Exact Statistics of a Markov Chain trhough Reduction in Number of States: A Satellite On-board Switching Example

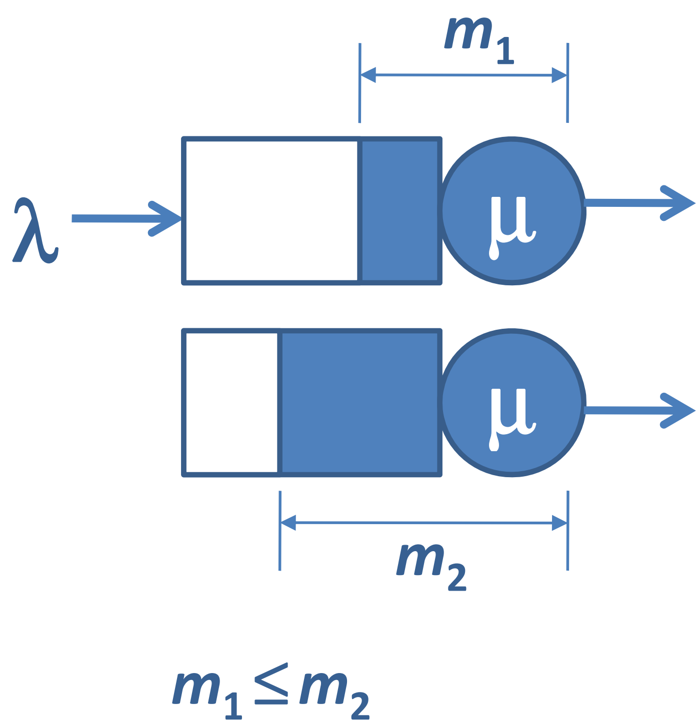
Marco Aurelio Alzate Monroy
Universidad Distrital Francisco José de Caldas

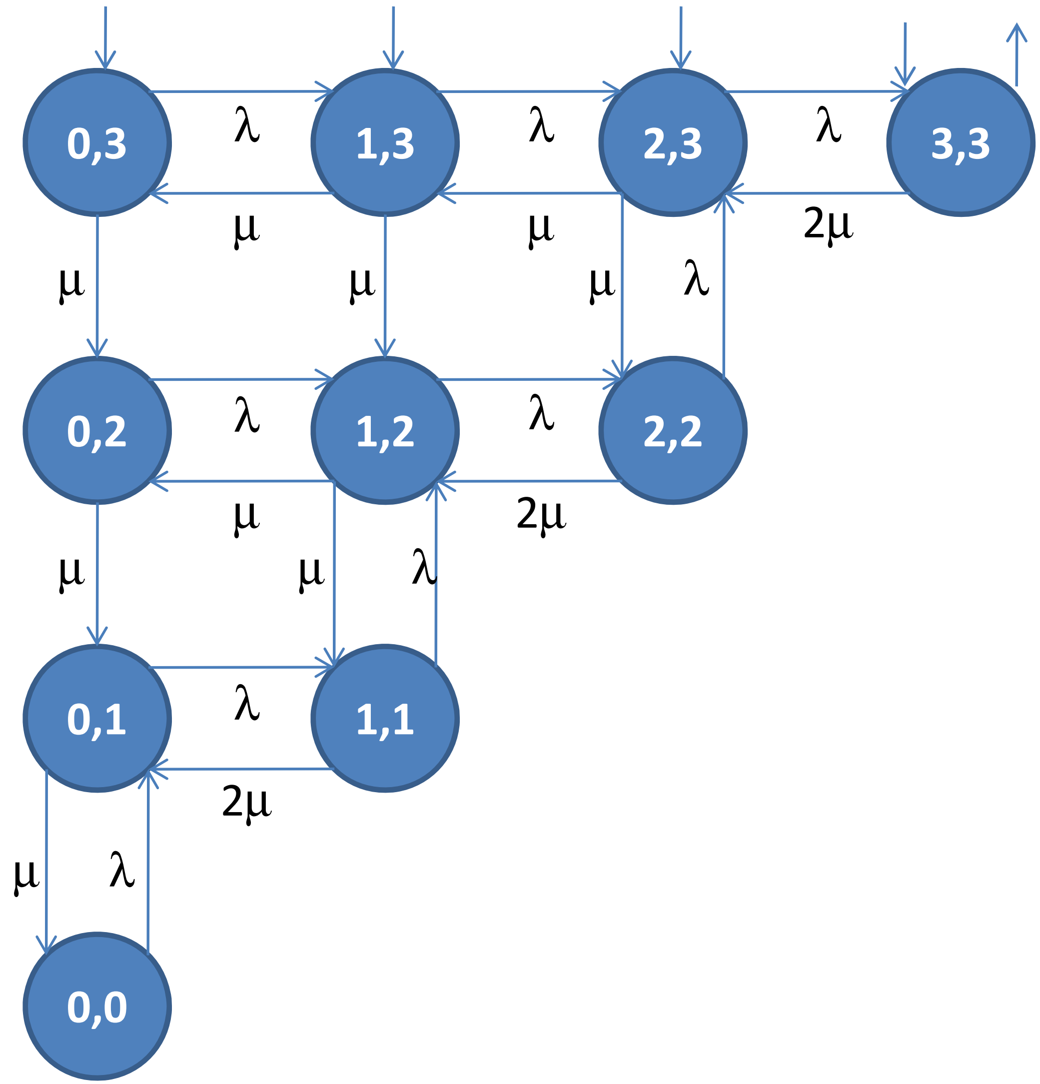
The general setting

- An irreducible discrete time Markov chain
  - States, $\boldsymbol{m} = (m_1, m_2, \ldots m_N)^{\mathrm{T}} \in \mathscr{M} \subseteq \mathbb{Z}^{\mathrm{N}}$
  - Transition probabilities $p(\boldsymbol{m}, \boldsymbol{n})$
  - Equilibrium distribution $\pi_p(\boldsymbol{m})$
- Some performance characteristic

$$g_p = \sum_{\boldsymbol{m} \in \mathscr{M}} c_p(\boldsymbol{m}) \pi_p(\boldsymbol{m})$$

