

Queueing analysis of peer-to-peer swarms: stationary distributions and their scaling limits

Andrés Ferragut*, Fernando Paganini

Universidad ORT Uruguay, Montevideo, Uruguay

Abstract

In this paper we analyze the dynamics of a P2P file exchange swarm from a queueing standpoint. In such systems, the service rate a peer receives 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 in the case of a fixed population of servers; the result behaves like a combination of $M/G/1$ and $M/G/\infty$ queues. We apply scaling limits to this queue and identify two limiting regimes, depending on whether the server or peer contribution becomes dominant. For the latter, more important case we refine the fluid limit description of the download profile with a suitable functional Gaussian approximation. We also analyze the case of a slowly varying population of servers, extending the fixed case through a quasi-stationary analysis. For practical validation we offer comparisons with detailed packet simulations.

Keywords: P2P, Processor-sharing queues, General job sizes, Fluid limits.

1. Introduction

In recent years, peer-to-peer (P2P) file sharing systems such as BitTorrent [1] have become widespread, representing an important portion of Internet traffic. Content is disseminated by subdividing it into small pieces (*chunks*), and enabling peers to exchange such units bidirectionally. The power of P2P as a means of content distribution lies on the fact that downloading peers simultaneously contribute by uploading pieces to others, so the supply of service capacity for a certain content scales with the corresponding demand.

The set of peers exchanging a given content file is often referred to as a *swarm*, in which one distinguishes two classes: some peers (*seeders* in BitTorrent parlance) already own the entire file and act like servers to others, while the downloading peers (*leechers*) act simultaneously as clients and servers. Such swarms evolve in time subject to peer arrivals and departures, and thus it is natural to analyze their population dynamics with tools of queueing theory.

A first queueing model of this nature was the seminal paper [2], where the authors proposed a continuous-time Markov chain with two M/M queues in tandem, respectively representing leecher and seeder populations. Poisson arrivals feed the leecher queue; upon service completion peers transit to a seeder queue, stay some exponential time and eventually leave the system. The defining feature of P2P is that the service capacity that controls the inner transition is state-dependent, growing with the total peer population. This Markovian model assumes implicitly that download requirements for individual peers are independent exponentials, a coarse assumption since peers are sharing the same content file. Despite this simplification, the resulting chain does not admit an analytic solution: [2] studies it numerically. This work and other relevant references are reviewed in Section 2.

*Corresponding Author

Email address: ferragut@ort.edu.uy

Postal address: Cuareim 1451, 11100

Tel./Fax: (+598) 29021505 / 29081370

In this paper we are interested in overcoming both the above limitations: we seek *analytical* characterizations of the resulting stationary distributions; and, we would also like to allow *general* job-size distributions for peers (the domain of M/G queues), covering in particular the important practical case of a deterministic job sizes, where all peers want to download the same shared file.

It turns out that both aims are within reach if one focuses the model on the *leecher* swarm, assuming seeders are a fixed amount. This is arguably a closer model of real P2P systems, where leechers rarely stay beyond their own download, and the seeder function is sustained by a set of long-term committed server peers. In the M/M case the leecher population becomes a one-state birth and death process, whose stationary distribution can be solved explicitly, a combination of the $M/M/1$ and $M/M/\infty$ queues. Going to the M/G case requires making a more detailed description of how the service capacity is distributed between peers in the system; we adopt here a *processor sharing* model, suitable for systems with *homogeneous* peer parameters, as further justified below. In this situation we can exploit insensitivity properties [3] to characterize the M/G stationary distribution. These models are covered in Section 3.

The next objective of our paper is to obtain simpler descriptions of the stationary distributions in the limit of large population sizes. Here we exploit a characterization [4, 5] of the M/G -state in terms of a point measure on the positive real-line, which captures both population and residual work. In Section 4, taking this state descriptor through a scaling limit, we use analytical tools from point processes to prove convergence to certain limiting laws, identifying two cases. If the server component is dominant we have essentially an $M/G/1$ queue. In the more important case where file sharing is dominant we obtain, through a law-of-large numbers type scaling, a deterministic fluid limit describing the steady-state population profile of leechers as a function of their residual work. In Section 5 we further apply a central-limit type scaling to study deviations from the limit profile, obtaining a functional Gaussian characterization of this variation process.

In Section 6 we move back closer to the two-queue scenario of [2] by allowing (in the M/M case) the seeder population to vary *slowly*, fed by a thinned version of the leecher termination process. Slowing down this seeder process, a quasi-stationary analysis yields expressions for the limiting joint distribution as a suitable mixture of the fixed seeder laws seen before.

Simulations that validate the results as relevant to practical BitTorrent networks are presented in Section 7. Finally, conclusions and lines of future work are presented in Section 8. Partial results leading up to this paper were presented in [6].

2. Background and related work

We provide first more detail on the Markov chain model of [2, 7]: leechers arrive as a Poisson process of intensity λ , $x(t)$ denoting the amount of leechers in the system at time t , while $y(t)$ 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. We note again that despite its simplicity, analytical results do not readily follow since this Markov chain is not reversible.

This analytical difficulty motivated [8] to consider an ordinary differential equation model which describes the populations as real variables, but otherwise has the same structure (some refinements are added as well). In this case the equilibrium is readily characterized, as well as its stability. Another early reference in this matter is [9], in which heterogeneous download rates are considered. Still, by treating the populations as state variables, there is an implicit memoryless assumption, with no detail provided on the residual workloads that would apply to the general job-size distribution case. Motivated by this limitation we developed in [10] a partial differential equation (PDE) model that tracks population as a function of residual work for a P2P swarm with general file distribution. That work is complementary to the present paper: there we describe the swarm dynamics, including transients, but only in the fluid sense; here we give steady-state results and develop the bridge between the queues and their fluid limits in equilibrium. Other papers that model download progress through multiple tandem queues are [11, 12], however their service rates do not correspond to a P2P system with file sharing across all classes of peers.

The above models do not attempt to track individual file pieces, but rather just describe aggregate quantities of download, assuming the microscopic chunk mechanisms enable a fair sharing of aggregate bandwidth. At the other extreme of detail are Markov models which consider the piece-level detail by writing a population variable for each subset of possessed chunks [13, 14]; the number of state variables is thus huge, exponential in the number of chunks

(which is commonly in the hundreds). A successful application of such models has been to identify, via Foster-Lyapunov techniques, certain population instabilities that may arise when new peer arrival rates exceed the ability of seeders to inject new pieces [15, 16]. While interesting, this is a rather extreme regime. In a stable situation it is difficult to obtain much information from this Markov model of enormous state dimension; in particular most subsets of chunks will have no population for very long times, beyond the span of realistic service times.

Another aspect of P2P that has received attention is the resource allocation between downloading peers. A high-level discussion of this design space is given in [17], see also [18, 19] for more recent references. In the case of BitTorrent, this allocation is indirectly determined by the tit-for-tat reciprocity mechanism introduced in [1] to avoid free-riding (see [20, 21] for a game-theoretic analysis of these rules). An alternative proportional reciprocity mechanism, more amenable to analytic studies, was studied in [7] and implemented in [22]; theoretical results on the resulting allocation under these strategies are given in [23]. A common conclusion of both these theoretical papers and of empirical studies of BitTorrent [24] is that the amount of bandwidth one receives from other leechers is in proportion to the peer's upload capacity; in a *homogeneous* situation as considered in this paper, an egalitarian (processor sharing) allocation of the bandwidth is indeed a very good approximation.

3. A queueing model for fixed seeders and its stationary distribution

We begin our study by considering the model from [2, 7], but focusing on the evolution of the leecher population. To this end we assume that the population of seeders is fixed at y_0 , and leechers leave the system immediately upon completion of their downloads. This case of unconditionally “generous” seeders and totally “selfish” leechers is a simplification, but not far from typical exchange situations. Besides, it provides a first step towards analyzing situations of slowly varying seeder populations.

Let $x(t)$ denote as before the leecher population in this system, and assume as before that transition times are exponential. Leechers arrive as a Poisson process of intensity λ and finish download at a rate $R_{up} := \mu(x + y_0)$, i.e. the total uplink bandwidth available. Here μ is the individual bandwidth contributed by each peer in the system, in files per second. We assume here that these are the effectively contributed bandwidths discounted for possible protocol overheads, and that they are homogeneous to the population.

