Steady-State Behavior of CTMCs

Suppose we want to look at the long-term behavior of a CTMC.

A solution can be derived from the differential equation $\frac{d}{dt}P(t) = P(t)Q.$

Recall that

$$\pi(t) = \pi(0)P(t).$$

Taking the limit as $t \to \infty$,

$$\lim_{t\to\infty} \pi(t) = \lim_{t\to\infty} \pi(0)P(t),$$

and differentiating

$$\lim_{t\to\infty} \tfrac{d}{dt}\pi(t) = \lim_{t\to\infty} \tfrac{d}{dt}\pi(0)P(t)$$

so $\quad 0 = \lim_{t\to\infty} \pi(0)\tfrac{d}{dt}P(t)$

$$0 = \tfrac{d}{dt}\pi(0)\lim_{t\to\infty} P(t) = \pi(0)\lim_{t\to\infty}\tfrac{d}{dt}P(t) = \lim_{t\to\infty}\pi(0)P(t)Q = \lim_{t\to\infty}\pi(t)Q = \pi^{*}Q$$

Steady-State Behavior of CTMCs

Suppose we want to look at the long-term behavior of a CTMC.

A solution can be derived from the differential equation $\frac{d}{dt}P(t) = P(t)Q.$

In steady state, $\frac{d}{dt}P(t) = 0.$ To see this, recall

$$\pi(t) = \pi(0)P(t).$$

Taking the limit as $t \to \infty$,

$$\lim_{t\to\infty}\pi(t) = \lim_{t\to\infty}\pi(0)P(t),$$

and differentiating

$$\lim_{t\to\infty}\tfrac{d}{dt}\pi(t) = \lim_{t\to\infty}\tfrac{d}{dt}\pi(0)P(t)$$

so $\quad 0 = \lim_{t\to\infty}\pi(0)\tfrac{d}{dt}P(t)$

$$0 = \tfrac{d}{dt}\pi(0)\lim_{t\to\infty}P(t) = \pi(0)\lim_{t\to\infty}\tfrac{d}{dt}P(t) = \lim_{t\to\infty}\pi(0)P(t)Q = \lim_{t\to\infty}\pi(t)Q = \pi^{*}Q$$

Steady-State Behavior of CTMCs via Flow Equations

Another way to arrive at the equation $\pi^* Q = 0$, where $\pi^* = \lim_{t \to \infty} \pi(t)$, is to use the flow equations. The global flow rate of transitions into a state i must equal the global flow rate of transitions out of state i.

The global flow rate from state i to state j is simply $\pi_i q_{ij}$, which is the probability of being in state i times the rate at which transitions from i to j take place.

Flow into state i : $\sum_{\substack{j=1 \\ j \neq i}}^{n} \pi_j q_{ji}$

Flow out of state i :

(1) $\sum_{\substack{j=1 \\ j \neq i}}^{n} \pi_i q_{ij} = \pi_i \sum_{\substack{j=1 \\ j \neq i}}^{n} q_{ij} = \pi_i (-q_{ii})$

(2) $\sum_{\substack{j=1 \\ j \neq i}}^{n} \pi_j q_{ji} = \pi_i (-q_{ii})$

Flow in equals flow out

Re-arrange and combine sums

(3) $\sum_{\substack{j=1 \\ j \neq i}}^{n} \pi_j q_{ji} + \pi_i q_{ii} = 0$

(4) $\sum_{j=1}^{n} \pi_j q_{ji} = 0$

In matrix form, for all i, we get $\pi Q = 0$.

Steady-State Behavior of CTMCs, cont.

This yields the elegant equation $\pi^* Q = 0$, where $\pi^* = \lim_{t \to \infty} \pi(t)$, the steady-state probability distribution. If the CTMC is irreducible, then π^* can be computed with the constraint that $\sum_{i=1}^{n} \pi_i^* = 1$.

If the CTMC is not irreducible, then more complex solution methods are required.

Notice that for irreducible CTMCs, the steady-state distribution is independent of the initial-state distribution.

Steady-State Solution Methods

It is convenient to write the linear equations for the steady-state solution as $\pi^* Q = 0$.

