Backlog-Based Random Access in Wireless Networks

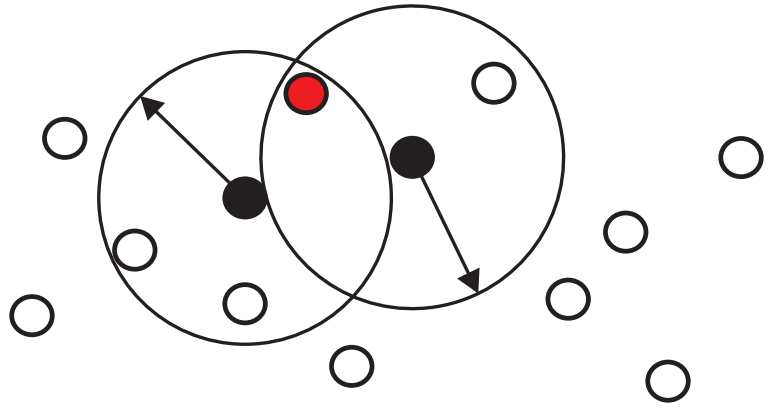
Johan van Leeuwaarden

Eindhoven University of Technology and EURANDOM

joint work with
Niek Bouman, Sem Borst, Alexandre Proutiere

workshop in honour of Frank Kelly

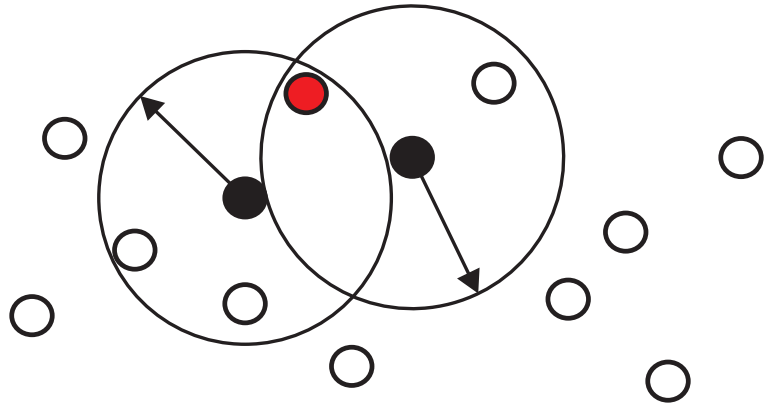
Large-scale wireless networks



- covering large areas, huge numbers of nodes
- centralized control is infeasible
- nodes operate autonomously, and share medium in distributed fashion

Nodes do not just use the network, they are the network

Large-scale wireless networks



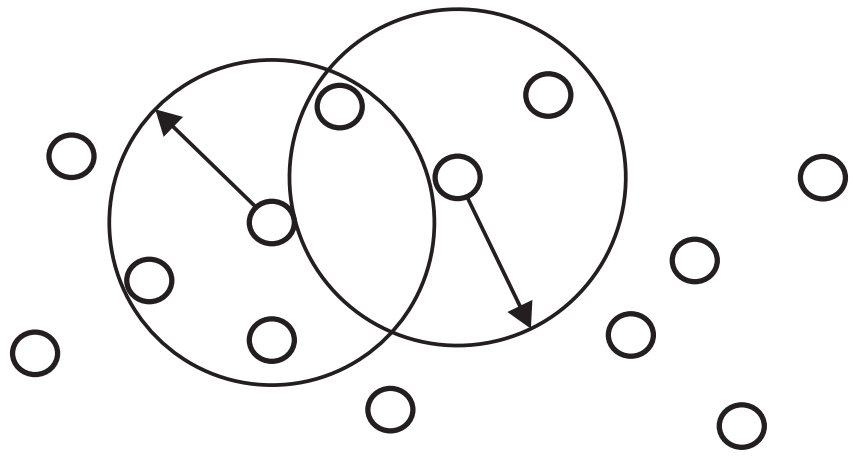
Randomized algorithms provide popular mechanism for distributed medium-access control

CSMA (Carrier-Sense Multiple-Access) protocol

- nodes sense their surroundings for ongoing transmissions
- a node will activate only if all nearby nodes are silent
- low implementation complexity, but highly complex behavior on macroscopic level

