Content Dynamics in P2P Networks from Queueing and Fluid Perspectives.

Andres Ferragut and Fernando Paganini
Universidad ORT Uruguay, Montevideo, Uruguay

***Abstract*—In this paper we analyze the dynamics of P2P file exchange networks, considering both queueing and fluid models. In such systems, the service rate depends on one mostly fixed component (servers or seeders), and another that scales with the number of peers present. We analyze a class of M/G Processor Sharing queues that describe populations and residual workloads in this situation, characterizing its stationary regime. It is shown that, under a law of large numbers scaling, the system behaves as a $M/G/1$ or a shifted $M/G/\infty$ queue, depending on whether the server or peer contribution becomes dominant. We also consider fluid models for populations and residual workloads in the form of a partial differential equation, and establish connections with the queueing approach. This method provides broadly applicable results on stability, variability and transient performance, which we validate against packet simulations, showing improvement with respect to earlier models.**

I. Introduction

In a peer-to-peer (P2P) file-sharing system, *seeders* wishing to disseminate some content can leverage the upload capacity of the swarm of *leechers* that are downloading. This fundamental difference with the traditional client-server architecture is essential for its scalability. From a queueing perspective, while client-server systems have a fixed capacity shared among clients, P2P systems superimpose an infinite server component whose capacity grows with queue occupation.

Several attempts have been made at modeling population evolution of P2P systems. A first queueing model was proposed in [1], under memoryless distribution assumptions for peer arrival, completion and departure times. The ensuing Markov chain is not reversible and hence limited analytical results are available for it. Another limitation of this model is that independent exponential times do not capture the fact that the P2P swarm is often sharing a common content.

An ordinary differential equation (ODE) model for the populations was subsequently proposed by [2], in essence a fluid version of the dynamics in [1]. This model also adds a download capacity limit, and shows there are two cases for the resulting equilibrium, depending on whether the download or upload capacity is acting as bottleneck. Global asymptotic stability results for these equilibria were established in [3]. The model however does not capture general file sizes.

Some other relevant related works are the following: in [4] a fluid model is considered, where heterogeneous clients are allowed. Also, [5]–[7] analyze the evolution of content in detail, considering a model where the exact chunks each peer possess is accounted for in the state. These models become harder to analyze due to the size of the state space. In [8] a simple model for transient analysis is given, based on a fixed number of peers arriving at random times. With similar ideas, download progress is considered in [9], [10]. A related approach using fluid and diffusion approximations is [11].

This paper has three contributions. First in Section II we reconsider queueing models of P2P under the simplifying assumption of a fixed number of seeders, a restrictive but arguably practical scenario. This queue can be solved exactly, and the stationary population law is insensitive to the workload distribution, thus giving results of general applicability. We perform a scaling limit of this process for large numbers of peers and find two asymptotic regimes, depending on whether the seeders can sustain the load on their own, or on the contrary the P2P contribution is essential. This special case thus provides valuable insights on P2P behavior within a classical queueing perspective.

Moving beyond this case, in particular dealing with a varying number of seeders, has motivated the use of fluid models; in our recent work [12], [13] we proposed a Partial Differential Equation (PDE) that describes the evolution of populations and residual content, and applied control theoretic methods to study these dynamics, extending the ODE method to general file size distributions. How do these models relate to their M/G queueing counterparts? A partial answer to this question is the second contribution of this paper, presented in Section III. We show that the mean value of an appropriately chosen stochastic process satisfies the corresponding PDE, in the case of fixed service rates. We further show that for the fixed seeder case, both approaches give consistent results for the steady-state download advance profiles, and the steady-state variance of the population.

The third objective of this paper is to validate the fluid approach beyond its connection to queueing models. In Section IV we perform direct comparisons with packet-level simulations of BitTorrent [14], which include low-level protocol details such as chunk availability and exchange rules. Despite the fact that these details are not represented in the PDE model, we show that the large scale behavior of the system is captured with high fidelity, outperforming the previous ODE models. Conclusions and lines of future work are discussed in Section V, and an Appendix contains some of the proofs.

This work was supported by AFOSR-US under grant FA9550-09-1-0504. The authors would also like to thank Fabian Kozynski for developing the simulation code. E-mail:ferragut@ort.edu.uy.

II. Queueing analysis of a class of P2P systems under general workload sizes

In a P2P system, *content* is disseminated by subdividing it into small *chunks*, and enabling peers to exchange such units bidirectionally. Thus every peer present is a server; those who are also clients are referred to as *leechers*, whereas *seeders* are those peers present in the system only to altruistically distribute content. There could be a common file of interest to all peers (e.g. a swarm around a single torrent), in which case the download is of fixed size. More generally we could think of "content" as a larger entity (e.g. multiple torrents) and different peers could have interest in only a portion, which introduces variability in the job size required by each leecher.

To analyze the dynamics of the population of seeders and leechers, [1] proposed the following queueing model: leechers arrive as a Poisson process of intensity λ, $x(t)$ denoting the amount of leechers in the system at time t. $y(t)$ is the number of seeders. Leechers turn into seeders at an exponential rate $\mu(\eta x+y)$, where η is an efficiency parameter, and μ represents the upload rate of a single peer, in files per second. Seeders stay for an $\exp(\gamma)$ time. Despite the simplicity of this model, the resulting Markov chain is not reversible, hence it is not easy to obtain analytic results for it; [1] studies it numerically.

Another limitation of this model lies in assuming independent exponential times for download completion, which does not seem natural since workloads involve a common content. For instance a determinitic job size queue would be more appropriate for swarms downloading a single file.

It turns out, as we will now show, that both limitations of the model disappear if we consider a simple variant: namely, that the population of seeders remains fixed at y_0, and leechers leave the system upon completion. This assumption is admittedly restrictive, but note that leechers are often selfish and do not remain after completion. For these cases, disseminating a file requires the "generous" contribution of a set of unconditional seeders. Also, we shall see that the fixed seeder case provides substantial insights. It is also a first step toward analyzing slowly varying seeder populations.

We make two final assumptions before introducing our queueing model. Both are supported by our packet simulation studies in the case of BitTorrent [15].