Solving $\pi^* Q = 0$ will not give a unique solution; we must add the constraint that $\sum_{i=1}^{n} \pi_i^* = 1$ to guarantee uniqueness. This leads to two approaches:

- Replace i^{th} column of Q with $(1, 1, \ldots, 1)^T$ to form $\tilde{Q}$ and solve $\pi^* \tilde{Q} = e_i^T$ to guarantee a unique solution. This typically leads to worse numerical properties.
- Find any solution to $\pi^* Q = 0$ and then normalize the results. Not all solution methods can find a non-unique solution.

Direct Methods for Computing $\pi^* Q = 0$

A simple and useful method of solving $\pi^* Q = 0$ is to use some form of Gaussian elimination. This has the following advantages:

- Numerically stable methods exist
- Predictable performance
- Good packages exist
- Must have a unique solution

The disadvantage is that many times Q is very large (thousands or millions of states) and very sparse (ones or tens of nonzero entries per row). This leads to very poor performance and extremely large memory demands.

Stationary Iterative Methods

Stationary iterative methods are solution methods that can be written as

$$\pi^{(k+1)} = \pi^{(k)} M,$$

where M is a constant (stationary) matrix. Computing $\pi^{(k+1)}$ from $\pi^{(k)}$ requires one vector-matrix multiplication, which is one *iteration*.

Recall that for DTMCs, $\pi^* = \pi^* P$.

Converting this to an iterative method, we write

$$\pi^{(k+1)} = \pi^{(k)} P.$$

This is called the *power method*.

- Simple, natural for DTMCs
- Gets to an answer slowly
- Can find a non-unique solution

Power method is an iterative technique to find solutions to

$$\underset{\text{DTMC}}{\pi = \pi P} \quad \text{and} \quad \underset{\text{CTMC}}{\pi Q = 0}$$

DTMC

- simple intuition from $\pi(k+1) = \pi(k) \cdot P$

  if $\lim_{k \to \infty} \pi(k) = \pi^*$ exists, then

  repeating $\pi(k+1) = \pi(k) \cdot P$ for many k will get us there

Simplest example of power method

repeated application of vector - matrix (or matrix - vector) product

Background

Let A be an $n \times n$ matrix, λ is an **eigenvalue** and x_λ is an **eigenvector** for λ if

$$Ax_\lambda = \lambda \cdot x_\lambda$$

Eigenvectors are very important in analysis of linear algebra.

A may have multiple eigenvalues and hence multiple eigenvectors

Enumerate the eigenvalues so that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$

interested in case where λ_1 strictly dominates others

① For a stochastic matrix (e.g. transition probability matrix)
$|\lambda_1| = 1$

Definition: An eigenvector V is normalized if the coordinate of largest magnitude $= 1$

NOTE: Given $v = [v_1, v_2, \ldots, v_n]^T$ and $v_j = c$ has largest magnitude

then $\tilde{v} = \frac{1}{c}[v_1, v_2, \ldots, v_n]^T$ is normalized

THEOREM Suppose $n \times n$ matrix A has n eigenvalues ordered by

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$

IF x_0 is chosen "appropriately" then the sequences

$$\{ z_k = [x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}]^T \} \text{ and } c_k$$

generated recursively by

$$y_k = A z_k \quad \text{and} \quad z_{k+1} = \frac{1}{c_{k+1}} \cdot y_k$$

where c_{k+1} is component of z_k with largest magnitude

then

$$\lim_{k \to \infty} z_k = v_1 \quad \text{and} \quad \lim_{k \to \infty} c_k = \lambda_1$$

In the DTMC case we recognize that

$$P^T \cdot \Pi^{*T} = \Pi^{*T}$$

means Π^{*T} is an eigenvector of P^T, with eigenvalue 1

and we are told by experts that the dominent eigenvalue of P is 1.

$$\Rightarrow P^T \cdot \Pi(k)^T = \Pi(k+1)^T$$

will converge under assumptions of Power method theorem

We can use technique to get at stationary distribution of CTMC state occupancy vector

BUT NEED TO MAKE A SHORT DETOUR FIRST

Embedded in <u>any</u> CTMC is a DTMC that describes transitions, when they occur

