Analysis of System Performance

IN2072

Chapter 5 – Analysis of Non Markov Systems

Dr. Alexander Klein
Prof. Dr.-Ing. Georg Carle

Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Non Markov Systems

- Content:
  - (Non Memory-less) systems
    - Embedded markov chain
    - General distributed service times
  - Waiting system M/GI/1-s (**Infinite** number of sources)
    - State probabilities
    - Transition probabilities
    - Waiting time distribution
    - Impact of variance of the service process on system performance
  - Waiting system GI/M/1-s (**Infinite** number of sources)
    - State probabilities
    - Transition probabilities
    - Waiting time distribution

Non Markov Systems

- Markov Systems:
  - Arrival process is memory-less.
  - Service process is memory-less.

⇨ System is memory less at any given point in time.

- Non Markov Systems:
  - Have one component which is memory-less AND
  - one component which is NOT memory-less.

⇨ System becomes memory-less either at the time of an arrival or at the time a service is completed.

⇨ M / GI / 1 and GI / M / 1

- Idea:

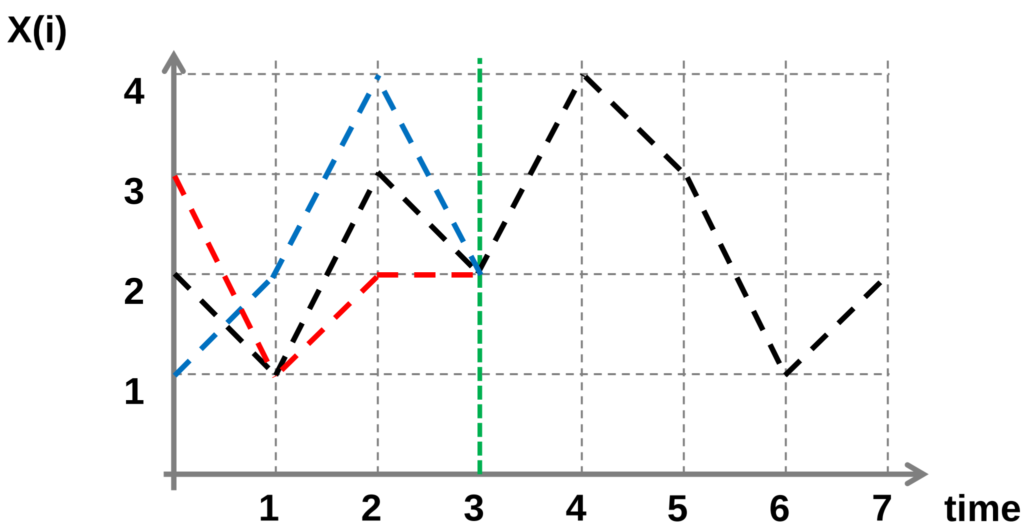
  Analyze the system when it is memory-less.

Embedded Markov Chain

Assumption:

State discrete stochastic process $\{X(t), t > 0\}$ which is memory-less (markovian) at time $\{t_n, n = 0,1,\ldots\}$.

$$\Rightarrow \quad P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, \ldots, X(t_0) = x_0\} = P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\}, t_0 < t_1 < \ldots < t_n < t_{n+1}.$$



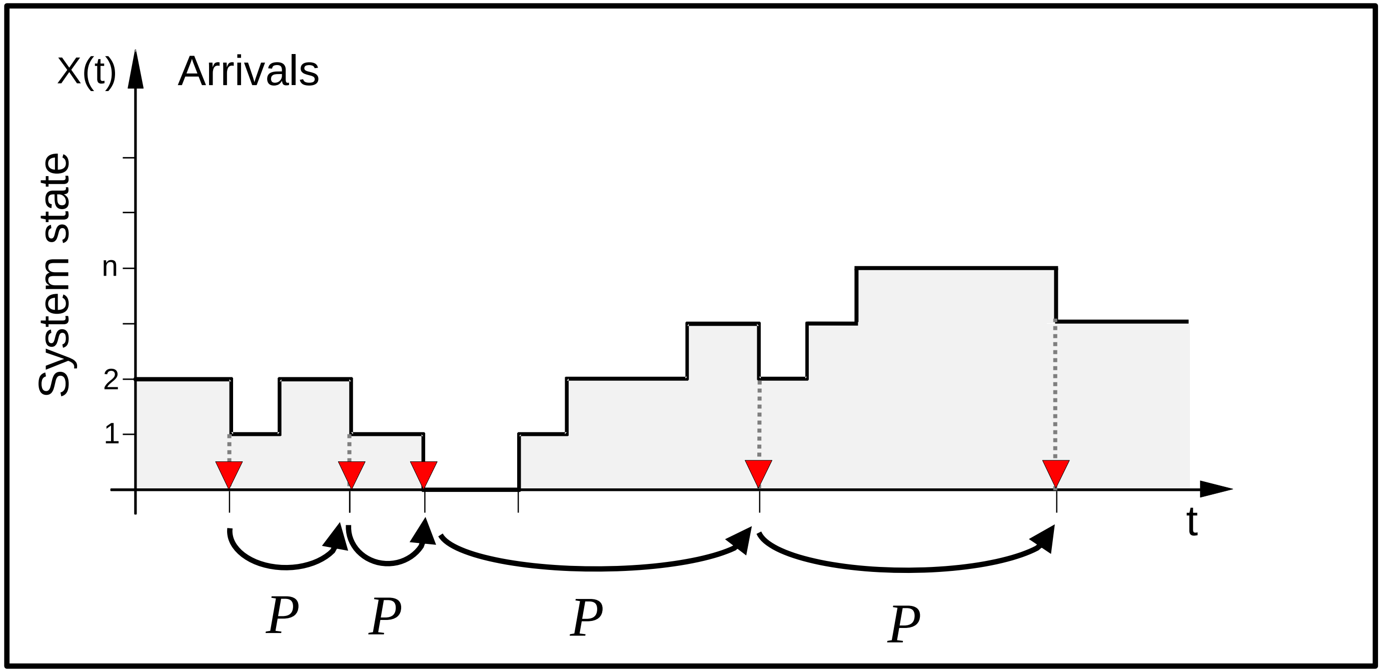
Embedded Markov Chain

Characteristics:

- The future development of the process only depends on its current state.
- Knowledge about the current state $X(t_n)$ at time t_n is sufficient to calculate its consecutive states $X(t_{n+1}), X(t_{n+2}), \ldots$
- Due to the state discrete nature of the process, the states $\{X(t_n)\}$ represent a chain.
- The chain is a markov chain according to our assumption which states that the process is memory-less at time $\{t_n, n = 0,1,\ldots\}$.