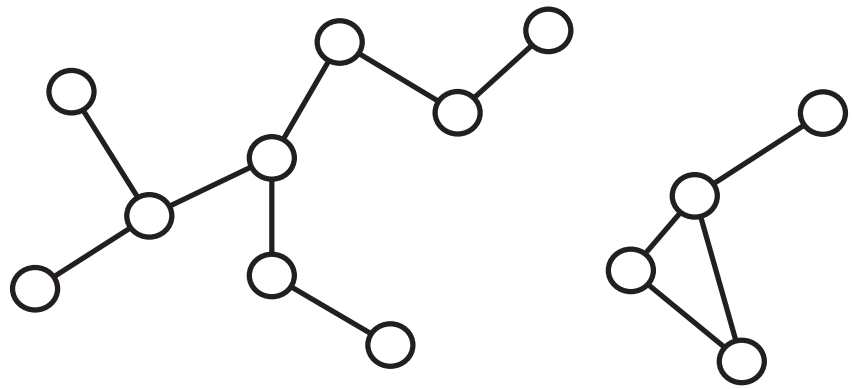
Interference graph

Consider a network of N nodes on an undirected interference graph $G(V, E)$, where a transmitting node blocks all its neighbors in the graph



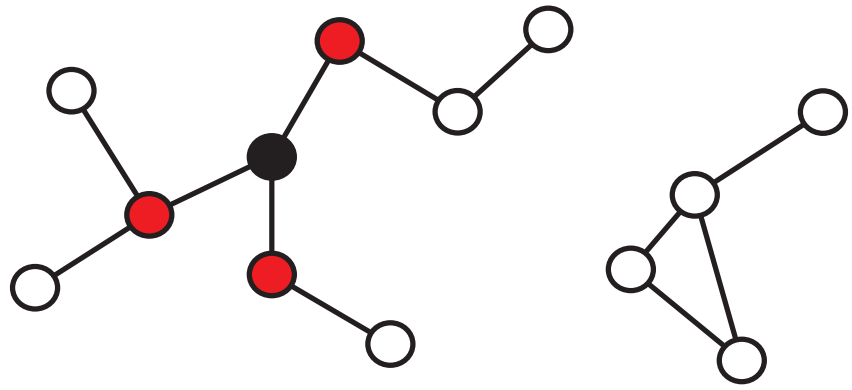
Interference graph

Consider a network of N nodes on an undirected interference graph $G(V, E)$, where a transmitting node blocks all its neighbors in the graph



Interference graph

Consider a network of N nodes on an undirected interference graph $G(V, E)$, where a transmitting node blocks all its neighbors in the graph



- if not blocked, node i activates after exponential back-off with mean ν_i^{-1}
- exponential transmissions with mean μ_i^{-1}
- all nodes are saturated
- denote $\sigma_i = \nu_i/\mu_i$

[Boorstyn *et al.* (1980), Wang & Kar (2005), Durvy & Thiran (2006)]

Stationary distribution

$X(t) = (X_1(t), \ldots, X_N(t)) \in S$: activity states at time t

$(X(t))$ is a reversible Markov process on

$$S := \{x \in \{0,1\}^N : x_i + x_j \leq 1 \forall i,j \in E\}$$

(collection of independent sets in G) with stationary distribution

$$\pi(x) = Z^{-1} \prod_{i=1}^{N} \sigma_i^{x_i}, \qquad x \in S,$$

with $\sigma_i := \nu_i/\mu_i$ potential activity factor ('offered' load)

Throughput ('carried' load) of node i is

$$\theta_i = \sum_{x \in S} \pi(x) \mathrm{I}_{\{x_i=1\}}$$

Loss networks and insensitivity

So far we assumed back-off periods and transmission durations to be exponentially distributed

Connection with loss networks readily implies that stationary distribution is in fact insensitive to distribution of transmission durations

Partial balance properties show that stationary distribution is insensitive to distribution of back-off periods as well, irrespective of whether or not back-off process is frozen during activity of neighbors (with Peter van de Ven)

Outline

- Queueing dynamics
- Backlog-based CSMA
- Single node and delay
- Metastability and mixing
- Fluid limits

Queueing dynamics

So far we assumed saturated buffer conditions, where nodes always have packets pending for transmission