– ignores time in each state

Given Q, create P_Q

$$P_Q = -\frac{1}{q_{11}}\begin{bmatrix} 0 \\ q_{12} \\ q_{13} \\ \vdots \\ q_{1n} \end{bmatrix} - \frac{1}{q_{22}}\begin{bmatrix} q_{21} \\ 0 \\ q_{23} \\ q_{24} \\ \vdots \\ q_{2n} \end{bmatrix} - \cdots - \frac{1}{q_{n,n}}\begin{bmatrix} q_{n,1} \\ q_{n,2} \\ \vdots \\ q_{n,n-1} \\ 0 \end{bmatrix}$$

what's this?? * normalize each transition rate, turning into a probability

$q_{i,j} \longrightarrow \Pr\{\text{going to state } j \mid \text{leaving state } i\}$

N.B. DTMC never stays in same state

Let α be the solution to $\alpha = \alpha P_Q$, i.e., the state occupancy probability vector of the embedded DTMC.

Now for the magic part

Let m_i be the mean holding time of the CTMC in state i

$$m_i = \frac{1}{q_{i,i}}$$

note: no solutions of anything involved

$$\lim_{t \to \infty} \Pi_i(t) = \frac{\alpha_i \cdot m_i}{\sum_{j=1}^{n} \alpha_j \cdot m_j}$$

Think about — α_i is fraction of DTMC transitions departing i
m_i is mean time in i, fraction is relative proportion of time in i

So? 1. From Q, create P_Q

2. Can use any solution method to get

$$P_Q \cdot \alpha = \alpha$$

but particularly, the power method

3. Compute each $m_i = -\frac{1}{q_{i,i}}$

4. Compute each

$$\pi_i = \frac{m_i \cdot \alpha_i}{\sum_{j=1}^{n} m_j \cdot \alpha_j}$$

DONE!

Convergence

iterative method converges if

$$\lim_{k \to \infty} \| \pi^{(k)} - \pi^* \| = 0$$

Understanding convergence very important.

if λ_1 is the largest eigenvalue for M, and λ_2 is next largest, theory shows that the convergence rate is $\left| \frac{\lambda_2}{\lambda_1} \right|$ which means, approximately that the error

$$\| \pi^{(k)} - \pi^* \| \leq \| \pi^{(k-1)} - \pi^* \| \cdot \left(\frac{\lambda_2}{\lambda_1} \right)$$

in a matrix associated with a Markov chain, $\lambda_1 = 1$, so convergence described by λ_2

Gauss-Seidel method

iterative solution for solving $Ax = b$

when A is $n \times n$ matrix

Defined by $$L \cdot x^{(k+1)} = b - U x^{(k)}$$

where L is lower triangular part of A, U is strictly upper triangular

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ & & \cdots & & \\ a_{n,1} & a_{n,2} & \cdots & & a_{n,n} \end{bmatrix}}_{L} + \underbrace{\begin{bmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & 0 & a_{23} & \cdots & a_{2,n} \\ & & & \cdots & \\ 0 & 0 & \cdots & & 0 \end{bmatrix}}_{U}$$