In this context, $x(t)$ is a continuous time Markov chain with state space \mathbb{N} and the following transition rates q_{ij} :

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

The above Markov chain corresponds to a simple birth-death process, which can be readily solved.

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

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

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

Note that this system can be thought of as a combination of the $M/M/1$ and $M/M/\infty$ queues. If we disregard the contribution of leechers, it reduces to an $M/M/1$ 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 $M/M/\infty$ queue, and the system is stable for all ρ .¹ Note also that in the case $\rho > y_0$ the leecher contribution is essential to maintain stability.

In the above analysis, we have assumed that the bandwidth contributed by leechers and seeders is the same. If these were different, the total bandwidth available would be $R_{up} = \mu x + \mu_0 y_0$, where μ_0 is the individual seeder contribution, with possibly $\mu \neq \mu_0$. In fact, the original model of [2] considers $\mu \leq \mu_0$, attributing the inequality to possible inefficiencies in the exchange of leechers.

¹Of course, disregarding the seeders is an extreme situation where the model is not accurate, due to leechers possibly not having the complete file to redistribute, but it serves as a limiting case for our comparison.

We remark, however, that in practice the file-sharing algorithms are highly efficient: if a peer has spare bandwidth someone will find something to download from it, and thus leechers fully utilize their upload, except in the very short period when they acquire the initial pieces. Such efficiency is observed in empirical studies [24, 25] for the case of BitTorrent, and supported via an analytical model in [8]. In our simulations of Section 7 we find this as well.

Still, for completeness, let us discuss the case where $\mu \neq \mu_0$. In that case, the invariant distribution can be characterized in similar terms to (2), with a Gamma function replacing the factorial:

$$\pi_\eta(n) = K \frac{\rho^n}{\Gamma\left(n + \frac{y_0}{\eta} + 1\right)} \quad \text{for } n \geq 0; \quad (3)$$

here $\rho = \lambda/\mu$ as before, $\eta := \mu/\mu_0$ and K is a suitable normalization constant. Note that when $\mu \leq \mu_0$, we have $\eta \leq 1$, coinciding in this case with the parameter introduced in [2].

The reason for the above modification is that y_0/η , which is the effective number of seeders relative to the leecher contribution, may no longer be an integer. Nevertheless, it is easy to show the following bound:

Proposition 2. *Let $X \sim \pi_\eta$ represent the random population of leechers in the system in equilibrium. Consider X_l and X^u the equilibrium populations of a system with $\lceil y_0/\eta \rceil$ and $\lfloor y_0/\eta \rfloor$ seeders respectively. Then we have:*

$$X_l \leq_{st} X \leq_{st} X^u,$$

where \leq_{st} denotes stochastic ordering.

The above proposition enables us to approximate the behavior of the queue in equilibrium when $\eta \neq 1$ by simply bounding it by an appropriate system with an equivalent number of seeders and $\eta = 1$. Using this approximation, the Theorems we present in Section 4 extend to the case $\eta \neq 1$ replacing y_0 by y_0/η , so there is no loss of generality in considering $\mu = \mu_0$ in what follows.

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:

$$G(z) = z^{-y_0} \frac{\sum_{m=y_0}^{\infty} (\rho z)^m / m!}{\sum_{m=y_0}^{\infty} \rho^m / m!}. \quad (4)$$

Equation (4) 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 3. *For $y_0 > 0$, the average number of leechers in the system is given by:*

$$\bar{x} = \rho - y_0 + \frac{\rho^{y_0-1} / (y_0 - 1)!}{\sum_{m=y_0}^{\infty} \rho^m / m!}. \quad (5)$$

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 $T = \frac{\bar{x}}{\lambda}$. We remark also that other performance measures such as the variance $Var(x)$ can be obtained with similar methods.

3.1. The case of general job requests

The assumption of exponential service times in the previous system is not well suited to model P2P behavior, since typically job requests are not exponential. Indeed, peers exchanging a certain file are in principle after the same amount of work, so a deterministic job size appears more accurate. In other cases the content represents a bundle of files, of which individual peers may have interest in a subset: this motivates more general distributions of workload.

We thus consider that job requests follow a general distribution, given by a complementary cumulative distribution function (CCDF) $H(\sigma)$. More precisely, $H(\sigma) = P(\sigma_n > \sigma)$ where σ_n is the job requests of user n , assumed to be an independent sequence and independent of the arrival process. We also assume that $E[\sigma_n] = \int_0^{\infty} H(\sigma) d\sigma = 1$, a normalization condition which amounts to our choice of units for the upload rate.

In this M/G case, the total population $x(t)$ is not a Markov chain, due to non exponential transitions. To completely describe the dynamics requires taking into account the remaining services of current jobs at any given time. Furthermore, the queueing discipline becomes relevant to the analysis. In this respect, under our assumption of homogeneous bandwidth among peers, we will assume that upload bandwidth is equally shared among leechers present in the system, i.e. each leecher receives a service rate of $r = R_{up}/x$. This assumption also matches well with empirical studies of BitTorrent [24, 26]. We are now ready to fully specify our queueing system.

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

System 1 constitutes a *processor sharing* queue with state dependent rates. This class of systems is thoroughly analyzed in [27] under the broader class of *symmetric* policies. In particular, [27, Theorems 3.8 and 3.10] state that, in equilibrium, the total population of the system is *insensitive* to the detailed characteristics of the job size distribution except their mean. By applying this Theorem to our case we can conclude:

Proposition 4. *The population $x(t)$ of the P2P system 1 in equilibrium follows the distribution given by equation (2).*

The above result characterizes the total population in equilibrium in terms of the underlying Markov chain of the exponential setting. However, we would like to analyze also the download progress behavior, and how this download progress is distributed among peers present in the system in steady-state. To this end, we will rely on a characterization of the steady-state distribution of the queue given in [3, Theorem 1], and originally based on results from [28]. Introduce the distribution \bar{H} of residual life of a renewal process with inter-arrival times given by H , which has CCDF given by:

$$\bar{H}(\sigma) = \int_{\sigma}^{\infty} H(s) ds. \quad (6)$$

Note that the “residual life” here should be thought of as a residual workload rather than a time.

In this context, [3, Theorem 1] proves that the equilibrium distribution of an M/G PS queue with state dependent rates is obtained by a two step process: first the total number of jobs x following π as in the M/M case, i.e. the solution of the balance equations. Then, given that $x = n$, the residual workloads of the jobs present in the system are chosen as n iid realizations of the distribution \bar{H} .

We will use the above decomposition to obtain limit results on the large scale behavior of the system, focusing in particular in the download progress distribution. In this regard, the above result describes the stationary law through a two-stage description; to proceed further we would like to have a state descriptor that captures both population and residual work in a single mathematical object amenable to scaling. It turns out that the language of point processes is adequate for this purpose.

3.2. Point-process state descriptor

As mentioned before, in the general job size case the population $x(t)$ does not follow a Markov process. To obtain a Markov representation of the system dynamics it is necessary to incorporate the remaining workloads in the state. There are several ways to do this: for instance [3] represents the system state as a random size list of remaining services (here the list size is the population). However, a more amenable choice to store the state of PS systems is a random counting measure on \mathbb{R}^+ , with point masses at the remaining service times. This approach enables to characterize the system dynamics as a Markov process in a measure-valued space [5, Chapter 7], and stationary distributions can be computed. Moreover, the use of point measures is further explored in [4, 29] to derive fluid limits in terms of transport equations for the limiting measure. The full power of this approach is exploited in Theorem 2.

Let us now define the point-process state descriptor: 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* [30], defined by

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

for any $f \geq 0$ and bounded on \mathbb{R}^+ . We now apply it to the invariant distribution just described.

Proposition 5. *The stationary distribution of the measure valued process Φ_t is that of a random measure in \mathbb{R}^+ with Laplace functional:*

$$\mathcal{L}_\Phi[f] = G \left(\int_0^\infty e^{-f(\sigma)} \bar{H}(d\sigma) \right) \quad (7)$$

for any bounded $f \geq 0$, where $G(\cdot)$ is the pgf of π .

