 0.18, respectively. Fig. 14 shows that the branching model presents an upper bound of extinction time. The CDFs obtained from the branching approximation are very close to the numerical results at the early stage of file dissemination. Especially, for the case  $(Y_1(0) = 1, Y_2(0) = 1)$ , the extinction probability is well approximated by the product of the probabilities in the cases  $(Y_1(0) = 1, Y_2(0) = 0)$  and  $(Y_1(0) = 0, Y_2(0) = 1)$ .

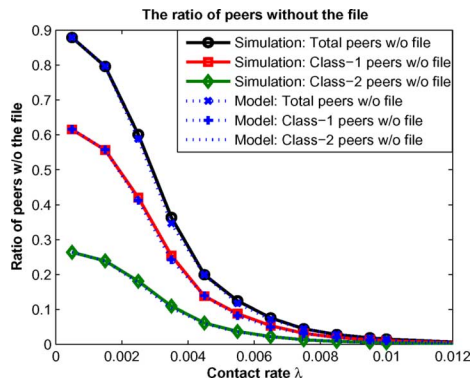


Fig. 15. Fraction of peers without the file as a function of  $\lambda$  (as time  $T$  is large enough).

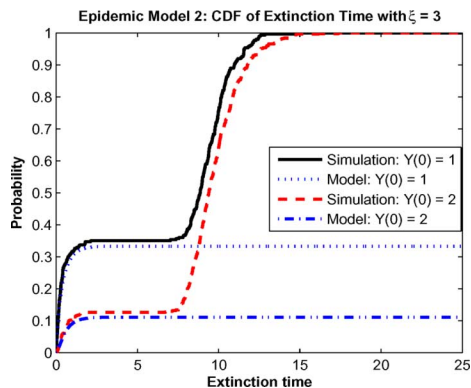


Fig. 16. CDF of extinction time of the model with fixed request rate per node:  $\lambda = 3$  and  $\mu = 1$ .

We evaluate the number of peers that do not receive the file after a long period  $T$  (e.g.,  $T \geq 50$  in the mean field model, while the file has disappeared in the experiments). The number of class-1 peers is 340, and that of class-2 peers is 160. In each class, 25 peers (or equivalently  $y_{1,0} = 0.05, y_{2,0} = 0.05, x_{1,0} = 0.63, x_{2,0} = 0.27$ ) obtain the file at time 0. The departure rates  $\mu_1$  and  $\mu_2$  are 1 and 2, respectively. The contact rate  $\lambda$  increases from  $5 \times 10^{-4}$  to  $1.2 \times 10^{-2}$ . Fig. 15 shows the ratio of all peers without the file and the ratios of peers in each class without the file. We observe that the mean-field ODEs can predict the file availability accurately. The percentage of peers without the file in class-1 is proportional to that in class-2. Fig. 15 also exhibits a similar phase transition of file availability, though without an analytical expression.

### E. Epidemic Model II: Extinction Time and Availability

We evaluate the transient behaviors of the epidemic model where a node contacts only one randomly selected peer in each time. There are 500 cooperative peers with the departure rate  $\mu = 1$  in this set of experiments. Fig. 16 compares the CDF of extinction time with  $Y(0) = 1$  and that with  $Y(0) = 2$  when  $\lambda = 3$ . Note that Fig. 16 exhibits similar transient behaviors as Fig. 3. The branching process model provides a good approximation at the early stage, and deviates from the numerical results as  $t$  becomes large. We next show the phase transition of file availability. Two scenarios are considered:  $x_0 = 0.6$  (i.e.,  $X(0) = 300$ ) and  $x_0 = 0.8$  (i.e.,  $X(0) = 400$ ). The departure

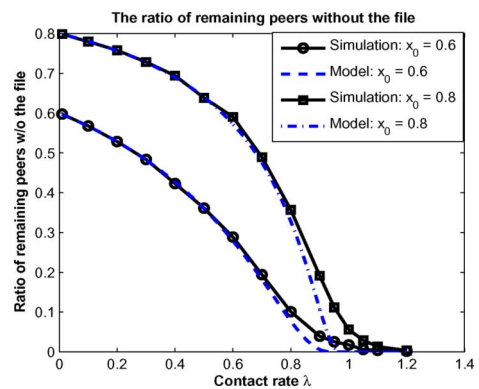


Fig. 17. Rate of peers without the file as a function of  $\lambda$  and  $x_0$ .

rate  $\mu$  is set to 1, and the contact rate increases from 0 to 1.2. Fig. 17 exhibits the phase transition of file availability in both experiments. In general, the mean-field approximation matches the numerical study well. However, when  $\xi = \lambda/\mu$  changes from 0.9 to 1.1, this model overestimates the file availability by up to 0.11 when  $x_0 = 0.8$ , and up to 0.025 when  $x_0 = 0.6$ .