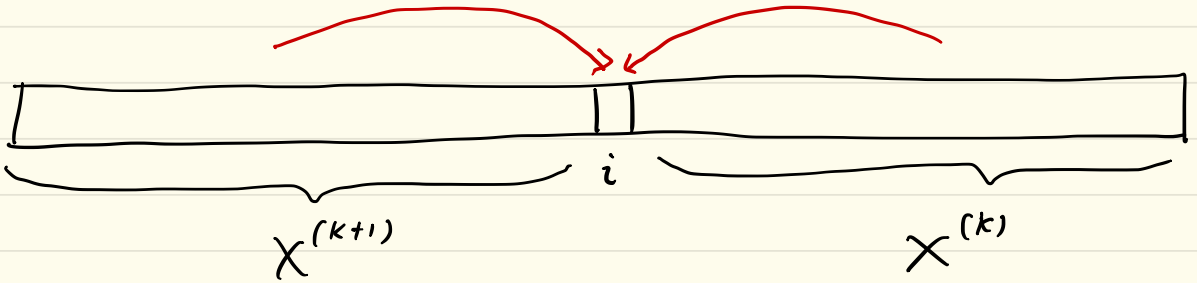
$$L\,x^{(k+1)} = b - U x^{(k)}$$

Gauss-Seidel solves left hand side of equation for $x^{(k+1)}$, given $x^{(k)}$

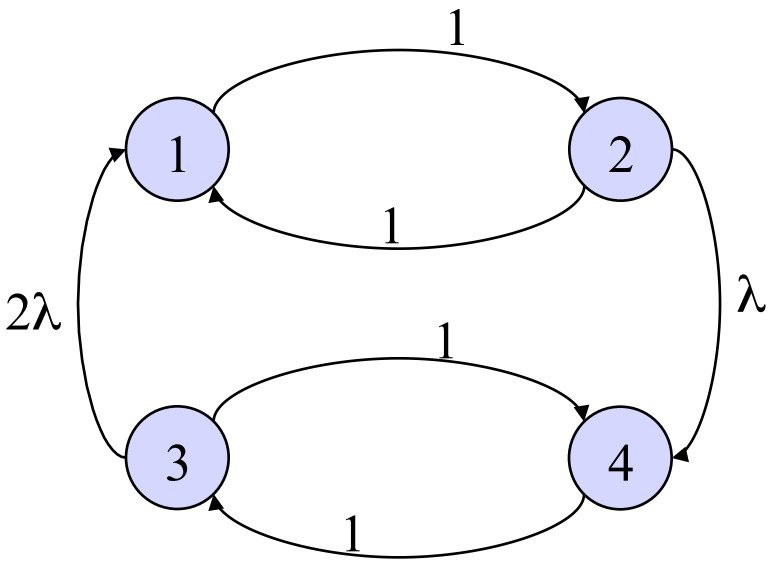
Using forward substitution

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{i,j}\, x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j}\, x_j^{(k)} \right)$$

Notice — we can transform $x^{(k)} \rightarrow x^{(k+1)}$ "in place"



Numerical Example



The above CTMC serves as an example of some of the issues involved in numerical solution, where $\lambda << 1$.

We illustrate an experiment of solving the CTMC using Gauss-Seidel, ranging λ over many values, and stopping when $\left\|\pi^{(k+1)} - \pi^{(k)}\right\| < 10^{-7}$.

Analysis

We range λ from 10^{-7} to 1.

The measure we examine for this model is the steady-state probability of being in state 1 or 2, i.e., $\pi_1^* + \pi_2^*$.

For $\lambda << 1$, it is easy to approximate by observing that it is twice as likely to be in states 1 and 2 than in 3 and 4, so the measure is approximately 2/3.

For our experiments, we used the $\left\| \pi^{(k+1)} - \pi^{(k)} \right\| < 10^{-7}$ stopping criterion.

Results

λ	number of iterations	computed solution	exact solution
10^{-7}	1	0.5000	0.6667
10^{-6}	305432	0.6000	0.6667
10^{-5}	107298	0.6600	0.6667
10^{-4}	18408	0.6660	0.6667
10^{-3}	2612	0.66648	0.6666
10^{-2}	342	0.66556	0.66556
10^{-1}	46	0.6562495	0.65625
1	10	0.600000	0.6

Discussion

- For $\lambda = 10^{-7}$, the first Gauss-Seidel iteration resulted in $\left\|\pi^{(2)} - \pi^{(1)}\right\| < 10^{-7}$, so it stopped immediately. 0.5 is the initial guess.

- For smaller λ, more iterations are required, and less accuracy is achieved.

- For larger λ, much fewer iterations are required, and very high precision is achieved.

- The property of the CTMC has more effect on the rate of convergence than the size of the CTMC does.

A Second Example: Multiprocessor Failure/Repair Model

System Description:

1) n processors, 1 needed for system to be up.
2) Each processor fails with rate λ.
3) Processors repaired (one at a time) with rate μ.
4) After failure, reboot with prob(1 - c), reconfigure with probability c
5) Reconfiguration time exponential (rate δ) results in system with one fewer processor.
6) Reboot (exponential, rate γ) results in fully functioning system.
7) Second failure during reconfiguration, or all processors failed, results in crash.
8) Crash repair (exponential, rate β) results in fully functioning system.
9) System unavailable if reboot or crash.