Proof. Consider the process Φ_t in steady-state (and drop the subscript t for simplicity) drawn using the above two-stage process. Then the event $\{x = n\}$ has probability $\pi(n)$ and conditioning on these events we have:

$$\mathcal{L}_\Phi[f] = \sum_{n=0}^{\infty} E \left[e^{-\int_0^\infty f(\sigma)\Phi(d\sigma)} \mid x = n \right] \pi(n). \quad (8)$$

The expectation term for $n > 0$ can be readily calculated using that, on the event $x = n$, the remaining workloads are *iid* realizations of \bar{H} and thus:

$$\begin{aligned} E \left[e^{-\int_0^\infty f(\sigma)\Phi(d\sigma)} \mid x = n \right] &= E \left[e^{-\sum_{i=1}^n f(\sigma_i)} \mid x = n \right] = E \left[\prod_{i=1}^n e^{-f(\sigma_i)} \mid x = n \right] = \prod_{i=1}^n E \left[e^{-f(\sigma_i)} \mid x = n \right] \\ &= \left(\int_0^\infty e^{-f(\sigma)} \bar{H}(d\sigma) \right)^n. \end{aligned}$$

The above expression is also trivially valid for $n = 0$ since in that case Φ is the null measure. Plugging this result in (8) we get:

$$\mathcal{L}_\Phi[f] = \sum_{n=0}^{\infty} \pi(n) \left(\int_0^\infty e^{-f(\sigma)} \bar{H}(d\sigma) \right)^n,$$

which is the desired result. \square

In conclusion, equations (4) and (7) characterize the steady state distribution of the total number of jobs and remaining workloads in the system. In what follows, we will analyze this distribution under suitable scalings that provide insight into the behavior of the system.

4. Steady state distribution under a large system scaling

Typical file sharing systems have a large number of peers. We now analyze the steady state distribution under a large system scaling and characterize the limit behavior of the system as the size 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 $L \rightarrow \infty$. With this scaling, the leecher load parameter ρ also grows linearly 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.

4.1. Seeder sustained case ($\rho < y_0$)

Whenever $\rho < y_0$ the seeders can cope with the demand. In the limit case when $L \rightarrow \infty$, the leecher's upload contribution becomes negligible and the system behaves as an $M/G/1 - PS$ queue served only by the seeders. More formally, we have the following:

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 $\rho_0 = \frac{\rho}{y_0} < 1$.*

Proof. We use the fact (see [30]) that convergence in law for such point processes is guaranteed by pointwise convergence of their Laplace functionals. Now, in the scaled system the Laplace functional takes the form

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

where dependence on the scaling L appears *only* in the pgf component $G_L(z)$, but not in the inner term. Thus it suffices to show that $G_L(z)$ converges pointwise to the pgf of the geometric distribution with parameter ρ/y_0 . Fix first $z < y_0/\rho$. Using (4), the pgf of the scaled system can be rewritten as:

$$G_L(z) = z^{-Ly_0} e^{L\rho(z-1)} \frac{P(S_L^{(\rho)} \geq Ly_0)}{P(S_L^{(\rho)} \geq Ly_0)}, \quad (9)$$

where $P(S_L^{(\xi)} \geq Ly_0)$ is the tail probability of a Poisson distribution with mean $L\xi$. We can interpret S_L as the sum of L independent copies of Poisson(ξ) 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 [31] for lattice distributions, we write:

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

Taking $\xi = \rho z$ and $\xi = \rho$, replacing the probabilities in (9) by its equivalent expressions, and canceling terms we have:

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

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

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

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

Since pointwise convergence of the pgfs also implies uniform integrability and hence, convergence of moments [32], we have:

Corollary 1. *When $\rho < y_0$, as $L \rightarrow \infty$, the average number of leechers \bar{x}^L verifies:*

$$\bar{x}^L \xrightarrow{L \rightarrow \infty} \frac{\rho/y_0}{1 - \rho/y_0} = \frac{\rho}{y_0 - \rho}.$$

As for the average download time, since the arrival rate is $L\lambda$ then $\bar{T}^L \approx \frac{1}{L\lambda} \frac{1}{y_0 - \rho} \rightarrow 0$ as $L \rightarrow \infty$, which is consistent with the fact that the number of servers is scaling with L and they are able to cope with the load.

4.2. Globally sustained case ($\rho > y_0$)

We now focus on the more interesting case $\rho > y_0$, where the contribution of leechers becomes crucial. Note that if we scale the system as in the previous subsection, the average number of peers present (5), 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:

$$\tilde{\Phi}^L = \frac{1}{L} \sum_{i=1}^{x^L} \delta_{\sigma_i^L}, \quad (10)$$

where as before x^L is the number of jobs in the processor sharing system with arrival rate $L\lambda$ and number of seeders Ly_0 , and σ_i^L are their remaining workloads, given that there are x^L customers present. The factor $1/L$ will normalize the number of leechers, leading to the following result:

Theorem 2. *If $\rho > y_0$, the equilibrium distribution of the scaled system $\tilde{\Phi}^L$ converges in law, as $L \rightarrow \infty$ 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 (6).*

Proof. The proof again relies on the pointwise convergence of the Laplace functionals. By analogous calculations to the ones in Proposition 5, we have that the Laplace functional of $\tilde{\Phi}^L$ is given by:

$$\mathcal{L}_{\tilde{\Phi}^L}[f] = G_L \left(\int_0^\infty e^{-\frac{f(\sigma)}{L} \bar{H}(d\sigma)} \right), \quad (11)$$

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:

$$z_L := \int_0^\infty e^{-\frac{f(\sigma)}{L} \bar{H}(d\sigma)} = 1 - \frac{1}{L} \int_0^\infty f(\sigma) \bar{H}(d\sigma) + o\left(\frac{1}{L}\right) \quad (12)$$

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

Recalling (9), we have:

$$\log G_L(z) = -Ly_0 \log z + L\rho(z - 1) + \log \left(P(S_L^{(\rho z)} \geq Ly_0) \right) - \log \left(P(S_L^{(\rho)} \geq Ly_0) \right). \quad (13)$$

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, $h_L(z^*) \rightarrow 1$ and thus $h_L(z) \rightarrow 1$ uniformly in $[z^*, \infty)$. Observe that $z_L > z^*$ for large enough L . Substituting z_L in (13) we get:

$$\log G_L(z_L) = -Ly_0 \log z_L + L\rho(z_L - 1) + \log \left(\frac{h_L(z_L)}{h_L(1)} \right).$$

Substituting z_L by the expression in (12) we get:

$$\log G_L(z_L) = -Ly_0 \log \left(1 - \frac{1}{L} \int_0^\infty f(\sigma) \bar{H}(d\sigma) + o\left(\frac{1}{L}\right) \right) - \rho \left(\int_0^\infty f(\sigma) \bar{H}(d\sigma) + o(1) \right) + \log \left(\frac{h_L(z_L)}{h_L(1)} \right).$$

As $L \rightarrow \infty$, the last term vanishes by the above argument, yielding:

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

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

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

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

where $Y \sim \text{Poisson}(L\rho)$. If $\rho > y_0$, for large L , $P(Y \geq Ly_0) \approx 1$ and therefore x , the number of jobs, behaves as a Poisson random variable shifted to the left by Ly_0 . Therefore, the average number of clients in the system is $L(\rho - y_0)$ and their remaining workloads are *iid* distributed according to $\bar{H}(d\sigma)$. As L grows, 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 contribution of the seeders is then to lower the average number of peers in the system, thus also contributing to the average download time. In fact, we have the following:

Corollary 2. *For $\rho > y_0$, as $L \rightarrow \infty$, 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. In fact, it can be shown by differentiating the pgf function G_L that, as $L \rightarrow \infty$, $\text{Var}(x^L) = L\rho + o(L)$.

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 $H(\sigma) = \mathbf{1}_{[0,1]}(\sigma)$, a deterministic distribution. 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 download profile was *assumed* in [8] to analyze the efficiency of the file sharing. Note that our model captures this effect, showing that it is in fact a *consequence* of the deterministic job sizes and the processor sharing model in a large scale setting. It will not hold, for instance, in the case of exponentially distributed file sizes.