Embedded Markov Chain

- Points of regeneration:



⇨ Embedded points in time when the system is memory-less (markovian).

Embedded Markov Chain

- The state probabilities at the embedded points t_n can be described by a state probability vector.

$$\mathrm{X}_n = \{x(i,n), \quad i = 0,1,\ldots\}$$

$$x(i,n) = P\{X(t_n) = i\}$$

State transition probability matrix P which describes the relation between any consecutive state probability vectors X_n and X_{n+1} at embedded points t_n and t_{n+1} is given by

$$P = \{p_{ij}\}$$

$$p_{ij} = P\{X(t_{n+1}) = j \mid X(t_n) = i\}, \quad i, j = 0,1,2,\ldots$$

$$\mathrm{X}_{n+1} = \mathrm{X}_n \cdot P$$ Relation of consecutive state transition vectors.

$$\mathrm{X} = \mathrm{X} \cdot P$$ Probability vector in stationary state.

Embedded Markov Chain

- Characteristics:
  - The probability vector of the markov chain in steady state is given by the left eigenvector of the transition probability matrix P of eigenvalue 1.
  - Embedded markov chain is usually applied if only one component is non-memory less.
  - Embedded points are located where this component becomes memory less.

- **M / GI / 1**
  - Service process is the only non-memory less component.
  - State process becomes memory less after service completion.

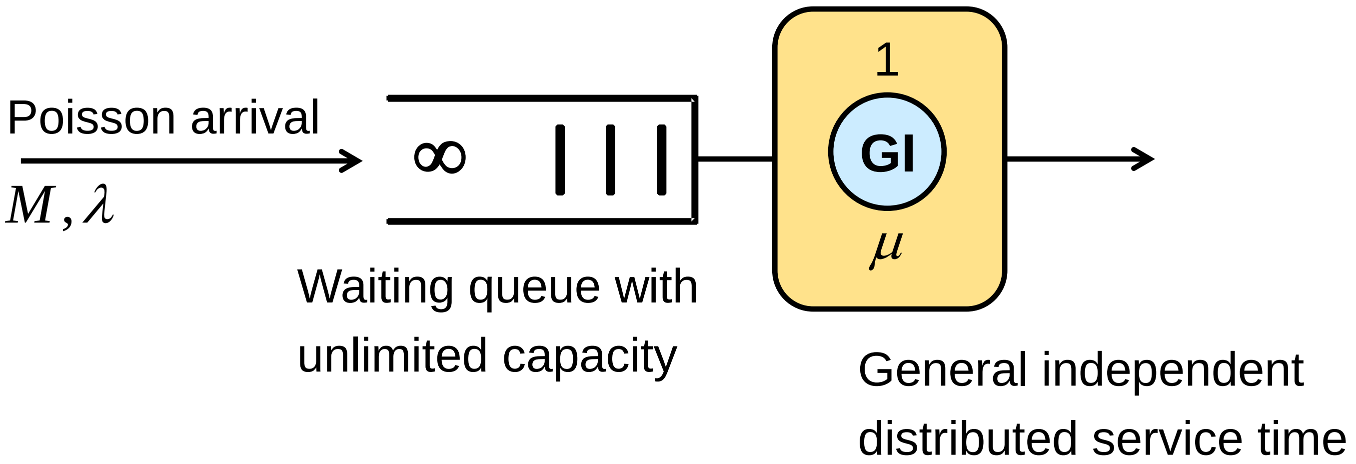
  ⇨ Embedded points are located directly after a service completion.

- **GI / M / 1**
  - Arrival process is the only non-memory less component.
  - State process becomes memory less after an arrival.

  ⇨ Embedded points are located directly before an arrival occurs.

M / GI / 1 – Waiting system

- Model:



- Model and parameter description:
  - M / GI / 1 – ∞ (No jobs are blocked!)
  - Arrival process is a Poisson process with an exponential distributed inter-arrival time A.
  - Service time B is general independent (GI) distributed.
  - Jobs that arrive at a point in time when all service units are busy, are queued and served in FIFO order as soon as a free serving unit is available.

M / GI / 1 – Waiting system

- Arrival process:

  Arrival rate λ

  Average number of arriving jobs per time unit.

$$A(t) = P(A \le t) = 1 - e^{-\lambda t}, \qquad E[A] = \frac{1}{\lambda}$$

- Service process:

  Service rate μ

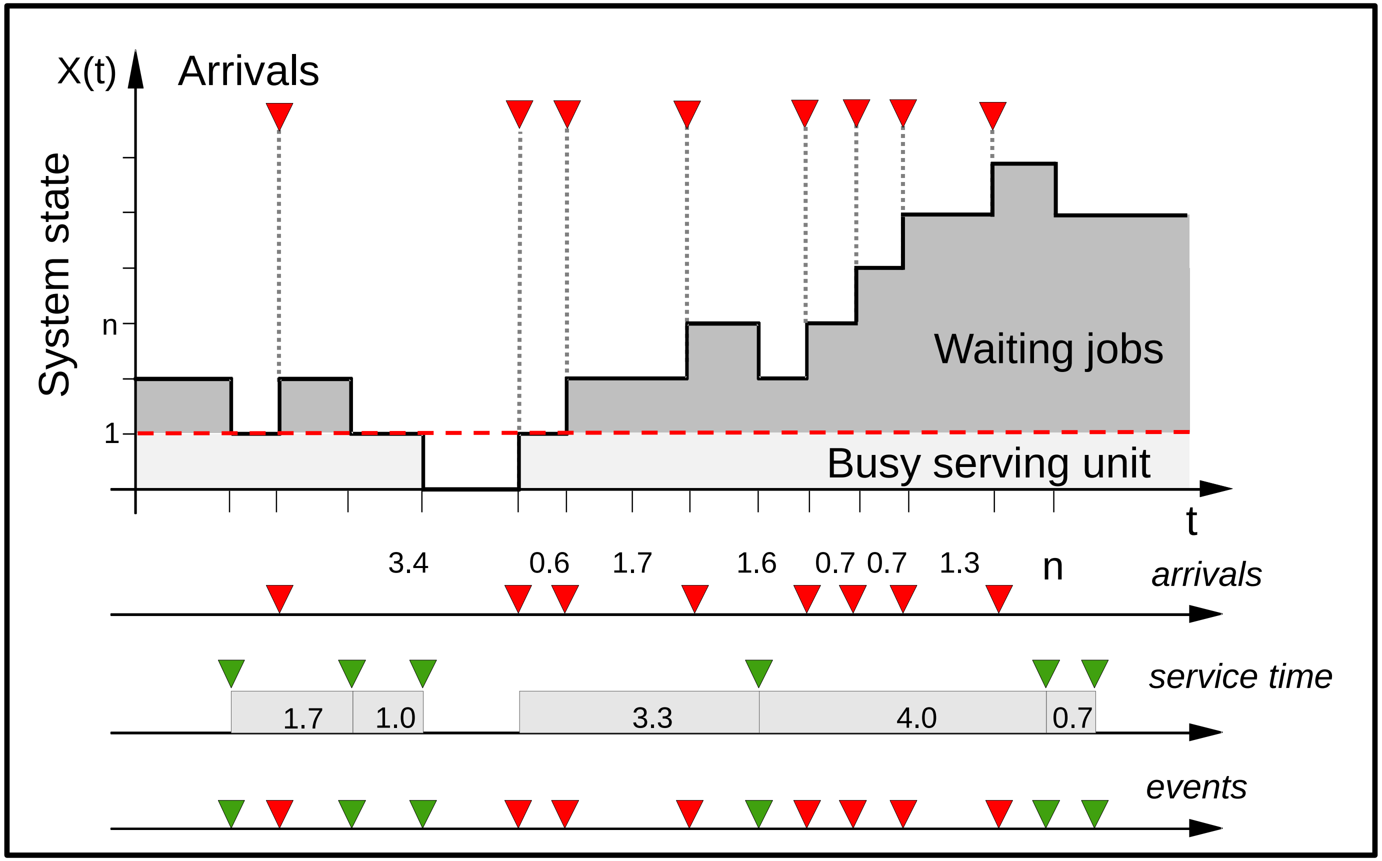
  Average number of service completions.
  (assuming the service unit only has two states – idle or busy)

$$B(t) = P(B \le t), \qquad E[B] = \frac{1}{\mu}$$

- System:
  - Waiting system
  - Waiting queue with unlimited capacity
  - Queuing strategy – First In First Out (FIFO)

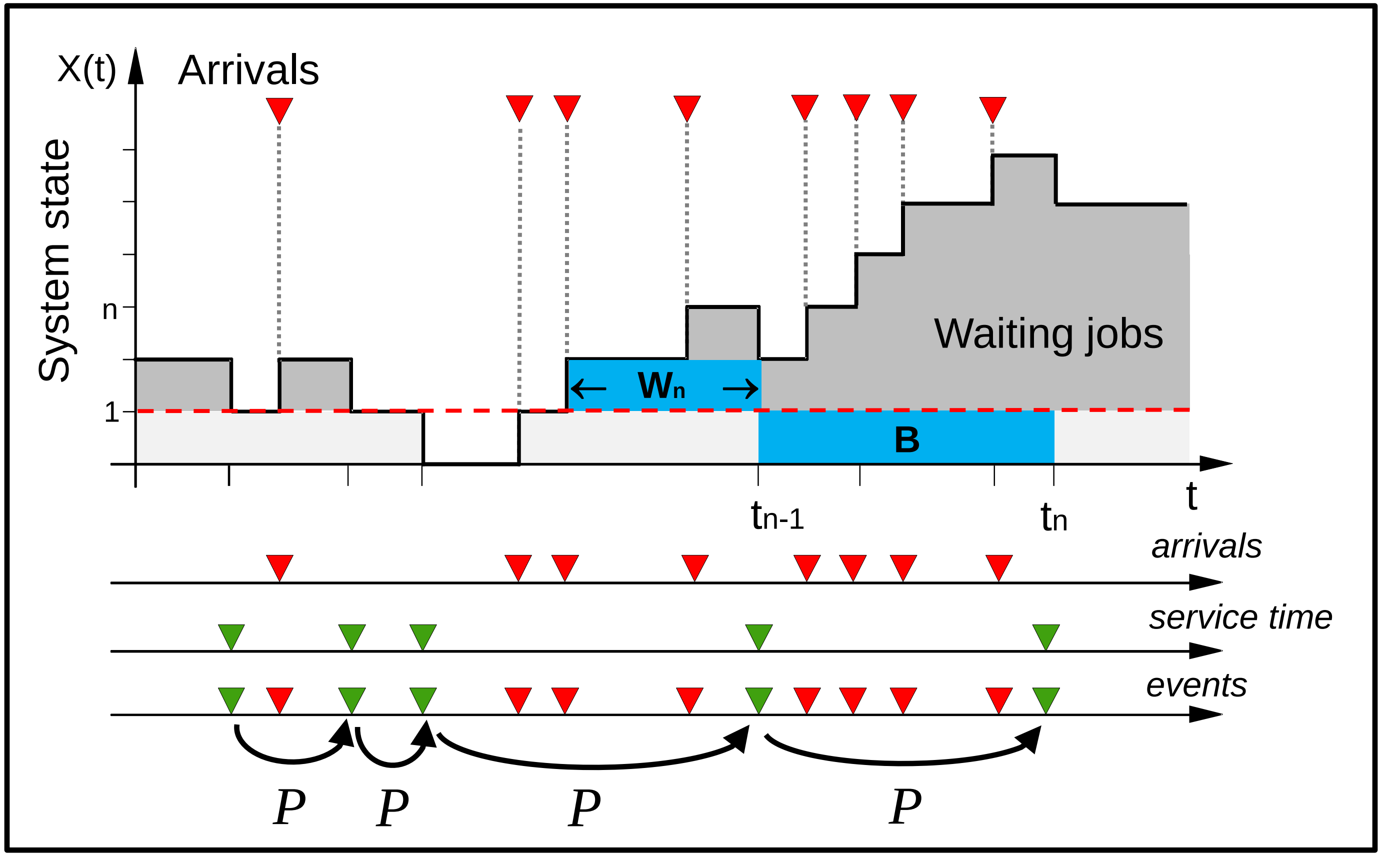
M / GI / 1 – Waiting system

- State space:
  - Random variable $X(t)$ describes the number of (waiting and currently served) jobs in the system.
  - State process is state discrete and time continuous stochastic process