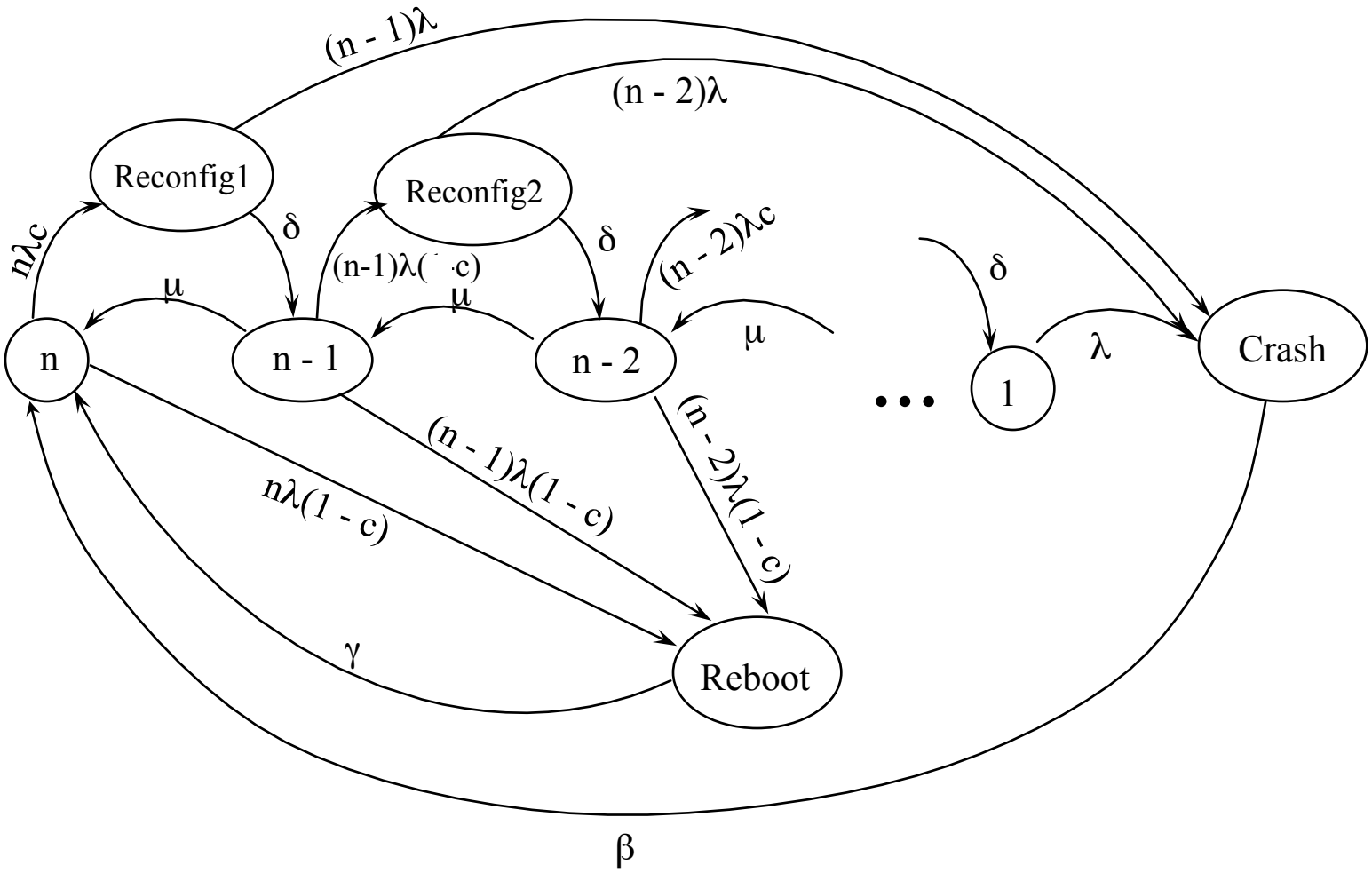
ECE/CS 541: Computer System Analysis. ©2006 William H. Sanders. All rights reserved. Do not copy or distribute to others without the permission of the author.