First simple example: mean shortest queue length





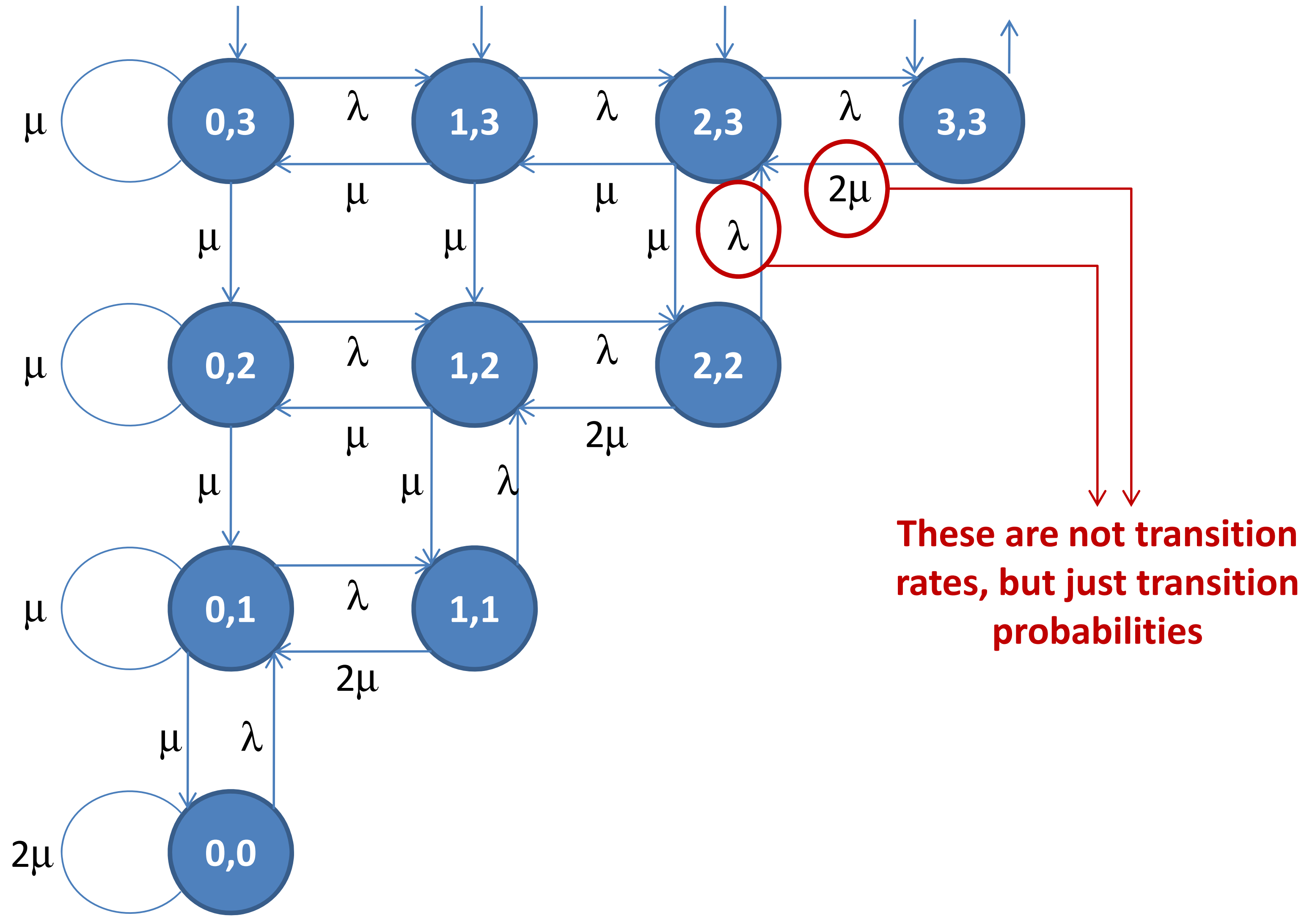
Discretization

- Normalize every transition rate with respect to $\lambda+2\mu$ and let

$$\mu \leftarrow \frac{\mu}{\lambda+2\mu}, \quad \lambda \leftarrow \frac{\lambda}{\lambda+2\mu}, \quad \Rightarrow \quad \lambda+2\mu=1$$

- An idle server attends a ficticious client, which does not depart, and is interrupted by real clients
- (just typical *normalization*: the outgoing transition rates for each state add up to one, and the mean time spent in a state is one)

Equivalent discrete time Markov Chain



The proposed procedure

- Redirect one or more outgoing transitions from state $\boldsymbol{m}$
  - If a transition from $\boldsymbol{m}$ to $\boldsymbol{n}_1$ is redirected to $\boldsymbol{n}_2$,

$$p(\boldsymbol{m},\boldsymbol{n}_2) \leftarrow p(\boldsymbol{m},\boldsymbol{n}_2) + p(\boldsymbol{m},\boldsymbol{n}_1)$$

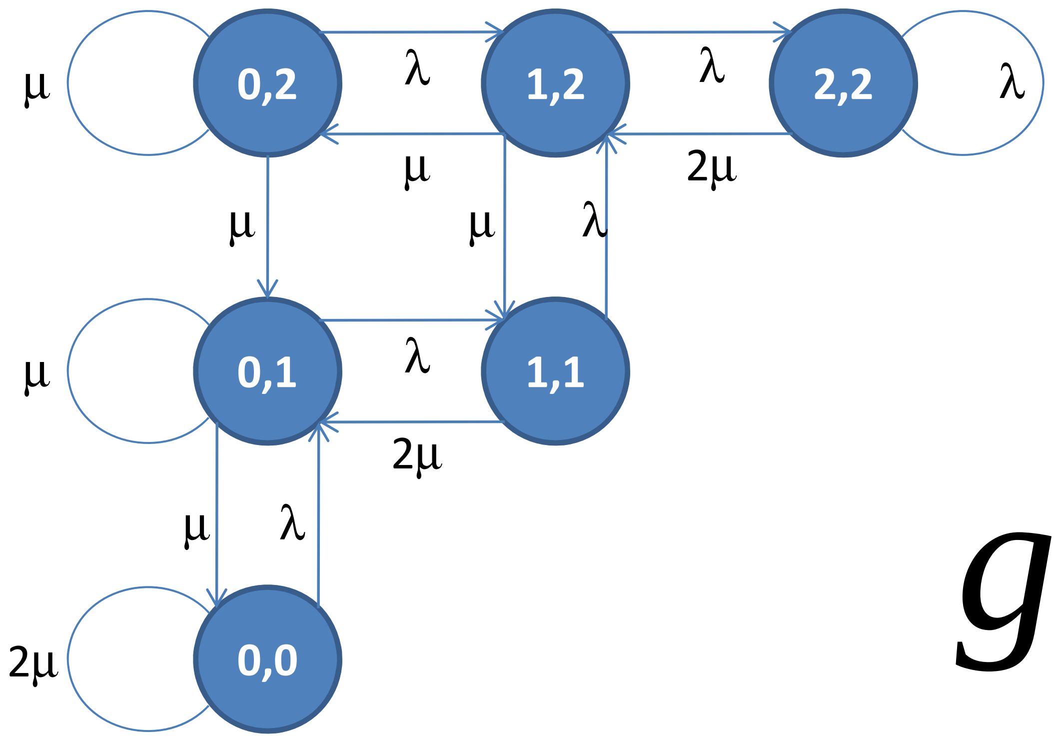
$$p(\boldsymbol{m},\boldsymbol{n}_1) \leftarrow 0$$

