Content Availability and Bundling in Swarming Systems

Presenter: Xuanjia Qiu
Apr. 14, 2010

- Best award paper in CoNEXT'09
- Reason of adopting it: a clever way of modeling the availability of P2P problems as ”busy period ” of a queue system

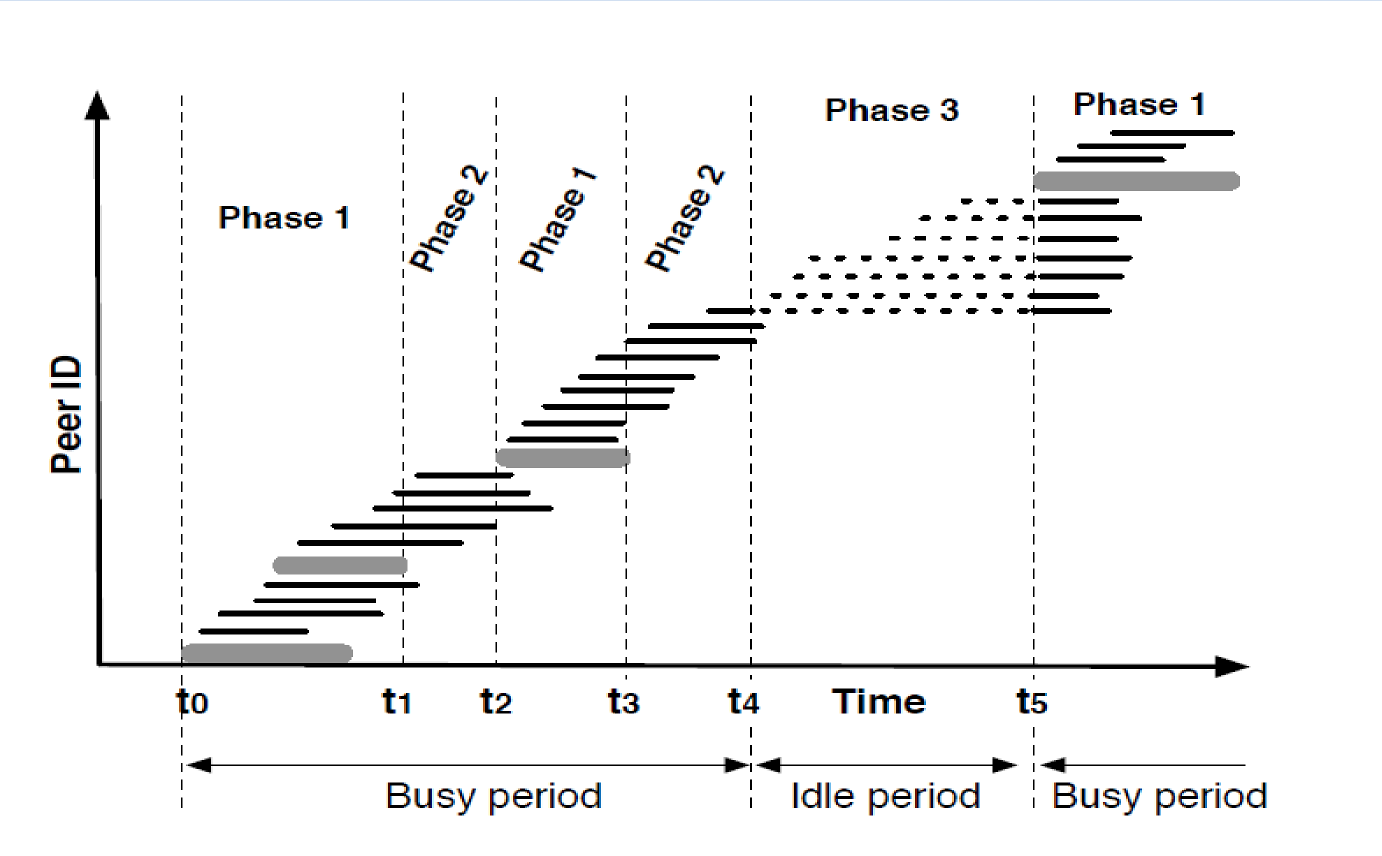
Queuing theory

- Queuing system
- $M/G/\infty$
- busy period

Outline

- Problem definition
- Model framework
- The evolution of the analysis of the model
- Some proof
- Conclusion