1) $\eta = 1$, i.e. the file-sharing is efficient. The entire upload bandwidth $R_{up} := \mu(x + y_0)$ is used for the service of leechers present; if a peer has spare bandwidth someone will find something to download from it. This is a natural situation if the number of chunks is not exceedingly small, as suggested by the analysis in [2].
2) The upload bandwidth is equally shared, so each leecher receives a service rate of $r = R_{up}/x$. This fact is not obvious, it depends on choices of exchange peers, themselves influenced by incentives such as tit-for-tat rules [14]; nevertheless, we have validated it empirically.

We are now ready to define our system.

Definition 1 (P2P queueing system with fixed seeders): The queueing system is defined by: a Poisson process of client arrivals, with intensity λ; each client requiring an independent service of size $\sigma > 0$ with complementary distribution function (CCDF) $H(\sigma)$ of (normalized) mean 1; and a state-dependent processor sharing discipline that, for an occupation state x, services each client at rate $r = \mu\frac{x+y_0}{x}$.

A. Queueing analysis for fixed seeders

If job sizes are exponentially distributed, $H(\sigma) = e^{-\sigma}$, the leecher queue $x(t)$ will behave as a Markov chain with state space $\mathbb{N}$ and the following transition rates q_{ij}:

$$q_{x,x+1} = \lambda, \quad q_{x,x-1} = \mu(x + y_0),\ x > 0. \tag{1}$$

The above is a specialization of the model in [1] to the case of fixed seeders. The advantage here is that this chain can be solved explicitly, using basic birth-death process results:

Proposition 1: If we denote $\rho = \lambda/\mu$, the equilibrium distribution for the number of leechers in the birth-death process (1) is:

$$\pi(n) = \left[\sum_{m=y_0}^{\infty} \frac{\rho^m}{m!}\right]^{-1} \frac{\rho^{n+y_0}}{(n+y_0)!} \quad \text{for } n \geqslant 0. \tag{2}$$

In particular, the system is stable for any λ, μ and y_0.

Moreover, by the insensitivity of PS queues (c.f. [16] and references therein), the invariant distribution of queue occupation will be independent of the job size distribution, thus holding also for the system of Definition 1. Note that the system is a combination of the $M/G/1$ and $M/G/\infty$ queues. If we disregard the contribution of leechers, it reduces to an $M/G/1$ PS system with load $\lambda/(\mu y_0)$, and would only be stable if $\rho < y_0$, which is natural since only the seeders must cope with the load. If instead we disregard the contribution of the seeders, the system becomes an $M/G/\infty$ queue, and the system is stable for all ρ[1]. Note that in the case $\rho > y_0$ the leecher contribution is essential to maintain stability.

In order to obtain some performance metrics, it is useful to compute the *probability generating function* (pgf) of π. Recall that if $X \sim \pi$, the pgf is given by $G(z) = E[z^X]$. By direct calculation, the pgf of (2) is:

$$G(z) = z^{-y_0} \frac{\sum_{m=y_0}^{\infty} (\rho z)^m/m!}{\sum_{m=y_0}^{\infty} \rho^m/m!}. \tag{3}$$

Equation (3) enables us to analyze the performance of the system, in particular, the average number of leechers and the average download completion time, which is the key performance metric:

Proposition 2: For $y_0 > 0$, the average number of leechers in the system is given by:

$$\bar{x} = \rho - y_0 + \frac{\rho^{y_0-1}/(y_0-1)!}{\sum_{m=y_0}^{\infty} \rho^m/m!}. \tag{4}$$

The proof of the above follows by direct evaluation of $G'(1)$. Through Little's law, we can deduce from here the average download time $T = \frac{\bar{x}}{\lambda}$. Other performance measures such as $Var(x)$ can be obtained with similar methods.

[1]Of course, this is an extreme situation where the model is not accurate, due to possibly missing chunks, but it serves as a limiting case.

B. Measure-valued state

A complete description of system evolution requires taking into account the remaining services of current jobs. Nevertheless, for the steady-state distribution, the distribution of residual work has been further characterized in [16], as follows. Introduce *residual lifetime* distribution $\bar{H}$ associated to H, which has CCDF given by:

$$\bar{H}(\sigma) = \int_{\sigma}^{\infty} H(s)\, ds. \tag{5}$$

Then [16, Thm. 1] states that the invariant distribution of a $M/G-PS$ system is obtained by choosing the number of jobs x following π, the solution of the balance equations, and given that $x = n$, choosing n copies of the remaining workloads as iid replications of the distribution $\bar{H}$.

The above result describes the stationary law through a two-stage description; to proceed further, and to cover also the transient process, we would like to have a state descriptor that captures both population and residual work. A natural choice in such PS systems (c.f. [17]) is a random measure on $\mathbb{R}^+$, which stores the remaining services. Specifically, if at a given time the number of customers is $x(t)$ and each one of them has a remaining service $\sigma_i(t)$, then the state of the system is

$$\Phi_t = \sum_{i=1}^{x(t)} \delta_{\sigma_i(t)},$$

where δ_σ is the Dirac measure concentrated on σ. This gives a unified description for the state regardless of the number of jobs present, and performance metrics of the system can be recovered by integration.

For further analysis of such random measures, a convenient characterization is the *Laplace functional* [18], defined by

$$\mathcal{L}_\Phi[f] = E\left[e^{-\int_0^\infty f(\sigma)\Phi(d\sigma)}\right]$$

for any $f \geqslant 0$ and bounded on $\mathbb{R}^+$. We now apply it to the invariant distribution of [16, Thm. 1].

Proposition 3: The stationary distribution of the measure valued process Φ_t is that of a random measure in $\mathbb{R}^+$ with Laplace functional:

$$\mathcal{L}_\Phi[f] = G\left(\int_0^\infty e^{-f(\sigma)}\, \bar{H}(d\sigma)\right) \tag{6}$$

for any bounded $f \geqslant 0$, where $G(\cdot)$ is the pgf of π.
The proof follows by conditioning on $x = n$ and noting that:

$$E[e^{-\int_0^\infty f(\sigma)\Phi(d\sigma)} \mid x = n] = \left(\int_0^\infty e^{-f(\sigma)}\, \bar{H}(d\sigma)\right)^n.$$

We now apply this result together with expression (3) to analyze the asymptotic behavior of the system in Definition 1.