- Denote the new transition probabilities by $q(\boldsymbol{m},\boldsymbol{n})$ and the new performance characteristic by

$$g_q = \sum_{\boldsymbol{m}\in\mathscr{M}} c_q(\boldsymbol{m})\pi_q(\boldsymbol{m})$$

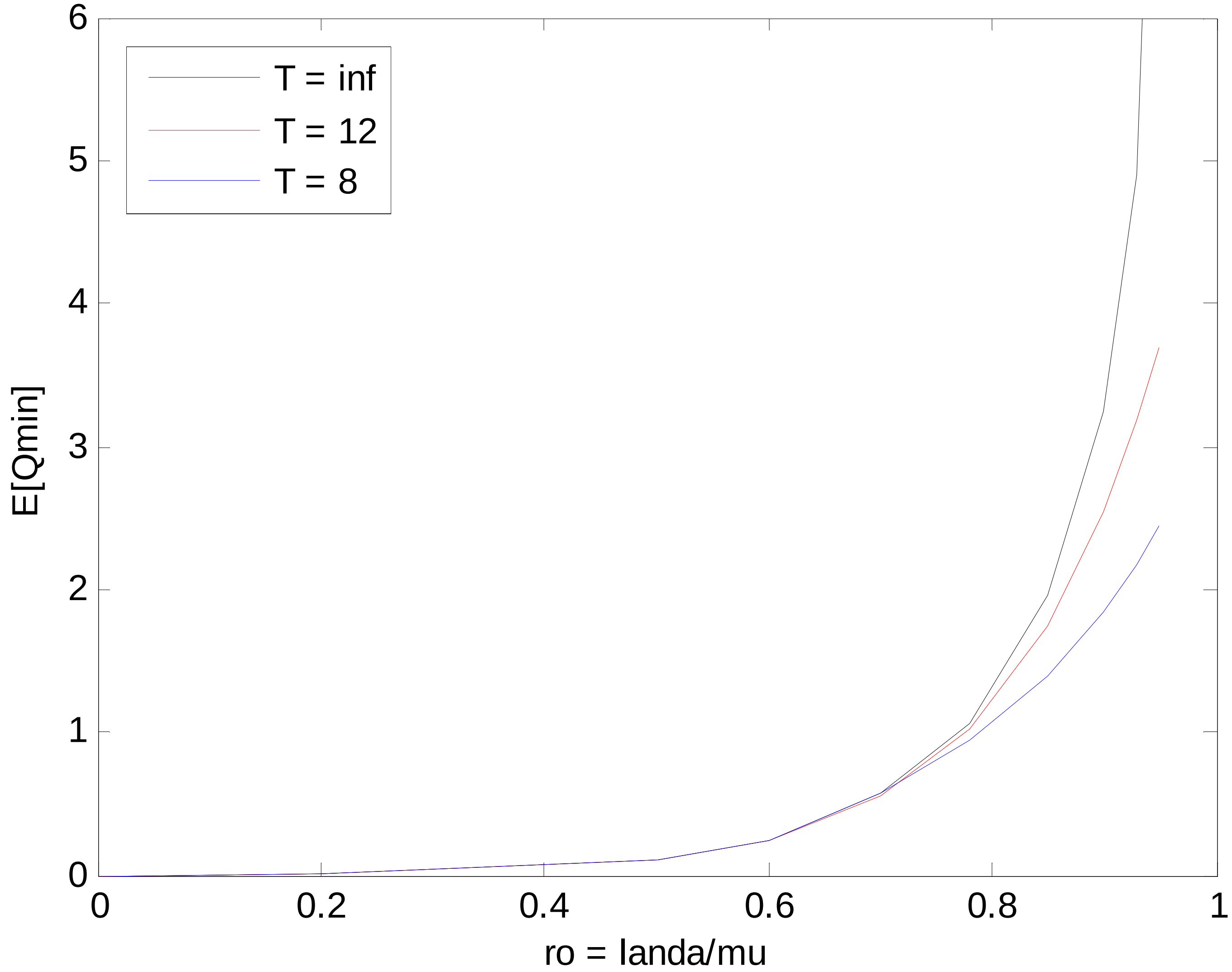
For example:

- Truncate the buffers at T clients
  - Redirect the transition from (T,T) to $(T,T+1)$ as a transition from (T,T) to itself (reject the new arrival)