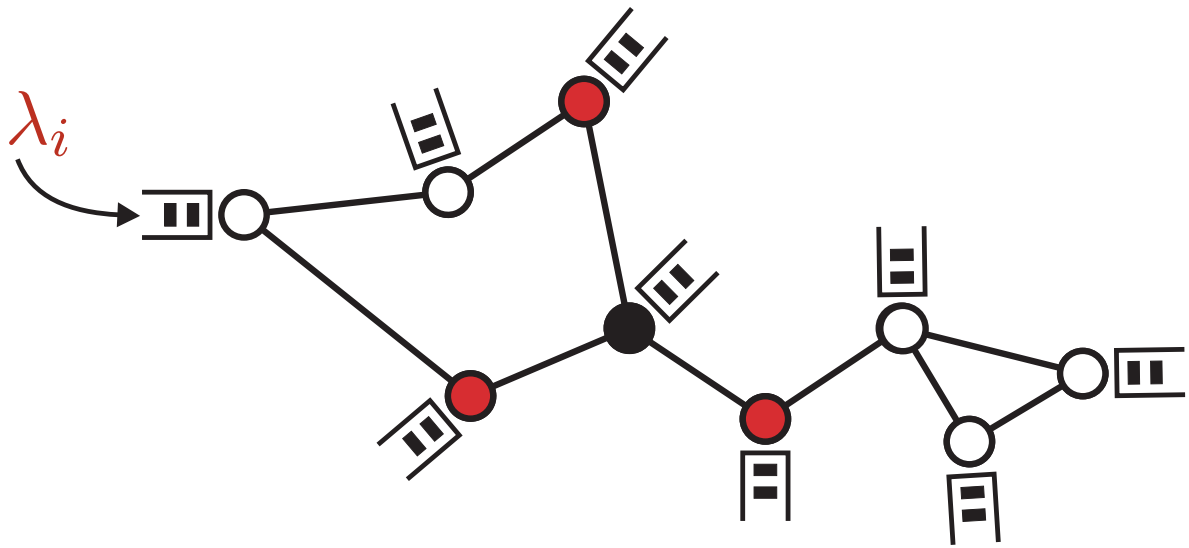
In reality, however, buffer contents fluctuate as packets are randomly generated and transmitted over time, giving rise to queueing dynamics

In particular, buffers may empty from time to time, and nodes may refrain from competition for medium during these periods

Queueing dynamics

Consider the same model as before, except

- packets arrive at node i according to a renewal process with mean interarrival time $1/\lambda_i$
- once a packet has been transmitted, it leaves the system



Backlog-based CSMA protocols

- Node i activates at exponential rate $f_i(Q_i(t))$ where $Q_i(t)$ denotes the number of packets at node i at time t.
- $f_i : [0, \infty) \mapsto [0, \infty)$ is called *activation function* of node i.
- Node i releases the medium at exponential rate $g_i(t) = p_i(Q_i(t))\mu_i$.
- $g_i(\cdot)$ is referred to as the *de-activation function*.
- $h_i(Q_i) = f_i(Q_i)/g_i(Q_i)$ is referred to as *activity function*

Backlog-based CSMA protocols

Under any of the aforementioned queue-based CSMA protocols, $(Q(t), X(t), t \geq 0)$ with $Q(t) = (Q_1(t), \ldots, Q_N(t))$ is continuous-time Markov process. Are there conditions on $\rho = (\rho_1, \ldots, \rho_N)$ guaranteeing ergodicity?

We are also interested in quantifying (mean) delays, depending on the functions $f_i(\cdot)$ and $g_i(\cdot)$.

Backlog-based CSMA protocols

For suitable choices of $f_i(\cdot)$ and $g_i(\cdot)$ (e.g., $f_i(n) \equiv 1$, $g_i(n) = \exp(-\log\log(n+1))$), system is stable as long as $\rho \in \mathrm{int}(\mathrm{conv}(S))$ [Liu *et al.* (2008), Rajagopalan *et al.* (2009)]

$f_i(n) = \log(n+1)$ and $g_i(n) \equiv 1$ also works

Jiang & Walrand (2008): The throughput function $\boldsymbol{\theta}(\boldsymbol{\sigma})$ is invertible on its range $\mathrm{int}(\mathrm{conv}(S))$. Solve inverse problems

$$\boldsymbol{\sigma} : \ \boldsymbol{\theta}(\boldsymbol{\sigma}) = \boldsymbol{\gamma} \in \mathrm{int}(\mathrm{conv}(S))$$

Single node

System is stable if $\rho < 1$ and $f(x) \to \infty$ or $g(x) \to 0$ as $x \to \infty$.

For any choice of the functions $f(\cdot)$ and $g(\cdot)$, $(Q(t), X(t), t \geq 0)$ is a continuous-time Markov process with state space $\{0, 1, \ldots\} \times \{0, 1\}$.