Problem definition

- Content availability
  - Publisher(seeds), Leecher(peer)
  - 

Bundling

- Combine several files together
- A user who is interested in one of the bundled files has to download all files
- Effects:
  - Increase the arrival rate of publishers
  - Increase the arrival rate of leechers
  - Increase the file size
  - Make the users stay longer
  - Benefits the unpopular swarms
  - Self-sustaining, may decrease download time

Problem definition

- Download time = waiting time + service time
- Availability → waiting time
- Availability → service time
- Who dominates? Need quantitive analysis

Model framework

- M/G/∞ model
  - Model the uninterrupted intervals during which the content is available as busy periods of the queue
  - In classic queuing models for P2P systems, ”server” is the peers which upload, and ”client” is the peers which download. But, here, the ”server” is ”nothing”, and the client is the stochasitic process which makes the content availability sustains.

Notation table

Variable	Description (units)
λ_k	peer arrival rate (1/s)
$\Lambda = \sum_{i=1}^{K} \lambda_k$	bundled peer arrival rate (1/s)
s_k	file size (bits)
$S = \sum_{i=1}^{K} s_k$	bundle size (bits)
μ	mean download rate of peers (bits/s)
r_k	arrival rate of publishers (1/s)
R	arrival rate of publishers for the bundle (1/s)
u_k	mean publisher residence time (s)
U	mean bundled publisher residence time (s)

Metric	Description (units)
P_k	unavailability
$\mathcal{P}$	unavailability of bundle
T_k	download time (s)
$\mathcal{T}$	bundle download time (s)

Simplest model

- Availability of an individual swarm
  - Avarage busy period
  -
  -
  -

$$E[B_k] = \frac{e^{r_k u_k} - 1}{r_k}$$

  - The prob. That a peer arrives to swarm k to find the content unavailable

$$P_k = \frac{1/r_k}{E[B_k] + 1/r_k}, \qquad k = 1, \ldots, K$$

Simplest model

- Availability of a bundled swarm
-
- $$E[\mathcal{B}] = \frac{e^{K^2 r u} - 1}{K r}$$
- $$\mathcal{P} = e^{-K^2 r u}$$
- We see
- $$-\log P_k = \Theta(1) \text{ and } -\log \mathcal{P} = \Theta(K^2)$$
- So,
- Bundling reduces content unavailability by a factor of $e^{-\Theta(K^2)}$

Download Time

- $$E[T] = \frac{s}{\mu} + \frac{1}{r}P$$
-
-
- When service time dominates, bundling can increase the download time by up to a factor K
- When wait time dominates, mean download time is decreased by a factor O(1/R) which grows unbounded as R->0

More detailed models

- The residence time of all other customers, X, takes the form of one of two exponentially distributed random variables, X1 or X2

$$E[B] = \theta + \sum_{i=1}^{\infty} \frac{\beta^i}{i!} \sum_{j=0}^{i} \binom{i}{j} \frac{q_1^j q_2^{i-j} \alpha_1^{1+j} \alpha_2^{1-j+i} \theta}{\alpha_1 \alpha_2 + j\theta\alpha_2 + \theta\alpha_1 i - \theta\alpha_1 j}$$

- Threshold
  - $B(n,m)$ Is the expected length of a residual busy period that begins with n leechers and ends as soon as the population size reaches m.
  -

LEMMA 3.3. *For all n,*

$$B(n,0) = \sum_{i=1}^{n} \frac{s}{i\mu} + \frac{s}{\mu} \sum_{i=1}^{\infty} \left(\frac{s\lambda}{\mu}\right)^{i} \frac{(n+i)! - n!i!}{i!(n+i)!i} \qquad (12)$$

For $m < n$, $B(n,m)$ is obtained using the recursion $B(n,m) = B(n,0) - B(m,0)$.