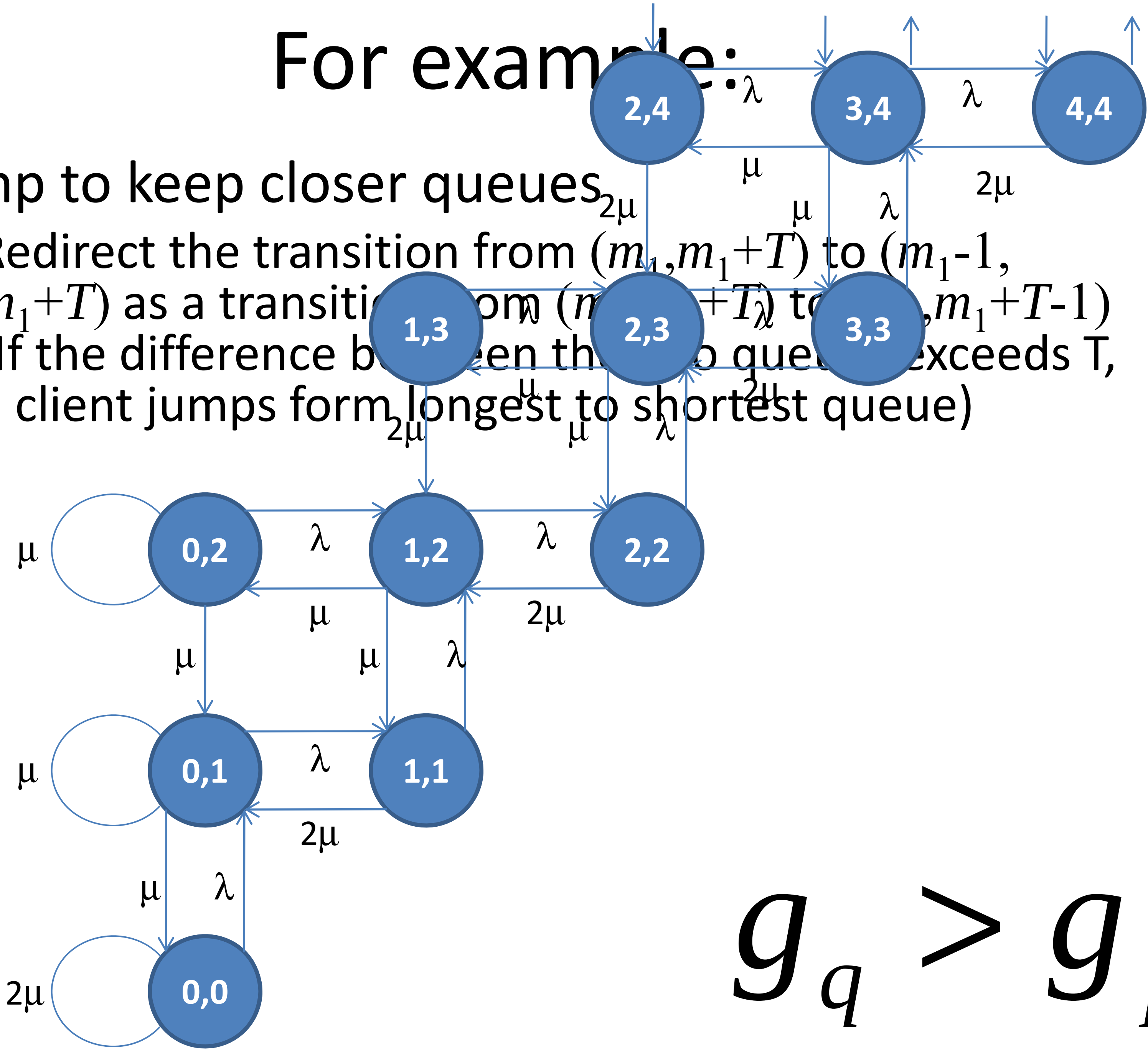
$$g_q < g_p$$

Truncate the buffers at T clients



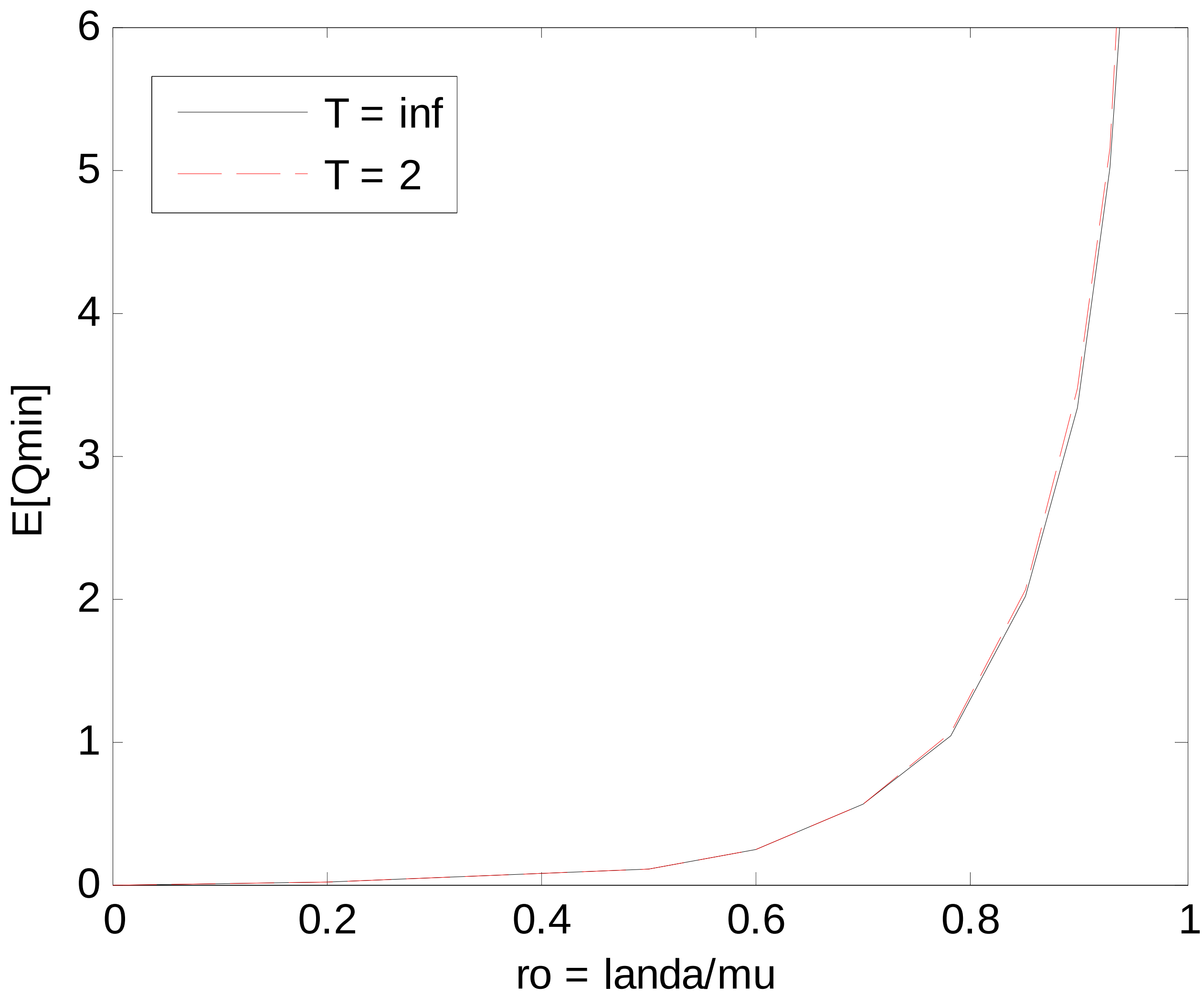
For example:

- Jump to keep closer queues
  - Redirect the transition from (m_1, m_1+T) to (m_1-1, m_1+T) as a transition from (m_1, m_1+T) to (m_1, m_1+T-1) (If the difference between the two queues exceeds T, a client jumps form longest to shortest queue)



$$g_q > g_p$$

Jump to keep closer queues



Among the many possibilities,…

- Which transitions should be redirected?
- Do they lead to an upper or lower bound?
- Can we prove it?
- How tight is this bound?

The proposed method

- Define $v(t,\boldsymbol{m})$ as the total cost over a time interval of t frame periods, starting in state $\boldsymbol{m}$

$$v(0,\boldsymbol{m}) = 0$$

$$v(t+1,\boldsymbol{m}) = c(\boldsymbol{m}) + \sum_{\boldsymbol{\mu}} p(\boldsymbol{m},\boldsymbol{\mu}) \cdot v(t,\boldsymbol{\mu})$$

- And, assuming ergodicity

$$g \overset{a.s.}{=} \lim_{t\to\infty} \frac{1}{t} v(t,\boldsymbol{m})$$

for every initial state $\boldsymbol{m}$

Precedence among states

- We say that state $\boldsymbol{m}$ is more attractive than state $\boldsymbol{\mu}$ (written $\boldsymbol{m} \to \boldsymbol{\mu}$) if $v(t,\boldsymbol{m}) \leq v(t,\boldsymbol{\mu})$ for all $t \geq 0$.
- Then, for a lower bound, redirect to more attractive states; and, for an upper bound, redirect to less attractive states!
- Why?