M / GI / 1 – Waiting system

- State space:
  - Random variable $X(t)$ describes the number of (waiting and currently served) jobs in the system.
  - State process is state discrete and time continuous stochastic process



M / GI / 1 – Waiting system

- **Embedded points:**
  - State process becomes memory less at the time of a service completion.
  - Embedded points of the markov chain are located directly after service completion.

- **Embedded Markov Chain:**
  - The point in time of the n_{th} embedded point corresponds to the n_{th} service completion.
  - The sequence of the process states

$$\{X(t_0), X(t_1), \ldots, X(t_n), X(t_{n+1}), \ldots\}$$

at these points represent the embedded markov chain.

M / GI / 1 – Waiting system

- Analysis:
  - Introduce a random variable Γ which describes the number of arrivals during a service duration.

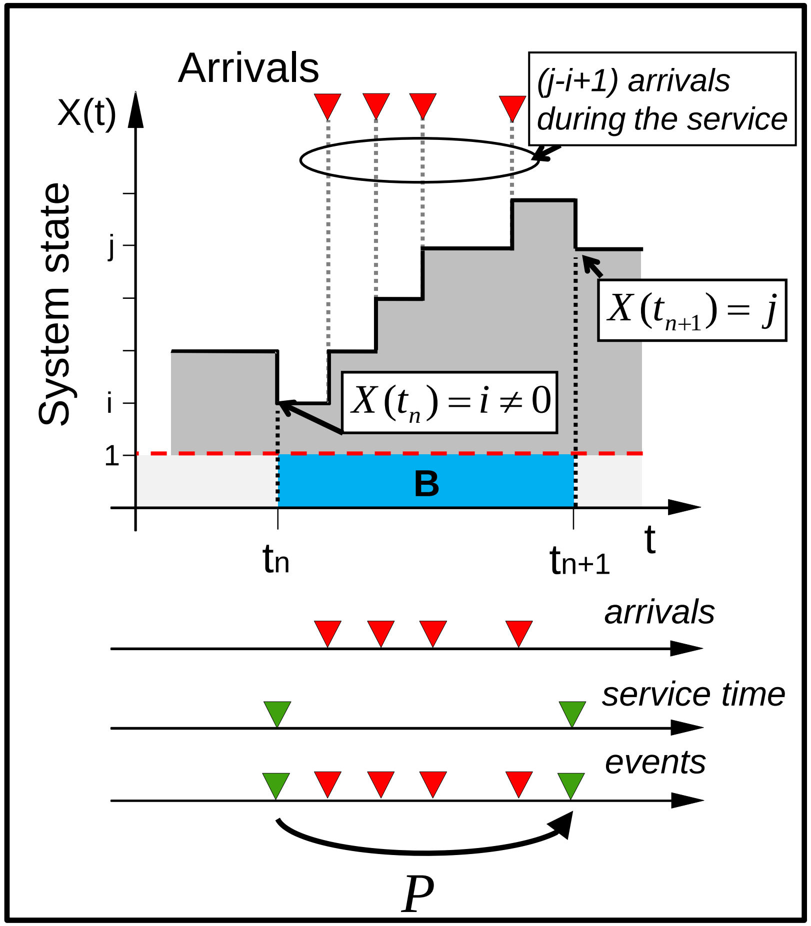
$$\gamma(i) = P(\Gamma = i)$$

with $$\Gamma_{GF}(z) = \sum_{i=0}^{\infty} \gamma(i) \cdot z^i$$ Generation Function

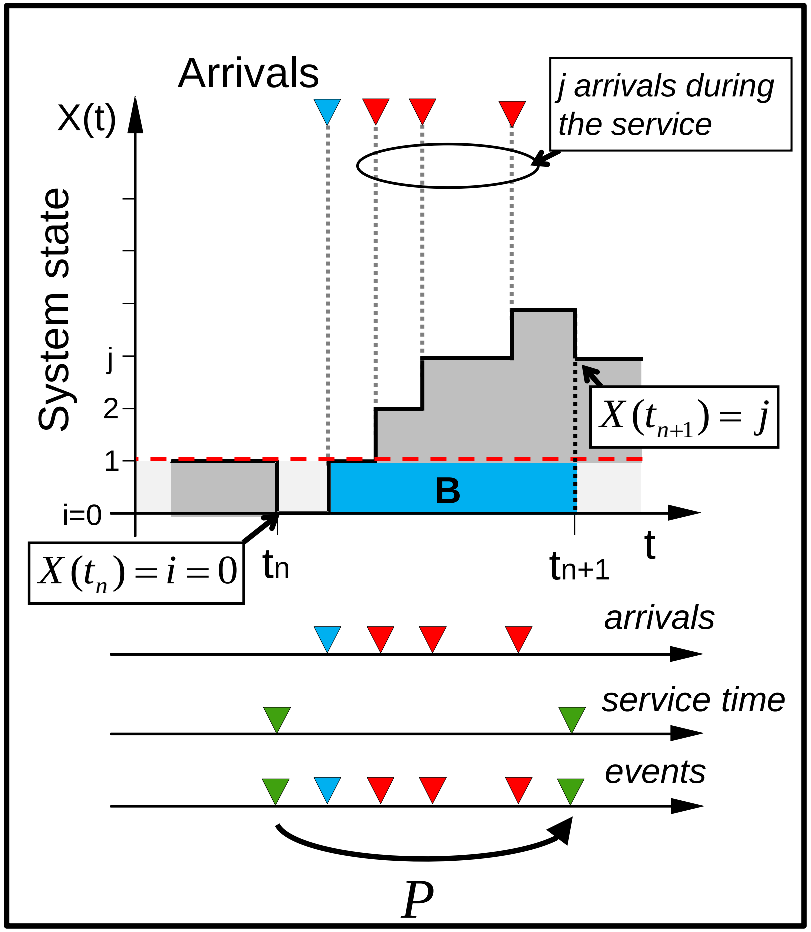
$$\Rightarrow \quad E[\Gamma] = \left. \frac{d\Gamma_{GF}(z)}{dz} \right|_{z=1} = \lambda \cdot E[B] = \rho$$