C. Asymptotic analysis

Typical file sharing systems have a large number of peers. We would like to use the scale of the system in order to simplify (2) through the study of its asymptotic behavior, as the size of the system grows.

For this purpose, consider a family of systems as in Definition 1 with arrival rate $L\lambda$, where $L > 0$ is a scaling parameter, and we will let $L \to \infty$. With this scaling, the load grows as $L\rho$. As the number of arriving leechers scales, we also enlarge the number of fixed seeders as Ly_0. Let π_L and G_L denote the invariant distribution and its pgf for the scaled system. We distinguish two cases, depending on whether the seeders alone can cope with the demand or not.

1) Seeder sustained case ($\rho < y_0$): Whenever $\rho < y_0$ the seeders can cope with the demand. In the limit case when $L \to \infty$, the system behaves as an $M/G/1-PS$ queue. Formally, we have the following theorem, proved in the Appendix:

Theorem 1: If $\rho < y_0$, the equilibrium distribution of the scaled system converges in law to the equilibrium distribution of an $M/G/1-PS$ queue with load $\nu = \frac{\rho}{y_0} < 1$.
In particular we conclude:

Corollary 1: When $\rho < y_0$, as $L \to \infty$, the average number of leechers $\bar{x}$ in the system is given by:

$$\bar{x} = \frac{\rho/y_0}{1-\rho/y_0} = \frac{\rho}{y_0 - \rho}.$$

Of course, this is a limit value as $L \to \infty$. As for the average download time, since the arrival rate is $L\lambda$ then $\bar{T} \approx \frac{1}{L\mu}\frac{1}{y_0-\rho} \to 0$ as $L \to \infty$, which is consistent with the fact that the number of servers is scaling with L.

2) Globally sustained case ($\rho > y_0$): We now focus on the more interesting case $\rho > y_0$, where the contribution of leechers becomes crucial. If we scale the system as before, the average number of peers present (4), will also grow with L. We will employ a law of large numbers type of scaling to obtain a non-trivial limit. Consider then the family of processes:

$$\tilde{\Phi}^L = \frac{1}{L}\sum_{i=1}^{x^L} \delta_{\sigma_i^L},$$

where as before x^L is the number of jobs present in the system with arrival rate $L\lambda$ and Ly_0 seeders, and σ_i^L are their remaining workloads. The factor $1/L$ normalizes the total remaining workload leading to the following result, proved in the Appendix:

Theorem 2: If $\rho > y_0$, the equilibrium distribution of the scaled system $\tilde{\Phi}^L$ converges in law to the deterministic measure $(\rho - y_0)\bar{H}(d\sigma)$ on $\mathbb{R}^+$, where $\bar{H}(d\sigma)$ is the measure with CCDF defined in (5).

We now give an intuitive explanation of Theorem 2. Recalling (2), the invariant distribution for the number of jobs in the scaled system can be rewritten as:

$$\pi_L(n) = P(Y = Ly_0 + n \mid Y \geqslant Ly_0),$$

where $Y \sim \text{Poisson}(L\rho)$. If $\rho > y_0$, for large L, $P(Y \geqslant Ly_0) \approx 1$ and therefore x behaves as a shifted Poisson random variable of average $L(\rho - y_0)$ and their remaining workloads are iid distributed according to $\bar{H}(d\sigma)$. As the average number of jobs grows large with L, the rescaled process $\tilde{\Phi}^L$ is then an empirical estimator of the distribution $\bar{H}(d\sigma)$, scaled by $\rho - y_0$. On average, the process will behave as an $M/G/\infty$ queue with

load ρ, shifted y_0 units to the left. The seeders contribute to lower the average number of peers in the system, thus also contributing to the average download time. More formally:

Corollary 2: For $\rho > y_0$, as $L \to \infty$, the average number of leechers and download time in the scaled system verify:

$$\bar{x}^L = L(\rho - y_0) + o(L); \quad \bar{T}^L = \frac{1}{\mu}\left(1 - \frac{y_0}{\rho}\right) + o(1).$$

Note, however, that the *variance* of the number of leechers is not changed by the shift. By analogous calculations we have that, as $L \to \infty$, $Var(x^L) = L\rho + o(L)$.

The above result holds for general file sizes, however it is worth specializing it to the case of all peers downloading the same content, in which case $H(\sigma) = \mathbf{1}_{[0,1)}(\sigma)$. The residual job size CCDF is then $\bar{H}(\sigma) = 1 - \sigma$, $\sigma \in [0,1]$, i.e. the residual job sizes are uniformly distributed in $[0,1]$. The rescaled process then converges to a measure with total mass $(\rho - y_0)$ uniformly distributed in $[0,1]$. This uniform advance profile is a consequence of the processor sharing model and the fact that jobs are deterministic in size, and would not hold on exponentially distributed file sizes. We shall also recover this profile from the fluid analysis of the following section.

III. Fluid models and connections to queues

We have obtained closed-form results in the previous section with M/G queue models for a very special case; if instead we allow variability in the number of seeders, the relevant queues are no longer easily solved, even in the exponential case. This motivated [2] to analyze ODE models for this problem. In the general workload distribution case, the natural fluid model takes the form of a PDE, as was recently proposed in [12], [13]. We summarize the approach here.

The state in such a fluid model is a real-valued function $F(t,\sigma)$ that counts the population of leechers present at time t that have residual workload larger than σ. This function is monotonically decreasing, satisfying $F(t,\infty) = 0$ and $F(t,0) = x(t)$, the total number of leechers, and is assumed to be smooth. The following incremental analysis provides a heuristic motivation for the choice of state evolution:

$$F(t+dt,\sigma) = \lambda H(\sigma)dt + F(t,\sigma + rdt). \tag{7}$$

Given the current state $F(t,\sigma)$, its value after a small time interval dt is determined in (7) by two components:

- The number of new arrivals in $(t, t+dt]$ with workload larger than σ. Of the λdt total arrivals, a fraction $H(\sigma)$ has workload larger than σ.
- The number of jobs present at time t with residual work larger than $\sigma + rdt$. These jobs process rdt units of service, thus remaining above σ at time $t + dt$.

