Modelling the population dynamics and file availability of a P2P file sharing system

Riikka Susitaival, Samuli Aalto and Jorma Virtamo

Networking Laboratory

Helsinki University of Technology

Introduction

1. BitTorrent most popular p2p system now

   - In October 2004, bitTorrent accounted for 21 % of outgoing and 44 % of incoming traffic in Turku University network (all p2p protocols accounted for 62 % and 85 % of traffic)
   - Idea: File divided into *chunks*
     - peers download and upload concurrently
     - chunk is typically 256 KB
   - Measurements have shown three phases in life time of a file: *flash crowd*, *steady state*, and *end* phase

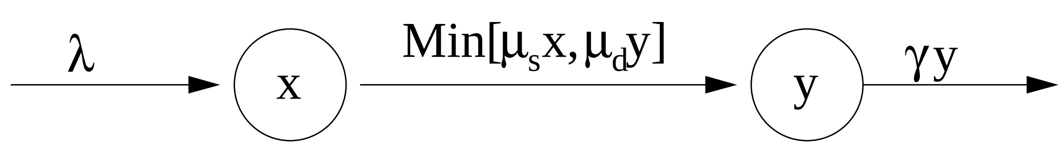
2. Some analytical models of BitTorrent-like p2p system done so far:
   - Analysis of transient regime by a branching process and steady state time regime by a Markov model
   - Performance modelling by a stochastic fluid model
   - Modelling the network level latencies by delay of a single class open queueing network and peer level latencies by delay of M/G/1/K processor sharing queues.

Purpose of our study

- Modelling of bitTorrent-like P2P system by stochastic models
  - Population dynamics; evolution of number of downloaders and seeds
  - Life time of file sharing process when availability of file is not guaranteed
  - Effect of flash crowd to the population dynamics
  - Single chunk versus multiple chunks

Sharing of a single chunk

- Basic assumptions
  - New requests for the chunk arrive with rate λ
  - A peer downloads the chunk with rate μ_d
  - After download the peer can upload the chunk with rate μ_s
  - Seeds leave the system with rate γ, also original one. If all seeds are left, system dies.
  - Let $x(t)$ be the number of downloaders and $y(t)$ the number of seeds at time t



1. Deterministic fluid model

- Evolution of $x(t)$ and $y(t)$ can be described by a deterministic fluid model:

$$\begin{aligned} \frac{dx(t)}{dt} &= \lambda - \min\{\mu_d x(t), \mu_s y(t)\}, \\ \frac{dy(t)}{dt} &= \min\{\mu_d x(t), \mu_s y(t)\} - \gamma y(t), \end{aligned} \tag{1}$$

- If $\mu_s \geq \gamma$, the steady state solution is $\bar{x} = \lambda/\mu_d$ and $\bar{y} = \lambda/\gamma$, otherwise $x \to \infty$ and $y = 0$.

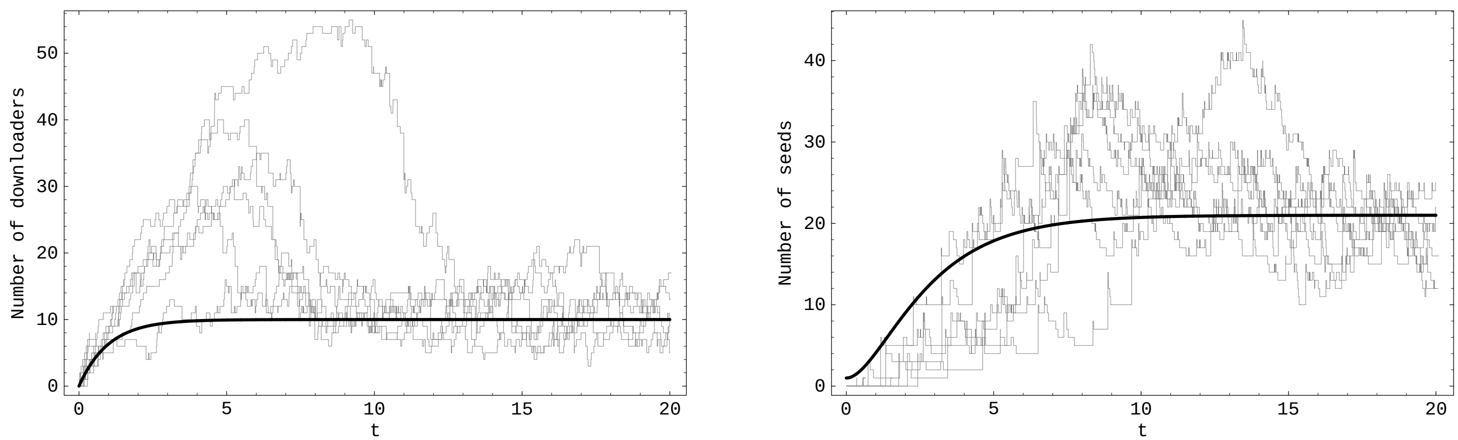


Figure 1: The number of downloaders and seeds as a function of time, when $\lambda/\mu = 10$ and $\lambda/\gamma = 20$. Solid line: fluid model, Gray lines: simulation.