Nominal System Parameter Values

$\lambda = 1/(6000 \text{ hours})$	- processor failure rate
$\mu = 1/(1 \text{ hour})$	- processor repair rate
$\delta = 1/(.01 \text{ hours})$	- processor reconfiguration rate
$\gamma = 1/(.5 \text{ hours})$	- reboot rate
$\beta = 1/(20 \text{ hours})$	- crash repair rate
$c = .99$	- coverage probability

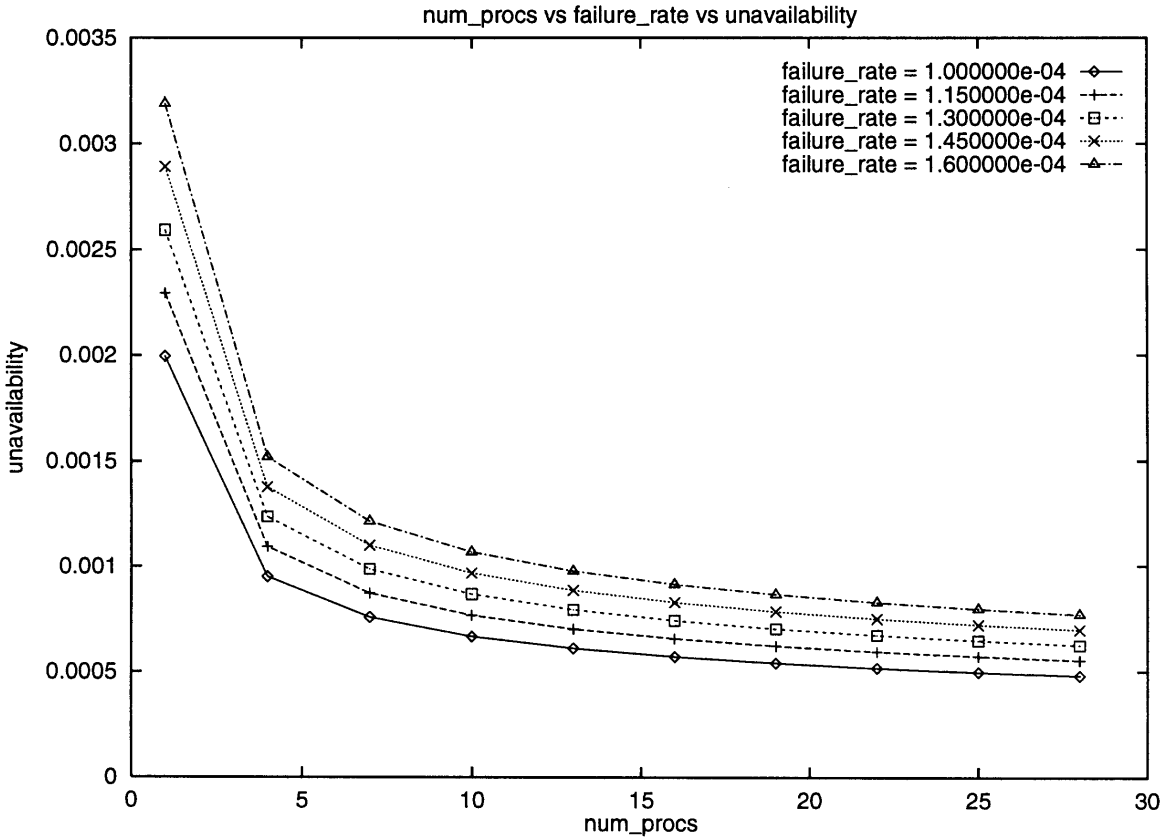
ECE/CS 541: Computer System Analysis. ©2006 William H. Sanders. All rights reserved. Do not copy or distribute to others without the permission of the author.