By subtracting $F(t,\sigma)$, dividing by dt and letting $dt \to 0$ we have the following evolution for the state, in the form of a transport PDE:

$$\frac{\partial F}{\partial t} = \lambda H(\sigma) + r(F,y,\sigma)\frac{\partial F}{\partial \sigma}. \tag{8}$$

Here r is the service rate, which in general can depend on the network state, the number of seeders y (not necessarily fixed) and possibly the download advance.

A. PDE model for the mean of an M/G/∞ queue.

To connect both approaches, let us focus on the case where the service rate r is taken to be constant, and hence the dynamics (8) is linear. In the P2P setting this can arise when download capacity is the bottleneck. Instead of the measure-valued state Φ_t considered before, take as a state its complementary cumulative distribution function:

$$\varphi(t,\sigma) = \int_\sigma^\infty \Phi_t(du). \tag{9}$$

This piecewise constant, step-decreasing function satisfies $\varphi(t,0) = x(t)$, $\varphi(t,\infty) = 0$.

Proposition 4: Let $F(t,\sigma) := E[\varphi(t,\sigma)]$ be the expectation of the random process defined in (9). Then $F(t,\sigma)$ satisfies the PDE (8).

Proof: Denote by $\{T_n\}$ the arrival instants and $\{\sigma_n\}$ the job sizes. For any fixed times $s < t$ we have the following evolution equation for the state:

$$\varphi(t,\sigma) = \varphi(s,\sigma + r(t-s)) + \sum_n \mathbf{1}_{\{T_n \in (s,t]\}}\mathbf{1}_{\{\sigma + r(t-T_n) < \sigma_n\}},$$

where the first term accounts for the service of jobs already in the system at time s, and the second term accounts for new arrivals and their corresponding service, $\mathbf{1}$ denoting the indicator function. Define $F(t,\sigma) := E[\varphi(t,\sigma)]$, then by taking expectations in the above equation and applying Campbell's formula [18] to the independently marked arrival process $\{T_n, \sigma_n\}$ we get:

$$F(t,\sigma) = F(s,\sigma + r(t-s)) + \lambda \int_s^t H(\sigma + r(t-\tau))d\tau.$$

Assuming[2] $F(0,\sigma)$ differentiable in σ, then $F(t,\sigma)$ is differentiable, and a solution of:

$$\frac{\partial F}{\partial t} = \lambda H(\sigma) + r\frac{\partial F}{\partial \sigma}. \tag{10}$$

■

Thus, the PDE dynamics (8) above allows us to track the average state of the system, for the case of constant service rates. Unfortunately, this method does not easily extend to a general $r(F,y,\sigma)$; in that case a connection must be sought through scaling limits of the stochastic process. A full derivation is outside the scope of this paper, we refer the reader to [17] for fluid limit results on processor sharing systems with the measure-valued state descriptor, as well as [19] for the relation between fluid limits and PDE models. Nevertheless, in what follows we show further evidence of consistency between the queueing and fluid approaches, for the fixed seeder case.

[2]This is not a restrictive hypothesis, it holds if the initial condition is picked using the residual job size distribution $\bar{H}$, which always has a density.

B. Fluid equilibrium of a globally-sustained torrent in the fixed seeder case.

Note that (8) allows the modeling of different bandwidth sharing disciplines, depending on the choice of $r(F, y, \sigma)$. We now specialize it to the processor sharing discipline for fixed seeders of the previous section $r = \frac{x+y_0}{x}$, for which we have

$$\frac{\partial F}{\partial t} = \lambda H(\sigma) + \mu \left(\frac{x + y_0}{x} \right) \frac{\partial F}{\partial \sigma}. \tag{11}$$

Let us analyze the equilibrium of the above model. Denote by * the values of the system at equilibrium. Setting $\partial F/\partial t = 0$ and integrating in the positive real line we have:

$$\lambda \int_0^\infty H(\sigma)\, d\sigma + \mu \left(\frac{x^* + y_0}{x^*} \right) \int_0^\infty \frac{\partial F^*}{\partial \sigma}\, d\sigma = 0.$$

Recall that $\int_0^\infty H(\sigma)\, d\sigma = 1$, the (normalized) average job size. Also by the hypothesis on F we have

$$\int_0^\infty \frac{\partial F^*}{\partial \sigma}\, d\sigma = -F^*(0) = -x^*.$$

We conclude that the number of leechers in equilibrium is:

$$x^* = \rho - y_0,$$

provided $\rho = \frac{\lambda}{\mu} > y_0$. Thus, in the globally sustained case analyzed in Section II-C, the equilibrium of the PDE model reflects the correct average number of leechers obtained in the asymptotic analysis.[3]

Substituting the value of x^* in the equilibrium condition we get, using the boundary condition $F^*(\infty) = 0$:

$$F^*(\sigma) = (\rho - y_0) \int_\sigma^\infty H(s)\, ds = (\rho - y_0)\bar{H}(\sigma),$$

where we have used the definition of $\bar{H}$ in (5). The equilibrium distribution in the PDE model when $\rho > y_0$ with the deterministic limit found in Theorem 2.

C. Variance around equilibrium for fixed seeders.

We now focus on the case of deterministic content, i.e. $H(\sigma) = \mathbf{1}_{[0,1)}(\sigma)$, where we can simplify (11) to:

$$\frac{\partial F}{\partial t} = \lambda + \mu \left(\frac{x + y_0}{x} \right) \frac{\partial F}{\partial \sigma}, \tag{12}$$

for $0 \leqslant \sigma \leqslant 1$, with $F(t, \sigma) \equiv 0$ for all $\sigma \geqslant 1$, since no peers can have more remaining workload than the file size.

