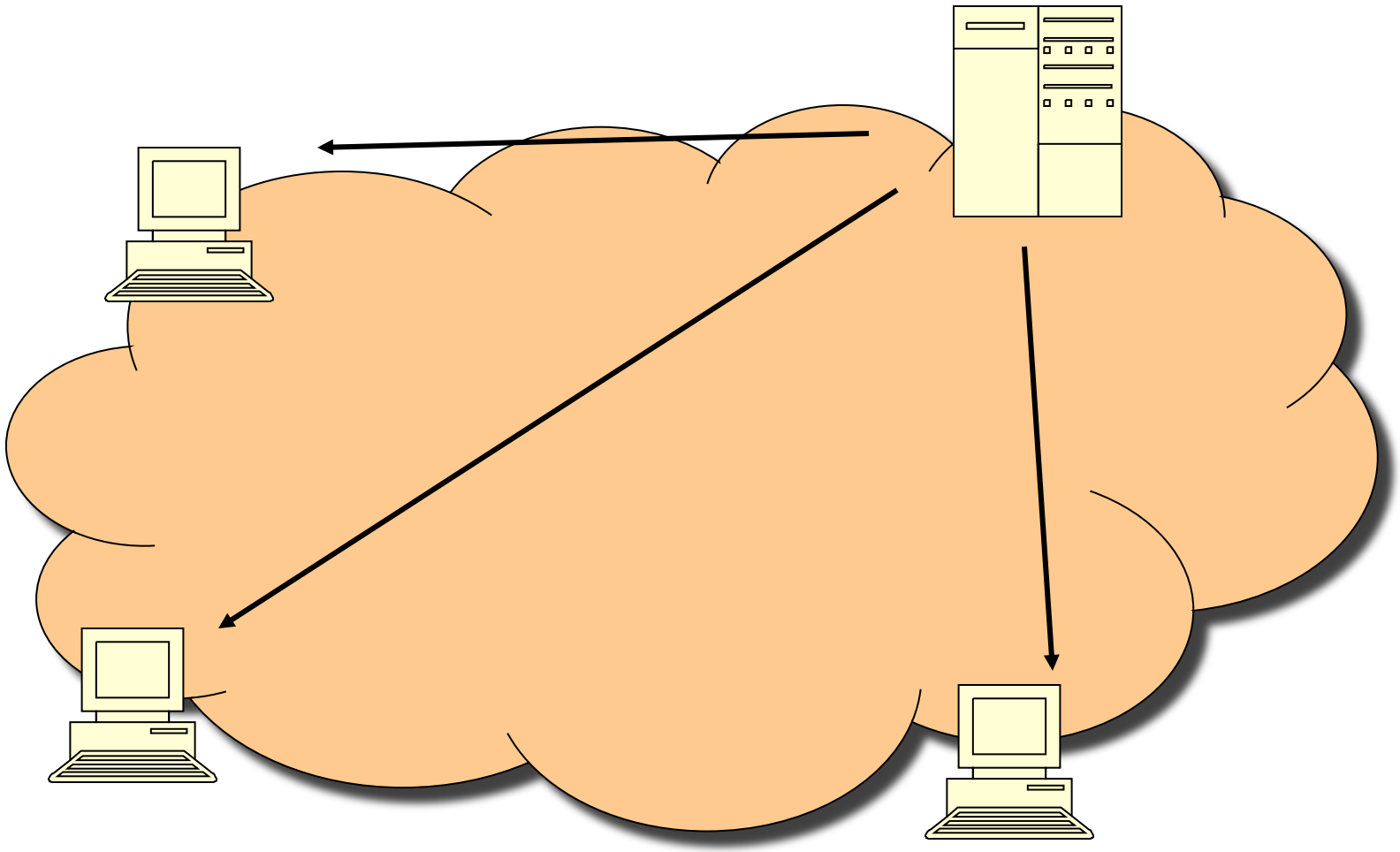
Collecting Coupons with Continuous Arrivals of Collectors

Outline of presentation

- Overview
- Fluid analysis (after Massoulie and Vojnovic)
- Direct stochastic analysis (work with Ji Zhu)
- Discussion

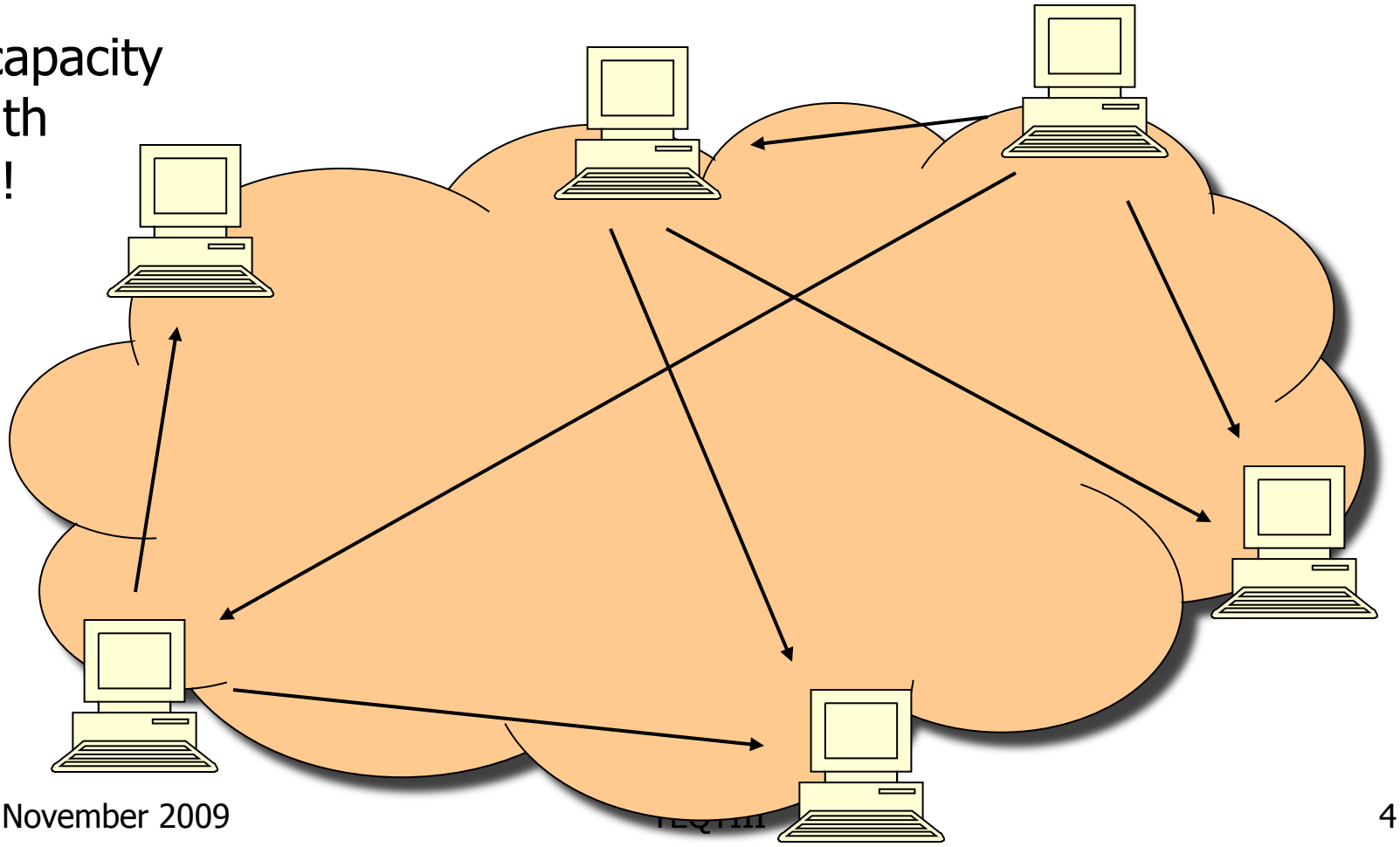
Traditional File Service



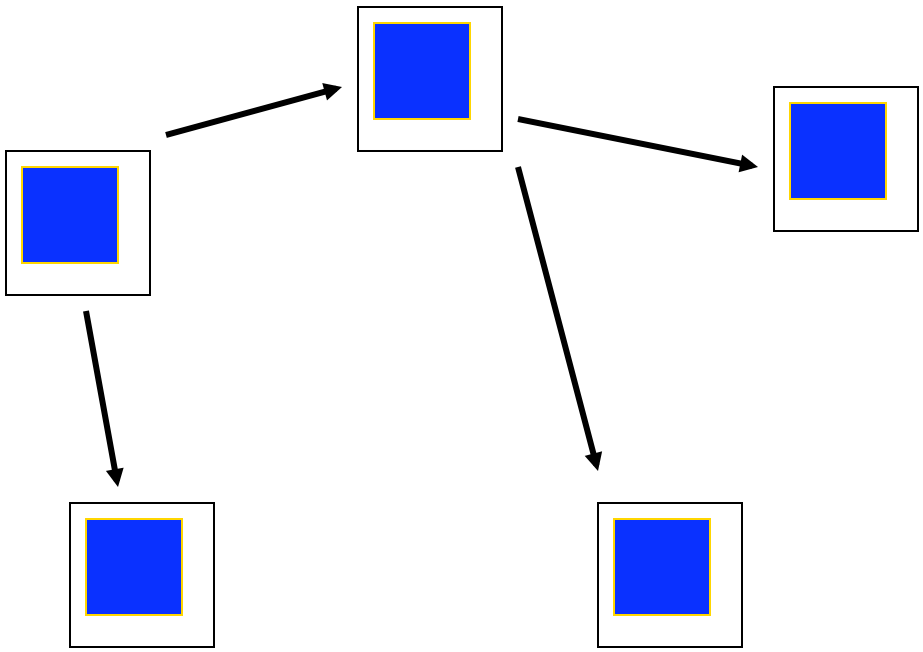
Peer-to-peer File Service

Service capacity scales with demand !