M / GI / 1 – Waiting system

- Transition behavior:



State transition i→j with i≠0



State transition i=0 →j

M / GI / 1 – Waiting system

- State transition:

  State transition between consecutive embedded points t_n and t_{n+1}.

$$\Rightarrow \quad p_{ij} = P\{X(t_{n+1}) = j \mid X(t_n) = i\}$$

Case 1: System not empty (i $\neq$ 0) at time

- At time t_n are i jobs in the system.
- The service of the next job starts directly after t_n.
- j jobs remain in the system at time t_{n+1}.

$\Rightarrow$ (j-i+1) jobs have to arrive during the interval $[t_n; t_{n+1}]$ which corresponds (in this case) to the service duration.

$$\Rightarrow \quad p_{ij} = \gamma(j - i + 1), \quad i = 1, 2, \ldots, \quad j = i - 1, i, \ldots$$

M / GI / 1 – Waiting system

Case 2: System is empty (i = 0) at time t_n

- At time t_n are i=0 jobs in the system.
- The service of the next job starts directly after the arrival of the job.
- j jobs remain in the system at time t_{n+1}.

⇨ j jobs have to arrive during the service duration.

⇨ $p_{0j} = \gamma(j), \qquad j = 0,1,\ldots$

⇨ $$P = \{p_{ij}\} = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \cdots \\ \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \cdots \\ 0 & \gamma(0) & \gamma(1) & \gamma(2) & \cdots \\ 0 & 0 & \gamma(0) & \gamma(1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

M / GI / 1 – Waiting system

- The state probabilities at the embedded points t_n can be described by a state probability vector.

⇨ $$\mathrm{X}_n = \{x(0,n), x(1,n), \ldots, x(j,n), \ldots \qquad j = 0,1,\ldots\}$$

⇨ $$x(j,n) = P\{X(t_n) = j\}$$

State transition probability matrix P which describes the relation between any consecutive state probability vectors X_n and X_{n+1} at embedded points t_n and t_{n+1} is given by

⇨ $$\mathrm{X}_{n+1} = \mathrm{X}_n \cdot P$$ Relation of consecutive state transition vectors.

A start vector X_0 is sufficient to calculate the future state probability vectors $\mathrm{X}_n, n = 1,2,\ldots$.

⇨ This method allows us to evaluate systems in overload or during transient phase which are typical issues in communication networks.

M / GI / 1 – Waiting system

- Stationary state equation:

A system is called stable if it state probability vector does not further change. (c.f. Chapter 3)

$$\Rightarrow \quad X_n = X_{n+1} = \ldots = X$$

$$\Rightarrow \quad X = \{x(0), x(1), \ldots, x(j), \ldots\}$$

$$\Rightarrow \quad X = X \cdot P$$ Probability vector in stationary state.

$$\Rightarrow \quad x(j) = x(0)\gamma(j) + \sum_{i=1}^{j+1} x(i)\gamma(j-i+1) \qquad j = 0,1,\ldots$$

State probabilities can be calculated by using the distribution of RV Γ and the probability vector.

- In general the process or system is in an instationary state X_0 at the beginning.
- The stationary state vector is typically determined by numerical method.

M / GI / 1 – Waiting system

- Power method:

  Robust numerical method to determine the steady probability state vector.

  ⇨ Calculate the general state probability equation $X_{n+1} = X_n \cdot P$ until a certain abortion criteria is reached.

  ⇨ It is assumed that the statistical equilibrium is reached if the following abortion criteria holds true:

  ⇨ $\left| E[X(t_{n+1})] - E[X(t_n)] \right| < \varepsilon = 10^{-6}$

M / GI / 1 – Waiting system

- Complementary waiting time distribution function in M / GI / 1:

  In the following the complementary waiting distribution function of a M / GI / 1 Waiting system depending on ist utilization ρ and the coefficient of variation c_B of the service time distribution is shown.

- Coefficient of variation (Variationskoeffizient)
  - The coefficient of variation is a normalized measure of dispersion of a probability distribution
  - It is a dimensionless number which does not require knowledge of the mean of the distribution in order to describe the distribution

$$c_X = \frac{\sigma_X}{E[X]}, \qquad E[X] > 0$$

  - $c = 0$ $\rightarrow$ deterministic (constant service duration)
  - $c < 1$ $\rightarrow$ variance lower than exponential distribution
  - $c > 1$ $\rightarrow$ variance higher than exponential distribution

 Picture taken from Tran-Gia, Einführung in die Leistungsbewertung und Verkehrstheorie