5. Central limit approximation of the download profile

The results of Section 4 provide, for the case $\rho > y_0$, a law of large numbers characterization for the download progress in the system. It is interesting to refine the above approximation by providing a central limit theorem result. For this purpose, rather than working with the random measure state Φ_t , it is more convenient to introduce its complementary CDF:

$$F(t, \sigma) = \int_{\sigma}^{\infty} \Phi_t(ds).$$

The measure-valued process Φ_t converges as $t \rightarrow \infty$ to the steady state described by (7). This in turn defines a steady-state process in the variable σ . We will state our further asymptotic results in this function space: incorporating the scaling L in the parameters and in the measure as done in (10), the corresponding steady-state process will be denoted by $\bar{F}_L^*(\sigma)$, alternatively written as

$$\bar{F}_L^*(\sigma) = \frac{1}{L} \sum_{i=1}^{x_L} \mathbf{1}_{\{\sigma_i^L > \sigma\}}. \quad (14)$$

where the population x_L and residual workloads σ_i^L are drawn from the steady-state law for the scaled system, described in the previous section. In this notation, Theorem 2 can be restated as:

$$\bar{F}_L^*(\sigma) \xrightarrow[L]{w} \bar{F}(\sigma) := (\rho - y_0)\bar{H}(\sigma). \quad (15)$$

where \Rightarrow^w denotes convergence in distribution, in this case to a deterministic limit.

Equation (15) can be thought as a Glivenko-Cantelli type of result: for large L , the empirical distribution of the remaining workloads of jobs in the system converges to the CCDF of the residual lifetime distribution. A central-limit refinement can be obtained by using a result of Nikitin [33] from the theory of empirical processes with random population size, which we reproduce here for ease of reference, rewritten to use complementary CDFs.

Theorem 3 (Y. Nikitin, [33]). *Let ξ_1, \dots, ξ_{x_L} be a sample of random volume x_L and average size $\bar{x}_L := E[x_L]$ from a uniform distribution on $[0, 1]$, with x_L independent of the ξ_i . Let:*

$$F_L^*(u) = \frac{1}{\bar{x}_L} \sum_{i=1}^{x_L} \mathbf{1}_{\{\xi_i > u\}}$$

be the modified Kac empirical distribution function.² Assume that $\bar{x}_L \rightarrow \infty$ with $\text{Var}(x_L) < \infty$ for each L and that:

$$\frac{x_L - \bar{x}_L}{\sqrt{\text{Var}(x_L)}} \xrightarrow[L]{w} N(0, 1),$$

i.e. x_L is asymptotically Gaussian. If additionally $\lim_L \frac{\text{Var}(x_L)}{\bar{x}_L} = \beta < \infty$, then the rescaled empirical process satisfies:

$$\zeta_L(u) := \sqrt{\bar{x}_L} (F_L^*(u) - (1 - u)) \xrightarrow[L]{w} B_{\beta}(u)$$

²The main difference from traditional empirical distributions is normalization by the mean number of points instead of the sample size, which in this case is random.

where $B_\beta(u)$ is centered Gaussian process in $[0, 1]$ with correlation function $r(u, v) = \min(u, v) - uv + \beta(1 - u)(1 - v)$, $u, v \in [0, 1]$.

Note that the above Theorem generalizes the classical result on empirical distributions when $\text{Var}(x_L) = 0$, i.e. a deterministic sample size, and in this case $\beta = 0$ and $B_0(u)$ is the standard Brownian bridge.

Remark 1. The process $B_\beta(u)$ admits the following decomposition, which follows directly from the form of $r(u, v)$:

$$B_\beta(u) \stackrel{d}{=} B_0(u) + \sqrt{\beta}Z(1 - u),$$

where Z is a standard normal random variable, independent of B_0 . This means that the limit process is a standard Brownian bridge perturbed by a straight line of random slope, related to the variability of the sample size.

We apply the above theorem to obtain a refined estimate of the deviations of the scaled state $\bar{F}_L^*(\sigma)$ around the mean. We begin by studying the case of deterministic job sizes, where $\bar{H}(\sigma) = 1 - \sigma$, $\sigma \in [0, 1]$ is the uniform distribution.

We will apply Theorem 3 with $x_L \sim \pi_L$, the random population in steady state, being the random sample size. Note that, by the results on Section 4.2:

$$\lim_L \frac{\text{Var}(x_L)}{\bar{x}_L} = \frac{\rho}{\rho - y_0} < \infty. \quad (16)$$

Moreover we have the following Lemma, proved in the Appendix:

Lemma 1. Let x_L be a random variable with distribution π_L , the steady state distribution of the scaled system. Then:

$$Z_L = \frac{x_L - L(\rho - y_0)}{\sqrt{L\rho}} \xrightarrow{w} N(0, 1),$$

i.e. the population in equilibrium for the scaled system is asymptotically Gaussian.

Introduce the following scaling for the function state in (14):

$$v_L(\sigma) := \sqrt{L}(\bar{F}_L^*(\sigma) - (\rho - y_0)(1 - \sigma)), \quad (17)$$

We are now ready to prove the main result of this section.

Theorem 4. In the globally sustained P2P system with deterministic job sizes, the rescaled download profile process $v_L(\sigma)$ to a Gaussian process in $[0, 1]$, i.e.:

$$v_L(\sigma) \xrightarrow{w} \sqrt{\rho - y_0}B_0(\sigma) + \sqrt{\rho}Z(1 - \sigma)$$

where B_0 is a standard Brownian bridge and Z is a standard normal variable independent of B_0 .

Proof. By Lemma 1, the population size is asymptotically Gaussian, and equation (16) ensures that we can apply Theorem 3 with $\beta = \frac{\rho}{\rho - y_0}$. Therefore:

$$\zeta_L(\sigma) = \sqrt{\bar{x}_L} \left[\frac{1}{\bar{x}_L} \sum_{i=1}^{x_L} \mathbf{1}_{\{\sigma_i^t > \sigma\}} - (1 - \sigma) \right] \xrightarrow{w} B_\beta(\sigma). \quad (18)$$

By routine manipulations, $\zeta_L(\sigma)$ can be rewritten as:

$$\zeta_L(\sigma) = a_L \frac{v_L(\sigma)}{\sqrt{\rho - y_0}} + b_L(1 - \sigma),$$

with:

$$a_L = \sqrt{\frac{L(\rho - y_0)}{\bar{x}_L}}, \quad b_L = \frac{L}{\sqrt{\bar{x}_L}} \left((\rho - y_0) - \frac{\bar{x}_L}{L} \right)$$

and $v_L(\sigma)$ as in (17).

By using the estimation in Corollary 2, it is easy to see that $a_L \rightarrow 1$. A more involved calculation using Stirling approximations for the exact expression of \bar{x}_L in (5) shows that $b_L \rightarrow 0$. Using (18) we conclude that:

$$\frac{v_L(\sigma)}{\sqrt{\rho - y_0}} \Rightarrow_L^w B_\beta(\sigma)$$

from which the result follows using the expression for $B_\beta(\sigma)$ in Remark 1. \square

The case of generally distributed job sizes, where the scaled CCDF converges to $(\rho - y_0)\bar{H}(\sigma)$ can be reduced to the uniform distribution case, by noting that $\xi_i = 1 - \bar{H}(\sigma_i)$ follow a uniform distribution in $[0, 1]$. With this change of variables we can conclude:

Corollary 3. *In the globally sustained P2P system with general job sizes, the rescaled download profile process*

$$v_L(\sigma) := \sqrt{L}(\bar{F}_L^*(\sigma) - (\rho - y_0)\bar{H}(\sigma)),$$

converges as $L \rightarrow \infty$ to a Gaussian process in \mathbb{R}^+ , i.e.:

$$v_L(\sigma) \Rightarrow_L^w \sqrt{\rho - y_0}B_0(\bar{H}(\sigma)) + \sqrt{\rho}Z\bar{H}(\sigma)$$

with B_0 and Z as in Theorem 4.