We still assume $\rho > y_0$. The equilibrium then becomes $x^* = \rho - y_0$ and $F^*(\sigma) = (\rho - y_0)(1 - \sigma) = x^*(1 - \sigma)$ for $\sigma \in [0, 1]$. In equilibrium, the download progress of leechers is uniformly distributed in $[0, 1]$, which corresponds to the residual lifetime distribution for deterministic jobs. The equilibrium rate per leecher is

$$r^* = \mu \left(\frac{x^* + y_0}{x^*} \right) = \mu \frac{\rho}{\rho - y_0},$$

and we will denote $\tau := 1/r^*$, which can be interpreted as the average download time in equilibrium.

[3]If $\rho < y_0$, no equilibrium with positive x^* exists, and the solution of the PDE approaches 0 as $t \to \infty$.

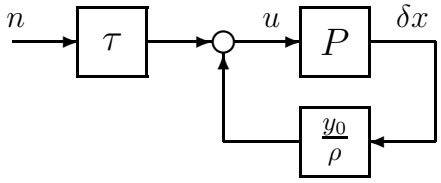

Figure 1. Linearized dynamics as a feedback loop, with injected noise.

We wish to analyze the local variability around equilibrium using the fluid model, for which we employ the classical techniques of linearizing the dynamics and representing the source of variability as injected noise. Since we are treating deterministic job sizes and fixed seeders, the only source of randomness are the Poisson arrivals. This corresponds to introducing a white noise source $n(t)$ on the right-hand side of equation (12). More formally, Poisson arrivals in an interval $[0, t]$ can be approximated by $\lambda t + \sqrt{\lambda} W(t)$, $W(t)$ being unit Brownian motion; the classical white noise model identifies this with $\int_0^t (\lambda + n(\tau)) d\tau$, with $n(t)$ a stationary process of power spectral density $S_n(\omega) \equiv \lambda$. Since the dynamics exhibits feedback, characterizing the noise response requires evaluation of the corresponding closed-loop transfer functions.

We now perform the linearization of (12), using incremental variables $x = x^* + \delta x$, $r = r^* + \delta r$ and $F = F^* + f$, and introducing the injected noise $n(t)$. The linearized dynamics at the equilibrium become:

$$\frac{\partial f}{\partial t} = r^* \frac{\partial f}{\partial \sigma} + \frac{\partial F^*}{\partial \sigma} \delta r + n = r^* \frac{\partial f}{\partial \sigma} - x^* \delta r + n.$$

Noting that $\delta r = \mu \delta \left(1 + \frac{y_0}{x}\right) = -\mu \frac{y_0}{x^{*2}} \delta x$, we arrive at:

$$\begin{aligned} \frac{\partial f}{\partial t} &= \mu \frac{\rho}{\rho - y_0} \frac{\partial f}{\partial \sigma} + \mu \frac{y_0}{\rho - y_0} \delta x + n \\ &= \frac{1}{\tau} \frac{\partial f}{\partial \sigma} + \frac{1}{\tau} \left(\frac{y_0}{\rho} \delta x + \tau n \right). \end{aligned} \tag{13}$$

The above dynamics can be written as a feedback loop of the infinite dimensional system

$$P : \begin{cases} \frac{\partial f}{\partial t} = \frac{1}{\tau} \frac{\partial f}{\partial \sigma} + \frac{1}{\tau} u, \\ \\ \delta x = f(t, 0). \end{cases} \tag{14}$$

with the static feedback equation $u := \frac{y_0}{\rho} \delta x + \tau n$ as depicted in Figure 1. The above dynamics are now analyzed via transfer functions in the Laplace domain.

Proposition 5: The transfer function of system P is:

$$\hat{P}(s) = \frac{1 - e^{-\tau s}}{\tau s}, \tag{15}$$

and the closed loop transfer function between the input n and the output δx:

$$\hat{Q}(s) = \frac{\tau \hat{P}(s)}{1 - \frac{y_0}{\rho} \hat{P}(s)}, \tag{16}$$

with the latter being stable (analytic in $Re[s] \geq 0$).

Proof: Let $\hat{f}(s,\sigma)$ denote the Laplace transform in the time variable of $f(t,\sigma)$. For zero initial conditions, (14) yields

$$s\hat{f}(s,\sigma) = \frac{1}{\tau}\frac{\partial \hat{f}}{\partial \sigma} + \frac{1}{\tau}\hat{u}(s);$$

this is now an ordinary differential equation in σ, with constant coefficients. Noting that $\hat{f}(s,1) = 0$, we have the solution:

$$\hat{f}(s,\sigma) = \frac{1}{\tau s}(1 - e^{\tau s(\sigma-1)})\hat{u}(s).$$

Evaluating at $\sigma = 0$ gives $\widehat{\delta x}(s)$, the Laplace transform of the output. Therefore we obtain the transfer function $\hat{P}(s)$ in (15) for the block, as claimed. Moreover, it is easily checked that $\|\hat{P}(s)\|_\infty = \sup_{\omega\in\mathbb{R}} |\hat{P}(j\omega)| = 1$, achieved at $\omega = 0$. It follows that the feedback loop has gain:

$$\left\| \hat{P}(s)\frac{y_0}{\rho} \right\|_\infty = \frac{y_0}{\rho} < 1,$$

and therefore from the small-gain theorem [20], the closed-loop transfer function $\hat{Q}(s)$ in (16) is analytic in the closed right half-plane. ∎

As a stationary process, the output variations δx will thus have power spectral density $S_x(\omega) = \lambda\ |\hat{Q}(j\omega)|^2$, with $\hat{Q}$ in (16). It turns out that

$$\|\hat{Q}\|_2^2 = \int_{-\infty}^{\infty} |\hat{Q}(j\omega)|^2 \frac{d\omega}{2\pi} = \frac{1}{\mu},$$

therefore the steady state variance of δx is given by

$$\int_{-\infty}^{\infty} S_x(\omega)\frac{d\omega}{2\pi} = \frac{\lambda}{\mu} = \rho.$$

Referring back to Section II-C, we had argued that for a large value of the scaling L, the stationary distribution of the number of leechers had mean $\approx (\rho - y_0)L$, and variance $\approx \rho L$. If we denote this process by x^L, we will thus have

$$\frac{x^L - (\rho - y_0)L}{\sqrt{L}}$$

approaching a law of variance ρ; so again we have consistency with the computed variance for the fluid limit.