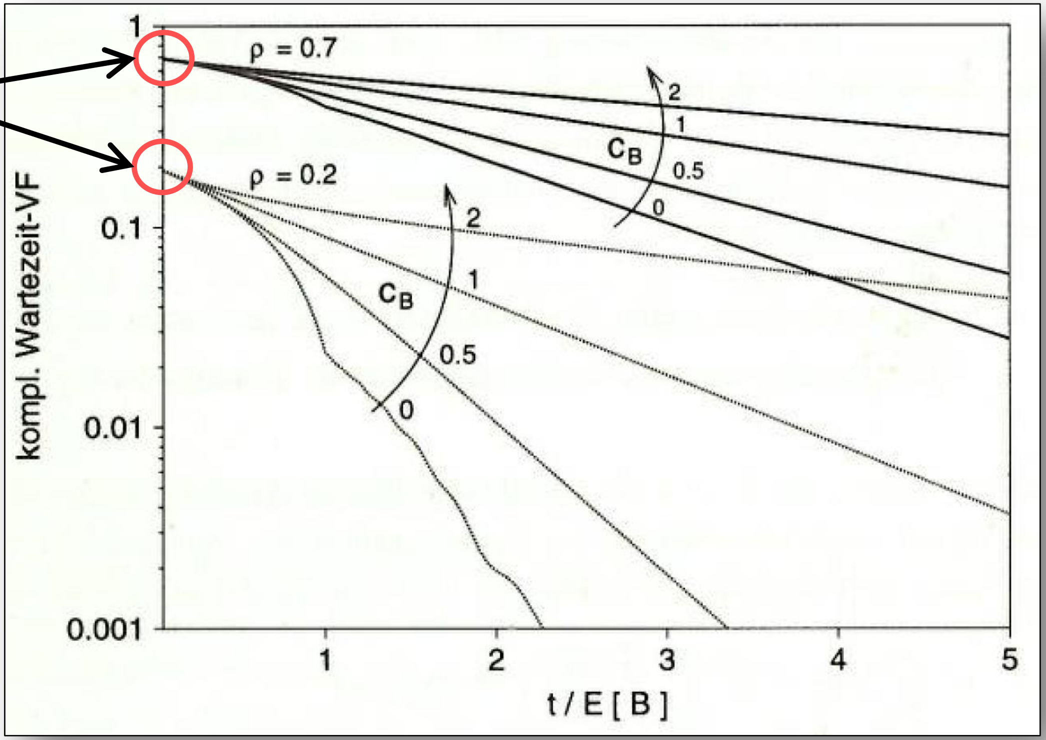
M / GI / 1 – Waiting system

- Complementary waiting time distribution function in M / GI / 1:

Probability that the system is busy.

⇨ $x(0) = 1 - \rho$

⇨ $\rho = 1 - x(0)$

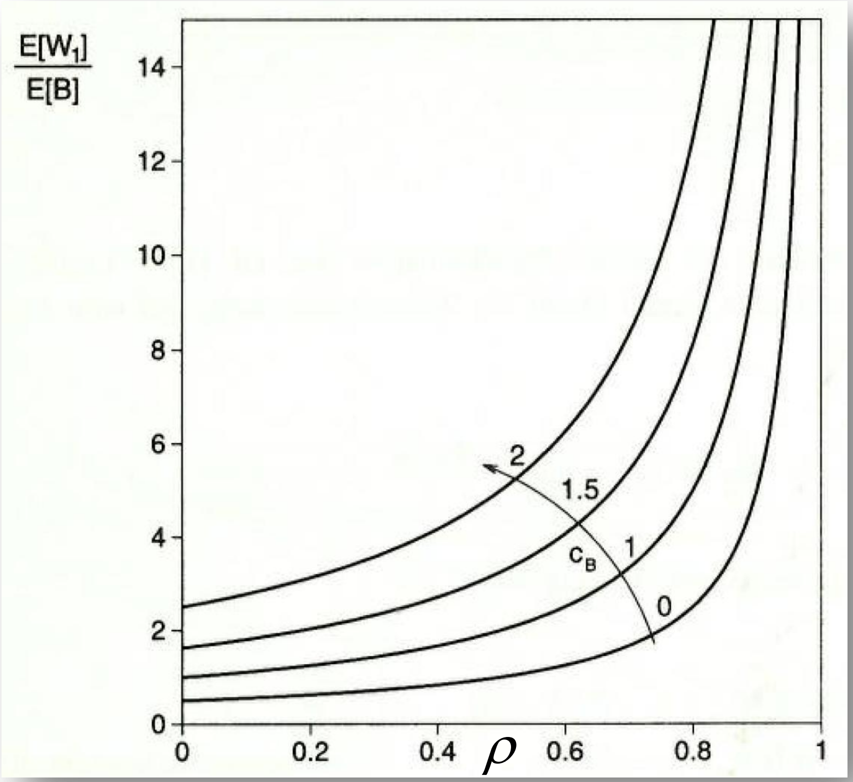


⇨ Waiting duration increases with higher variance of the service process!

⇨ Variance of the service process becomes the dominating factor.

M / GI / 1 – Waiting system

- Average waiting time of waiting jobs:



⇨ High utilization should be avoided if the service duration has a high coefficient of variation!

⇨ Best performance is achieved by systems with constant service times.

M / GI / 1 – Waiting system

- State probabilities at random observation points:

⇨ $x^*(i), \quad i = 0,1,\ldots,n$ State probabilities at a randomly chosen observation point t*.

⇨ $x_A(i), \quad i = 0,1,\ldots,n$ State probabilities at points of arrival.

⇨ State probabilities of a M / G / 1 – Waiting system are valid at any randomly chosen observation point t*.

⇨ $x(i) = x^*(i) = x_A(i), \quad i = 0,1,\ldots$

M / GI / 1 – Waiting system

- Idea:
  - Observe the state process over an interval of length T.
  - Focus on a single state $[X = i]$ and count the following state transitions:

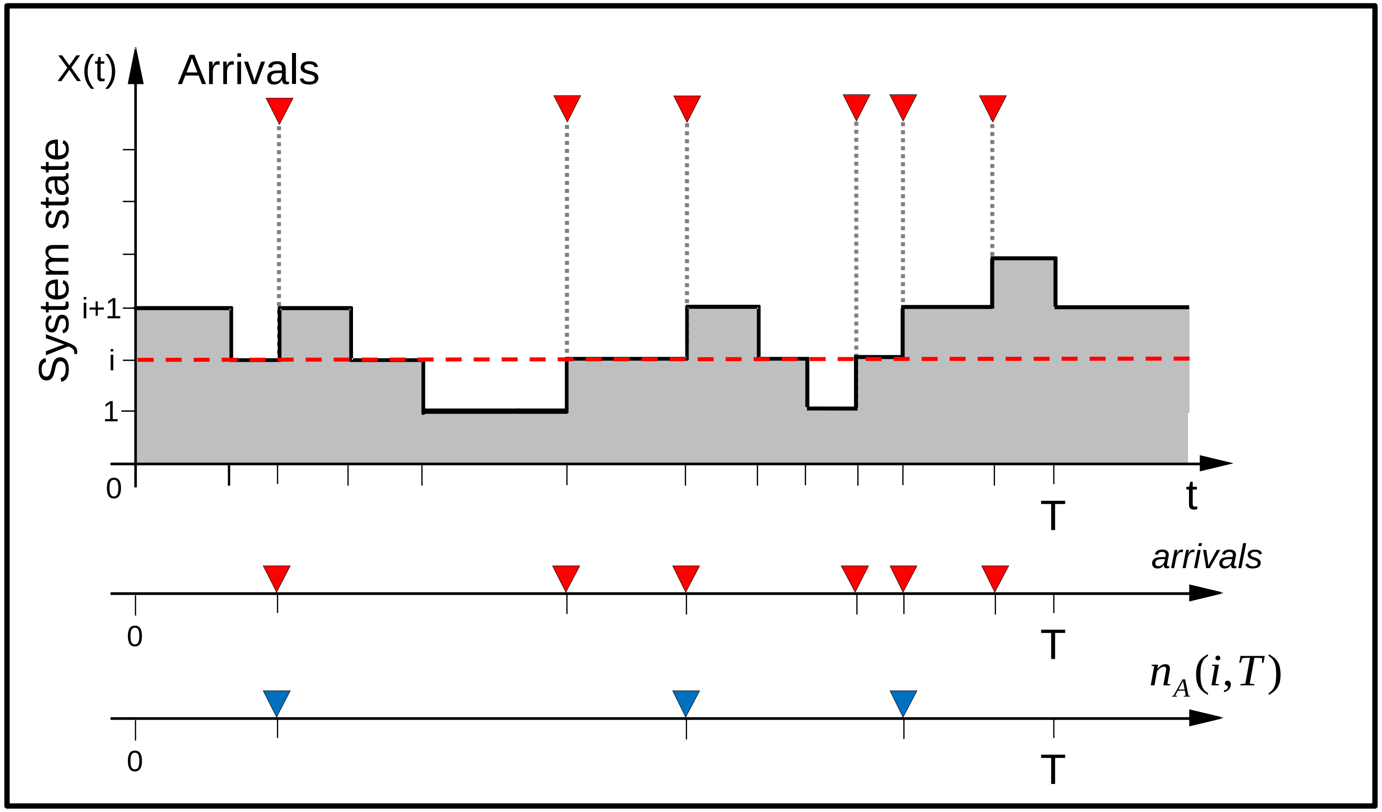
⇒ Arrival event:
- State transition $[X = i] \rightarrow [X = i + 1]$
- $n_A(i, T)$ number of these state transitions during interval T.

⇒ Departure event:
- State transition $[X = i + 1] \rightarrow [X = i]$
- $n_D(i, T)$ number of these state transitions during interval T.

M / GI / 1 – Waiting system

- State probabilities at random observation points:

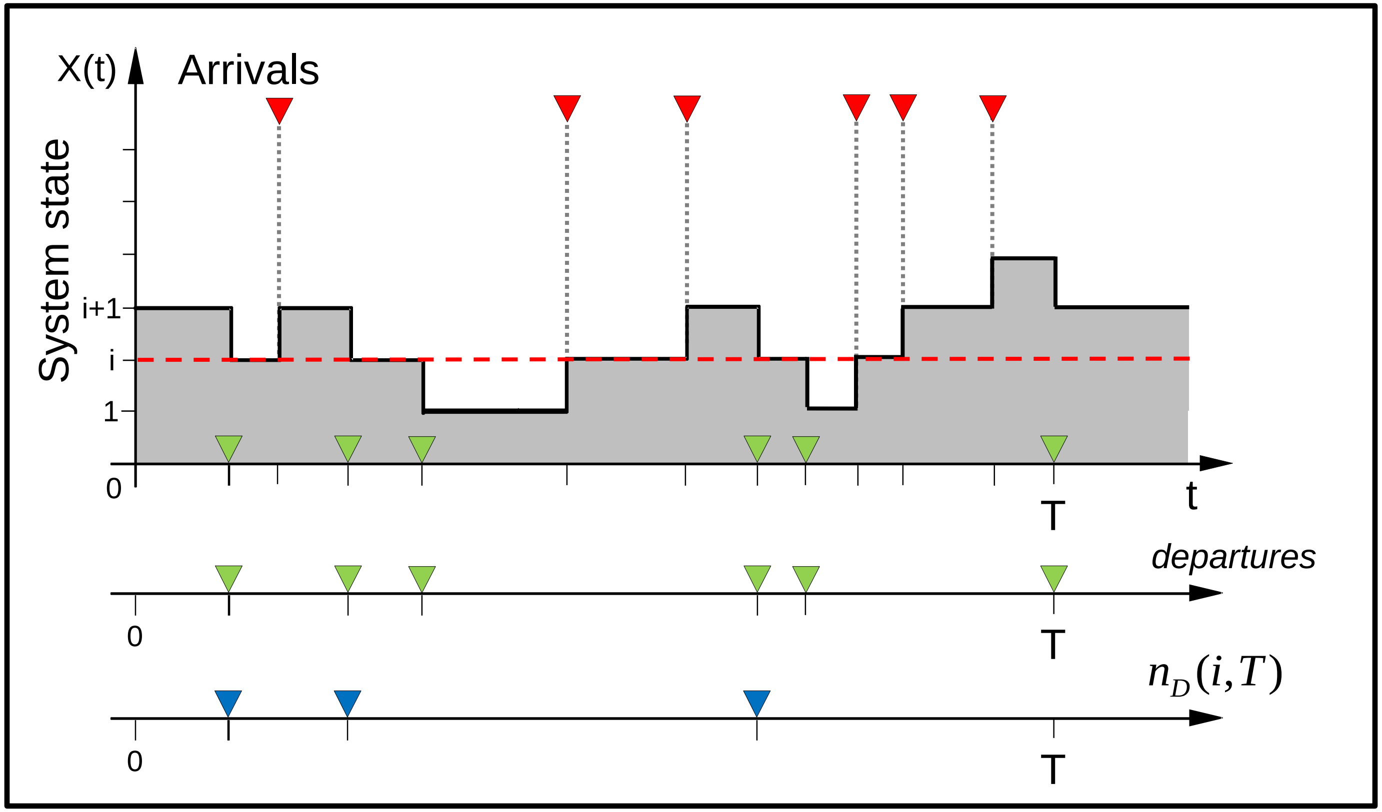
  Number of arrivals during interval T while the system was in state i.