Deleting less attractive states

- Let us redirect all transitions leading to state $\boldsymbol{k}$ as transitions leading to state $\boldsymbol{j}$, where $\boldsymbol{j} \to \boldsymbol{k}$. Then $q = p$ except for

$$q(\boldsymbol{m}, \boldsymbol{k}) = 0$$

$$q(\boldsymbol{m}, \boldsymbol{j}) = p(\boldsymbol{m}, \boldsymbol{j}) + p(\boldsymbol{m}, \boldsymbol{k}) \qquad \forall \boldsymbol{m} \in \mathscr{M}$$

- So we just made state $\boldsymbol{k}$ a transient state

Then we can prove that

$$v_q\left(t,\boldsymbol{m}\right)\le v_p\left(t,\boldsymbol{m}\right)\ \ \forall t\ge 0\quad (*)$$

- Indeed, (*) holds for $t=0$ because

$$v_q\left(0,\boldsymbol{m}\right)=v_p\left(0,\boldsymbol{m}\right)=0$$

- But, if it holds fot t, it also holds for $t+1$ because

$$v_q(t+1,\boldsymbol{m})=c(\boldsymbol{m})+\sum_{\boldsymbol{\mu}}q(\boldsymbol{m},\boldsymbol{\mu})\cdot v_q(t,\boldsymbol{\mu})\le c(\boldsymbol{m})+\sum_{\boldsymbol{\mu}}q(\boldsymbol{m},\boldsymbol{\mu})\cdot v_p(t,\boldsymbol{\mu})$$

by hypothesis

$$v_q(t+1,\boldsymbol{m}) = c(\boldsymbol{m}) + \sum_{\boldsymbol{\mu}} q(\boldsymbol{m},\boldsymbol{\mu}) \cdot v_q(t,\boldsymbol{\mu}) \le c(\boldsymbol{m}) + \sum_{\boldsymbol{\mu}} q(\boldsymbol{m},\boldsymbol{\mu}) \cdot v_p(t,\boldsymbol{\mu})$$

- But

$$q(\boldsymbol{m},\boldsymbol{\mu}) = \begin{cases} p(\boldsymbol{m},\boldsymbol{\mu}) & \text{if there were no redirections involving } \boldsymbol{\mu} \\ 0 & \text{if transitions leading to } \boldsymbol{\mu} \text{ were redirected to some other state} \\ p(\boldsymbol{m},\boldsymbol{\mu}) + p(\boldsymbol{m},\boldsymbol{k}) & \text{if transitions leading to } \boldsymbol{k} \text{ were redirected to } \boldsymbol{\mu} \end{cases}$$

- So

$$v_q(t+1,\boldsymbol{m}) \le c(\boldsymbol{m}) + \sum_{\mu} p(\boldsymbol{m},\boldsymbol{\mu}) \cdot v_p(t,\boldsymbol{\mu}) - p(\boldsymbol{m},k) \cdot \left[v_p(t,\boldsymbol{k}) - v_p(t,\boldsymbol{j}) \right]$$

$$\le c(\boldsymbol{m}) + \sum_{\mu} p(\boldsymbol{m},\boldsymbol{\mu}) \cdot v_p(t,\boldsymbol{\mu}) = v_p(t+1,\boldsymbol{m})$$

Transition (***m***,***k***) were redirected as transition (***m***,***j***), with ***j*** → ***k***

- So (*) holds for every *t*

So we have shown that

$$g_q = \lim_{t\to\infty} \frac{1}{t} v_q\left(t, \boldsymbol{m}\right) \leq \lim_{t\to\infty} \frac{1}{t} v_p\left(t, \boldsymbol{m}\right) = g_p$$

- Redirecting transitions to more attractive states lead to a lower bound for the performance characteristic
- Similarly, redirecting transitions to less attractive states lead to an upper bound

The trick is how to determine the attractiveness relation between states!

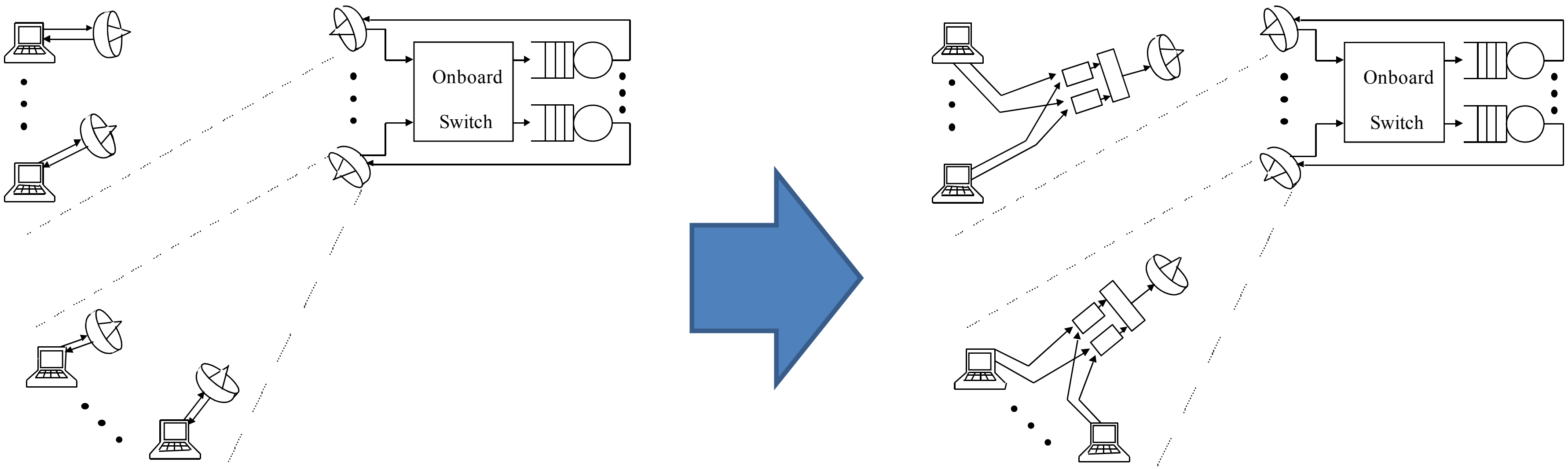
- Use the same procedure above (without redirecting transitions)
  - $v_p(0,\boldsymbol{j}) = v_p(0,\boldsymbol{k}) = 0$
  - Assume $v_p(t,\boldsymbol{j}) \leq v_p(t,\boldsymbol{k})$ for some t
  - Check from the recursion if it also holds for $t+1$

$$v_p\left(t+1,\boldsymbol{j}\right) = c\left(\boldsymbol{j}\right) + \sum_{\mu} p(\boldsymbol{j},\boldsymbol{\mu}) \cdot v_p\left(t,\boldsymbol{\mu}\right)$$

$$v_p\left(t+1,\boldsymbol{k}\right) = c\left(\boldsymbol{k}\right) + \sum_{\mu} p(\boldsymbol{k},\boldsymbol{\mu}) \cdot v_p\left(t,\boldsymbol{\mu}\right)$$