## IX. RELATED WORK

There has been a number of works on the mathematical studies of structured and unstructured P2P-based content distribution. A seminal work can be found in [25]. The authors propose a continuous-time branching process to analyze service capacity (i.e., maximum rate of downloading) and a coarse-grain Markov model to characterize the steady state of downloading rate. In [22], Qiu and Srikant propose a fluid model composed of ordinary differential equations to describe the dynamics of BitTorrent systems. Authors in [27] further propose a novel fluid model based on stochastic differential equations. This new model also extends [22] to a multiclass system and is able to describe chunk availability. Munding *et al.* [29] propose a deterministic scheduling algorithm to achieve the optimal makespan for a structured system that requires global knowledge. A coupon model is put forward in [28] to investigate the effectiveness of a generic P2P file-sharing system. Authors in [30] present an improved model with tighter bounds subsequently.

Recently, the process of file dissemination has attracted a lot of attention. Clévenot *et al.* adopt a hybrid approach (fluid and stochastic) to analyze Squirrel, a P2P cooperative Web cache in [31]. In [13], Queija *et al.* study the scaling law of mean broadcasting time in a closed P2P swarm with constant request rate. Authors in [33] formulate a ball-and-urn model to characterize the “flash crowd” effect in a closed P2P networks. The content provided by P2P networks such as music, movies, and software are usually unauthorized. Content provided are therefore inclined to combat illegal downloading/uploading via technical solutions. Authors of [1] and [2] propose an M/G/ $\infty$  queueing model to access the efficiency of noncooperative measures against unauthorized downloading. Authors in [32] also model the P2P service as an M/G/ $\infty$  queue. They investigate the impact of file-bundling strategy on the efficiency of file downloading as well as the file availability. Our general model is inspired by the one in [13]. However, it differs from [2], [13], [22], and [25] in four ways.

1) We are studying the transient behavior.

- 2) A peer can initiate a number of random contacts, instead of one, with other peers.
- 3) We observe several phase transitions in response to system parameters.
- 4) We adopt Markov branching process and mean-field approaches to characterize the file dissemination model comprehensively.

## X. CONCLUSION

In this paper, we have proposed to use the theory of continuous-time 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. This is used for analyzing the impact of measures to decrease nonauthorized Internet access to copyrighted files. We identify 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 present two counteractions against unauthorized file sharing in the presence of illegal publishers. Our methodology can be extended to analyze file bundling that serves as a positive action of file dissemination.

## APPENDIX

### A. Proof of Proposition 1

*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)

$$\frac{dz(t)}{dt} \leq f(z(t)) \quad (47)$$

where  $z(t) := E[Y(t)]/N \in (0, 1]$ . According to the Gronwall's lemma, the solution of (47) is upper-bounded by the ODE when the equality holds. By using (4) and the definition of  $z(t)$ , we obtain  $E[Y(t)] \leq Ny_0/(y_0 + (1 - y_0)e^{-\beta t})$ . ■

### B. Proof of Proposition 2

*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  $\mathbf{V} = \{V(t), t \geq 0\}$ ,  $U_1(t), U_2(t), V(t) \in \mathbb{N}$ , by generating transitions as follows. Assume that a transition has occurred 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  $\tau$  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  $\mu U_1(t+)/r$ ,  $U_1(t+) = U_1(t) - 1$ , and  $U_2(t+) = U_2(t)$ ,  $V(t+) = V(t) - 1$ ; 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 [17] 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.  $\mathbf{Y}_b$ ). ■