IV. Validation of the PDE model by simulation

In the previous Section we showed the consistency between the queueing and fluid approaches in special cases. The main interest of the fluid model, however, is that it allows us to easily go beyond these cases and still deliver analytical predictions. In this section we review some generalizations, introduced in [13], and offer simulation experiments that validate their accuracy. All simulations were performed using the network simulator `ns2` with the BitTorrent library [21], which closely mimics the behavior of the BitTorrent protocol, including the chunk availability, the tit-for-tat rules and the transport layer connections. The file size of interest if of 100Mbytes and is subdivided in 400 small size chunks. The uplink bandwidth of clients is 256kbps, which accounting for protocol overheads gives an average upload time $1/\mu \approx$ 1hr.

A. The case of variable, endogenously generated seeders.

Let us now include in the dynamics the seeder variability. Assume that leechers that finish download become seeders, and may stay in the system departing at an individual rate γ. The dynamics of the system accounting for the number of seeders $y(t)$, under the processor sharing discipline follows:

$$\frac{\partial F}{\partial t} = \lambda H(\sigma) + \mu\left(\frac{x+y}{x}\right)\frac{\partial F}{\partial \sigma}, \tag{17a}$$

$$\dot{y} = \mu\left(\frac{x+y}{x}\right)\left.\frac{\partial F}{\partial \sigma}\right|_{\sigma=0} - \gamma y. \tag{17b}$$

Now consider the case of deterministic file sizes, as before. We also assume that $\gamma > \mu$, which corresponds to seeders departing faster than the upload time of a copy of the file, leading to the globally sustained equilibrium [12]:

$$x^* = \lambda\left(\frac{1}{\mu} - \frac{1}{\gamma}\right), \quad F^*(\sigma) = x^*(1-\sigma), \quad y^* = \frac{\lambda}{\gamma}.$$

In [13], the local analysis of the previous section was carried out for this dynamics, injecting noise terms n_1, n_2 in leecher arrivals and departures. The closed loop transfer function from noise sources to δx is shown to take the form:

$$\widehat{\delta x}(s) = \frac{\hat{P}(s)}{1 - \hat{P}(s)\hat{P}_2(s)}[\hat{W}_1(s)\hat{n}_1(s) + \hat{W}_2(s)\hat{n}_2(s)],$$

where $\hat{P}(s)$ is exactly of the form (15) with $\tau = (\gamma-\mu)/\gamma$, $\hat{P}_2 = \frac{s+\mu}{s+\gamma}$, and $\hat{W}_1(s)$, $\hat{W}_2(s)$ are stable transfer functions that we omit for brevity. It is shown in [13] that the feedback loop has also the small-gain property and hence is stable.

This leads to a means of computing the steady-state variance through a frequency integral, as in the previous section. Here, however, we do not have closed-form expressions and the integral is only evaluated numerically. Similar calculations yield the steady-state variance of δy, the seeder population.

We now simulate this scenario, with an arrival intensity of $\lambda = 1.8$ arrivals per minute. After finishing download, peers stay in the system as seeders for an exponentially time with average $\approx$ 18 min, and thus $\gamma > \mu$. Results are shown in Figure 2. For comparison purposes, the dashed lines indicate the theoretical equilibrium $x^* = 77.8$, $y^* = 33.3$. We also calculated the variance of x and y by numerically integrating the noise input-output transfer functions, with noise terms of power λ. The dash-dot lines are based on two standard deviations ($\approx$ 95% confidence interval) predicted by our model, where we can see that indeed the system behaves as expected. Furthermore, comparing in Table I our predictions of variance with those of the ODE model in [2]. Note that our model provides a closer estimate.

B. Transient analysis

Consider now a different scenario. Suppose a given number of initial seeders y_0 would like to propagate a certain content to a given number of leechers, with no new arrivals into the system. Assume that the initial number of leechers is x_0, and their remaining workload is distributed in the positive half line. These leechers download the content and leave immediately

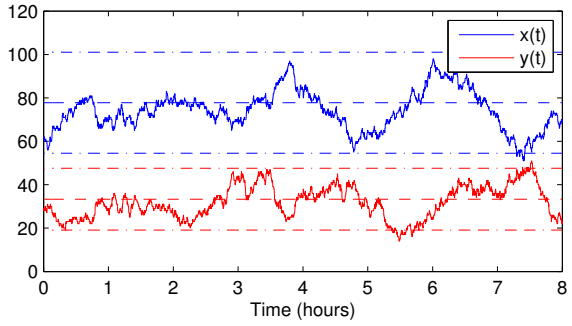

Figure 2. Evolution of the number of leechers and seeders in a P2P system which operates saturated by uplink capacity.

	ODE model	PDE model	Empirical
$\sigma(x)$	7.2	11.8	10.55
$\sigma(y)$	5.7	7.3	7.9

Table I
PREDICTED AND EMPIRICAL STANDARD DEVIATIONS.

after they finish. This case is a typical situation in torrents nowadays, with the main relevant performance metric being the completion time, i.e. the time needed to finish service of all the initial leechers. The corresponding PDE model is given by (11) with no arrivals, i.e. :

$$\frac{\partial F}{\partial t} = \mu \frac{x + y_0}{x} \frac{\partial F}{\partial \sigma}. \tag{18}$$

As for the initial condition, assume that $F(0, \sigma) = \phi(\sigma)$, a strictly decreasing differentiable function of σ, and constrain it to satisfy $\phi(0) = x_0$ (initial number of leechers), and $\phi(\sigma) \to 0$ when $\sigma \to \infty$. We have the following characterization of the completion time, which is a slightly generalized version of a result proved in [13]:

Proposition 6: The time needed to empty a processor-sharing P2P system with y_0 servers and starting from an initial condition $\phi(\sigma)$ is given by:

$$T = \frac{1}{\mu} \int_0^\infty \frac{\phi(\sigma)}{\phi(\sigma) + y_0} d\sigma. \tag{19}$$

We note that the above result encompasses many distributions of the initial pending workload; if we choose an exponential $\phi(\sigma) = x_0 e^{-\sigma}$, then the above integral evaluates to

$$T_{\text{exp}} = \frac{1}{\mu} \log \left(1 + \frac{x_0}{y_0} \right).$$

This result coincides with the predictions of time to completion that use an ODE model of the dynamics as in [2], see [13].