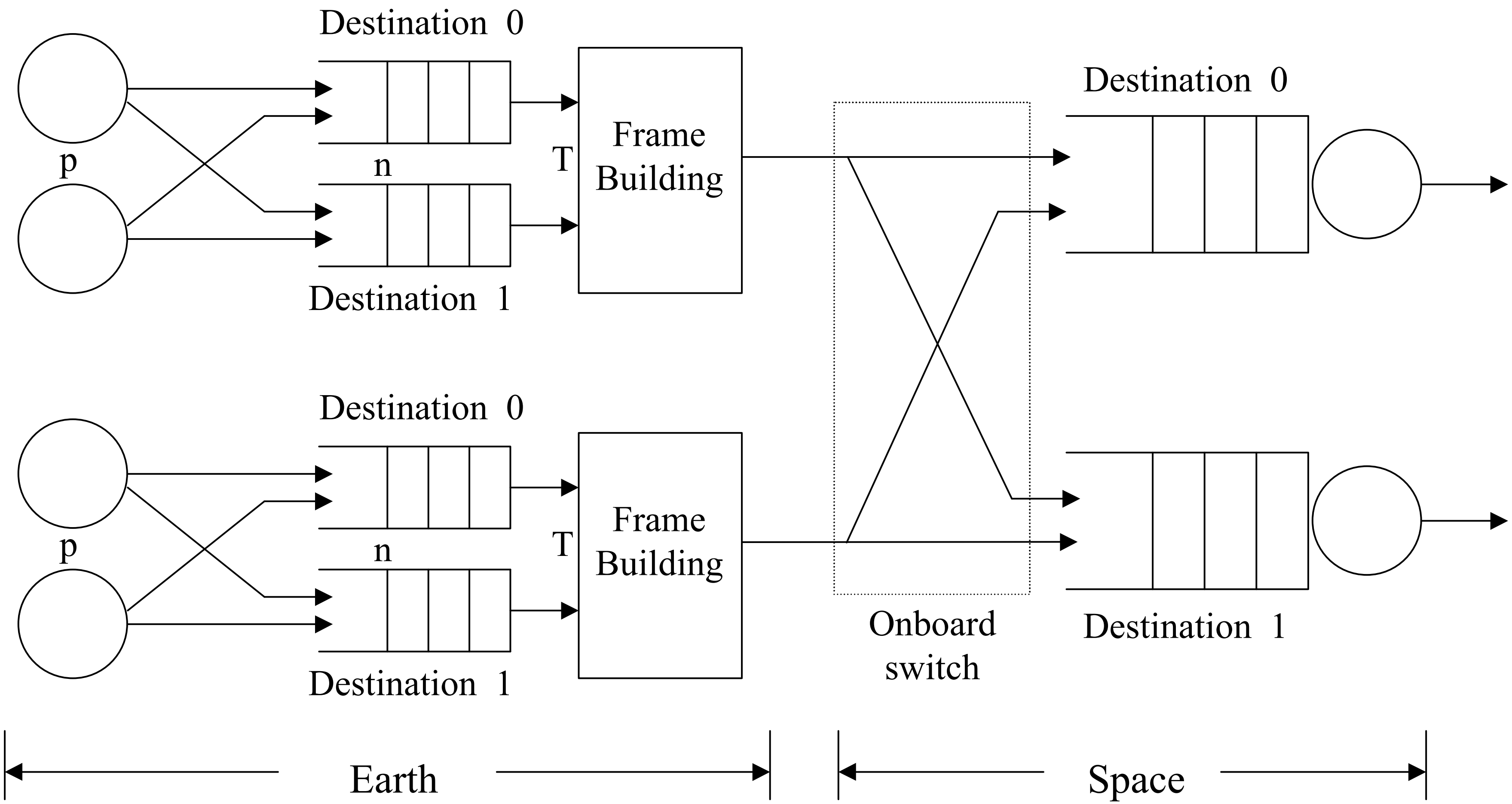
We made all this because…

- We were interested in reducing the number of switching operations onboard a satellite



- The frame builder looks at these queues every T slots. If at least one of the queues has 1 or more packets, a frame is constructed with up to T packets taken from the longest queue.

The queueing model becomes…



In an earth station the Markov chain is

$$p\left(\left(j_0, j_1\right),\left(k_0, k_1\right)\right) = pA_{k_0 - j_0 + f_0(j_0, j_1), k_1 - j_1 + f_1(j_0, j_1)}$$

- Where

$$\boldsymbol{f}\left(m_0, m_1\right) = \begin{cases} \left(\min\left(T, m_0\right) \quad 0\right)^{\mathrm{T}} & \text{if } m_0 > m_1 \\ \left(0 \quad \min\left(T, m_1\right)\right)^{\mathrm{T}} & \text{if } m_0 < m_1 \\ \text{any of the above with equal probability} & \text{if } m_0 = m_1 \end{cases}$$