Scalable
Robust



1st Generation P2P Systems



e.g. Gnutella, KaZaa

Entire file uploaded in one shot

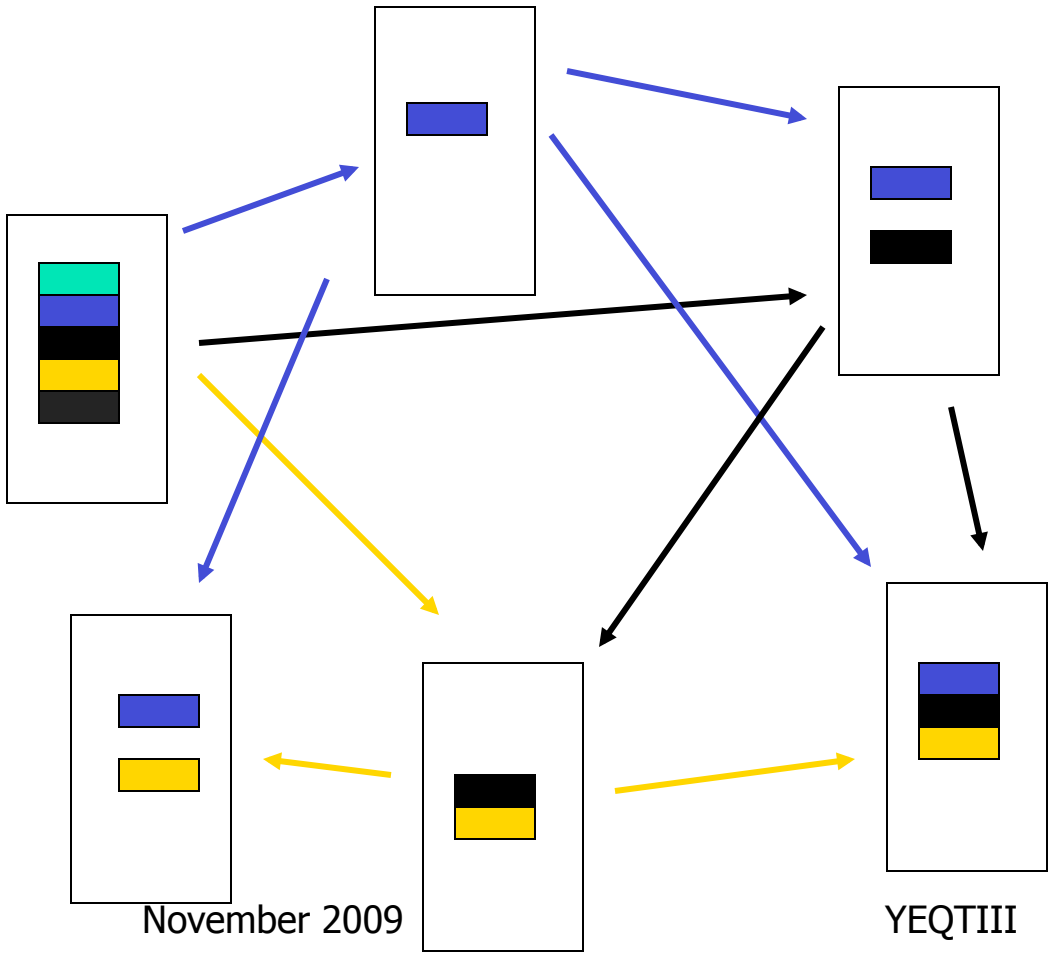
Users wait till entire file is received before relaying

BUT if file size is large,

1. Long delay before users become useful.

2. If users depart mid-way, a lot of upload time wasted.

2nd Generation P2P Systems



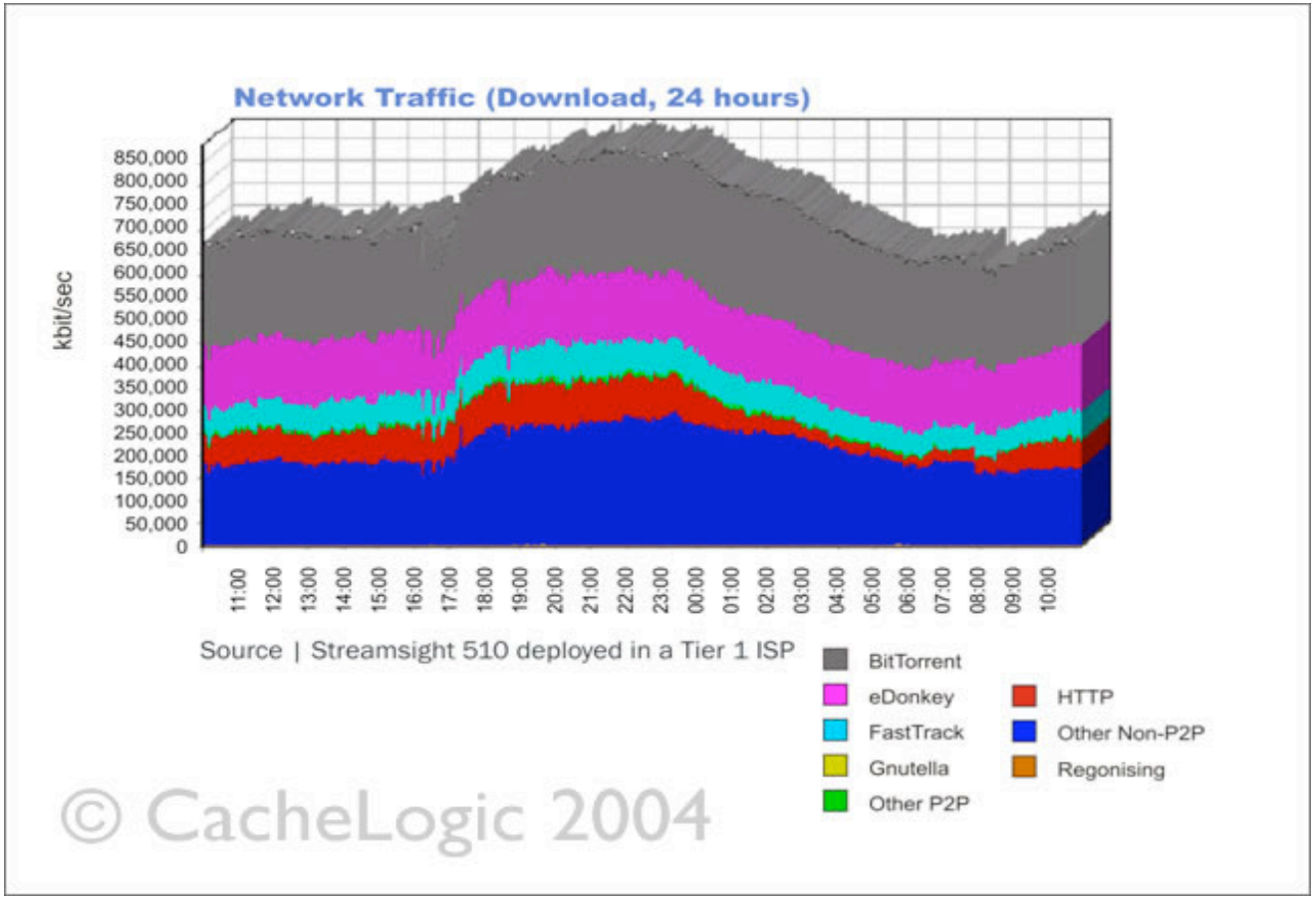
File split into pieces

Peers can start serving pieces without waiting for entire file.

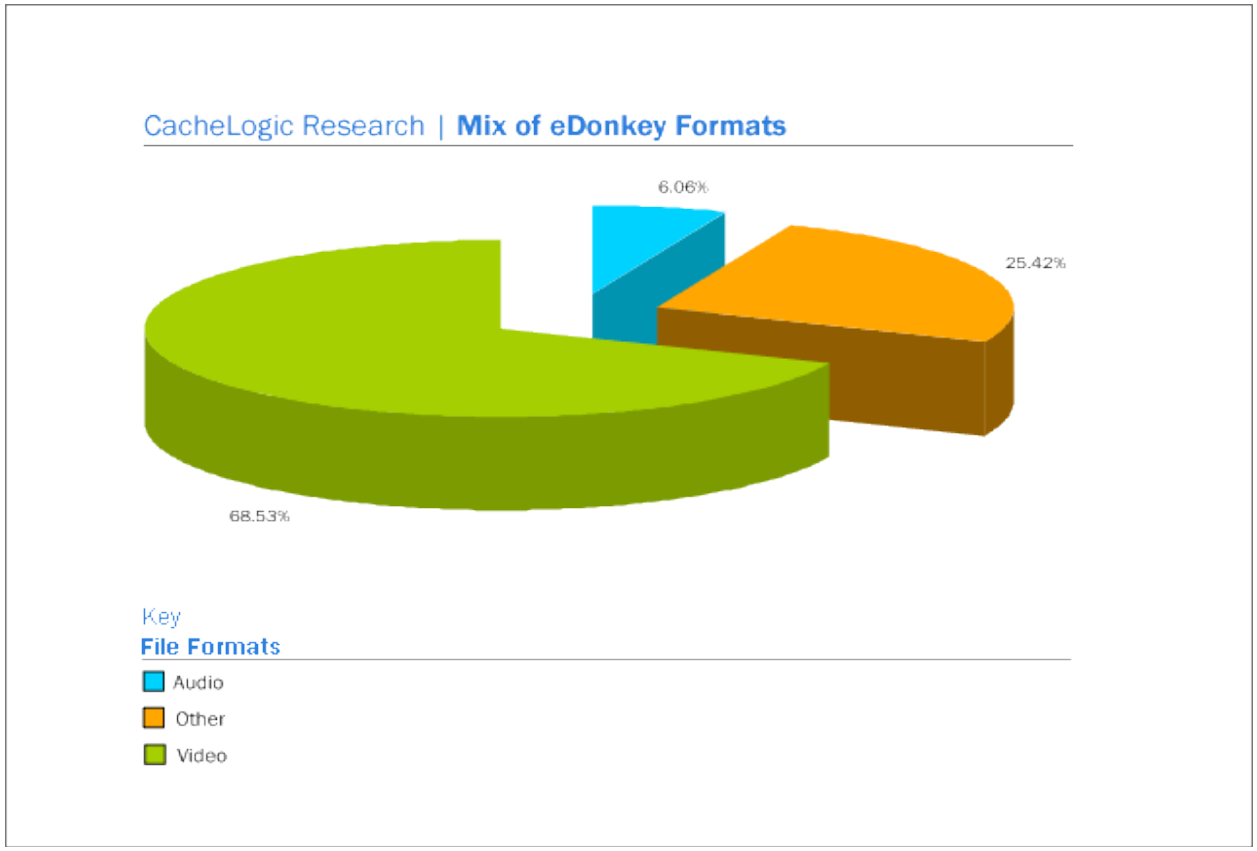
More effective at serving large files.