If, instead, the initial pending workload approaches $\phi(\sigma) = x_0 \mathbf{1}_{[0,1)}(\sigma)$ (i.e. all leechers want to download the same content of unit size), then the time becomes:

$$T_D = \frac{1}{\mu} \frac{x_0}{x_0 + y_0}. \tag{20}$$

The previous expressions have dramatically different behavior as x_0/y_0 grows; T_{exp} diverges logarithmically, while T_D remains bounded by $1/\mu$, the time to upload a copy of the file. This bound holds regardless of the initial number of leechers!

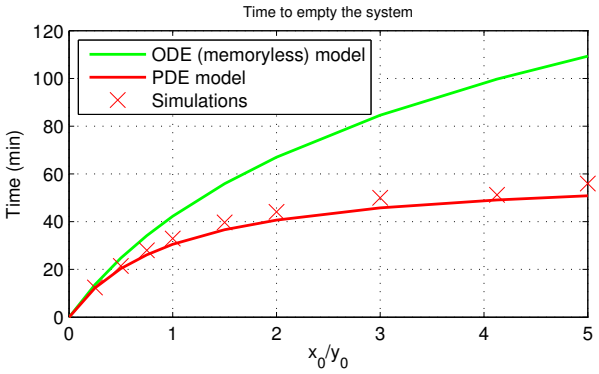

Figure 3. Time to finish service in a P2P system with varying initial leechers.

We simulate this scenario, starting the P2P system x_0 leechers with no initial content, and y_0 seeders. The time to finish service of the initial leechers is given by equation (20) which in particular in this case is bounded by $1/\mu \approx 60$min. In Figure 3 we plot the results for several initial values of the ratio between leechers and seeders x_0/y_0. The time to finish download predicted by our model is compared with simulation results, showing good fit. We also plot the time T_{exp} predicted by the simpler ODE model, which is pessimistic.

These results again emphasizes the scalability of P2P file exchange mechanisms: when the demand is large, so is the available supply. Moreover, these kind of results cannot be obtained with the previous ODE models, since they lack information about the download progress.

V. CONCLUSIONS AND FUTURE WORK

In this paper we analyzed models for content propagation dynamics in a P2P file exchange network. We first looked at queueing models in the form of an $M/G-$ Processor Sharing queue, which yield tractable analytical results for the case of a fixed number of seeders. We characterized the stationary distribution an showed that under a large network asymptotic the system can be approximated either by a $M/G/1$ or a shifted $M/G/\infty$ queue, depending on whether the server or peer contribution becomes dominant. We then switched to a fluid PDE model, and showed basic connections with the queueing system. In particular in the fixed seeder case we showed that the equilibrium advance distributions and population variance are mutually consistent. The PDE model applies more generally to the dynamics with seeder variability, and for studies of transient performance; the latter results are validated through packet-level simulations, showing improvement with respect to previously available models.

As future work, from a theoretical perspective, the analysis of the fluid limits and strong convergence results should be pursued. From a practical point of view, we believe PDE fluid models provide a versatile tool that could be applied to other scenarios of P2P or content dissemination systems.

APPENDIX

Proof of Theorem 1: To prove the Theorem, we shall show that the pgf $G_L(z)$ converges pointwise to the pgf of the geometric distribution with parameter ρ/y_0. Fix first $z < y_0/\rho$. Using (3), the pgf of the scaled system can be rewritten as:

$$G_L(z) = z^{-Ly_0} e^{L\rho(z-1)} \frac{P(S_L^{(z\rho)} \geqslant Ly_0)}{P(S_L^{(\rho)} \geqslant Ly_0)}, \qquad (21)$$

where $P(S_L^{(\eta)} \geqslant Ly_0)$ is the tail probability of a Poisson distribution with mean $L\eta$. We can interpret S_L as the sum of L independent copies of Poisson(η) random variables. Since $y_0 > \rho z$ and $y_0 > \rho$, both probabilities will go to 0 by the law of large numbers. Using the Bahadur-Rao large deviations asymptotic [22] for lattice distributions, we write:

$$P(S_L^{(\eta)} \geq Ly_0) \sim \frac{e^{-L\left(y_0 \log \frac{y_0}{\eta} + \eta - y_0\right)}}{\left(1 - \frac{\eta}{y_0}\right)\sqrt{2\pi L y_0}}.$$

Taking $\eta = \rho z$ and $\eta = \rho$, replacing the probabilities in (21) by its equivalent expressions, and cancelling terms we have:

$$G_L(z) \to_{L\to\infty} \frac{1 - \frac{\rho}{y_0}}{1 - \frac{\rho z}{y_0}},$$

for all $z < \rho/y_0$, and the right hand side is the pgf of the geometric distribution. For $z > \rho/y_0$, $P(S_L^{(\rho z)} \geqslant Ly_0)$ remains bounded away from 0, since now $[y_0, \infty)$ includes the mean. Meanwhile $P(S_L^{(\rho)} \geqslant Ly_0) \to 0$ as before. Substituting again the equivalent for the denominator term we arrive at:

$$G_L(z) \sim O(\sqrt{L}) e^{L\left(y_0 \log \frac{y_0}{\rho z} + \rho z - y_0\right)} \to \infty$$

when $L \to \infty$ for $z > y_0/\rho$, which concludes the proof. ■

Proof of Theorem 2: The proof relies on the pointwise convergence of the Laplace functionals. By analogous calculations to the ones in Proposition 3, we have that the Laplace functional of $\tilde{\Phi}^L$ is given by:

$$\mathcal{L}_{\tilde{\Phi}^L}[f] = G_L\left(\int_0^\infty e^{-\frac{f(\sigma)}{L}} \bar{H}(d\sigma)\right),$$

where the term $f(\sigma)/L$ comes from the fact that we have scaled the measure Φ^L by $1/L$. Given f, using the series expansion of the exponential and the fact that $\bar{H}$ is a finite measure, we can write the following approximation:

$$z_L := \int_0^\infty e^{-\frac{f(\sigma)}{L}} \bar{H}(d\sigma) = 1 - \frac{1}{L}\int_0^\infty f(\sigma)\bar{H}(d\sigma) + o\left(\frac{1}{L}\right)$$

where the term $o(1/L)$ may depend on f, but not on ρ or y_0. Note that $z_L \to_{L\to\infty} 1$.