## REFERENCES

- [1] S. Wong, E. Altman, and M. Ibrahim, "P2P networks: The interplay between legislation and information technology," in *Proc. NAEC Italy*, Riva del Garda, Italy, Oct. 8–11, 2009.
- [2] E. Altman, S. Wong, and J. Rojas-Mora, "P2P business and legal models for increasing accessibility to popular culture," in *Proc. Int. Conf. Digital Bus.*, London, U.K., 2009, vol. 21, Lecture Notes of the Institute for Computer Science, Social Informatics and Telecommunications Engineering, pp. 130–138.
- [3] E. Altman, S. Wong, J. Rojas-Mora, and A. Aram, "Cooperative solutions to unauthorized download of copyrighted creations," 2009.
- [4] S. Wong and E. Altman, "Restricting Internet access: Ideology and technology," in *Proc. 2nd COMSNETS*, Bangalore, India, Jan. 5–9, 2010, pp. 1–8.
- [5] S. Wong, E. Altman, and M. Ibrahim, "Internet access: Where law, economy, culture and technology meet," *Comput. Netw.*, vol. 55, no. 2, pp. 470–479, Feb. 2011.
- [6] A. Huygen, P. Rutten, S. Huvencers, S. Limonard, J. Poort, J. Leenheer, K. Janssen, N. van Eijk, and N. Helberger, "Ups and downs: Economic and cultural effects of LES sharing on music, LM, and games," TNO, SEO, IViR, Tech. Rep. 34782, 2009.
- [7] F. Oberholzer-Gee and K. Strumpf, "The effect of file sharing on record sales: An empirical analysis," *J. Polit. Econ.*, vol. 115, no. 1, pp. 1–42, 2007.
- [8] V. Hugo, "Congres litteraire international," in *Jan Baetens, Le Combat, Op. Cit.*, Jun. 25, 1878, p. 164.
- [9] M. Jayasuriya, J. Pearlman, R. Topolski, M. Weinberg, and S. Siy, "Forcing the net through a sieve: Why copyright filtering is not a viable solution for U.S. ISPs," Public Knowledge, White Paper, 2009 [Online]. Available: <http://www.publicknowledge.org/pdf/pk-filtering-whitepaper-200907.pdf>
- [10] 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," in *Proc. USENIX HotSec*, San Jose, CA, Jul. 29, 2008, pp. 90–97.
- [11] J. G. Steffensen, "Deux problèmes du calcul des probabilités," *Ann. l'Inst. Henri Poincaré*, vol. 3, no. 3, pp. 319–344, 1933.
- [12] D. Mollison, *Epidemic Models: Their Structure and Relation to Data*, D. Mollison, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1995.
- [13] R. Nunez-Queija and B. J. Prabh, "Scaling laws for file dissemination in P2P networks with random contacts," in *Proc. 16th IWQoS*, Enschede, The Netherlands, Jun. 2–4, 2008, pp. 75–79.
- [14] T. G. Kurtz, "Solutions of ordinary differential equations as limits of pure jump Markov processes," *J. Appl. Probab.*, vol. 7, no. 1, pp. 49–58, Apr. 1970.
- [15] J. K. Hale, *Ordinary Differential Equations*. Huntington, NY: Krieger, 1980, Pure and Appl. Mathematics XXI.
- [16] R. Groenevelt, P. Nain, and G. Koole, "The Message delay in mobile ad hoc networks," *Perform. Eval.*, vol. 62, no. 1–4, pp. 210–228, Oct. 2005.
- [17] T. Kamae, U. Krengel, and G. L. O'Brien, "Stochastic inequalities on partially ordered spaces," *Ann. Probab.*, vol. 5, no. 6, pp. 899–912, 1977.
- [18] T. E. Harris, *The Theory of Branching Processes*, 1963 ed. New York: Dover, 1989.
- [19] P. Narayan, "On the extinction time of a continuous time Markov branching process," *Australian J. Statist.*, vol. 24, no. 2, pp. 160–164, 1982.
- [20] D. Heinzmann, "Extinction times in multitype Markov branching processes," *J. Appl. Probab.*, vol. 46, no. 1, pp. 296–307, 2009.
- [21] A. Chaintreau, J.-Y. Le Boudec, and N. Ristanovic, "The age of gossip: Spatial mean field regime," in *Proc. ACM SIGMETRICS*, Seattle, WA, Jun. 15–19, 2009, pp. 109–120.
- [22] D. Qiu and R. Srikant, "Modeling and performance analysis of BitTorrent-like peer-to-peer networks," in *Proc. ACM SIGCOMM*, Portland, OR, Aug. 30–Sep. 3, 2004, pp. 367–378.

[23] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Comput. Netw.*, vol. 51, no. 10, pp. 2859–2891, 2007.

[24] R. Darling and J. Norris, "Differential equation approximations for Markov chains," *Probab. Surveys*, vol. 5, pp. 37–79, 2008.

[25] X. Yang and G. D. Veciana, "Service capacity of peer to peer networks," in *Proc. IEEE INFOCOM*, Hong Kong, China, Mar. 7–11, 2004, vol. 4, pp. 2242–2252.

[26] E. Altman, P. Nain, A. Shwartz, and Y. D. Xu, "Predicting the impact of measures against P2P networks on the transient behaviors," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 10–14, 2004, pp. 1440–1448.

[27] B. Fan, D. M. Chiu, and J. C. S. Lui, "On the performance and availability of bittorrent-like P2P file sharing systems," in *Proc. 14th IWQoS*, New Haven, CT, Jun. 19–21, 2006, pp. 30–39.

[28] L. Massoulié and M. Vojnović, "Coupon replication systems," in *Proc. ACM SIGMETRICS*, Banff, AB, Canada, Jun. 6–10, 2005, pp. 2–13.

[29] J. Munding, R. Weber, and G. Weiss, "Optimal scheduling of peer-to-peer file dissemination," *J. Sched.*, vol. 11, no. 2, pp. 105–120, 2008.