e.g. BitTorrent, eDonkey

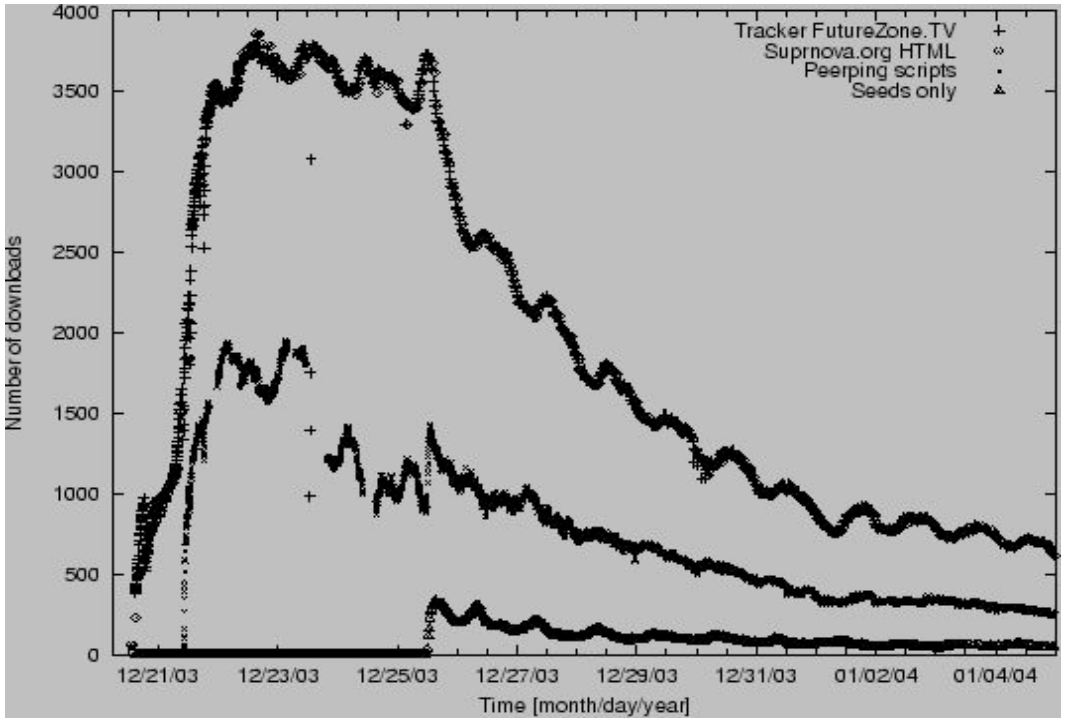
Relative Popularity



File Types in eDonkey



Flash Crowds



Users requesting LoTR III on BitTorrent

Flash Crowd: large number of users start requesting *same* file *simultaneously*

Scenario where P2P system design tested most severely.

Our Motivation

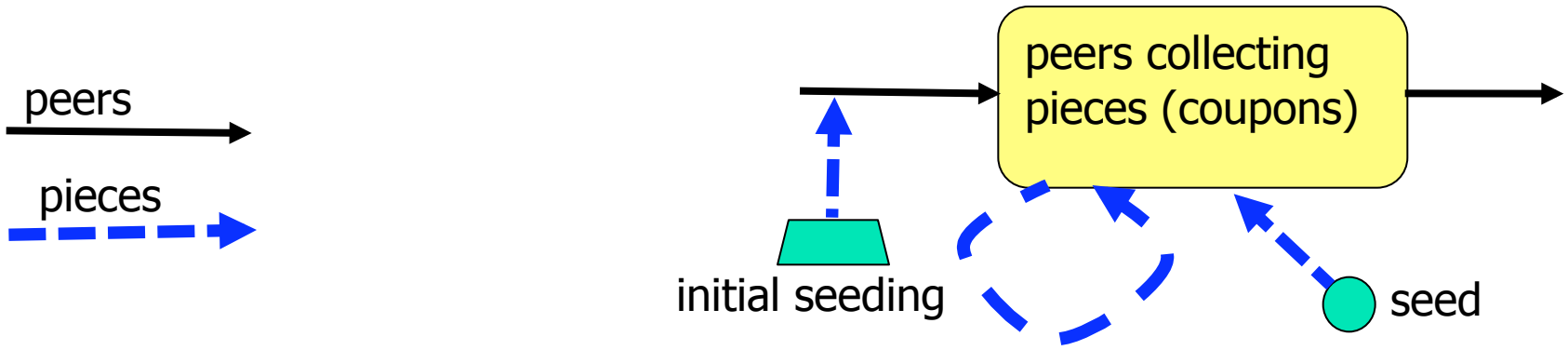
To understand second-generation peer-to-peer systems

Quantitatively study how splitting a file speeds its dissemination to large numbers of simultaneous users

Use understanding to provide design principles/algorithms

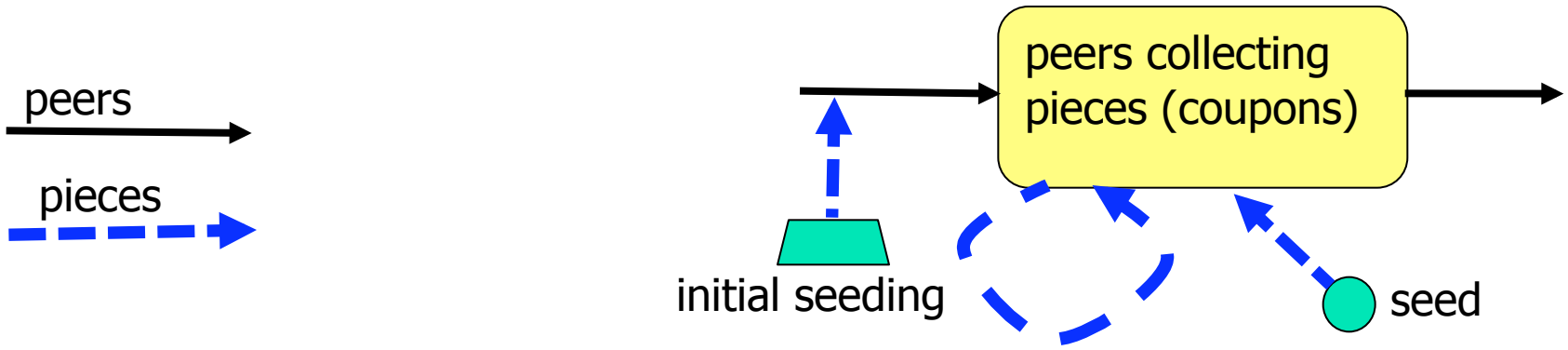
Develop decentralized piece selection algorithms which can give good speedup

Next few slides closely follow Massoulié and Vojnovíc, *Coupon replication systems*, 2006/2008



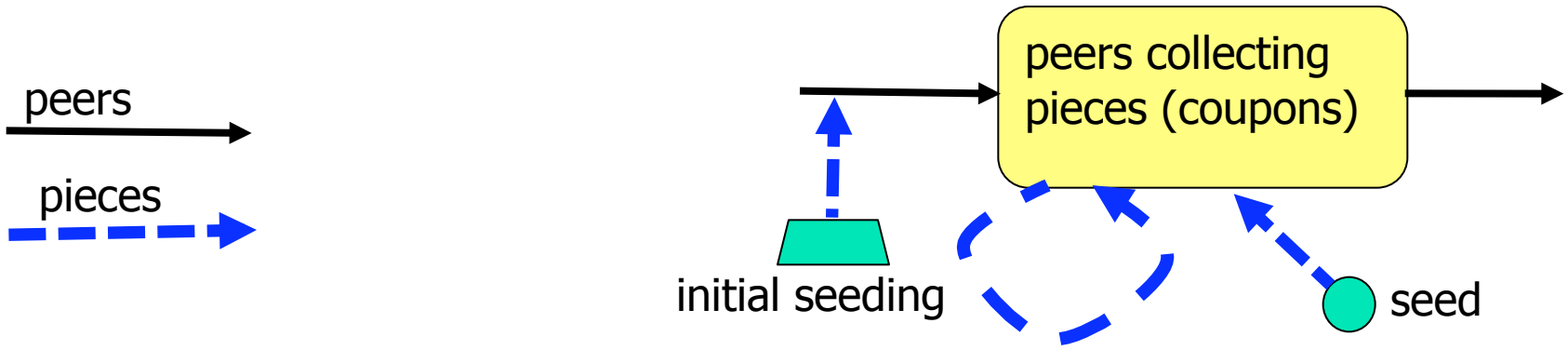
Markov model: (fixed seed/uniform contacts/random useful piece)