Balance equations:

$$\begin{aligned}
\lambda\pi(0,0) &= \mu\pi(1,1), \\
(\lambda+\mu)\pi(1,1) &= f(1)\pi(1,0) + (\mu - g(2))\pi(2,1), \\
(\lambda + f(n))\pi(n,0) &= \lambda\pi(n-1,0) \\
&+ g(n+1)\pi(n+1,1), \quad n \geq 1, \\
(\lambda+\mu)\pi(n,1) &= \lambda\pi(n-1,1) + f(n)\pi(n,0) \\
&+ (\mu - g(n+1))\pi(n+1,1), n \geq 2.
\end{aligned}$$

Set $\mu = 1$ and $\rho = \lambda$. Denote by N the stationary number of packets in the system, i.e.

$$\mathbb{P}\{N = n\} = \lim_{t \to \infty} \mathbb{P}\{N(t) = n\} = \pi(n, 0) + \pi(n, 1),$$

Introduce $G_0(z) = \sum_{n=0}^{\infty} \pi(n, 0) z^n$ and $G_1(z) = \sum_{n=1}^{\infty} \pi(n, 1) z^n$ and observe that

$$\mathbf{E}\{z^N\} = G_0(z) + G_1(z)$$

and

$$G_1(z) = \frac{\rho z}{1 - \rho z} G_0(z).$$

Theorem

If $f(n) = n$ *and* $g(n) = 1$,

$$G_0(z) = (1-\rho)\left(\frac{1-\rho}{1-\rho z}\right)^{\rho} \mathrm{e}^{(z-1)\rho}. \tag{1}$$

If $f(n) = 1$ *and* $g(n) = \frac{1}{n}$,

$$G_0(z) = \frac{(1-\rho)^2}{1-\rho z}. \tag{2}$$

In both cases,

$$\mathbf{E}\{N\} = \frac{2\rho}{1-\rho}. \tag{3}$$

Assume $g(n) = 1$.

Theorem

For $f(\cdot)$ strictly increasing and concave function,

$$\mathbf{E}\{N\} \geq \frac{\rho}{1-\rho} + f^{-1}\left(\frac{\rho}{1-\rho}\right). \tag{4}$$

For $f(\cdot)$ strictly increasing, continuous and convex function,

$$\mathbf{E}\{N\} \leq \frac{\rho}{1-\rho} + f^{-1}\left(\frac{\rho}{1-\rho}\right). \tag{5}$$

Note that, when $f(n) = n$ the inequalities are in fact equalities.

For concave sub-linear functions, the mean number of packets in the system is always larger than the mean number of packets in the system with a linear activation function.

The bounds are asymptotically sharp.

In heavy traffic as $\rho \uparrow 1$, $\mathbf{E}\{N\}$ grows for a logarithmic activation rate like

$$\exp(\frac{\rho}{1-\rho})$$

for a linear activation rate like

$$\frac{2\rho}{1-\rho}$$

and for an exponential activation rate like

$$\frac{\rho}{1-\rho}$$

We thus see that more aggressive activation rates improve the delay performance.

Theorem

(i) *If* $f(n) = n$, *then*

$$(1-\rho)N \xrightarrow{d} E_2(1) \text{ as } \rho \uparrow 1 \tag{6}$$

(ii) *If* $f(n) = n^{\alpha}$, $0 < \alpha < 1$, *then*

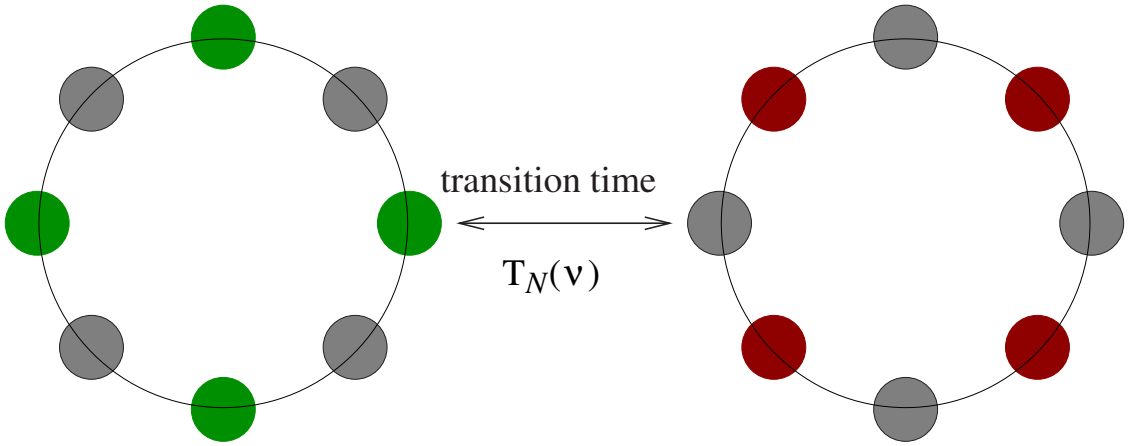
$$\frac{N}{\mathbf{E}\{N\}} \xrightarrow{d} 1 \text{ as } \rho \uparrow 1 \tag{7}$$