2. Markov chain model

- To analyze evolution of x and y further a continuous time Markov process with state pair (x, y) and transition matrix Q constructed.
- The state transitions in the matrix Q are:

$$\begin{aligned} q((x,y),(x+1,y)) &= \lambda, \\ q((x,y),(x-1,y+1)) &= \min\{\mu_d x, \mu_s y\}, \quad \text{if } x > 0, \\ q((x,y),(x,y-1)) &= \gamma y, \quad \text{if } y > 0. \end{aligned} \tag{2}$$

- The states (x, y) with $y = 0$ are absorbing
- The mean life time of file sharing process: solve the absorbtion times of the system by a recursion

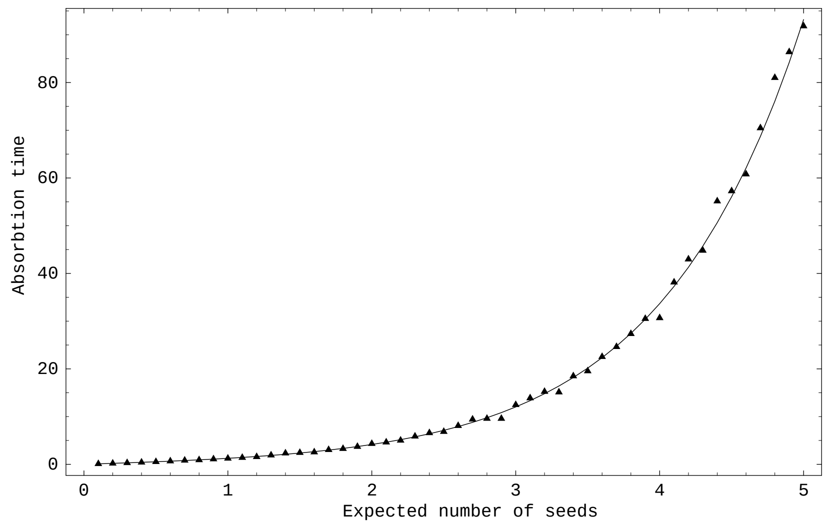


Figure 2: The mean absorbtion time of the system as a function of λ/γ. Solid line: truncated Markov process, dots: simulation of infinite system.

Flash crowd

- Some measurement indicate that the peer arrival rate decreases over time. Proposed traffic model for p2p networks:

$$\lambda(t) = \lambda_0 e^{\frac{-t}{\tau}},$$

where parameter τ describes the attenuation of the demand over time.

- As before, the system can be described by equations

$$\begin{aligned} \frac{dx(t)}{dt} &= \lambda_0 e^{\frac{-t}{\tau}} - \min\{\mu_d x(t), \mu_s y(t)\}, \\ \frac{dy(t)}{dt} &= \mu \min\{x(t), y(t)\} - \gamma y(t). \end{aligned} \tag{3}$$

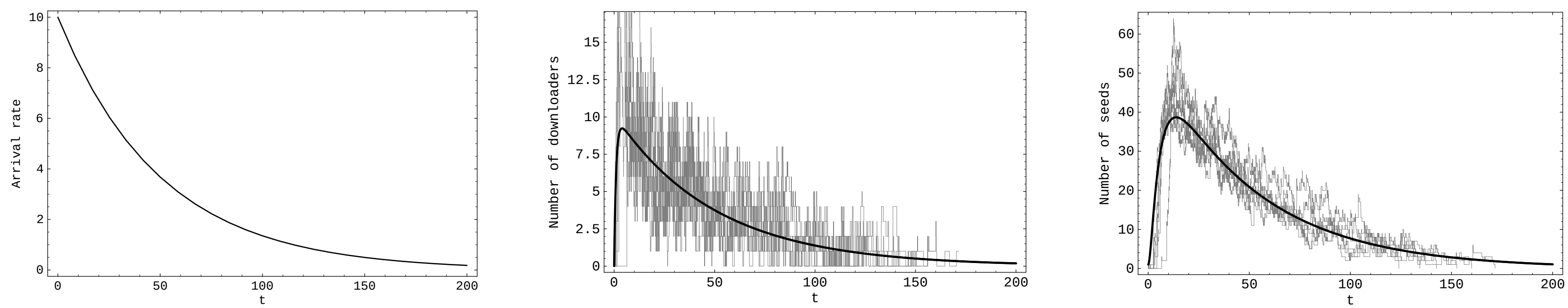
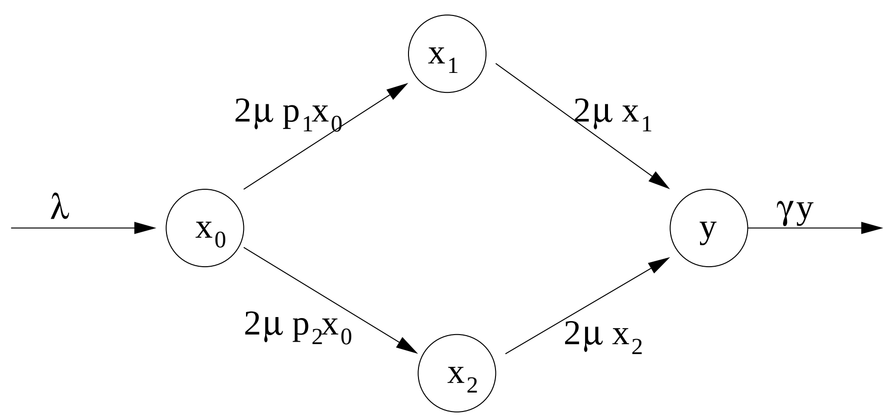


