Predicting the Impact of Measures Against Unauthorized Downloads on the Transient Behaviors of P2P Networks

Extended Abstracts

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

1 INRIA Sophia Antipolis, 2004 Route des Lucioles, France
2 Technion - Israel Institute of Technology, Haifa 32000, Israel
{eitan.altman, philippe.nain, yuedong.xu}@sophia.inria.fr; adam@ee.technion.ac.il

Abstract: This work has two objectives. The first is to study rigorously the transient behavior of some P2P networks where information is replicated and disseminated according to an epidemic type dynamics. The second is to use the insight gained in order to predict how efficient are the measures taken against peer to peer networks. Two abstract models are considered to describe the P2P file sharing based on random contacts. Phase transitions are exhibited in both cases.

Keywords: Epidemics, Branching Process, Mean Field Approximation, Phase Transition

1. Introduction

Along with the worldwide penetration of P2P application, a huge demand has appeared to copyrighted music and books that have been accessible for free over the Internet. While benefiting a very large internaut community, this unregulated access is not in the interests of content owners [4].

In this extended abstract we are interested in predicting the impact of measures against P2P file sharing, 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 formulate epidemic-like abstract models of a torrent in simplified P2P networks, where a large number of peers are interested in a file which is initially available at a small fraction of the population. We briefly state our contributions:

1. Modeling and approximating the transient behavior. We introduce a two-dimensional Markov process describing the system dynamics and use it to approximate the transient behavior. The first approximation yields a Markov branching process; the second approximation, of a mean-field type, holds when the number of peers goes to infinity and the initial state of the system scales linearly with the number of peers.

2. Analysis and identifying phase transitions. For both approximation models we show the existence of phase transitions: a small change in some parameters may cause a large change in the system behavior.

3. Application. We investigate two counteractive measures against unauthorized file sharing in the presence of illegal publishers.

2. Model

Assume there is a population N of mobile peers interested in a single file. We will consider two types of peers: *cooperative* peers and *free-riding* peers. Once a cooperative peer has acquired the file, it stays in the network for a random time distributed according to an exponential rv with parameter $1/\mu \geq 0$ and then leaves the network. During the lingering time of a cooperative peer with the file, it participates in the file dissemination. On the other hand, a free-rider leaves the network at once when it receives the file.

We model the system by a finite-state three dimensional Markov process $\mathbf{Y} = \{(Y(t), X_c(t), X_f(t))\}_t$, with $Y(t)$ denoting the number of *publishers* (i.e. the number of peers with the file) at time t, and $X_c(t)$ and $X_f(t)$ denoting the number of cooperative peers without the file and the number of free-riders (necessarily without the file) at time t, respectively. We assume that $Y(0) + X_c(0) + X_f(0) = N$. Let $N_c := X_c(0)$, $N_f := X_f(0)$ and $\rho := \lambda N_c/\mu$.

We consider an abstract P2P network in which file acquisition is done via random contact between peers: when a publisher meets a peer without the file the former transmits the file to the latter. We assume that the transmission is always successful and that the transmission time is negligible with respect to the peer intermeeting time and is taken to be zero. Successive contact times between any pair of peers is supposed to form a Poisson process with rate $\lambda > 0$[3]. All processes and rvs introduced so far are supposed to be mutually independent. Non-zero transition rates of $\mathbf{Y}$ are given by (with $e_c = (1,0)$ and $e_f = (0,1)$)

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

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \to \begin{pmatrix} Y(t)-1 \\ X(t) \end{pmatrix} \text{ with rate } \mu Y(t), \quad (2)$$

$$\begin{pmatrix} Y(t) \\ X(t) \end{pmatrix} \to \begin{pmatrix} Y(t) \\ X(t)-e_f \end{pmatrix} \text{ with rate } \lambda Y(t) X_f(t). \quad (3)$$

Note that all states of the form $(0,\cdot,\cdot)$ are absorbing states since the file has disappeared.