State Transition Rate Diagram



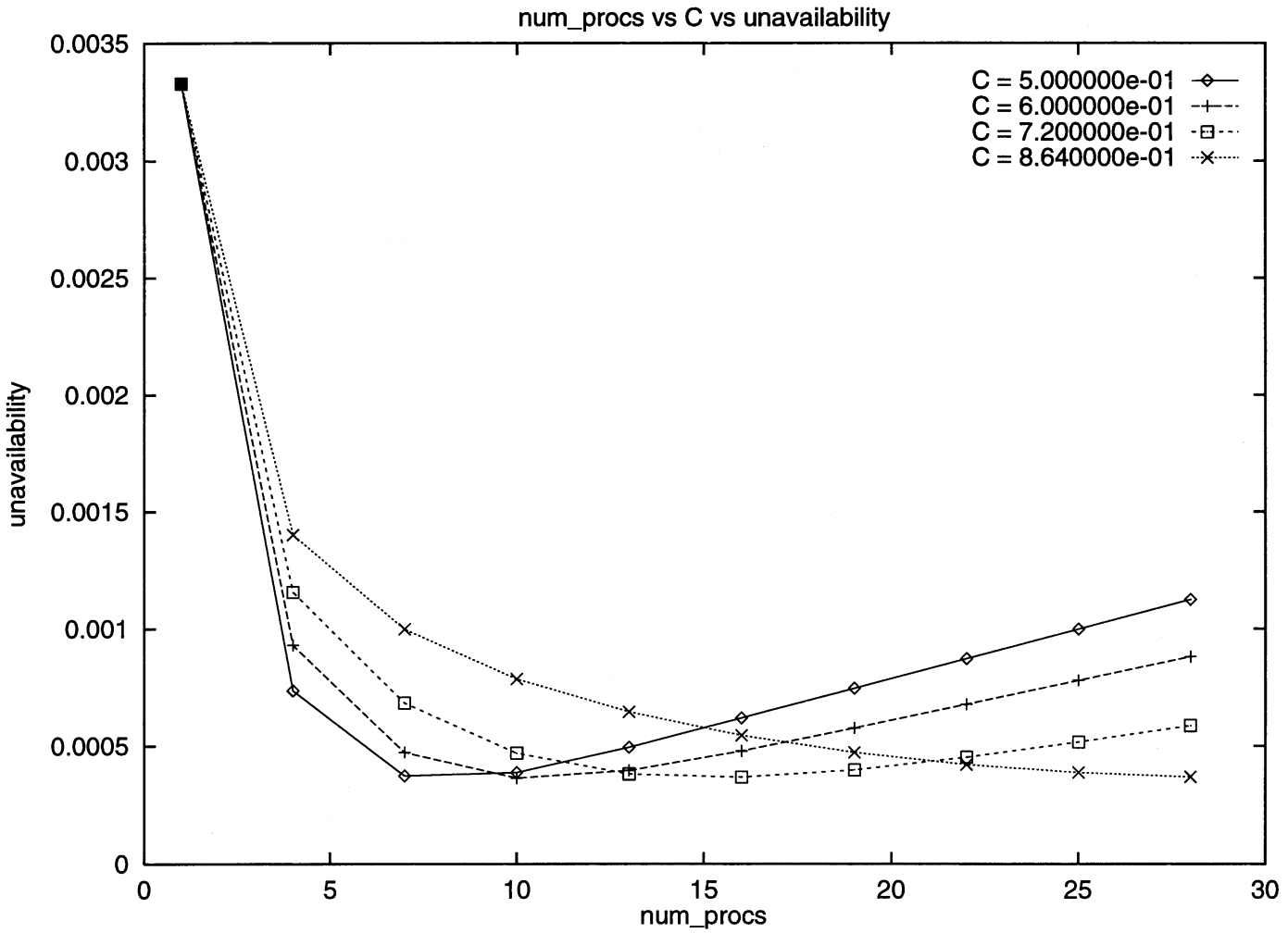
ECE/CS 541: Computer System Analysis. ©2006 William H. Sanders. All rights reserved. Do not copy or distribute to others without the permission of the author.

Unavailability vs. Number of Processors (varying failure rate)



ECE/CS 541: Computer System Analysis. ©2006 William H. Sanders. All rights reserved. Do not copy or distribute to others without the permission of the author.

Unavailability vs. Number of processors (varying coverage)



ECE/CS 541: Computer System Analysis. ©2006 William H. Sanders. All rights reserved. Do not copy or distribute to others without the permission of the author.

Review of State-Based Modeling Methods

- Random process
  - Classifications: continuous/discrete state/time
- Continuous-time Markov chain
  - Definition
  - Properties
  - Exponential random variable
    - Memoryless property
    - Minimum of two exponentials
    - Competing exponentials
    - Event/failure/hazard rate
  - State-transition-rate matrix
  - Transient solution
  - Steady-state behavior: $\pi^* Q = 0$
- Discrete-time Markov chain
  - Definition
  - Transient solution
  - Classification: reducible/irreducible
  - Steady-state solution
- Solution methods
  - Direct
  - Iterative: power, Gauss-Seidel
  - Stopping criterion
  - Example

ECE/CS 541: Computer System Analysis. ©2006 William H. Sanders. All rights reserved. Do not copy or distribute to others without the permission of the author.