- $\mathcal{C}$ = set of strict subsets of $\{1, \ldots, K\}$, where K is number of pieces
- A peer with set of pieces c is a *type c peer*
- A type c peer becomes a type $c \cup \{i\}$ peer if it downloads piece $i \notin c$.
- Downloads are modeled as being instantaneous
- Detailed Markov state: $\mathbf{x} = (x_c : c \in \mathcal{C})$, with x_c=number of type c peers
- Exogenous arrivals of type c per Poisson(λ_c) process, for $c \in \mathcal{C}$
- random, uniform contacts



Markov model (continued):

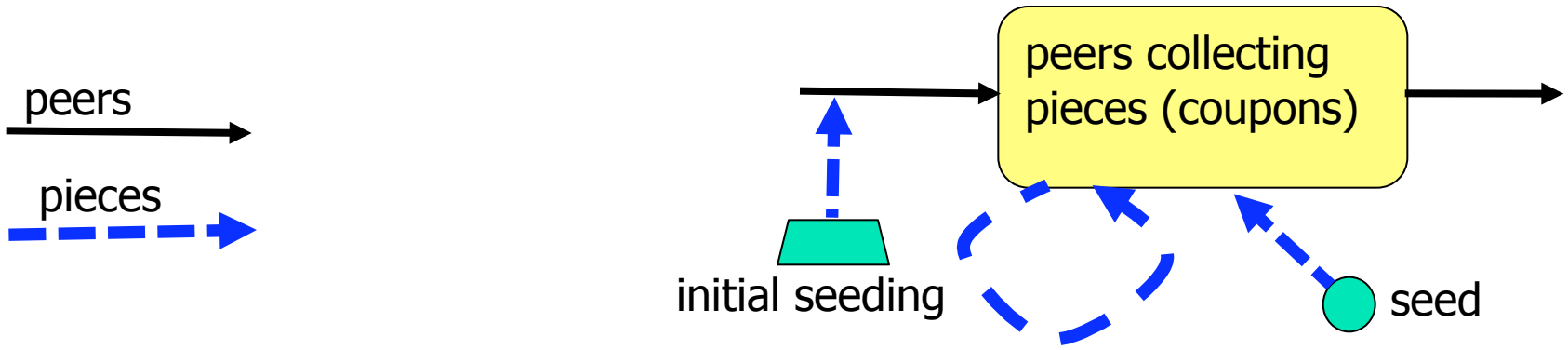
- Opportunities to upload to (download from) another peer occur at rate μ
- Random useful piece selection
- Peers operate among themselves in push mode or in pull mode – math is the same here)
- There is one seed, pushes pieces to peers at total rate U_s
- Peers depart upon completing a collection



Positive diagonal entries of $Q = (q(\mathbf{x}, \mathbf{x}') : \mathbf{x}, \mathbf{x}' \in \mathcal{C})$ are given by:

$$\mathbf{x} \to \quad \mathbf{x} + \mathbf{e}_c \quad \text{with rate } \lambda_c$$

$$\mathbf{x} \to \quad \mathbf{x} - \mathbf{e}_c + \mathbf{e}_{c+i} I_{\{c+i \in \mathcal{C}\}} \quad \text{with rate } \frac{x_c \left(\frac{U_s}{K - |c|} + \mu \sum_{s: i \in s} \frac{x_s}{|s - c|} \right)}{|\mathbf{x}|}$$

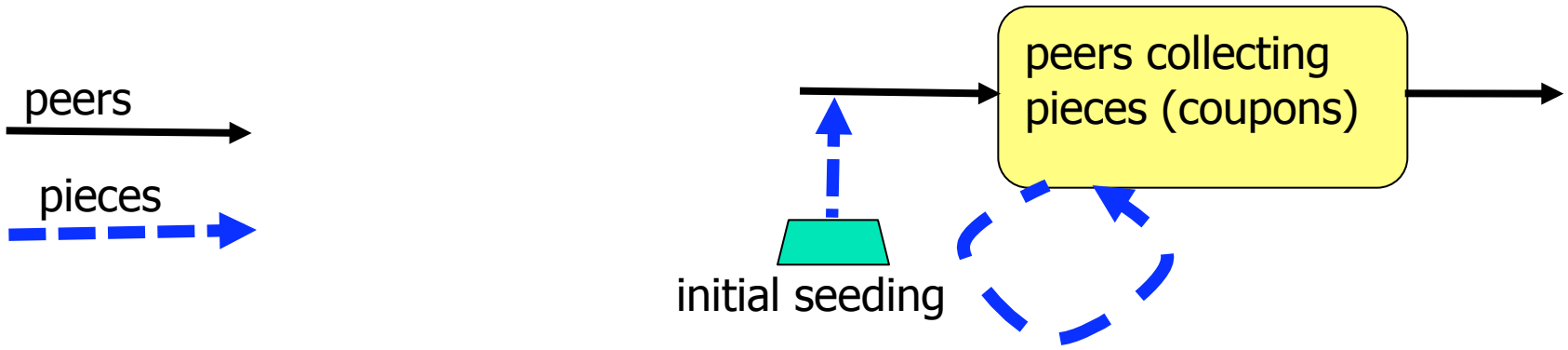


Fluid limit (Massoulié and Vojnovíc 2006/2008 applying Kurtz's theorem on density-dependent jump MPs)
Index processes by $N \to \infty$.
$\mathbf{X}^N$ has arrival vector $\lambda_c^N = N\lambda_c$
No seed: $U_s^N = 0$
If $N^{-1}\mathbf{X}^N \to \mathbf{x}(0)$, then

$$N^{-1}\mathbf{X}^N \to \mathbf{x}, \quad \text{in p. uniformly on bounded intervals}$$

where

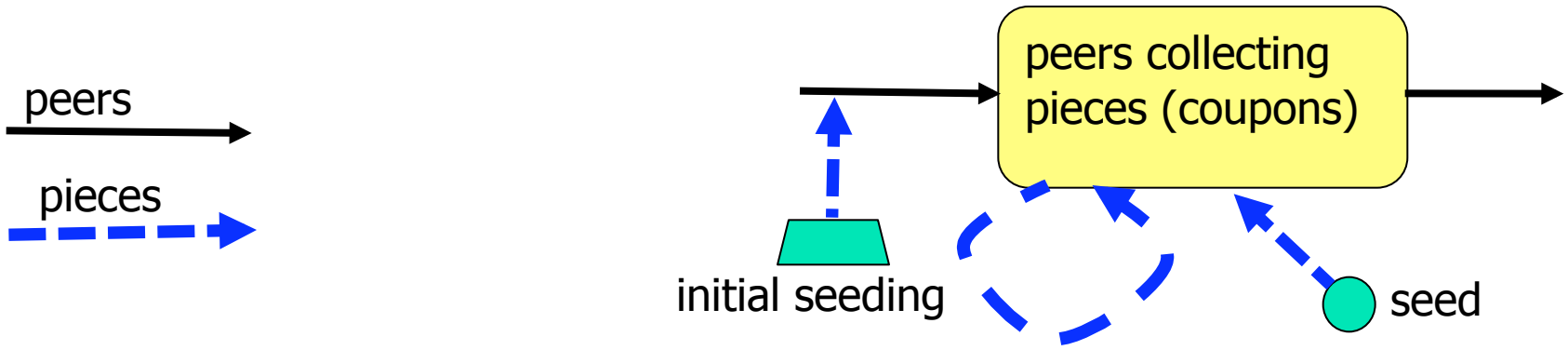
$$\dot{\mathbf{x}} = Q\mathbf{x}, \quad (\text{for } \mathbf{x}(0) \text{ given})$$



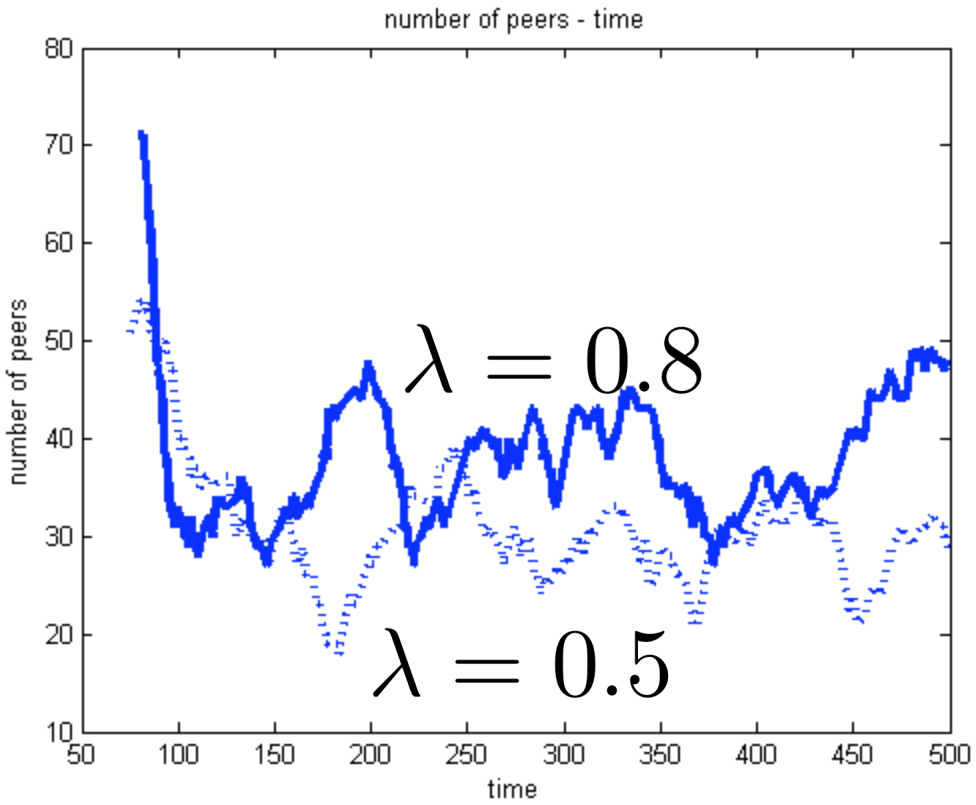
M&V analyzed the ODE for symmetric, one piece upon entry model
M&V identified resting point, and conjectured global asymptotic stability