6. Variable seeders and quasi-stationarity

The queueing analysis of Section 3 allowed us to obtain an exact expression for the equilibrium distribution of the number of leechers in the system, for the case of a fixed number of seeders. What if the number of seeders is also dynamic? In the case classical case of [2], with a seeder queue fed by leecher terminations, no exact analysis is available.

However there is a situation of endogenously generated seeders where the equilibrium distribution can be well approximated by explicit expressions: when only a small fraction of leechers choose to become seeders, and these stay in the system for a long period of time. This situation is quite typical of practical exchange scenarios.

To model this situation, consider the queueing network depicted in Fig. 1, which is a modification of the original model in [2] to take this thinning effect into account. Here $\alpha \in (0, 1]$ represents the probability a leecher becomes a seeder after finishing download, the case $\alpha = 1$ corresponding to the model in [2]. To simplify the analysis we shall return to the M/M case, assuming job sizes are exponentially distributed with mean 1, and that seeders stay in the system for an exponentially distributed time of parameter γ .

The transition rates of the Markov chain $(x(t), y(t))$ defined by the system in Fig. 1 are given by:

$$q_{(x,y),(x+1,y)} = \lambda, \tag{19a}$$

$$q_{(x,y),(x-1,y+1)} = \alpha\mu(x+y)\mathbf{1}_{\{x>0\}}, \tag{19b}$$

$$q_{(x,y),(x-1,y)} = (1-\alpha)\mu(x+y)\mathbf{1}_{\{x>0\}}, \tag{19c}$$

$$q_{(x,y),(x,y-1)} = \gamma y. \tag{19d}$$

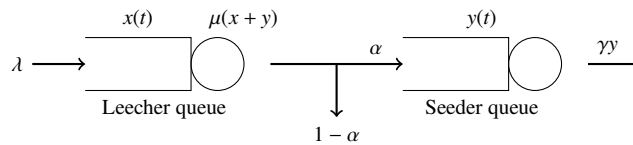


Figure 1: Queueing network for quasi-stationary analysis.

With the above assumptions, α and γ will be typically small, and therefore the second queue operates on a much larger timescale than the first queue. We shall show that under an appropriate limit, the analysis of the combined system decouples in these two timescales. Consider then a family of such systems, indexed by α , and let $\alpha \rightarrow 0$. We assume that γ also goes to zero with $\frac{\alpha}{\gamma} = \frac{1}{\nu}$ constant. In this situation, $1/\nu$ can be thought as the average time the leechers stay in the system as seeders, accounting for intermediate departures.

The above tandem system is a quasi birth-death process, and can also be thought as a denumerable singularly perturbed Markov chain, along the lines of [34]. As α and γ go to 0, the first queue evolves as the Markov chain analyzed in Section 3 with fixed seeders for each level y ; transitions in the seeder queue are observed at a slower timescale. This suggests that as $\alpha \rightarrow 0$, the equilibrium distribution of the system has a limit, and that this limit must be a mixture of the invariant distributions (2) found for fixed y .

We now prove that this quasi-stationary decomposition indeed holds and that in the limit the second queue behaves as an $M/M/\infty$ queue with load λ/ν . The main technical difficulty here is that the singularly perturbed chain (19) is not uniformizable, because the transition rates scale in x and y , and thus the results of [34] are not applicable to this case. We then follow a different approach, and begin by establishing two Lemmas:

Lemma 2. *The Markov chain with transition rates given by (19) is positive recurrent for every $\alpha \in (0, 1]$.*

Proof. The chain is clearly irreducible and using (19), the Lyapunov function $V(x, y) = x + y$ has drift:

$$\Delta V(x, y) = \lambda - (1 - \alpha)\mu(x + y)\mathbf{1}_{\{x > 0\}} - \gamma y$$

This drift is strictly negative provided $(x, y) \in \{x + y > \max\{\frac{\lambda}{(1-\alpha)\mu}, \frac{\lambda}{\gamma}\}\}$, whose complement is finite. Therefore, by Foster's criteria [5], the Markov chain is positive recurrent. \square

The above lemma ensures that, for every $\alpha \in (0, 1]$, there exists a strictly positive invariant probability distribution $\pi_\alpha(x, y)$. We have the following Lemma, proved in the Appendix:

Lemma 3. *The family $\{\pi_\alpha\}_{\alpha \in (0, 1]}$ of probability measures in $\mathbb{N} \times \mathbb{N}$ is tight. Moreover, if $(X_\alpha, Y_\alpha) \sim \pi_\alpha$, then the family of random variables $\{X_\alpha + Y_\alpha\}_{\alpha \in (0, 1]}$ is uniformly integrable.*

We are now ready to derive the main result.

Theorem 5. *The family of distributions $\{\pi_\alpha\}_{\alpha \in (0, 1]}$ has a limit $\pi_0(x, y)$ as $\alpha \rightarrow 0$ given by the following quasi-stationary decomposition:*

$$\pi_0(x, y) = \pi_0(y)\pi_0(x | y),$$

where $\pi_0(x | y)$ is the invariant distribution for the system with fixed seeders:

$$\pi_0(x | y) = \frac{1}{\sum_{m=y}^{\infty} \rho_x^m / m!} \frac{\rho_x^{x+y}}{(x+y)!} \quad \text{with } \rho_x := \frac{\lambda}{\mu}, \quad (20)$$

and

$$\pi_0(y) = e^{-\rho_y} \frac{\rho_y^y}{y!} \quad \text{with } \rho_y := \frac{\lambda}{\nu}. \quad (21)$$

Proof. By Lemma 3, the family $\{\pi_\alpha\}_{\alpha \in (0, 1]}$ is tight, and thus any subsequence π_{α_n} with $\alpha_n \rightarrow 0$ has a further subsequence $\pi_{\alpha'_n}$ which is convergent. We now prove that the limit π_0 along $\pi_{\alpha'_n}$ is unique, and therefore the whole family has a limit as $\alpha \rightarrow 0$.

The Markov chain is ergodic for $\alpha > 0$ by Lemma 2, its invariant distribution is $\pi_\alpha(x, y) > 0$ for all x, y and must satisfy the global balance equation:

$$\begin{aligned} \pi_\alpha(x, y)(\lambda + \mu(x + y) + \gamma y) = & \pi_\alpha(x - 1, y)\lambda + \pi_\alpha(x + 1, y - 1)\alpha\mu(x + y) \\ & + \pi_\alpha(x + 1, y)(1 - \alpha)\mu(x + y + 1) + \pi_\alpha(x, y + 1)\gamma(y + 1) \end{aligned} \quad (22)$$

for any $x, y > 0$ with appropriate modifications whenever x or y are 0.

If we take the limit (along an appropriate subsequence) on both sides of (22) as $\alpha \rightarrow 0$, $\gamma \rightarrow 0$ with $\alpha/\gamma = \nu$, we arrive at:

$$\pi_0(x, y)(\lambda + \mu(x + y)) = \pi_0(x - 1, y)\lambda + \pi_0(x + 1, y)\mu(x + y + 1). \quad (23)$$

For every fixed y , the above equation characterizes the behavior of the faster timescale. Given y , $\pi_0(x, y)$ should satisfy the balance equation for fixed seeders analyzed in Proposition 1, namely:

$$\pi_0(x, y) = K_y \frac{\rho_x^{x+y}}{(x+y)!} = \pi_0(y)\pi_0(x | y),$$

with $\pi_0(x | y)$ the probability distribution given by (20).