$$B(m) = \sum_{i=0}^{\infty} \frac{e^{-\frac{\lambda s}{\mu}} \left(\frac{\lambda s}{\mu}\right)^i}{i!} B(i, m)$$

THEOREM 3.3. *For a threshold coverage of m, the mean download time of a file when peers are patient is $s/\mu + P/r$ where*

$$P = \exp(-r(u + B(m))) \tag{14}$$

Evaluation

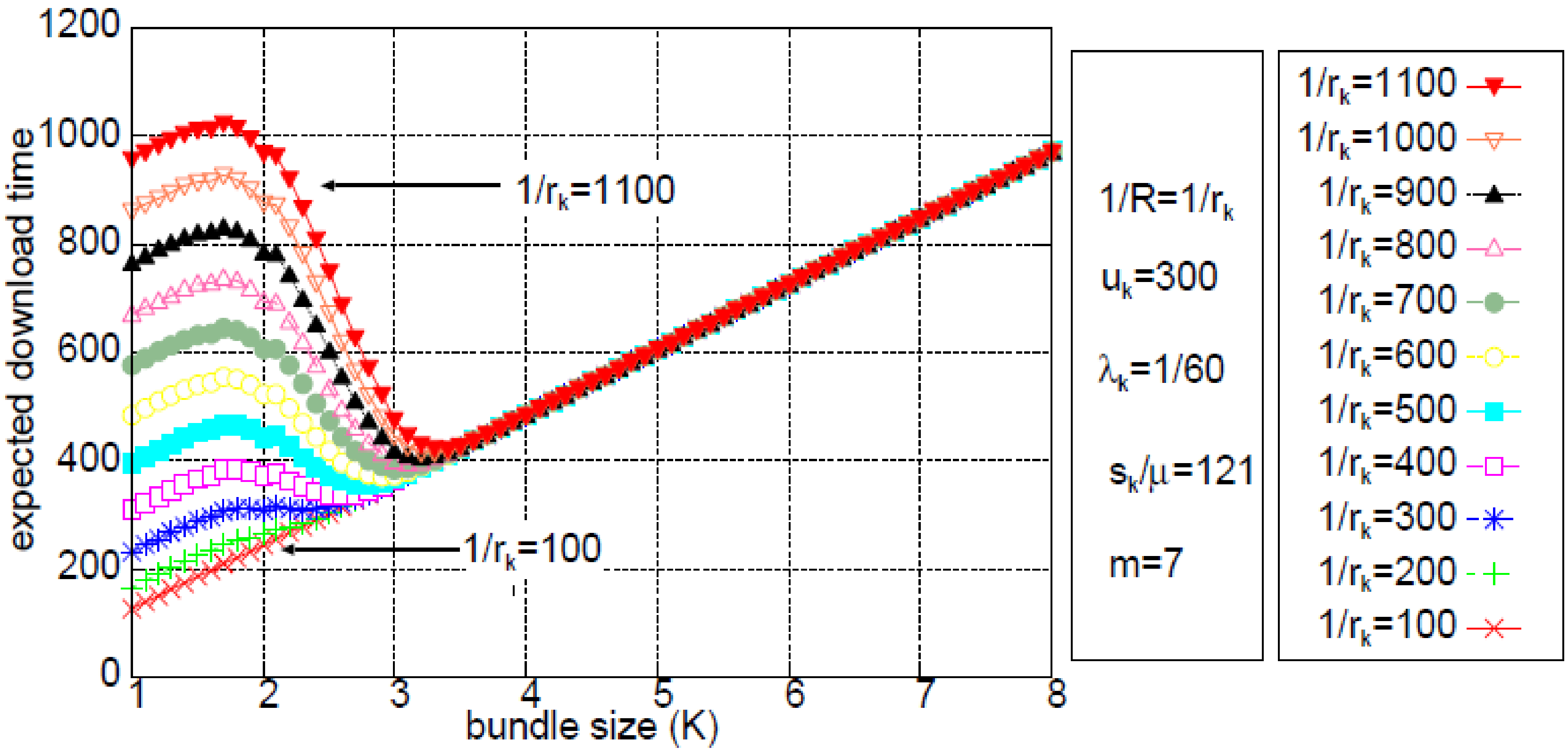


Figure 3: Bundles may reduce download time.