Is the number of packets taken away by the frame builder when it finds (m_0, m_1) packets in the queues, and

$$pA_{j,k} = \Pr\left[\boldsymbol{A} = \begin{pmatrix} j \\ k \end{pmatrix}\right] = \begin{pmatrix} 2 \cdot T \\ j+k \end{pmatrix} \cdot \begin{pmatrix} j+k \\ j \end{pmatrix} \cdots$$

$$p^{j+k} \cdot \left(1-p\right)^{2 \cdot T - j - k} \cdot \left(\frac{1}{2}\right)^{j+k} \qquad \begin{array}{l} j = 0, 1, \ldots, 2 \cdot T \\ k = 0, 1, \ldots, 2 \cdot T - j \end{array}$$

is the probability of (j,k) arrivals in T slots

The average number of packets in the terrestrial queues is

$$g = \sum_{\boldsymbol{m}} c(\boldsymbol{m}) \cdot \pi(\boldsymbol{m})$$

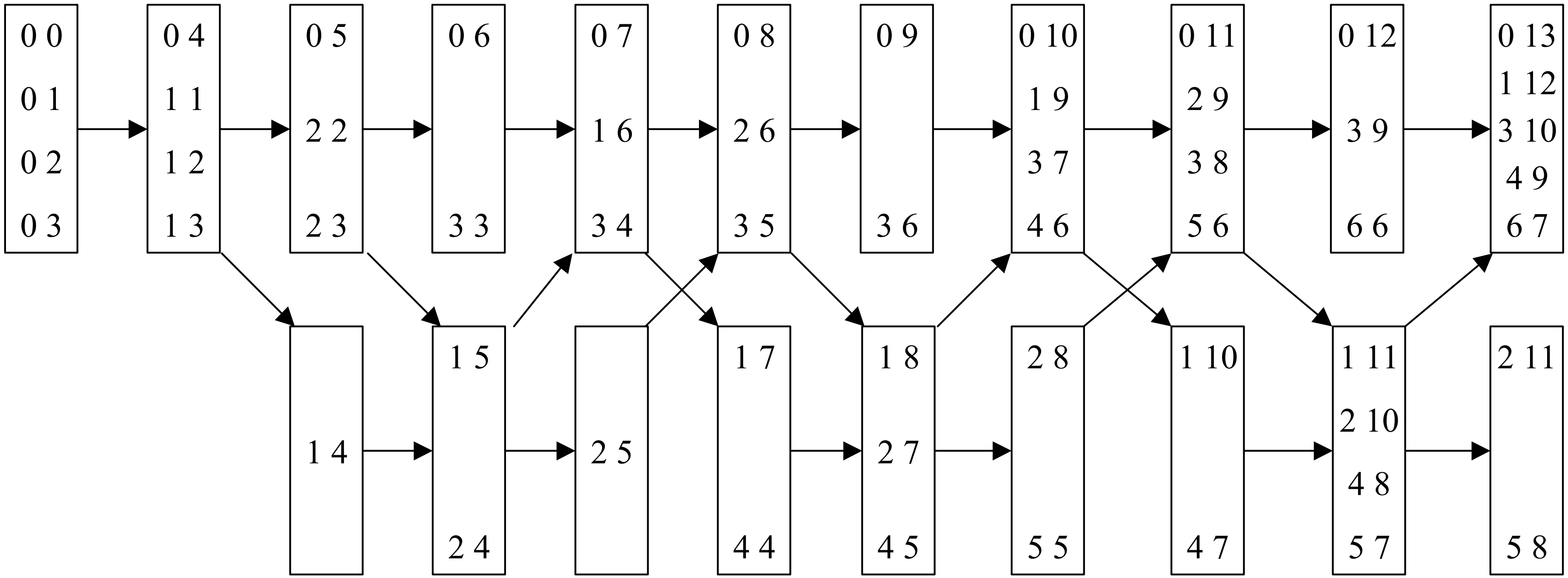
with

$$c(\boldsymbol{m}) = \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \left(\boldsymbol{m} - \boldsymbol{f}\left(\boldsymbol{m}\right) + E\left[\frac{1}{T} \cdot \sum_{t=0}^{T-1} (T-t) \cdot \boldsymbol{A}_t \right] \right) = \cdots$$

$$= m_0 + m_1 - \min\left(T, \max\left(m_0, m_1\right)\right) + \left(T+1\right) \cdot p$$

so all we need is to find attractiveness relations among states

There are also equivalent states



We say that state $\boldsymbol{m}$ is equivalent to state $\boldsymbol{\mu}$ (written $\boldsymbol{m} \leftrightarrow \boldsymbol{\mu}$) if $v(t,\boldsymbol{m}) = v(t,\boldsymbol{\mu})$ for all $t \geq 0$.

So these transitions lead to exact solutions

$$m_1 = \mu_0 + \mu_1 - m_0$$
$$m_0 = \min\left(\mathrm{mod}(\mu_0, T), \mathrm{mod}(\mu_1, T)\right)$$

if $m_1 < T$
    if $m_0 = 0$
        $m_1 = 0$
    else
        $m_1 = T + m_0$
        $m_0 = 0$
    end
end

Redirect transitions to (μ_0, μ_1) as transitions to (m_0, m_1)

The resulting system can be easily solved through Matrix Geometric Methods

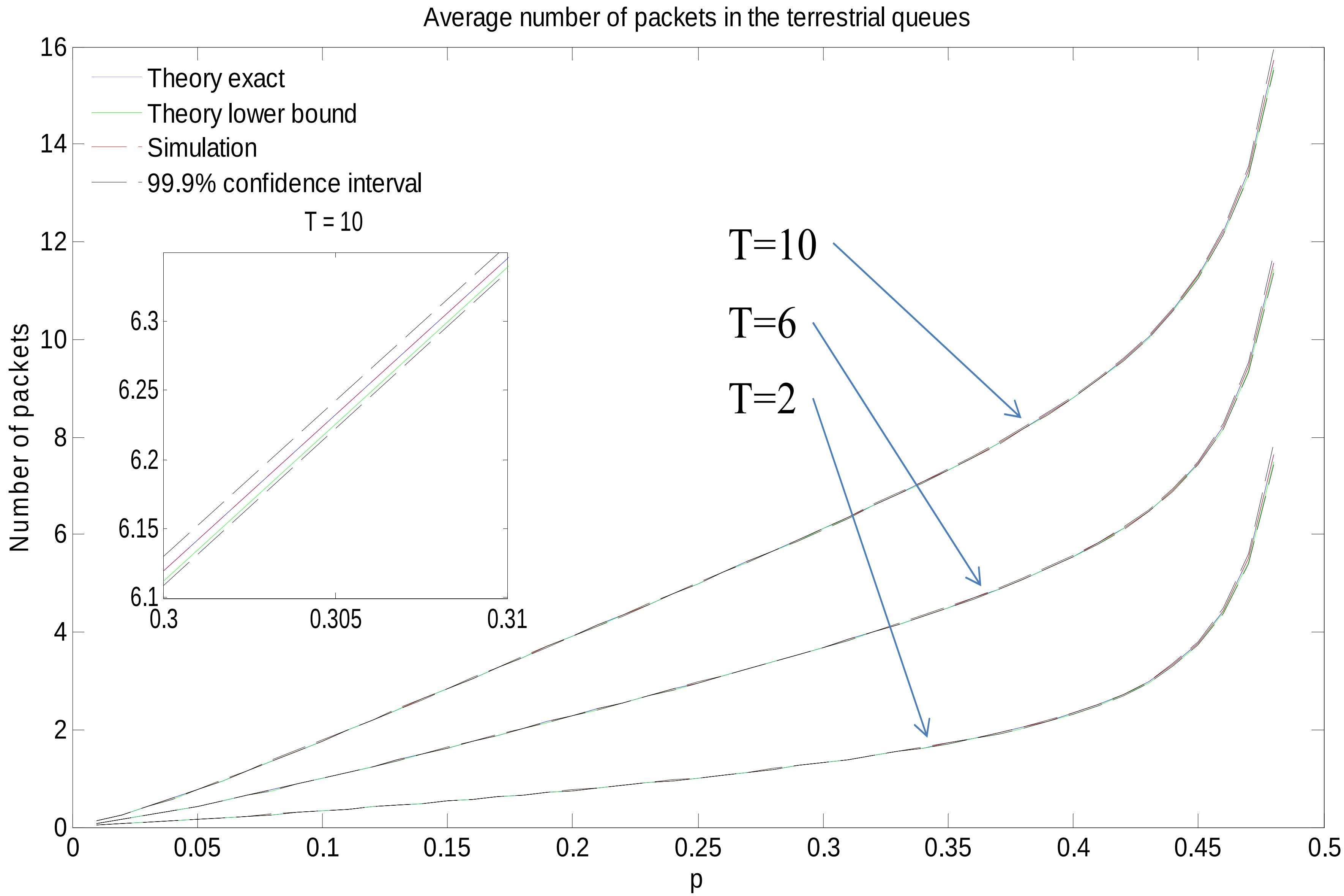
$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{A}_0 & 0 & 0 & \cdots \\ \boldsymbol{B}_1 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & 0 & \cdots \\ 0 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \cdots \\ 0 & 0 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Each submatrix becomes a $[T(T+1)/2]$ x $[T(T+1)/2]$ square one : Easily solved