We now develop two approximations of the process $\mathbf{Y}$. The first one will consist in replacing $X_c(t)$ by $N_c = X_c(0)$ in the transition rate (1), which will introduce a Markov branching process. Clearly, this (so-called) branching approximation will loose its accuracy as the ratio $X_c(t)/N_c$ changes. The second approximation will use an asymptotic argument as $N \to \infty$ based on a mean-field approximation of the process $\mathbf{Y}$. Such an approximation is accurate only if $Y(0)$ is of the order of N. Both approaches will allow us to approximate some key quantitative characteristics of the process $\mathbf{Y}$.

3. Branching approximation

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

$$Y_b(t) \to Y_b(t)+1 \qquad \text{with rate } \lambda Y_b(t) N_c \qquad (4)$$
$$Y_b(t) \to Y_b(t)-1 \qquad \text{with rate } \mu Y_b(t) \qquad (5)$$

where we recall that N_c is the number of cooperative peers without the file at time $t=0$. Using classical coupling arguments the following result can be proved:

Proposition 1 *If $Y(0) \le Y_b(0)$ then $Y(t) \le_{st} Y_b(t)$ for any $t>0$ ("st" stands for stochastic ordering).*

The linear state transition rates along with the independence of each member of $Y_b(t)$ imply that $\mathbf{Y}_b$ is a *Markov branching process* [1]. Conditioned on $Y_b(0)=k\ge 1$, the extinction probability is $q_k = (\min\{1,1/\rho\})^k = \min\{1,(1/\rho)^k\}$ [1]. The extinction will be certain iff $\rho \le 1$ or equivalently iff $\lambda N_c \le \mu$. The CDF $T_b(t) = P(T_b < t)$ of the extinction time $T_b := \min\{t>0 : Y_b(t)\}$ can be obtained from [1]. When $Y_b(0)=1$ we find $T_b(t) = e^{\mu(1-\rho)t} - 1/e^{\mu(1-\rho)t} - \rho$.

When $\rho < 1$ and $Y_b(0)=1$ the expected extinction time (denoted as T_b^1) is finite and given by

$$E[T_b^1] = -1/(\lambda N_c)\cdot \log(1-\rho). \qquad (6)$$

4. Mean-field approximation

The behavior of the process $\mathbf{Y}$ can be well approximated by a deterministic limit solution of ODEs, an approach known as mean-field approximation. We may evoke [2] to obtain that if $\lambda = \beta/N$ and if $\lim_N Y(0)/N = y_0 > 0$, $\lim_N X_c(0)/N = x_{c,0}$ and $\lim_N X_f(0)/N = x_{f,0}$ then the rescaled process $N^{-1}\mathbf{Y}$ converges in probability as $N\to\infty$, uniformly on all finite intervals $[0,T]$, to the solution of the system of ODEs

$$\frac{d}{dt}\begin{pmatrix} y \\ x_c \\ x_f \end{pmatrix} = \begin{pmatrix} y(\beta x_c - \mu) \\ -\beta y x_c \\ -\beta y x_f \end{pmatrix} \qquad (7)$$

with the initial conditions $(y(0), x_c(0), x_f(0)) = (y_0, x_{c,0}, x_{f,0})$. The solutions $y(t)$, $x_c(t)$, $x_f(t)$ of (7) provide an approximation of the fraction of available seeds at time t, of the fraction of cooperative peers without the file at time t and of the fraction of free-riders at time t, respectively. The accuracy of this approximation will increase with N, the total number of peers.

The first question we wish to ask is whether all (or almost all) peers interested in the file are able to obtain it or not. If the answer is no, then we shall be interested in computing the fraction of those that will never receive the file. Introduce $\theta := \frac{\beta}{\mu}$. The ODEs in (7) give $x_c + y = \frac{1}{\theta}\ln x_c + \phi(\theta)$ where $\phi(\theta) := x_{c,0} + y_0 - \frac{1}{\theta}\ln x_{c,0}$. Let y^{max} be the maximum ratio of cooperative peers with the file. According to the first equation of (7), y^{max} is reached when $x_c = \mu/\beta$ if $\beta > \mu$. That is, y^{max} is expressed as $y^{max} = -1/\theta(1+\ln\theta) + \phi(\theta)$. As t grows to infinity, y is approaching 0. Therefore, the ratio of cooperative peers that do not have the file is given by

$$x_c(\infty) - \theta^{-1}\ln(x_c(\infty)) - \phi(\theta) = 0. \qquad (8)$$

Similarly, we can find the ratio of free riders without the file by $x_f(t) = (x_{f,0}/x_{c,0})x_c(t)$.