(iii) *If* $f(\cdot)$ *is a strictly increasing, continuous and convex function with* $\lim_{x\to\infty} f^{-1}(x)/x = 0$, *then*

$$(1-\rho)N \xrightarrow{d} E_1(1) \text{ as } \rho \uparrow 1, \tag{8}$$

$$\tag{9}$$

Metastability and mixing



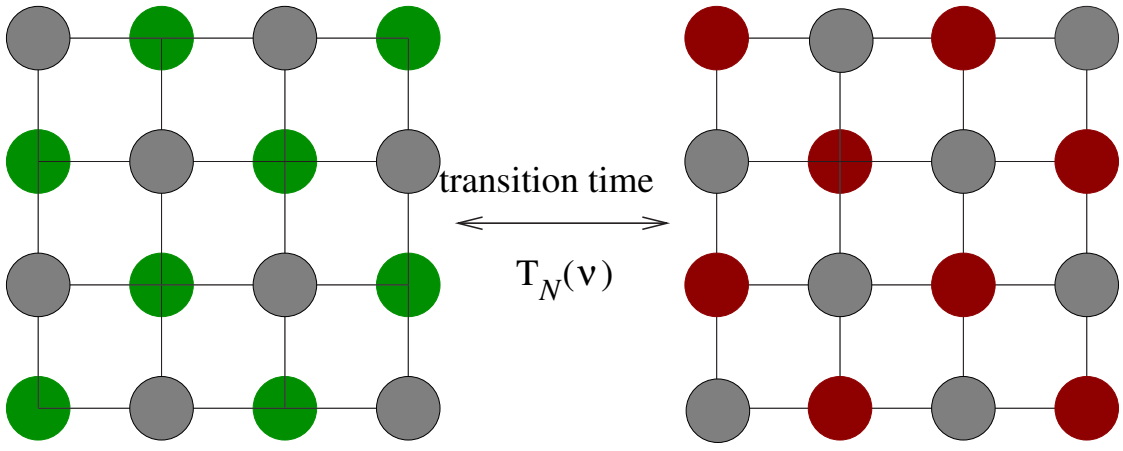
Consider the case $f(n) = 1$ and $g(n) = 1$. When all L green nodes are active, it will take long time for any of L red nodes to gain access to the medium

Once one of green nodes has turned off, adjacent green node must turn off as well before first one turns on again, in order for red node to gain access

Probability of latter event $O(1/\nu)$, and $T_N(\nu)$ is $O(\nu)$

Metastability and mixing

Consider $2L \times 2L$ grid network with 1-hop interference range



When all $2L^2$ green nodes are active, it will take long time for any of red nodes to gain access to the medium

Metastability and mixing

Fluid limits reflect interplay between metastability and queueing dynamics

Assume queue-dependent activation rate function $f(\cdot)$

Transition times between dominant activity states are $O(f(N)^H)$ when queue sizes are $O(N)$, where H depends on network structure, e.g.,

- Complete interference graph: $H = 0$
- Ring network: $H = 1$
- Grid network: $H = L$ or $H = 2L$
- Complete bipartite graph: $H = M - 1$

Fluid limits

Examine the dynamics of the Markov process $Z(t) = (Q(t), X(t))$ using fluid limits.

Consider a sequence of processes $Z^N(t)$, where the initial states satisfy $\sum_{i=1}^{M} Q_i(0) = N$ and $Q_i^N(0)/N \to q_i \geq 0$ as $N \to \infty$.

The process $\bar{Z}^N(t) = (\frac{1}{N} Q^N(Nt), X^N(Nt))$ is called the fluid-scaled version of the process $Z^N(t)$. Note that the activity process is scaled in time as well, but not in space.

Stochastic fluid limits

Unlike in most queueing systems where fluid limits follow deterministic trajectories described by a set of differential equations, our system may exhibit fluid limits that are stochastic processes.

$Z^N(t)$ has two interacting components, $Q^N(t)$ and $X^N(t)$.

On the one hand, the evolution of $Q^N(t)$ depends on the rate at which queues are served, and in turn depends on $X^N(t)$.

On the other hand, when queues $Q^N(t)$ are fixed, the process $X^N(t)$ is a reversible Markov process on the set of possible activation states whose transitions are functions of $Q^N(t)$.

Trichotomy

Fluid limits reflect interplay between metastability and queueing dynamics

As it turns out, we encounter different types of fluid limits depending on the *mixing* properties of the activity process $X^N(t)$. These properties depend on the choices of activity functions, $f_i(\cdot)$ and $g_i(\cdot)$.