Evaluation

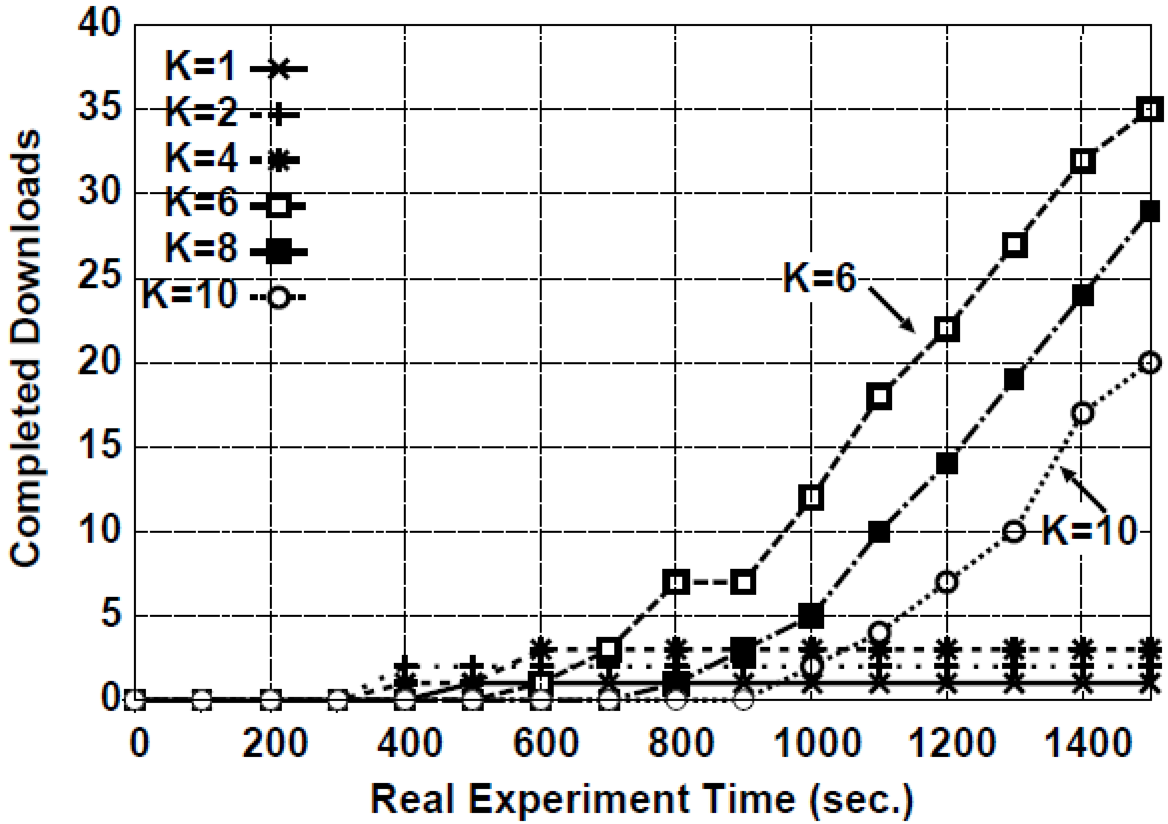


Figure 4: Availability of seedless swarms and the tradeoff in the choice of the bundle size.