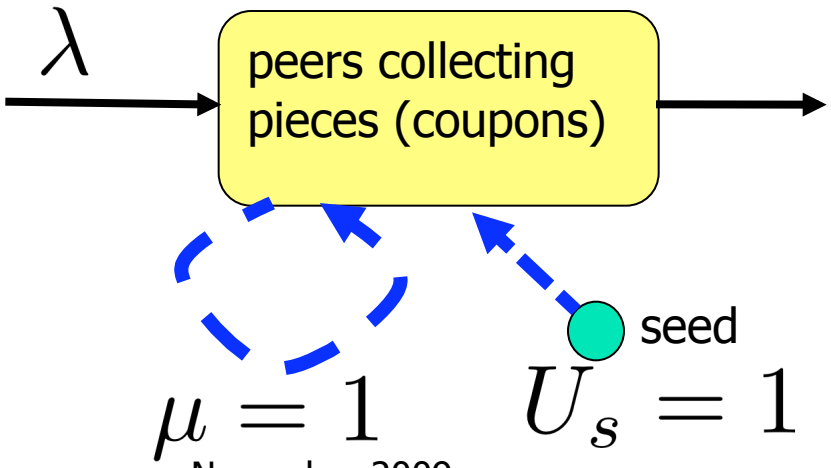
At the resting point, sojourn times in stages satisfy
$1 < T_1 < T_2 < \cdots < T_{K-1} < 2.$
Refines earlier 2-state model of Qui and Srikant (2004) and
Yang and de Vecianna (2004) which uses assumption
$T_1 = T_2 = \cdots = T_{K-1}.$

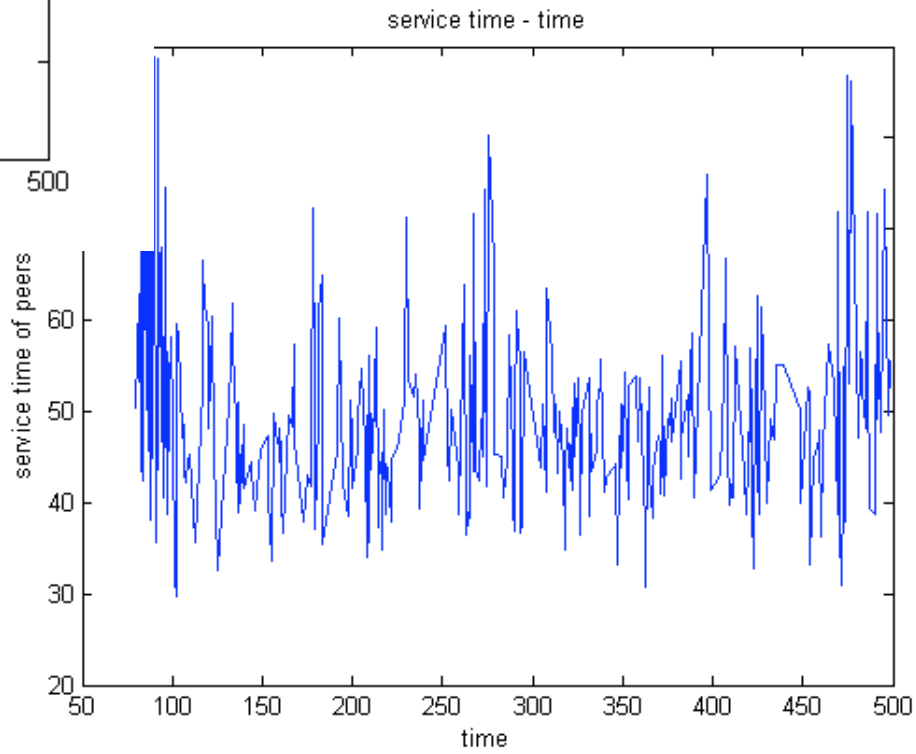
M&V results show earlier results not too far off.

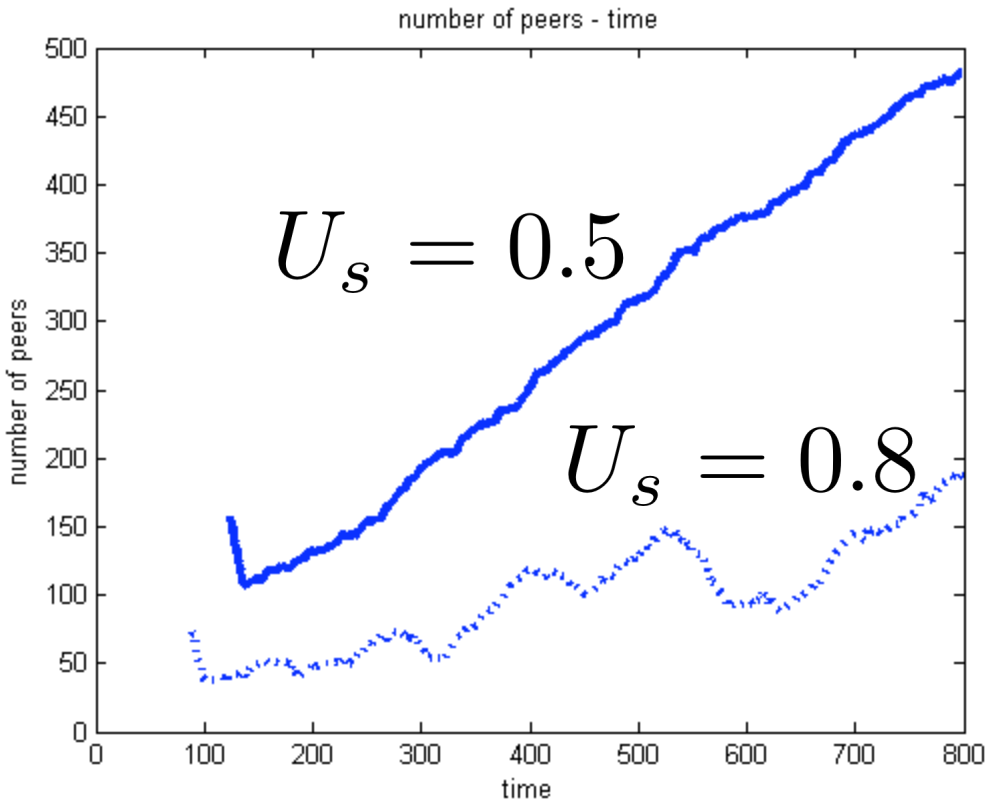


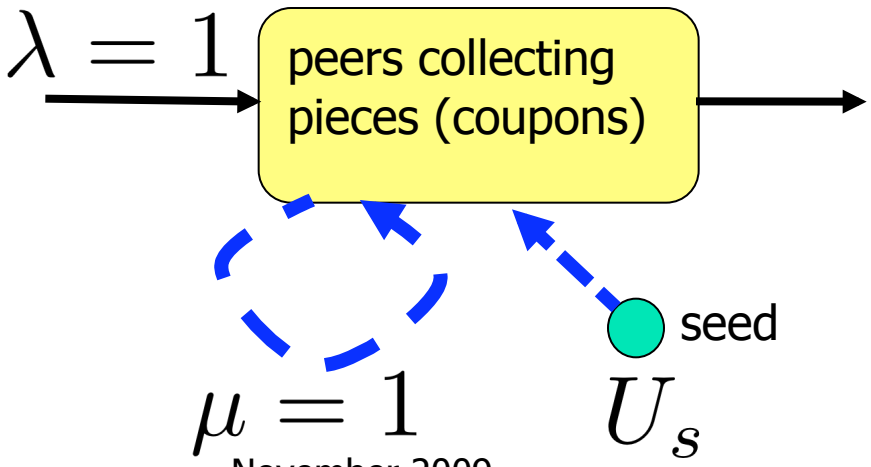
However, our stochastic, packet level simulations of this model showed poor performance. We focus here on a similar scenario. Next slides show simulations for no pieces upon entry but seed rate $U_s > 0$.