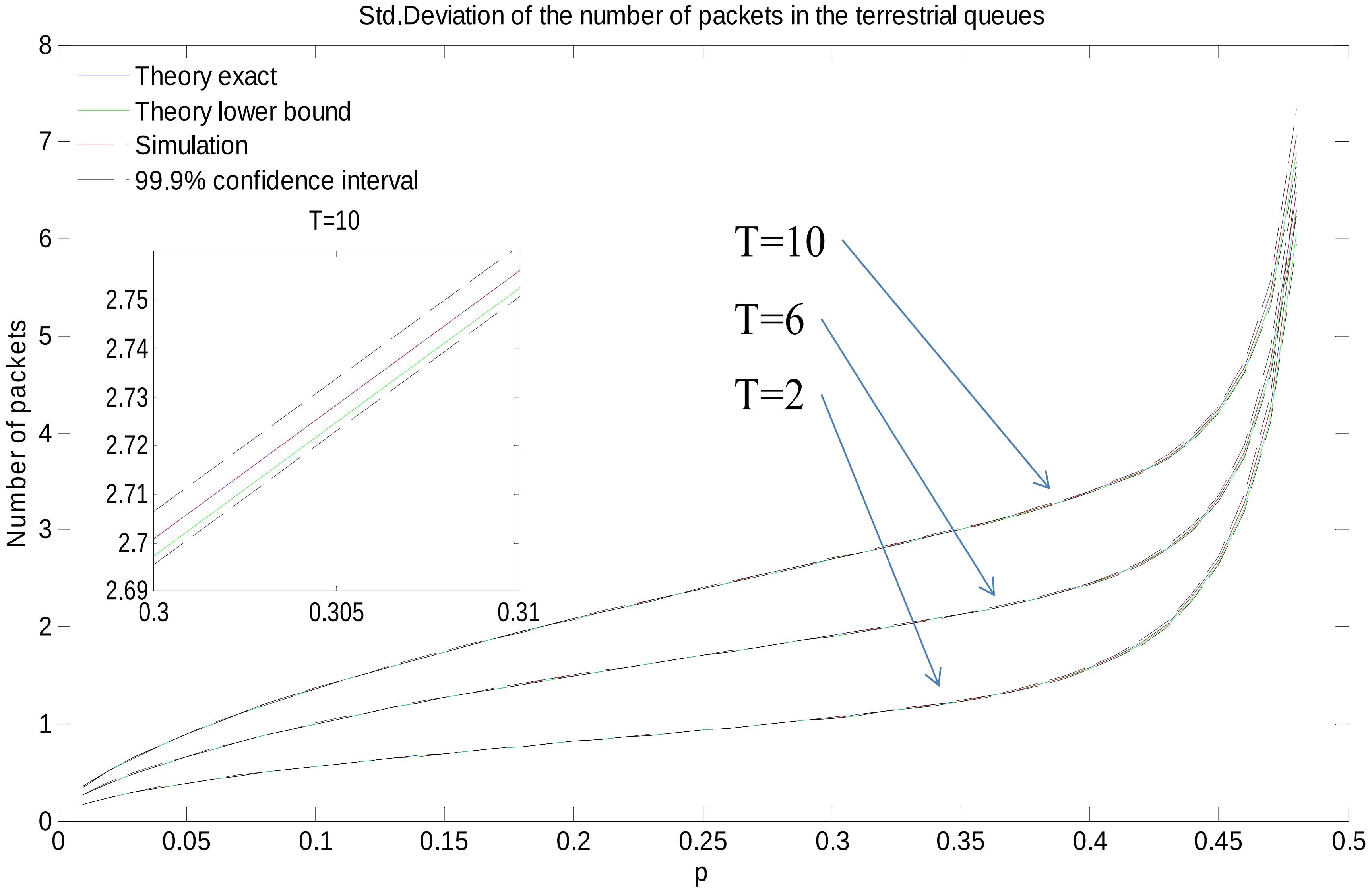
An easier to solve approximation

- Eliminate a queue!
  - If (m_0,m_1) has an equivalent state of the form $(0, \mu_1)$, redirect all transitions.
  - Otherwise, it has an equivalent states of the form
    $$\left\{(\mu_0, \mu_1) : \mu_1 = kT + b, \mu_0 \le b < T, k = 1, 2, 3, \ldots\right\}$$
  - redirect them to $(0, \mu_0+kT+b)$
  - (sometimes an upperbound, sometimes a lower bound… just an approximation)
  - The submatrix blocks are now just TxT

Average Number of Packets in an earth station



Variance of the number of packets in an earth station



Conclusions

- We find a systematic approach to find upper or lower bounds for the performance measures of impossible to solve highly dimensional infinite Markov chains.
- The method constructs a modified solvable Markov chain by redirecting some transitions between states in such a way that the bound can be guaranteed.
- We used the method to find exact analytical results for the average and the variance of the number of packets in the earth station queues of a satellite system that groups packets into frames for onboard switching.

GRACIAS! (THANK YOU ☺)

Questions?