Evaluation

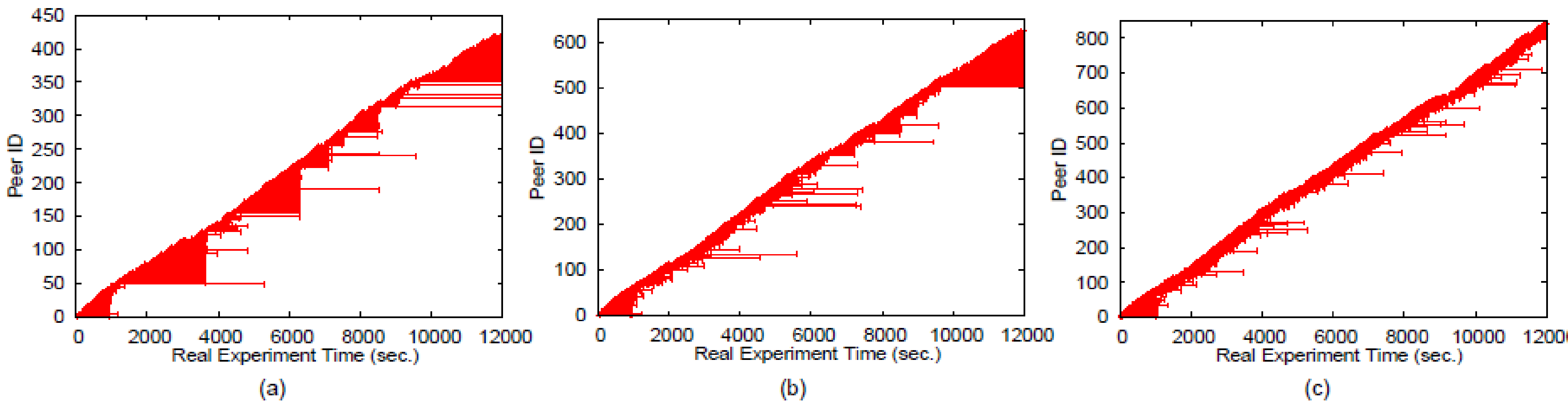


Figure 5: An intermittent publisher: (a) K=2; (b) K=3; (c) K=4. Each line starts when a peer arrives and ends when it leaves. As K increases, blocking probability decreases.

Evaluation

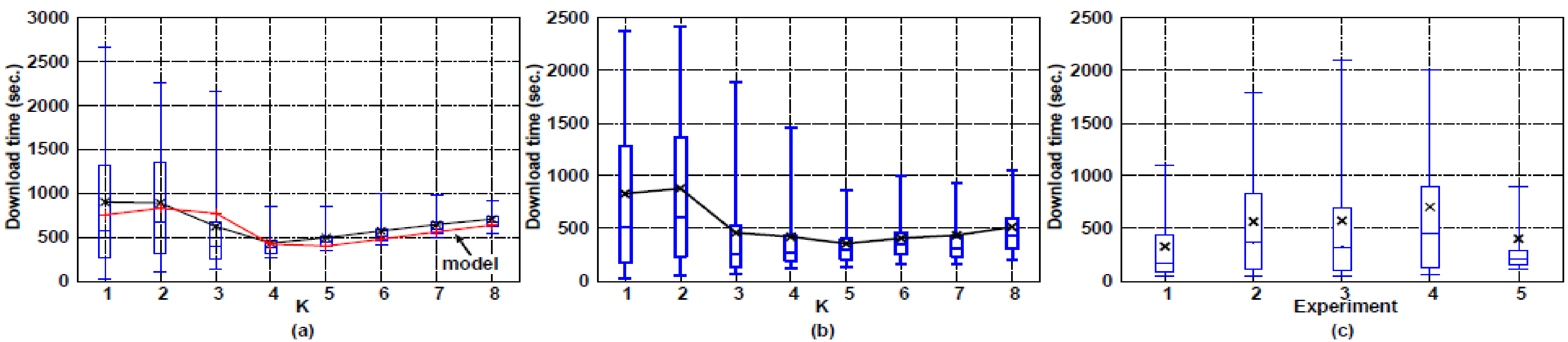


Figure 6: Download time versus bundling strategy. (a) exponential up and down times; (b) heterogeneous upload rates; (c) heterogeneous demand ($\lambda_i = \frac{1}{8i}$, $i = 1, \ldots, 4$), files bundled in experiment 5.

- Besides the mean, also consider the variance

Conclusiion

- Unsolved problems:
  - Side effect: increased traffic
  - When to apply bundling and when not to
  - May not hold if the mean arrival rate is not steady for a long enough duration of time
- Learn from this paper
  - Queuing models can be adapted to model problems that seem not to be able to modeled.
  - Progressive way of analysis