Fast mixing - Deterministic fluid limits

Transitions between the various activity states are not observed in the the fluid regime

In such cases, $X^N(t)$ evolves much faster than $Q^N(t)$ as N grows large, and to obtain the rate at which queues are served in the fluid regime, the activity process $X^N(t)$ is averaged.

$f(N)^H \ll N$: transitions occur on much faster time scale than $O(N)$.

Slow mixing - Inhomogeneous Poisson fluid limits

Transitions between the various activation states are observed in the fluid regime.

$f(N)^H \sim N$: transitions occur on time scale $O(N)$, and will be observed at fluid level, yielding piecewise linear fluid limit, with switching points governed by time-inhomogeneous Poisson process

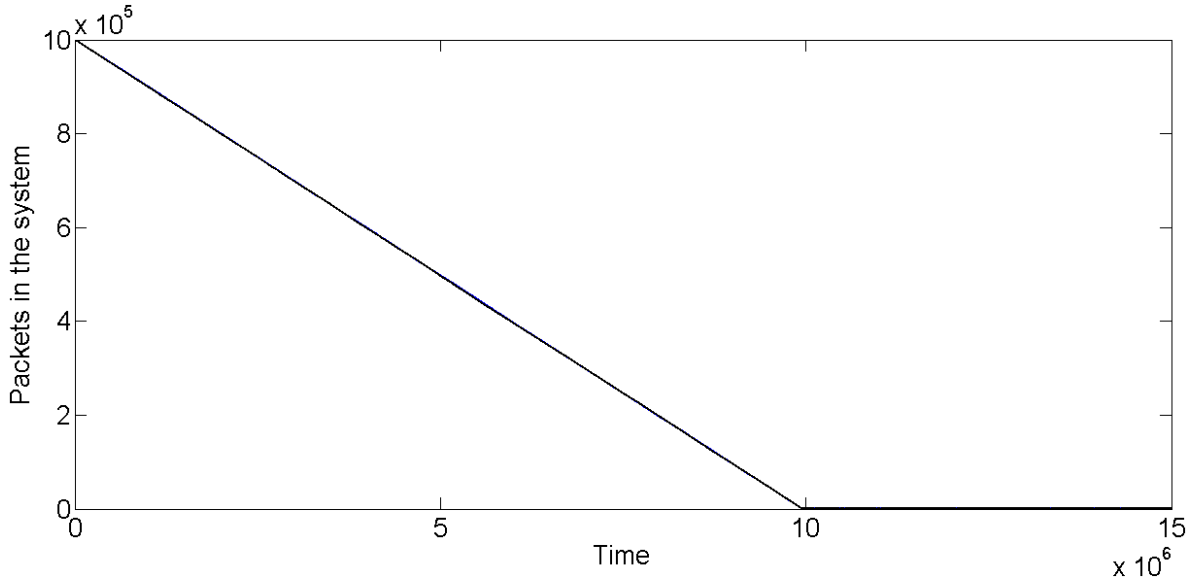
Torpid mixing - Pseudo-deterministic fluid limits

When the transitions between the various activation states occur on a time scale slower than N, the activation state seems to be frozen in the fluid regime.

$f(N)^H \gg N$: transitions occur on much slower time scale than $O(N)$, and will not manifest themselves at fluid level, yielding piecewise-linear fluid limit, pseudo-deterministic except for initial direction