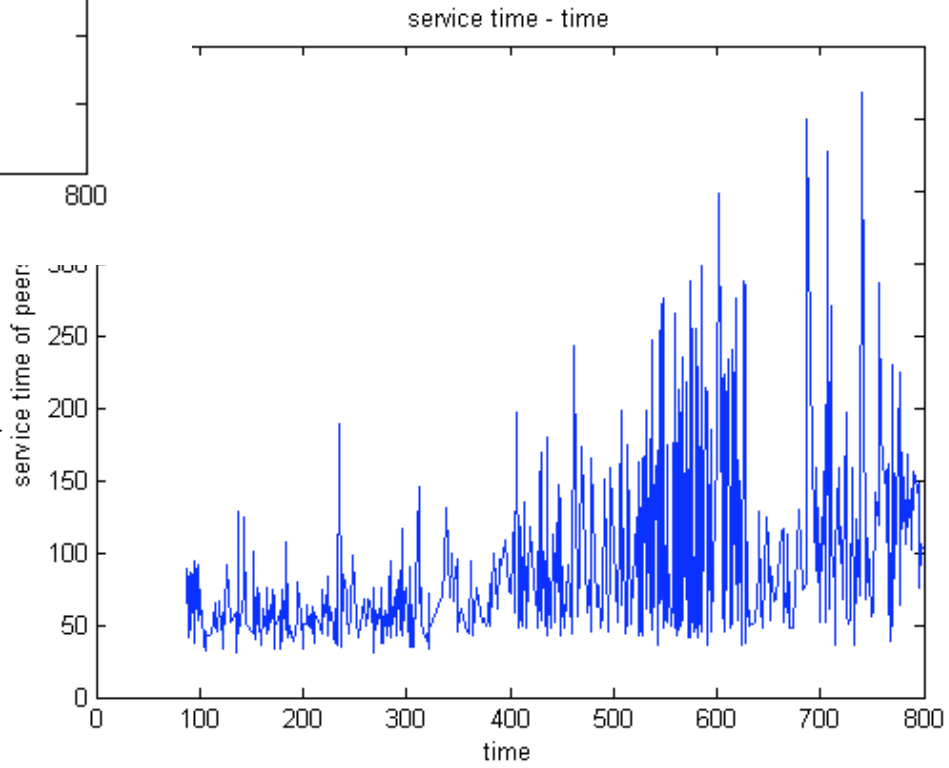

number of peers - time
number of peers
time
$\lambda = 0.8$
$\lambda = 0.5$


λ
peers collecting pieces (coupons)
seed
$\mu = 1$
$U_s = 1$

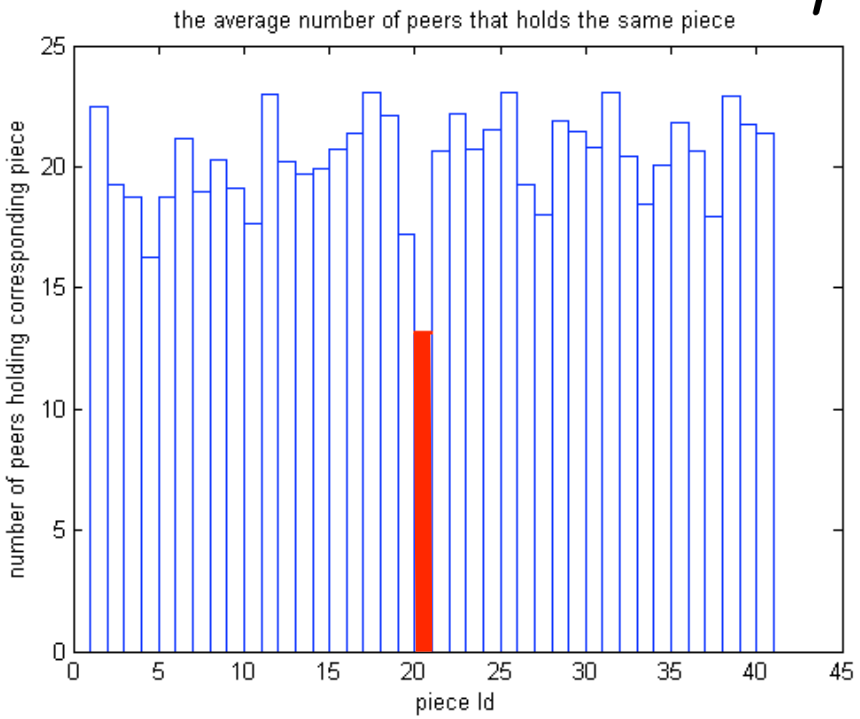

service time - time
service time of peers
time


number of peers - time
number of peers
time
$U_s = 0.5$
$U_s = 0.8$

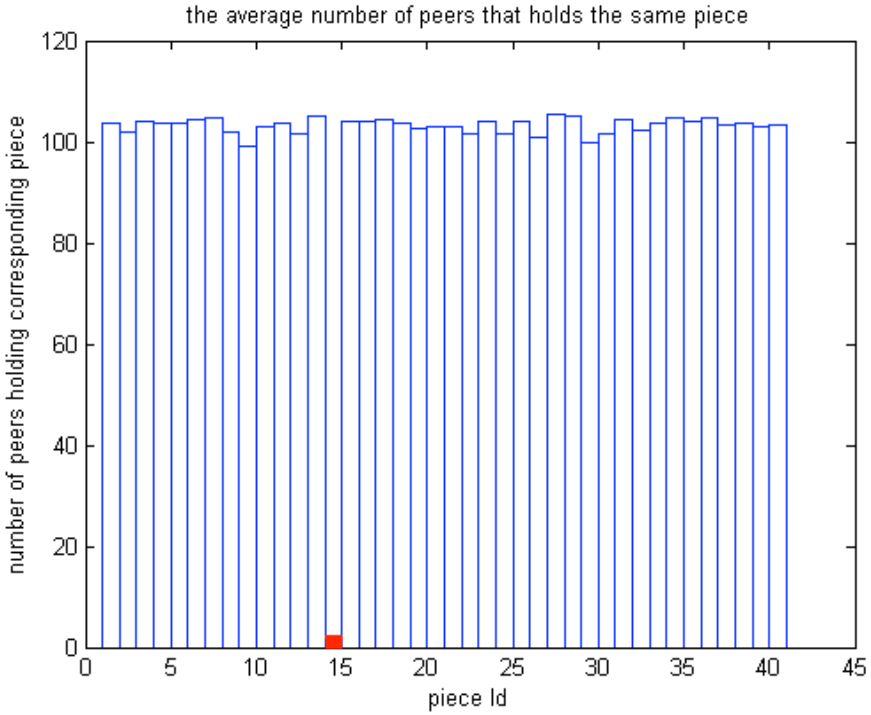

$\lambda = 1$
peers collecting pieces (coupons)
seed
$\mu = 1$
U_s


service time - time
service time of peer
time

$$\mu = 1$$

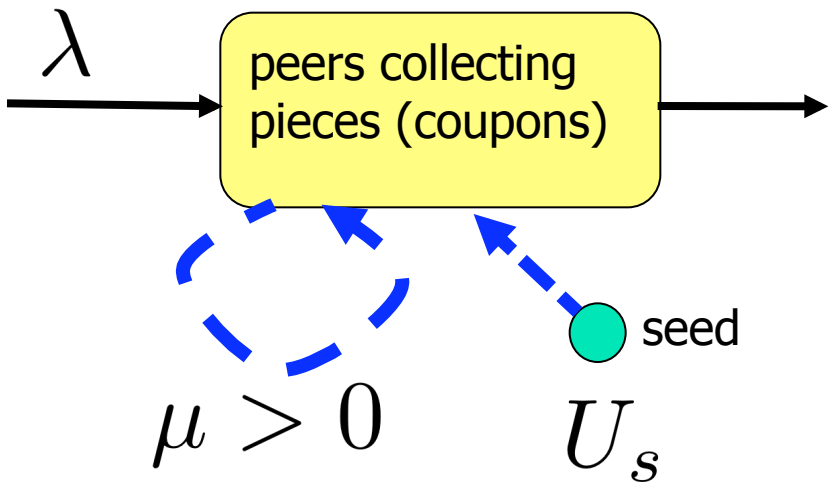


$$\lambda = 0.8$$
$$U_s = 1$$



$$\lambda = 1$$
$$U_s = 0.8$$

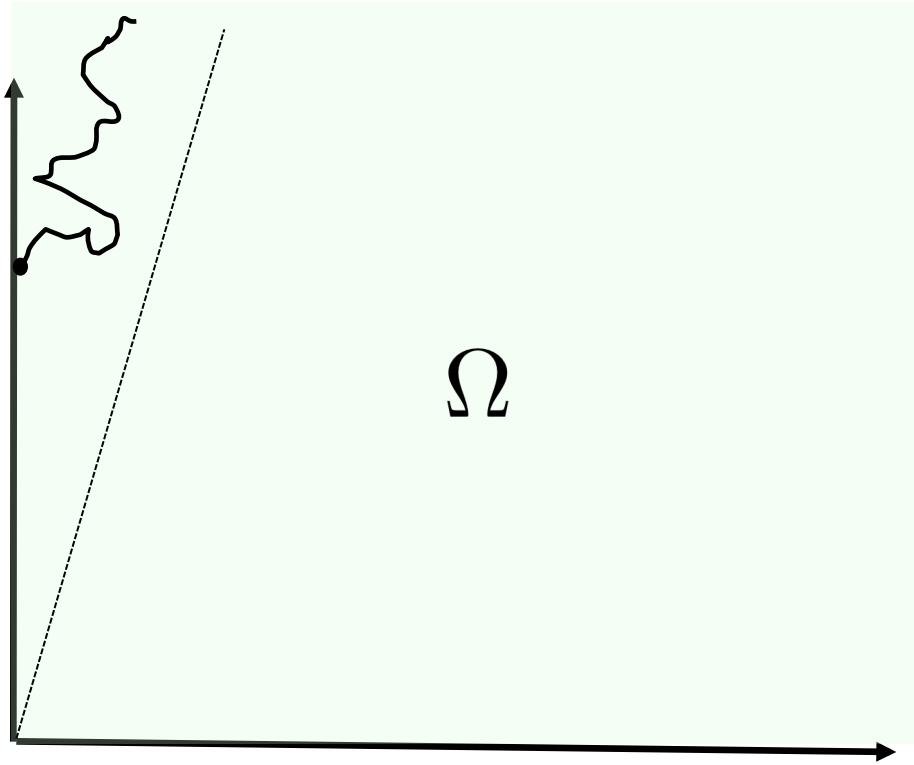
Proposition (Ji Zhu and H., forthcoming)
Suppose $\lambda_c = 0$ for $c \neq \emptyset$ and $\lambda_\emptyset = \lambda$.
The process is positive recurrent if $\lambda < U_s$ and transient if $\lambda > U_s$.



Next: outline of proof.