It remains to compute the limit $\pi_0(y)$. To do so, we go back to equation (22) and sum over x . Denote by:

$$\pi_\alpha(y) = \sum_{x=0}^{\infty} \pi_\alpha(x, y)$$

the marginal distribution of y and by:

$$r_\alpha(y) = \sum_{x=1}^{\infty} \mu(x + y)\pi_\alpha(x, y)$$

the average service rate of the first queue when the second queue is in state y . After summation, the balance equations become:

$$\sum_{x=0}^{\infty} \pi_\alpha(x, y) (\lambda + \mu(x + y)\mathbf{1}_{\{x>0\}} + \gamma y) = \pi_\alpha(y)\lambda + \alpha r_\alpha(y - 1) + (1 - \alpha)r_\alpha(y) + \pi_\alpha(y + 1)\gamma(y + 1).$$

After canceling terms we arrive at:

$$\alpha r_\alpha(y) + \pi_\alpha(y)\gamma y = \alpha r_\alpha(y - 1) + \pi_\alpha(y + 1)\gamma(y + 1).$$

Dividing the preceding equation by α and recalling $\gamma/\alpha = \nu$ we get:

$$r_\alpha(y) + \pi_\alpha(y)\nu y = r_\alpha(y - 1) + \pi_\alpha(y + 1)\nu(y + 1). \quad (24)$$

In order to take limits when $\alpha \rightarrow 0$, note that

$$\lim_{\alpha \rightarrow 0} r_\alpha(y) = \lim_{\alpha \rightarrow 0} \sum_{x=1}^{\infty} \mu(x + y)\pi_\alpha(x, y) = \lim_{\alpha \rightarrow 0} E \left[\mu(X_\alpha + Y_\alpha)\mathbf{1}_{\{Y_\alpha=y\}} \right],$$

where $(X_\alpha, Y_\alpha) \sim \pi_\alpha$. Since these variables are uniformly integrable, we can exchange the limit with the expectation to get:

$$\lim_{\alpha \rightarrow 0} r_\alpha(y) = \sum_{x=1}^{\infty} \mu(x + y)\pi_0(x, y) = \pi_0(y) \sum_{x=1}^{\infty} \mu(x + y)\pi_0(x | y) = \lambda\pi_0(y).$$

where the last equality follows by direct computation using (20). This is a consequence of the stability of the first queue in the fast timescale. Taking $\alpha \rightarrow 0$ in (24) and using the above we arrive at:

$$\pi_0(y)(\lambda + \nu y) = \pi_0(y - 1)\lambda + \pi_0(y + 1)\nu(y + 1), \quad (25)$$

which is simply the balance equation for an $M/M/\infty$ queue with load $\rho_y = \lambda/\nu$, as desired.

Since every sub-sequential limit must verify (20) and (21), which uniquely characterize the distribution, the limit is unique and the distribution converges weakly to the stated limit. \square

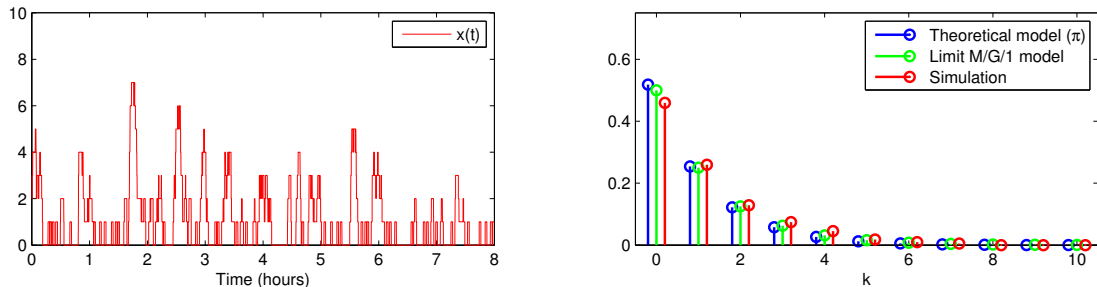


Figure 2: Time evolution of the number of leechers (left) and its steady state distribution (right) for the case $\rho < y_0$.

7. Simulation experiments

In order to validate our models and test the performance of a real P2P file exchange system, we now show some simulation scenarios.

All simulations were performed using the network simulator `ns2` [35] with the BitTorrent library [36], which closely mimics the behavior of the BitTorrent protocol, the dominant file sharing exchange mechanism in today's Internet. This library enables us to form an overlay network of P2P clients over a network topology, with the clients emulating the BitTorrent application. The file of interest has a size of 100 MBytes, which is subdivided in 400 small-size chunks. All peers have an uplink capacity of 256 kbps, corresponding to typical residential users.

In our simulation experiments, we measured the uplink bandwidth utilization of participating peers in order to determine the value of μ . We observed that leechers quickly reach an effective upload rate of approximately 90% of the link physical capacity, and this upload rate is steadily used throughout service. We also found that the effective upload rate of seeders is similar to leechers, justifying using the same value of μ for both type of peers. These results are consistent with the simulation results from [24] which also show $\approx 90 - 95\%$ uplink bandwidth utilization albeit in a slightly different simulation setting.

Accounting for these inefficiencies, the correct value of μ is determined to be:

$$\mu = \frac{\text{Eff. uplink bandwidth}}{\text{File size}} \approx 2.82 \times 10^{-4} \text{ files/min.}$$

Note that with the above value, an entire copy of the file takes $1/\mu \approx 1hr.$ to be uploaded. We now analyze several cases.

7.1. The seeder sustained case

The first simulation scenario is that of a system where the fixed seeders are able to cope with the demand, corresponding to the analysis of Section 4.1. The system is set up with a fixed quantity of $y_0 = 50$ seeders. New leechers arrive with no content as a Poisson process with intensity λ , and we choose this value such that $\rho = \lambda/\mu = 25 < y_0$. The system is empty of leechers at time 0.

A typical evolution of this system is shown in Figure 2. Note that the queue remains empty a significant portion of the time, consistent with the fact that the average demand is lower than the seeders' capacity alone. In Figure 2 we also plot the theoretical distribution predicted by the queueing model of Definition 1, which is given by π in equation (2). We also plot the predicted limit distribution given by Theorem 1, which in this case is Geometric(ρ/y_0) with $\rho/y_0 = 1/2$. Note that the scale of the system is large enough so that the limit provides a good estimation of the real invariant distribution. Finally, we show the empirical estimation of the probability of each state. Simulations closely agree with the predictions of our model. In particular, the average number of peers in the system $\bar{x} \approx 1.1$ which is similar to the predicted value of $\frac{\rho}{y_0 - \rho} = 1$.

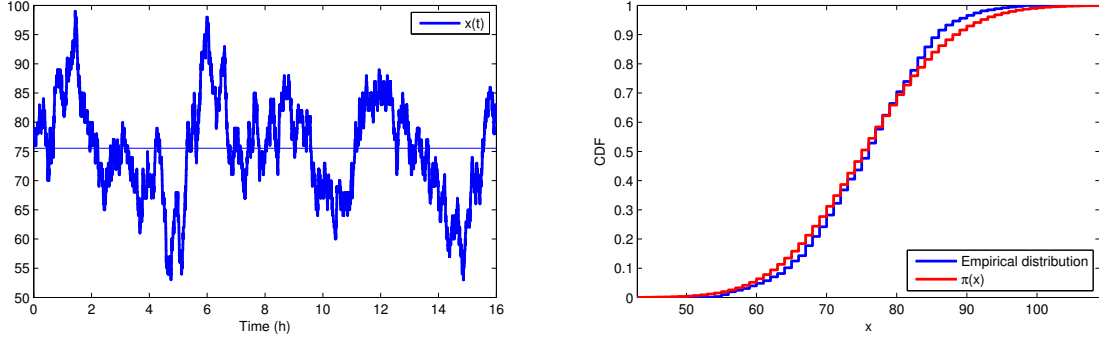


Figure 3: Time evolution of the number of leechers (left) and its steady state distribution (right) for the case $\rho > y_0$.

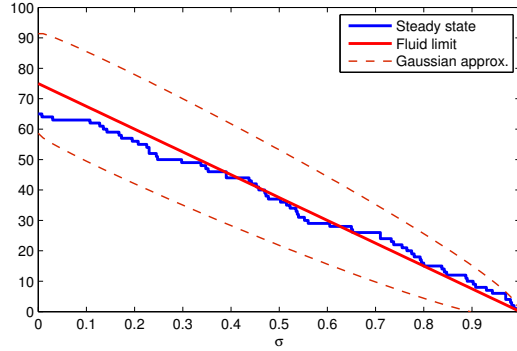


Figure 4: Steady state behavior of the download progress, and its fluid limit and Gaussian approximation.

7.2. The globally sustained case

Our second scenario corresponds to a system where leecher contribution is essential. This time we set up a fixed quantity of $y_0 = 25$ seeders. New leechers again arrive with no content into the system as a Poisson process with intensity λ , such that $\rho = \lambda/\mu = 100 > y_0$.