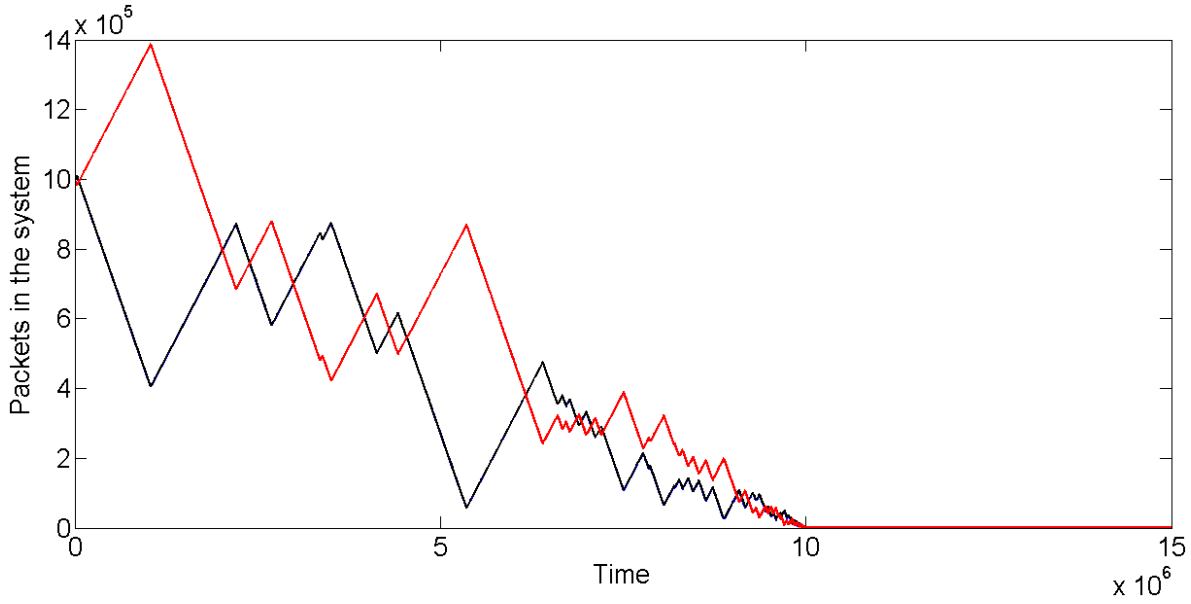
Example: 2-partite interference graphs

$M_1 = M_2 = 1$, $f(n) = n$, $g(n) = 1$



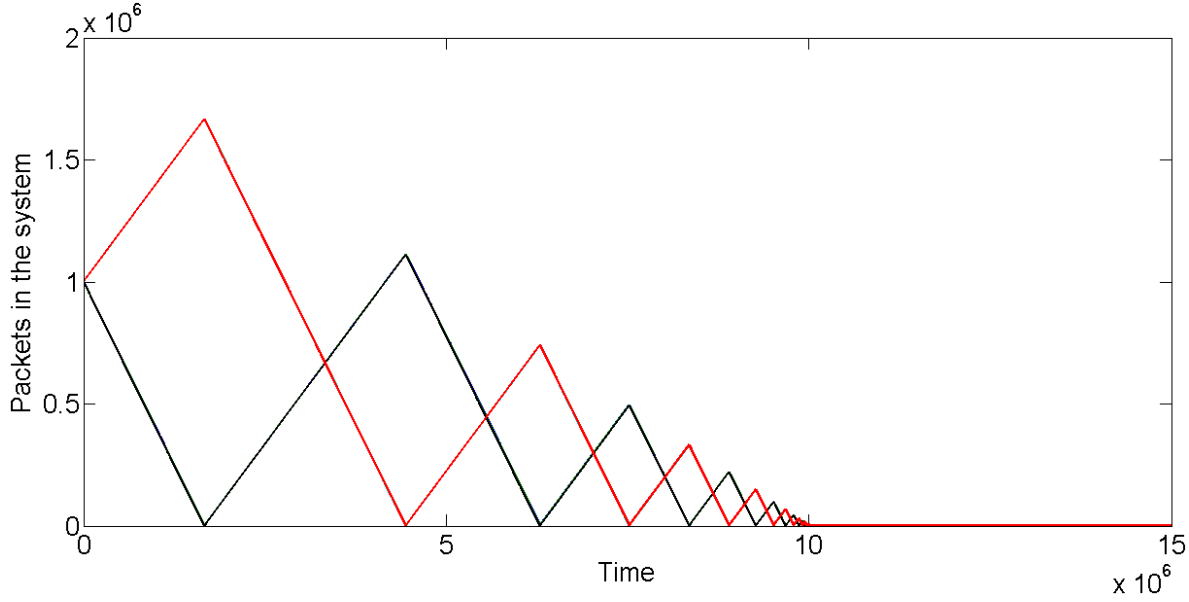
Example: 2-partite interference graphs

$M_1 = M_2 = 2$, $f(n) = n$, $g(n) = 1$



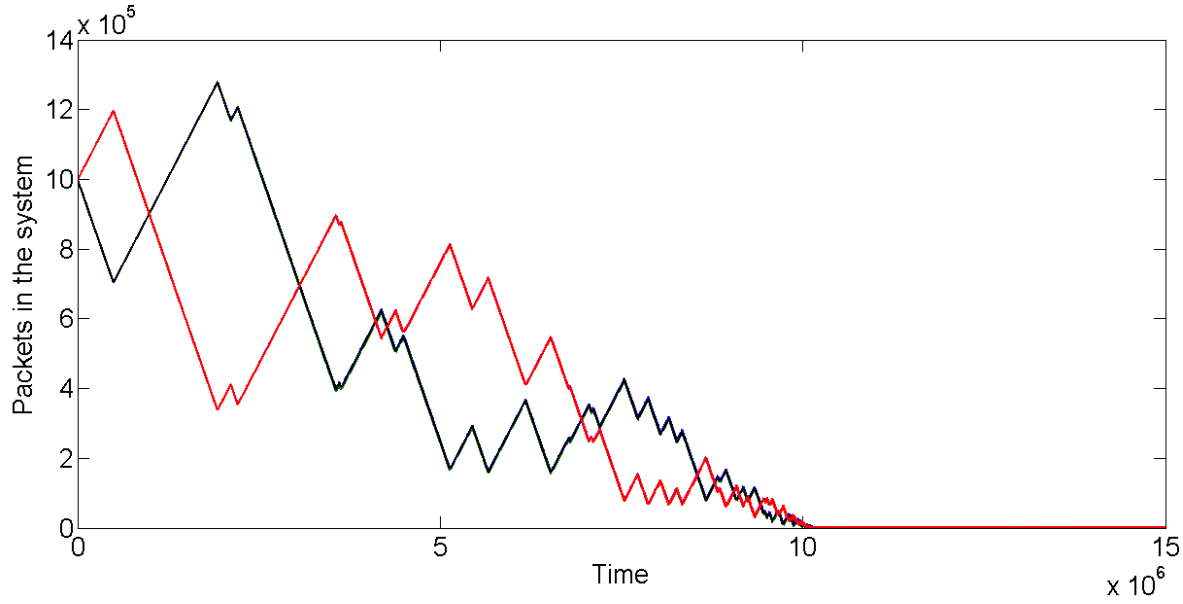
Example: 2-partite interference graphs

$M_1 = M_2 = 3$, $f(n) = n$, $g(n) = 1$



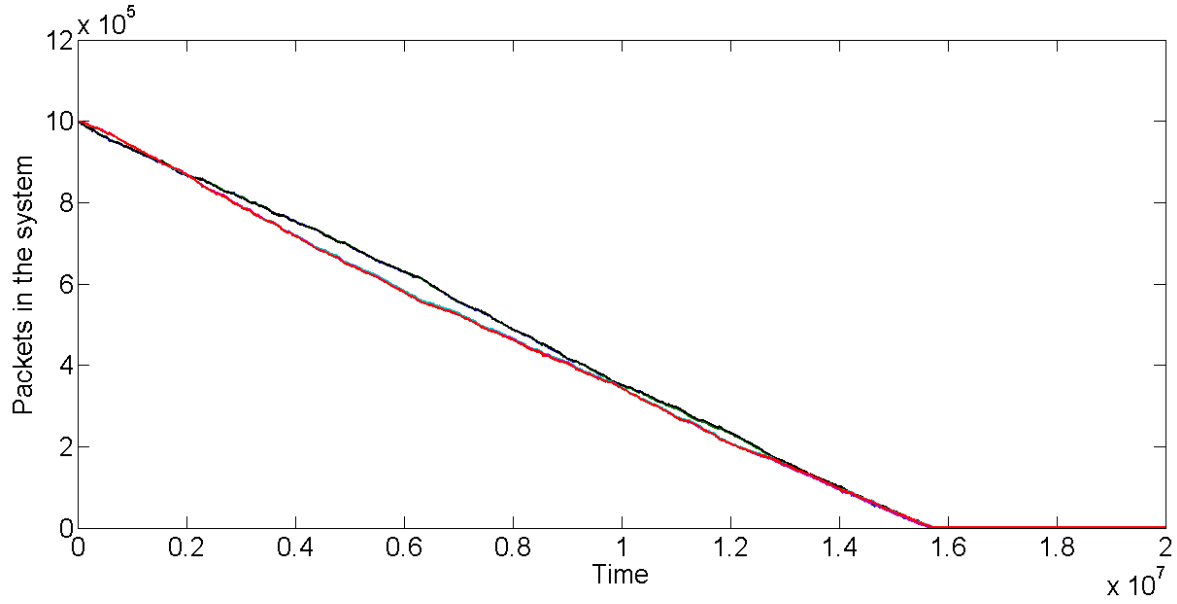
Example: 2-partite interference graphs

$M_1 = M_2 = 3,\ f(n) = \sqrt{n},\ g(n) = 1$



Example: 2-partite interference graphs

$M_1 = M_2 = 3$, $f(n) = \log(n+1)$, $g(n) = 1$



Summary

Various clever algorithms have been developed for finding the back-off rates that yield a particular *target throughput vector* [JW09]

In the same spirit, several powerful algorithms have been devised for adapting the transmission lengths based on backlog information, and been shown to guarantee *maximum stability* [JSSW10, RSS09].

Ghaderi & Srikant 2010 recently showed that activity functions can be used that are essentially *linear* in order to reduce the delays while preserving maximum-stability guarantees.

Can we say more (delays, time scales, heavy traffic, fluid limits)?