We are interested in whether there is an abrupt change in content availability (i.e. $x_c(\infty)$) as the parameter θ varies. To exhibit a phase transition, we approximate $\log(x_c(\infty))$ in (8) by using its Taylor extension at $x_{c,0}$ and obtain $x_c(\infty) \approx ((\frac{1}{\theta} - x_{c,0} - y_0) + \frac{1}{2\theta}(\frac{x_c(\infty)}{x_{c,0}} - 1)^2)/(\frac{1}{\theta x_{c,0}} - 1)$. Since the numerator is bounded, a phase transition happens to occur at $\theta = 1/x_{c,0}$.

5. The impact of measures against P2P networks

Hereafter we present numerical simulations to validate the "phase transitions" as well as the behavior of file extinction. Let $r := (Y(0)+X_c(0))/N$ be the ratio of cooperative peers at time $t=0$.

In Figure 5 we have set $Y(0)=1$, $\lambda = 6\times 10^{-3}$, $\mu=1$ and $N=400$. We display the CDF $T(x) = P(T<x)$ of the extinction time T obtained by simulation and from the "branching formula" respectively, for the cases when $r=1$ (no free rider) and $r=0.6$ (60% of the peers are cooperative). Consider the case when $r=1$ (bottom curves). There is an excellent match between the simulation and the branching approximation until $t=T_B$ after which the branching model is no longer accurate. Note the existence of a "plateau" between times $t=T_A$ and $t=T_B$ which can be interpreted as the existence of two types of extinction, the early one until time T_A and the late one after time T_B. Same comments hold for the case $r=0.6$ except that the branching model looses its accuracy earlier which is due to the fact that when $r=0.6$, N_c is smaller (equals to 239) than in the case $r=1$ ($N_c=399$).

In Figure 5 we investigate the impact of contact rate on the file availability under the mean field model. We assume that $\mu=1$, $N=300$ and $Y(0)=10$. This figure displays the ratio of peers without the file (i.e. $x(\infty)$) as a function of the contact rate λ when $r=1$ (lower curves) and $r=1/2$ (upper curves). For both values of r the match between the simulation results and the mean-field approximation is excellent across all values of λ. Note that Figure 5 also exhibits the existence of phase transitions.

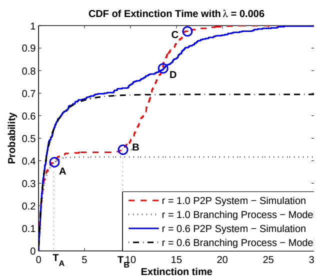

Figure 1: $Y(0)=1$ and $\lambda = 0.006$

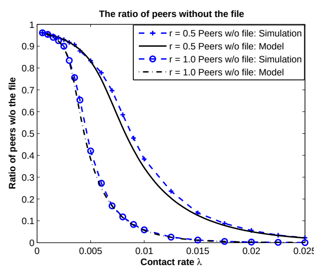

Figure 2: $x_c(\infty)$ versus λ

References

[1] T. E. Harris. *The Theory of Branching Processes.* Dover Publisher, 1963.

[2] T. G. Kurtz. *Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes.* Journal of Applied Probability, 1970.

[3] P. Nain R. Groenevelt and G. Koole. *Message Delay in Mobile Ad Hoc Networks.* Performance Evaluation, 2005.

[4] S. Wong and E. Altman. *Restricting Internet Access: Ideology and Technology.* Computer Networks, 2010.