[30] M. H. Lin, B. Fan, J. C. S. Lui, and D. M. Chiu, "Stochastic analysis of file swarming systems," in *Proc. IFIP Perform.*, Oct. 2007, pp. 856–875.

[31] F. Clévenot and P. Nain, "A simple fluid model for the analysis of the squirrel peer-to-peer caching system," in *Proc. IEEE INFOCOM*, Hong Kong, China, Mar. 7–11, 2004, vol. 1, pp. 86–95.

[32] D. S. Menasche, A. A. Rocha, B. Li, and D. Towsley, "Content availability and bundling in swarming systems," in *Proc. ACM CoNEXT*, Rome, Italy, Dec. 1–4, 2009, pp. 121–132.

[33] F. Simatos, P. Robert, and F. Guillemin, "A queueing system for modeling a file sharing principle," in *Proc. ACM SIGMETRICS*, Annapolis, MD, Jun. 2–6, 2008, pp. 181–192.

[34] X. Zhou, S. Ioannidis, and L. Massoulié, "On the stability and optimality of universal swarms," in *Proc. ACM SIGMETRICS*, San Jose, CA, Jun. 2011, pp. 341–352.



**Eitan Altman** (M'93–SM'00–F'10) received the B.Sc. degree in electrical engineering, the B.A. degree in physics, and the Ph.D. degree in electrical engineering from the Technion—Israel Institute, Haifa, Israel, in 1984, 1984, and 1990, respectively, and the B.Mus. degree in music composition from Tel-Aviv University, Tel-Aviv, Israel, in 1990.

Since 1990, he has been a Researcher with the National Research Institute in Computer Science and Control (INRIA), Sophia-Antipolis, France. He has been on the Editorial Boards of several scientific

journals: *Wireless Networks*, *Computer Networks*, *Computer Communications*, *Journal of Discrete Event Dynamic Systems*, *SIAM Journal of Control and Optimisation*, *Stochastic Models*, and *Journal of Economy Dynamic and Control*. His areas of interest include networking, stochastic control, and game theory.

Dr. Altman received the Best Paper Award in the Networking 2006, IEEE GLOBECOM 2007, and IFIP Wireless Days 2009 conferences. He is a coauthor of two papers that have received the Best Student Paper awards (at QoFis 2000 and at Networking 2002).



**Philippe Nain** received the Ph.D. degree in applied mathematics from the University of Paris XI, Orsay, France, in 1981.

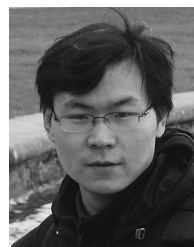
He is a Research Director with INRIA, Sophia Antipolis, France, where he has been the scientific leader of the MAESTRO research group devoted to the modeling of computer systems and telecommunications networks since 1998. He has held long-term visiting appointments with the University of Massachusetts, Amherst, from 1993 to 1994, the University of Maryland, College Park, in 1987, and North Carolina State University, Raleigh, in 1988. His research interests include modeling and performance evaluation of communication networks.

Dr. Nain is Editor-in-Chief of *Performance Evaluation*. He is past Associate Editor of the IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Performance Evaluation*, and *Operations Research Letters*. He was a Co-Program Chair of the ACM SIGMETRICS 2000 conference and the General Chair of the IFIP Performance 2005 conference. Since 1997, he has been a member of the IFIP Working Group 7.3 on Computer System Modelling (Vice-Chair since 2007).



**Adam Shwartz** (M'86–SM'89) received the B.Sc. degrees in physics and electrical engineering from Ben Gurion University, Beersheba, Israel, in 1979, and the M.A. degree in applied mathematics and Ph.D. degree in engineering from Brown University, Providence, RI, in 1982 and 1983, respectively.

Since 1984, he has been with the Technion—Israel Institute of Technology, Haifa, Israel, where he is a Professor of electrical engineering and holds the Julius M. and Bernice Naiman Chair in Engineering. He held visiting positions with the Institute for Systems Research (ISR), University of Maryland, College Park; the Mathematical Research Center, Bell Labs, Murray Hill, NJ; Rutgers Business School, New Brunswick, NJ; CNRS, Montreal, QC, Canada; Free University, Amsterdam, The Netherlands; and shorter visiting positions elsewhere. His research interests include probabilistic models—in particular, the theory of Markov decision processes, stochastic games, and asymptotic methods including diffusion limits and the theory of large deviations—and their applications to computer communications networks.



**Yuedong Xu** received the B.S. degree from Anhui University, Hefei, China, the M.S. degree from Huazhong University of Science and Engineering, Wuhan, China, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong.

Since late 2009, he has been with INRIA Sophia Antipolis, Sophia Antipolis, and Université d'Avignon, Avignon, France, as a Postdoc. His areas of interest include performance evaluation, control, optimization, and economic analysis of communication

networks.