A typical evolution of this system in steady state is shown in Figure 3. Now the leecher population oscillates around a nonzero equilibrium which should be $\approx \rho - y_0 = 75$ according to Corollary 2. We also show in Figure 3 the empirical distribution of the leecher population from the simulations, compared to the theoretical distribution of (2), and the limit distribution obtained in Section 4.2, i.e. a shifted Poisson(ρ) random variable.

We further illustrate the predictions of our models for the residual work profile. Figure 4 shows the empirical CCDF $F(\sigma)$ of remaining workloads for a certain steady-state sample. We show for comparison the fluid limit from Theorem 2, as well as the deviations predicted in Theorem 4; in the latter case we are showing a 90% confidence interval for the deviations around the fluid limit, based on the Gaussian approximation. We have again good fit with the packet-based simulation.

7.3. Slowly varying seeders

Our third scenario corresponds to a system where a portion of the leechers stay as seeders, as considered in Section 6. To allow time for slow variations of the seeder population we make the upload time-scale faster, with 512 kbps of upload capacity and $1/\mu \approx 30$ min. We doubled the rate of leecher arrivals so that the equilibrium population remains at $\rho_x \approx 100$. A fraction $\alpha = 0.1$ of those leechers stay as seeders after finishing download. γ is chosen such that approximately $\rho_y = 25$ is the average number of seeders present.

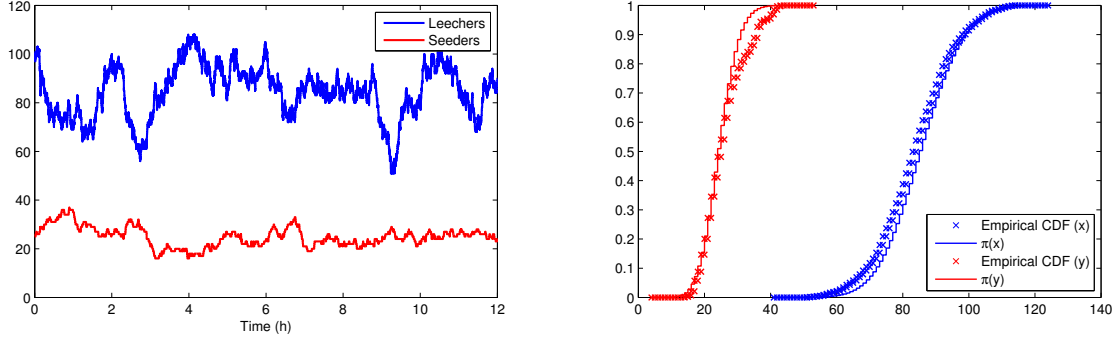


Figure 5: Time evolution of the number of leechers (above) and its steady state distribution (below) for the case $\rho > y_0$.

A typical evolution of this system in steady state is shown in Figure 5, where we see seeder populations vary more slowly than leechers'. In the second graph below we show the empirical distributions obtained from the simulations, compared with the theoretical distribution calculated using the quasi-stationary limit of Section 6. Although simulation file-sizes were deterministic (beyond the reach of the model of Section 6), there is quite close agreement between model and simulations, evidence of an (approximate) insensitivity.

8. Conclusions

In this paper we analyzed queueing models for a P2P file exchange network. In particular, we focused on a system with a fixed number of seeders, which allowed us to progress in the analysis from a queueing perspective with respect to earlier models. We identified a class of M/G -Processor Sharing queues which are tractable, characterized its stationary distribution, and showed that it is insensitive to job sizes, thus applicable to the deterministic jobs of P2P systems. We also analyzed the system under a large network asymptotic, showing that it can be approximated by a $M/G/1$ or a shifted $M/G/\infty$ queue, depending on whether the server or peer contribution becomes dominant. In the latter case, we also provide central limit theorem results for the download profile. Finally, we extended the results for fixed seeders to a case where the seeder population is slowly varying. All results were validated against detailed packet-level simulations.

Appendix

Proof of Lemma 1. By substituting $z = e^{-s}$ in (4), the Laplace transform of x_L satisfies:

$$\mathcal{L}_{x_L}(s) := E[e^{-sx_L}] = e^{sLy_0} \frac{\sum_{m=L y_0}^{\infty} (L\rho e^{-s})^m / m!}{\sum_{m=L y_0}^{\infty} (L\rho)^m / m!}.$$

Therefore:

$$\begin{aligned} \mathcal{L}_{z_L}(s) &:= E \left[e^{-s \left(\frac{x_L - L(\rho - y_0)}{\sqrt{L\rho}} \right)} \right] = e^{s \frac{\sqrt{L}(\rho - y_0)}{\sqrt{\rho}}} \mathcal{L}_{x_L} \left(\frac{s}{\sqrt{L\rho}} \right) \\ &= e^{\sqrt{L\rho}s} \frac{\sum_{m=L y_0}^{\infty} (L\rho e^{-s/\sqrt{L\rho}})^m / m!}{\sum_{m=L y_0}^{\infty} (L\rho)^m / m!} = e^{\sqrt{L\rho}s} \frac{e^{L\rho e^{-s/\sqrt{L\rho}}} P(S_L^{(\rho e^{-s/\sqrt{L\rho}})} \geq Ly_0)}{e^{L\rho} P(S_L^{(\rho)} \geq Ly_0)}. \end{aligned}$$

By the same arguments than those in Theorem 2, the probability terms go to 1 as $L \rightarrow \infty$, since $\rho > y_0$ and $\rho e^{-s/\sqrt{L\rho}} > y_0$ for large enough L , so the weak law of large numbers can be applied to the Poisson random variables involved.

To complete the proof, we have to show that, for each s :

$$g_L(s) = e^{\sqrt{L\rho}s + L\rho e^{-s/\sqrt{L\rho}} - L\rho} \xrightarrow{L \rightarrow \infty} e^{s^2/2}.$$

Consider $\log(g_L(s))$ and apply Taylor's formula for the remaining exponential term:

$$\begin{aligned} \log g_L(s) &= \sqrt{L\rho}s + L\rho \left[1 - \frac{s}{\sqrt{L\rho}} + \frac{s^2}{2L\rho} + o\left(\frac{1}{L}\right) \right] - L\rho \\ &= \frac{s^2}{2} + L\rho o\left(\frac{1}{L}\right) \xrightarrow{L \rightarrow \infty} \frac{s^2}{2}, \end{aligned}$$

as desired. \square

Proof of Lemma 3. We have to prove that, for any $\varepsilon > 0$, there exists a finite set K such that $\pi_\alpha(\{(x, y) \in K\}) > 1 - \varepsilon$. To do so, consider the auxiliary Markov process (\tilde{x}, \tilde{y}) with transition rates:

$$\begin{aligned} \tilde{q}_{(x,y),(x+1,y)} &= \lambda, \\ \tilde{q}_{(x,y),(x-1,y+1)} &= \alpha\mu x \mathbf{1}_{\{x>0\}}, \\ \tilde{q}_{(x,y),(x-1,y)} &= (1-\alpha)\mu x \mathbf{1}_{\{x>0\}}, \\ \tilde{q}_{(x,y),(x,y-1)} &= \gamma y. \end{aligned}$$

This Markov chain corresponds to a system similar to the one in Fig. 1 but where the seeder contribution is not present in the first queue. Having reduced the overall service capacity, it follows that the *total* number of clients in the system satisfies $\tilde{x}(t) + \tilde{y}(t) \geq x_\alpha(t) + y_\alpha(t)$ at all times, when starting from the same initial state. Thus

$$P(\tilde{x}(t) + \tilde{y}(t) > k) \geq P(x_\alpha(t) + y_\alpha(t) > k);$$

now both chains are ergodic so in steady state, the stationary distribution π_α must be stochastically dominated by $\tilde{\pi}$, stationary distribution of the auxiliary system. Observe that the auxiliary system is just a tandem connection of $/M/\infty$ queues, with Poisson arrivals at the first queue and abandonments in between. It is easy to check through the balance equations that the invariant distribution for this system is

$$\tilde{\pi}(x, y) = e^{-(\rho_x + \rho_y)} \frac{\rho_x^x \rho_y^y}{x! y!},$$

where:

$$\rho_x := \frac{\lambda}{\mu} \quad \rho_y := \frac{\lambda}{\nu},$$

with $\nu = \gamma/\alpha$ as before.