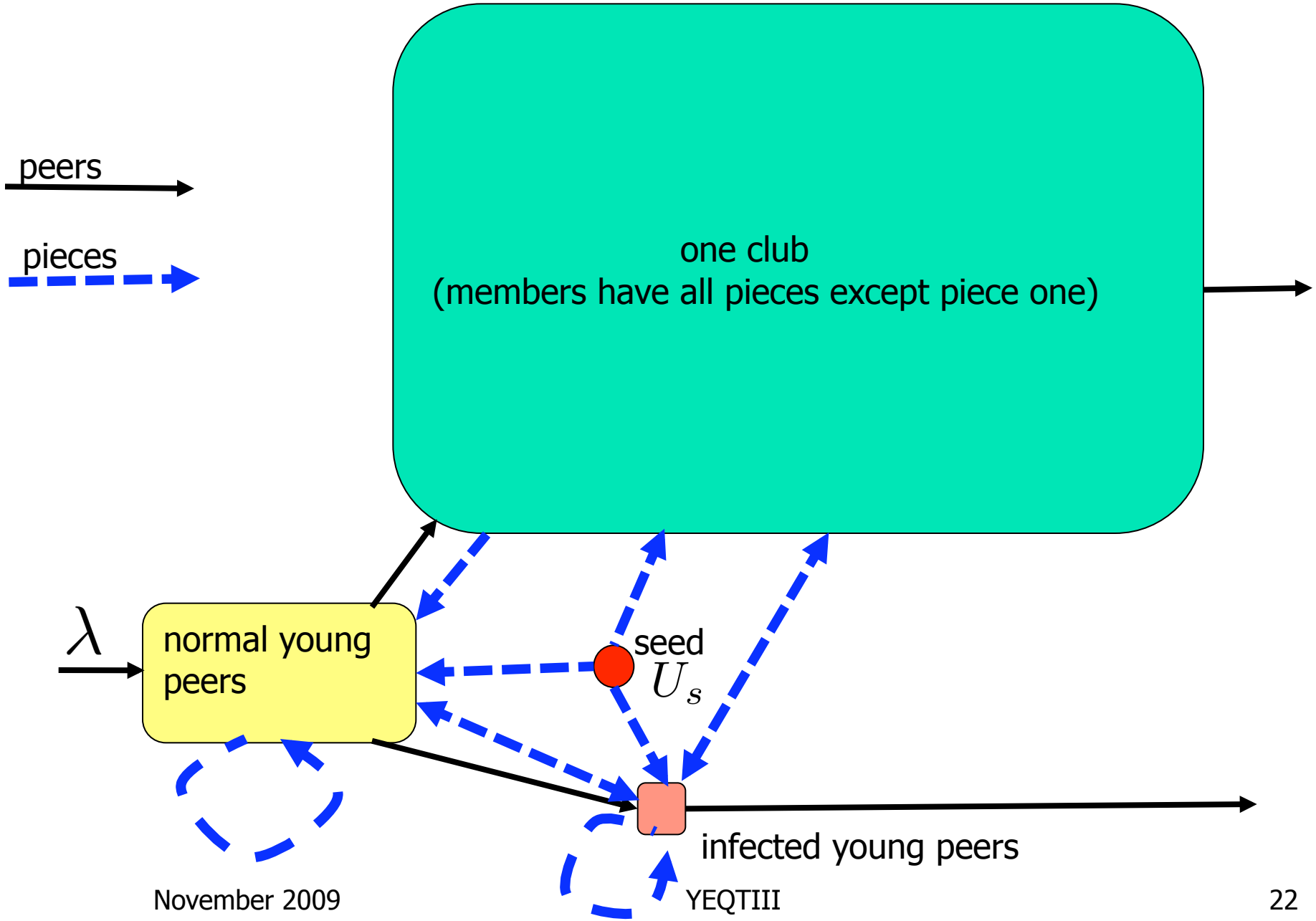
Proof of transience if $\lambda > U_S$

Select initial state with many peers, all in the *one club*. That is, having every piece except for piece one.

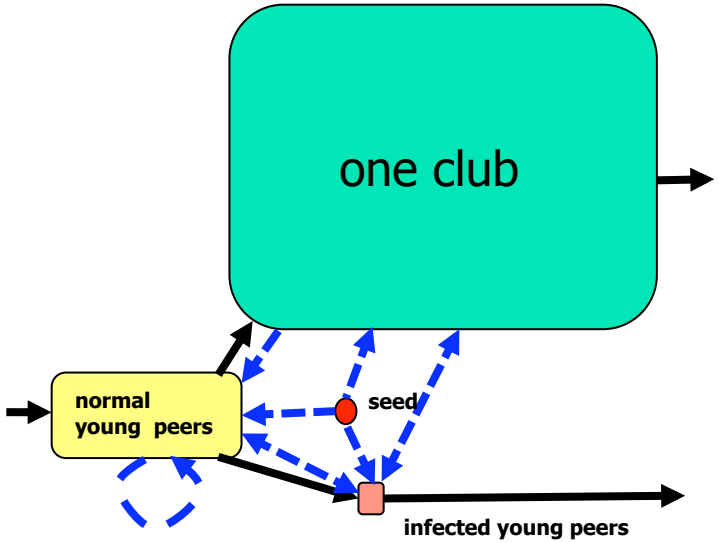


Localization:
Assume wlog that process restarts at
$T = \inf\{t : \text{fraction of peers in one club} \leq 1 - \xi\}$

Proof of transience if $\lambda > U_S$ (continued)

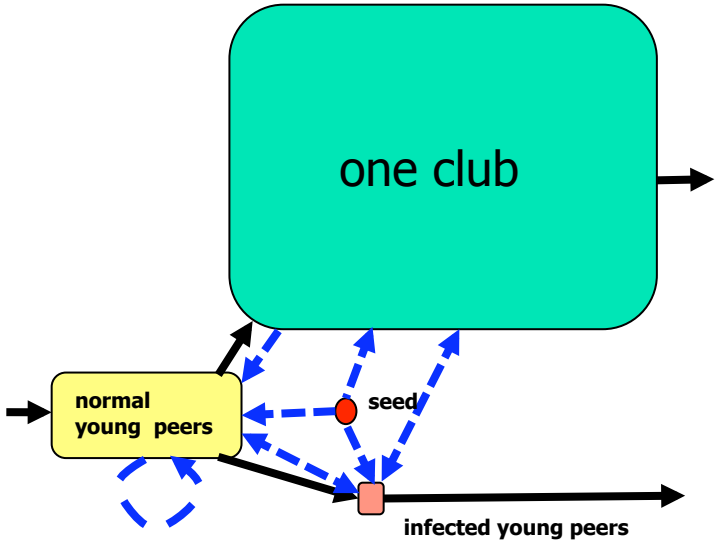


Proof of transience if $\lambda > U_S$ (continued)

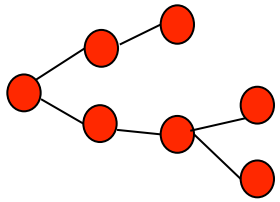


$$(\text{young peers}) \prec (M/GI/\infty \text{ queueing system})$$
$$\text{with } GI \leftrightarrow \text{Gamma}(K, \mu(1-\xi)) \text{ distribution}$$

Proof of transience if $\lambda > U_S$ (continued)



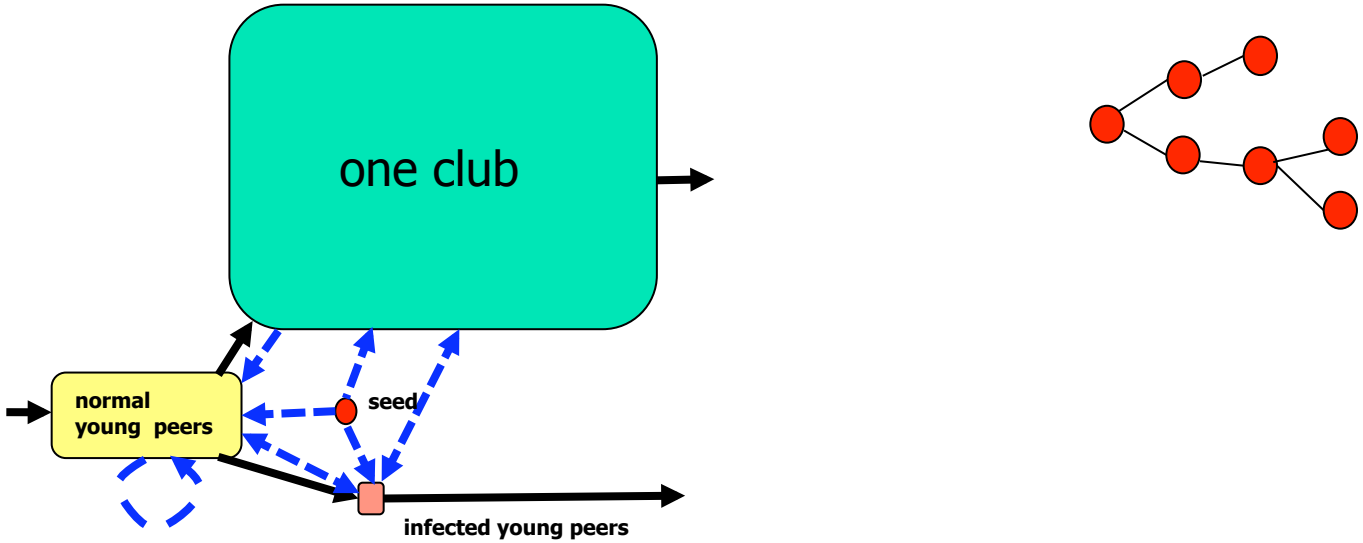
If seed creates a new infected peer, that peer can infect others.



The sum of the times the infected peers are in the system is stochastically smaller than a busy period of an $M/GI/1$ system with arrival rate $\mu\xi$ and same GI distribution.

Branching process is highly subcritical for small ξ.
The total number of one club members given piece one by the peers in a batch has finite mean and variance.