Let now $\eta_L(z) = \log G_L(z)$. Recalling (21), we have:

$$\eta_L(z) = -Ly_0 \log z + L\rho(z-1) + \log\left(P(S_L^{(\rho z)} \geq Ly_0)\right) - \log\left(P(S_L^{(\rho)} \geq Ly_0)\right). \qquad (22)$$

Choose now z^* such that $y_0/\rho < z^* < 1$. Then $h_L(z) := P(S_L^{(\rho z)} > Ly_0)$ is increasing in z. Since $\rho z^* > y_0$, by the weak law of large numbers, $h_L(z^*) \to 1$ and thus $h_L(z) \to 1$ uniformly in $[z^*, \infty)$. Observe that $z_L > z^*$ for large enough L. Substituting z_L in (22) we get:

$$\eta_L(z_L) = -Ly_0 \log z_L + L\rho(z_L - 1) + \log[h_L(z_L)/h_L(1)].$$

As $L \to \infty$, the last term vanishes by the above argument, and using the definition of z_L and taking limit we have:

$$\lim_L \eta_L(z_L) = -(\rho - y_0)\int_0^\infty f(\sigma)\bar{H}(d\sigma),$$

or equivalently, $\mathcal{L}_{\tilde{\Phi}^L}[f] \to e^{(\rho - y_0)\int_0^\infty f(\sigma)\bar{H}(d\sigma)}$, the Laplace transform of the deterministic measure, as desired. ■

REFERENCES

[1] X. Yang and G. De Veciana, "Service capacity of peer-to-peer networks," in *Proc. of IEEE Infocom, Hong Kong*, 2004.

[2] D. Qiu and R. Srikant, "Modeling and performance analysis of BitTorrent-like peer-to-peer networks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 367–378, 2004.

[3] D. Qiu and W. Sang, "Global stability of peer-to-peer file sharing systems," *Computer Communications*, vol. 31, no. 2, pp. 212–219, 2008.

[4] F. Clévenot, P. Nain, and K. Ross, "Multiclass P2P networks: Static resource allocation for service differentiation and bandwidth diversity," in *IFIP Performance, Juan-les-Pins, France*, 2005.

[5] G. Kesidis, T. Konstantopoulos, and P. Sousi, "Modeling file-sharing with bittorrent-like incentives," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Honolulu, USA*, 2007.

[6] L. Massoulié and M. Vojnović, "Coupon replication systems," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 603–616, 2008.

[7] B. Hajek and J. Zhu, "The missing piece syndrome in peer-to-peer communication," *Stochastic Systems*, vol. 1, no. 2, pp. 246–273, 2011.

[8] F. Simatos, P. Robert, and F. Guillemin, "Analysis of a queueing system for modeling a file sharing principle," in *Proc. of ACM Sigmetrics*, 2008.

[9] K. Leibnitz, T. Hossfeld, N. Wakamiya, and M. Murata, "Modeling of epidemic diffusion in peer-to-peer file-sharing networks," *Lecture Notes on Computer Science*, vol. 3853, pp. 322–329, 2006.

[10] L. Leskela, P. Robert, and F. Simatos, "Interacting branching processes and linear file-sharing networks," *Adv. in Applied Probability*, vol. 42, no. 3, pp. 834–854, 2010.

[11] G. Carofiglio, R. Gaeta, M. Garetto, E. Leonardi, and M. Sereno, "A fluid-diffusive approach for modelling P2P systems," in *Proc. of MASCOTS'06*, 2006.

[12] A. Ferragut, F. Kozynski, and F. Paganini, "Dynamics of content propagation in BitTorrent-like P2P file exchange systems," in *50th IEEE Conference on Decision and Control (CDC 2011), Orlando, USA*, 2011.

[13] F. Paganini and A. Ferragut, "PDE models for population and residual work applied to peer-to-peer networks," in *46th Annual Conference on Information Sciences and Systems*, 2012.

[14] B. Cohen, "Incentives build robustness in BitTorrent," in *1st Workshop on the Economics of Peer-2-Peer Systems, Berkeley*, 2003.

[15] F. Kozynski, A. Ferragut, and F. Paganini, "Reduction of oscilations in BitTorrent by preferential unchoking mechanisms," *Informatica na educacao*, vol. 14, no. 1, pp. 29–41, 2011.

[16] S. Zachary, "A note on insensitivity in stochastic networks," *Journal of Applied Probability*, vol. 44, pp. 238–248, 2007.

[17] H. Gromoll, A. Puha, and R. Williams, "The fluid limit of a heavily loaded processor sharing queue," *Annals of Applied Probability*, vol. 12, pp. 797–859, 2002.

[18] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes (Vol. II)*. NY: Springer, 2008.

[19] F. Paganini, K. Tang, A. Ferragut, and L. Andrew, "Stability of networks under general file size distribution and alpha fair bandwidth allocation," *IEEE Trans. on Automatic Control*, vol. 57, no. 3, pp. 579–591, 2012.

[20] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. NJ: Prentice-Hall, 1996.

[21] K. Eger, T. Hossfeld, A. Binzenhofer, and G. Kunzmann, "Efficient simulation of large scale p2p networks: packet-level vs. flow-level simulations," in *UPGRADE-CN '07, Monterey Bay, USA*, 2007.

[22] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. NY: Springer, 1998.