Given $\varepsilon > 0$, choose k such that $\tilde{\pi}(\{x, y : x + y > k\}) < \varepsilon$, and let $K = \{x, y : x + y \leq k\}$, which is finite. Since $x_\alpha + y_\alpha \leq_{st} \tilde{x} + \tilde{y}$ it is clear that, for every $\alpha \in (0, 1]$ we have $\pi_\alpha(K^c) \leq \tilde{\pi}(K^c) < \varepsilon$ and thus $\pi_\alpha(K) \geq 1 - \varepsilon$ uniformly in α .

Now consider a family of random vectors $(X_\alpha, Y_\alpha) \sim \pi_\alpha$. We have to prove that the family of random variables $\{X_\alpha + Y_\alpha\}_\alpha$ is uniformly integrable. To this end, note that:

$$E_{\pi_\alpha} \left[(X + Y)^{1+\delta} \right] \leq E_{\tilde{\pi}} \left[(X + Y)^{1+\delta} \right] < \infty$$

for every α by the same dominance argument, and $\delta > 0$ arbitrary. Since this last term does not depend on α , the family is uniformly bounded by an $L_{1+\delta}$ random variable and thus uniformly integrable [32]. \square

Acknowledgements

The authors would like to thank Fabian Kozynski for developing the first version of the simulation code, and Matthieu Jonckheere for his useful suggestions in the Proof of Theorem 5.

The authors were supported by AFOSR-US under grant FA9550-12-1-0398.

References

- [1] B. Cohen, Incentives build robustness in BitTorrent, in: Proc. of 1st Workshop on the Economics of Peer-2-Peer Systems, Berkeley, CA.
- [2] X. Yang, G. De Veciana, Service capacity of peer-to-peer networks, in: Proc. of IEEE Infocom, Hong Kong.
- [3] S. Zachary, A note on insensitivity in stochastic networks, *Journal of Applied Probability* 44 (2007) 238–248.
- [4] H. Gromoll, A. Puha, R. Williams, The fluid limit of a heavily loaded processor sharing queue, *Annals of Applied Probability* 12 (2002) 797–859.
- [5] P. Robert, *Stochastic networks and queues*, Springer-Verlag, Berlin, 2003.
- [6] A. Ferragut, F. Paganini, Content dynamics in P2P networks from queueing and fluid perspectives, in: Proc. of 24th International Teletraffic Congress, Krakow, Poland.
- [7] X. Yang, G. De Veciana, Fairness, incentives and performance in peer-to-peer networks, *Performance Evaluation* 63 (2006) 175–194.
- [8] D. Qiu, R. Srikant, Modeling and performance analysis of BitTorrent-like peer-to-peer networks, *ACM SIGCOMM Computer Communication Review* 34 (2004) 367–378.
- [9] F. Clévenot, P. Nain, K. Ross, Multiclass P2P Networks: Static Resource Allocation for Service Differentiation and Bandwidth Diversity, in: Proc. of IFIP Performance, Juan-les-Pins, France.
- [10] F. Paganini, A. Ferragut, PDE models for population and residual work applied to peer-to-peer networks, in: Proc. of 46th Annual Conference on Information Sciences and Systems.
- [11] K. Leibnitz, T. Hossfeld, N. Wakamiya, M. Murata, Modeling of Epidemic Diffusion in Peer-to-Peer File-Sharing Networks, *Lecture Notes on Computer Science* 3853 (2006) 322–329.
- [12] L. Leskela, P. Robert, F. Simatos, Interacting branching processes and linear file-sharing networks, *Adv. in Applied Probability* 42 (2010) 834–854.
- [13] G. Kesidis, T. Konstantopoulos, P. Sousi, Modeling file-sharing with BitTorrent-like incentives, in: Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Honolulu, HI.
- [14] L. Massoulié, M. Vojnović, Coupon replication systems, *IEEE/ACM Transactions on Networking* 16 (2008) 603–616.
- [15] B. Hajek, J. Zhu, The missing piece syndrome in peer-to-peer communication, *Stochastic Systems* 1 (2011) 246–273.
- [16] J. Zhu, B. Hajek, Stability of a peer-to-peer communication system, *IEEE Trans. on Information Theory* 58 (2012) 4693–4713.
- [17] B. Fan, J. C. Lui, D.-M. Chiu, The Design Trade-Offs of BitTorrent-Like File Sharing Protocols, *IEEE/ACM Trans. on Networking* 17 (2009) 365–376.
- [18] C. Aperijs, R. Johari, M. J. Freedman, Bilateral and Multilateral Exchanges for Peer-Assisted Content Distribution, *IEEE/ACM Trans. on Networking* 19 (2011) 1290–1303.
- [19] B. Q. Zhao, J. C. Lui, D. Chiu, A mathematical framework for analyzing adaptive incentive protocols in p2p networks, *IEEE/ACM Trans. on Networking* 20 (2012) 367–380.
- [20] R. T. Ma, S. C. Lee, J. C. Lui, D. K. Yau, Incentive and Service Differentiation in P2P Networks: A Game Theoretic Approach, *IEEE/ACM Transactions on Networking* 14 (2006) 978–991.
- [21] H. Zhang, G. Neglia, D. F. Towsley, G. L. Presti, Stability and efficiency of unstructured file sharing networks, *IEEE Journal on Selected Areas in Communications* 26 (2008) 1284–1294.
- [22] D. Levin, K. LaCurts, N. Spring, B. Bhattacharjee, BitTorrent is an Auction: Analyzing and Improving BitTorrent’s Incentives, *ACM SIGCOMM Computer Communication Review* 38 (2008) 243–254.
- [23] F. Wu, L. Zhang, Proportional response dynamics leads to market equilibrium, in: Proc. of the 39th Annual ACM Symposium on Theory of Computing, San Diego, CA.
- [24] A. Legout, N. Liogkas, E. Kohler, L. Zhang, Clustering and sharing incentives in bittorrent systems, in: *ACM SIGMETRICS Performance Evaluation Review*, volume 35, ACM, pp. 301–312.
- [25] F. Kozynski, A. Ferragut, F. Paganini, Reduction of oscillations in BitTorrent by preferential unchoking mechanisms, *Informatika na educacao* 14 (2011) 29–41.
- [26] A. Ferragut, F. Paganini, Fluid models of population and download progress in P2P networks, *IEEE Transactions on Control of Network Systems* (2015) to appear.
- [27] F. P. Kelly, *Reversibility and Stochastic Networks*, Wiley, Chichester, UK, 1979.
- [28] A. V. Pechinkin, An invariant queueing system, *Math. Operationsforsch. und Statist., ser. optimization* 14 (1983) 433–444.
- [29] L. Decreusefond, P. Moyal, A Functional Central Limit Theorem for the M/GI/∞ queue, *Annals of Applied Probability* 18 (2002) 2156–2178.
- [30] D. J. Daley, D. Vere-Jones, *An introduction to the theory of point processes (Vol. II)*, Springer, NY, 2008.
- [31] A. Dembo, O. Zeitouni, *Large deviations techniques and applications*, Springer, NY, 1998.
- [32] A. Gut, *Probability: a graduate course*, Springer-Verlag, NY, 2013.
- [33] Y. Y. Nikitin, Limit distributions and comparative asymptotic efficiency of the Smirnov – Kolmogorov statistics with a random index, *Journal of Soviet Mathematics* 16 (1981) 1042–1049.
- [34] E. Altman, K. Avrachenkov, R. Nunez-Queija, Perturbation analysis for denumerable markov chains with application to queueing models, *Advances in Applied Probability* 36 (2004) 839–853.
- [35] S. McCanne, S. Floyd, ns Network Simulator, <http://www.isi.edu/nsnam/ns/>, 2000.
- [36] K. Eger, T. Hossfeld, A. Binzenhofer, G. Kunzmann, Efficient simulation of large scale p2p networks: packet-level vs. flow-level simulations, in: Proc. of UPGRADE-CN ’07, Monterey Bay, USA.