Proof of transience if $\lambda > U_S$ (continued)

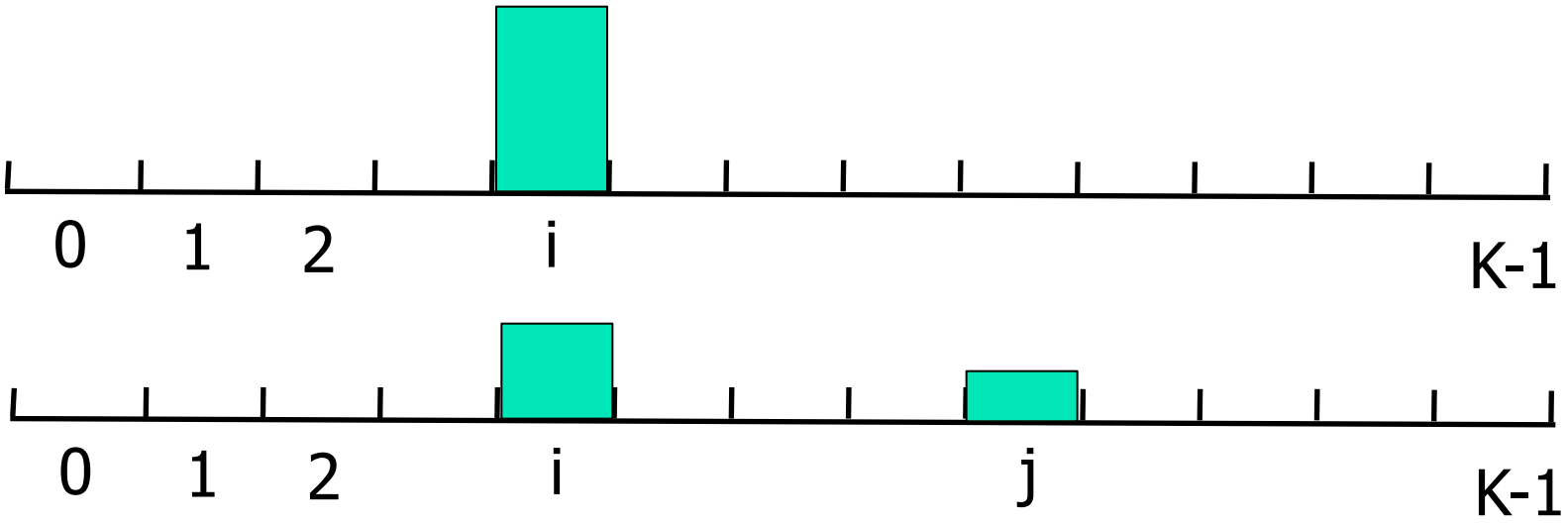


Uploads of piece one up to time t
bounded by Poisson process of rate U_s
(direct from seed) plus compound Poisson
process with batch rate $U_s\xi$.
Mean rate can be less than λ
Number of peers in system grow but number
of normal young peers is constrained.
So system avoids restart with large probability.

Proof of positive recurrence if $\lambda < U_S$

Idea: Seek Lyapunov function and use Foster-Lyapunov criterion.
Need $QV(\boldsymbol{x}) \leq -\epsilon$ or $QV(\boldsymbol{x}) \leq -c|\boldsymbol{x}|$ for $|\boldsymbol{x}|$ large

Perhaps a quadratic function?
Just depending on numbers of pieces peers have?

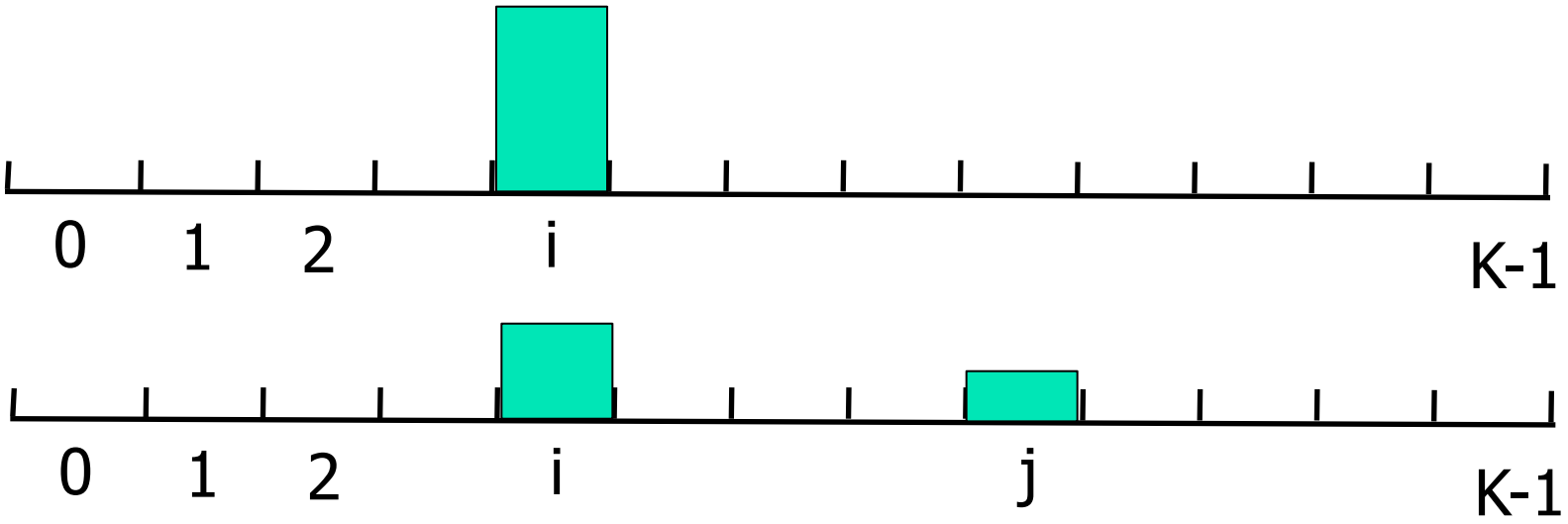


Proof of positive recurrence if $\lambda < U_S$ (continued)

Following works:

$$V(\mathbf{x}) = \sum_{i=1}^{M} c_i \times (\text{number of peers with } i \text{ or fewer pieces})^2$$

For an appropriate choice of $c_1 > c_2 > \cdots > c_{M-1}$ can show $QV(\boldsymbol{x}) \leq \epsilon|\boldsymbol{x}|$ for $|\boldsymbol{x}|$ large enough.

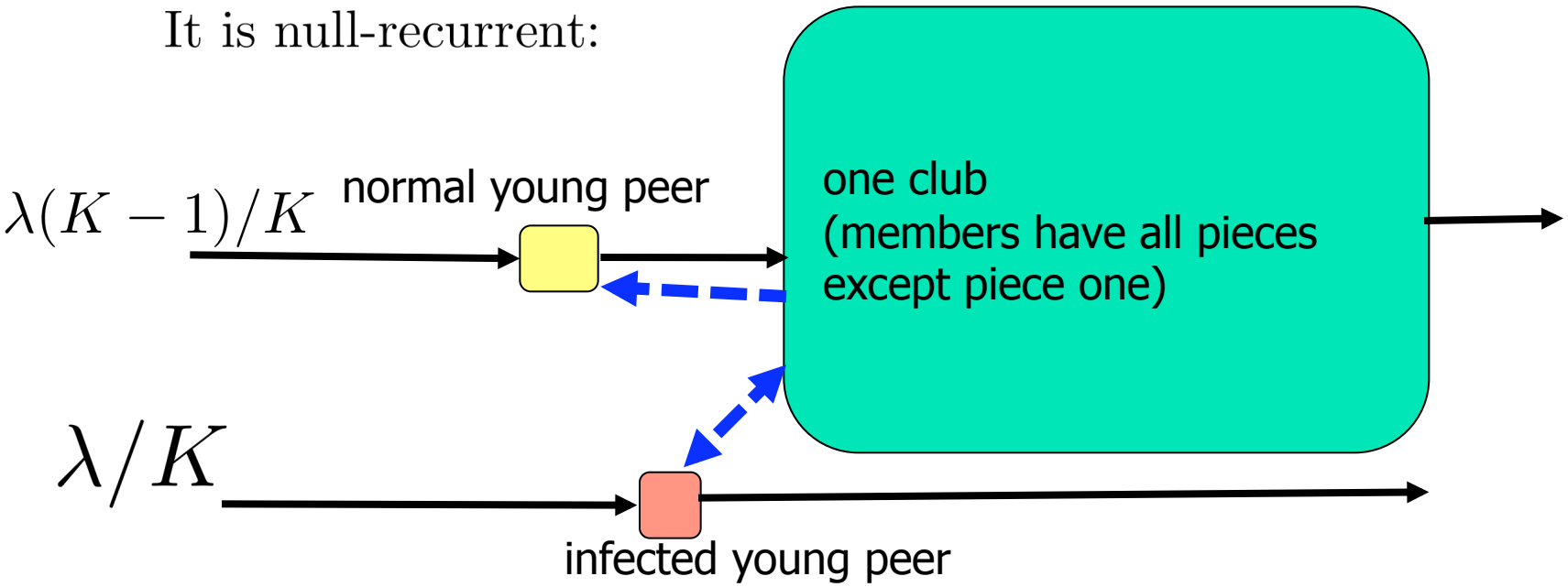


Discussion

One conclusion of results: Order of limits makes a difference.
M&V take $N \to \infty$ first. Then consider equilibria in time.
Perhaps M&V analysis is OK for flash crowds (large N, short time).

Discussion (continued)

Formation of a one club hampers the exact stochastic model studied by M&V. In fact, consider giving a random piece to each arriving peer. Consider the limiting form of the model obtained by letting $\mu \to \infty$. It is null-recurrent:



Infected peer uploads and downloads are rate μ
So on average, it causes $K-1$ peers to leave one club while collecting the other $K-1$ pieces. So on slow time scale, size of one club follows mean zero random walk.

Discussion (continued)

We expect the one club problem to show up as additional rest points or limits of the ODE equations of M&V, but nonsymmetric states must be considered even if model is symmetric.
Take intial state to correspond to large one club.

Note that a seed giving one piece per new peer does the same work on average as a seed constantly uploading at the critical rate $U_s = \lambda$ for the stochasic model we studied. In both cases, perhaps the seed at some time or another has to make up for the fact that new peers with no pieces would be downloading for some time before being able to upload.

Discussion (continued)

Three mechanisms are implemented or commonly discussed for enhancing P2P systems:

- Rarest first piece selection (rather than random useful)
- Network coding (peers and seed exchange linear combinations of original packets)
- Some peers remain in system for some time after completing collection

Q: Which one(s) of the above three defeats the one club problem?

A. Only the third one.

Thanks!