Figure 3: Arrival rate, number of downloaders and seeds as a function of time, when $\lambda_0 = 10$, $\tau = 50$, $\mu = 1$ and $\gamma = 1/5$. Solid line: fluid model, Gray lines: simulation.

Sharing of a file in multiple chunks

- Now the file of L bytes is divided into K chunks
  - When a peer has downloaded the first (random) chunk, it can upload it to other peers as a *leecher*.
  - Mean download time of a chunk is denoted by $1/(K\mu_d)$
  - If peer has all chunks it serves other peers as *seed*.
  - Seeds leave the system with rate γ, also original one.
  - If one of the chunks is missing, the system dies.

- Case $K = 2$
  - In the system we have four types of peers
    - $x_0(t)$ number of peers with no chunks
    - $x_1(t)$ number of peers with chunk 1
    - $x_2(t)$ number of peers with chunk 2
    - $y(t)$ number of peers with chunks 1 and 2



- The system can be described by a four-dimensional Markov chain, which state is quartet (x_0, x_1, x_2, y)
- States (x_0, x_1, x_2, y) with $y = 0 \cap (x_1 = 0 \cup x_2 = 0)$ are absorbing.
- The mean life can be calculated from the absorbtion times by a recursion
- $K > 2$: Only simulations

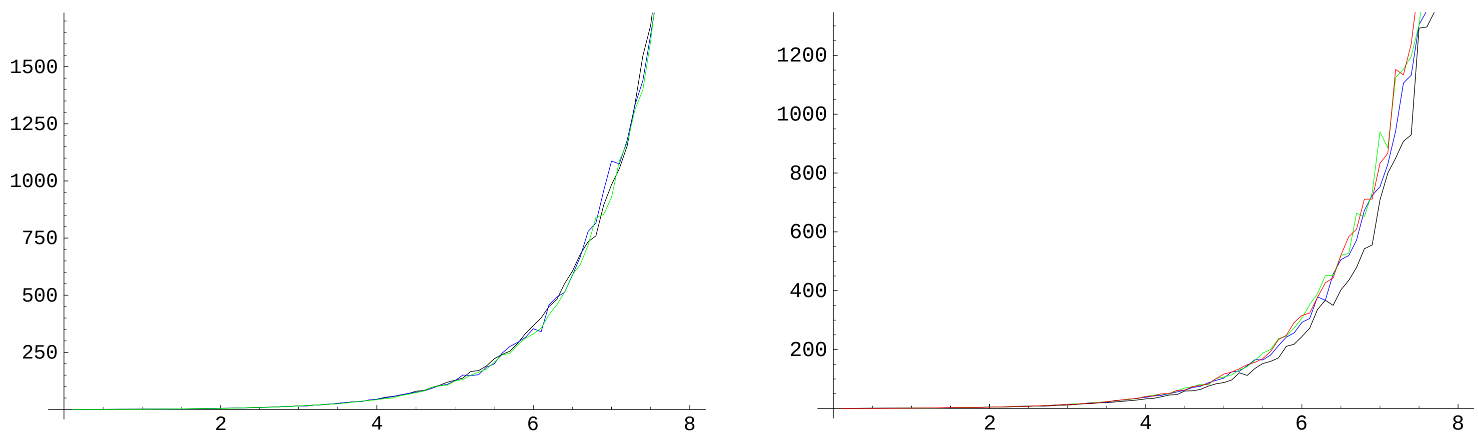


Figure 4: The mean life time of the system as a function of λ/γ. Left side: download constrained system, right side: capacity shared between downloaders.