M / GI / 1 – Waiting system

- State probabilities at random observation points:

  Number of departures during interval T which changed the system state from i+1 to state i.



M / GI / 1 – Waiting system

- Both events are alternating during the process development.
- During an observation interval T the following inequality holds true:

$$\Rightarrow \quad |n_A(i,T) - n_D(i,T)| \leq 1$$

- The total number of arrival and departure events during time interval T is given by:

$$\Rightarrow \quad n_A(T) = \sum_{i=0}^{\infty} n_A(i,T)$$ Total number of arrivals during interval T

$$\Rightarrow \quad n_D(T) = \sum_{i=0}^{\infty} n_D(i,T)$$ Total number of departure during interval T

- Start state is given by X(0) and the final state is given by X(T).

$$\Rightarrow \quad n_D(T) = X(0) - X(T) + n_A(T)$$ Total number of departure events during interval T

M / GI / 1 – Waiting system

- State probability at embedded points:

$$\Rightarrow \quad x(i) = \lim_{T\to\infty} \frac{n_D(i,T)}{n_D(T)} = \lim_{T\to\infty} \frac{n_A(i,T) + n_D(i,T) - n_A(i,T)}{n_A(i,T) + X(0) - X(T)}$$

$$= \lim_{T\to\infty} \frac{\dfrac{n_A(i,T)}{n_A(T)} + \dfrac{n_D(i,T) - n_A(i,T)}{n_A(T)}}{1 + \dfrac{X(0) - X(T)}{n_A(T)}} \qquad i = 0,1,\ldots$$

with

$$\Rightarrow \quad \lim_{T\to\infty} \frac{n_A(i,T)}{n_A(T)} = x_A(i), \qquad i = 0,1,\ldots$$

State transition at arrivals

and

$$\Rightarrow \quad \lim_{T\to\infty} \frac{\left|n_D(i,T) - n_A(i,T)\right|}{n_A(T)} = 0$$

M / GI / 1 – Waiting system

- State probability at embedded points:

Stationary system

$$\Rightarrow \quad \lim_{T\to\infty} \frac{|X(0) - X(T)|}{n_A(T)} = 0$$

$$\Rightarrow \quad x(i) = x_A(i), \qquad i = 0,1,\ldots$$

The arrival process is memory less. Thus, an arrival sees the system from the same perspective state than an independent observer.

PASTA

$$\Rightarrow \quad x^*(i) = x_A(i), \qquad i = 0,1,\ldots$$

q.e.d.

Questions

- What is an embedded chain?
- What is an embedded markov chain?
- When does a system become stable?
- What is the power method?
- Does the start probability vector have an impact on the power method?
- How can you analyze a M/GI/1 waiting system?
- What impact does the variation coefficient of the service distribution have on the system performance?
- What is the difference between waiting time of all jobs and waiting time of waiting jobs?
- How can you proof that arrival see a M/GI/1 system from the same perspective than an independent observer?

Analysis of System Performance

IN2072

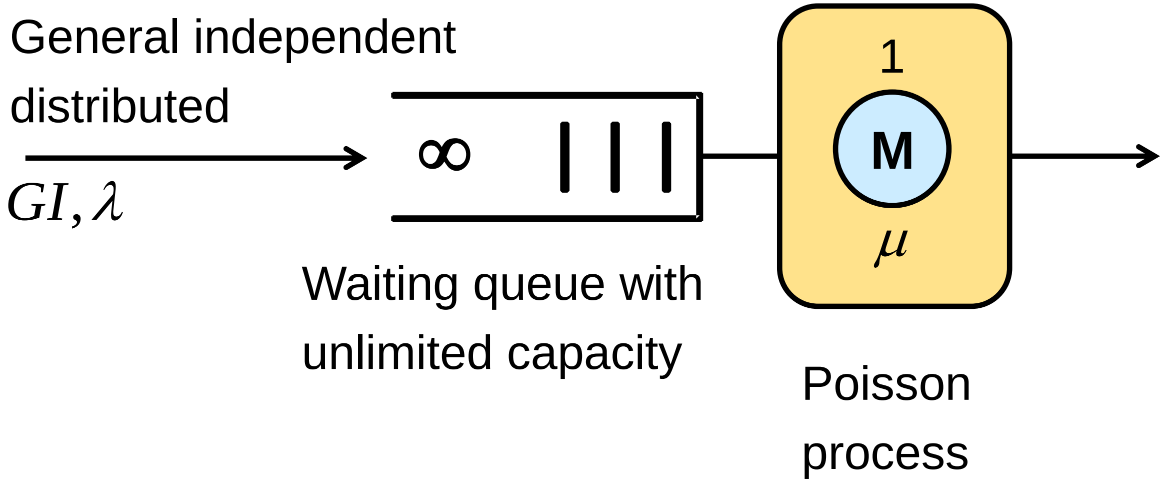
Chapter 5 – Analysis of

Non Markov Systems

GI / M / 1 – Waiting system

GI / M / 1 – Waiting system

- Model:



- Model and parameter description:
  - GI / M / 1 – ∞ (No jobs are blocked!)
  - Arrival process is general independent (GI) distributed.
  - Service time B is a Poisson process with an exponential distributed inter-arrival time A.
  - Jobs that arrive at a point in time when the service unit is busy, are queued and served in FIFO order as soon as the serving unit has served the current job.

GI / M / 1 – Waiting system

- **Arrival process:**

  Arrival rate λ

  Average number of arriving jobs per time unit.

$$A(t) = P(A \le t), \qquad E[A] = \frac{1}{\lambda}$$

- **Service process:**

  Service rate μ

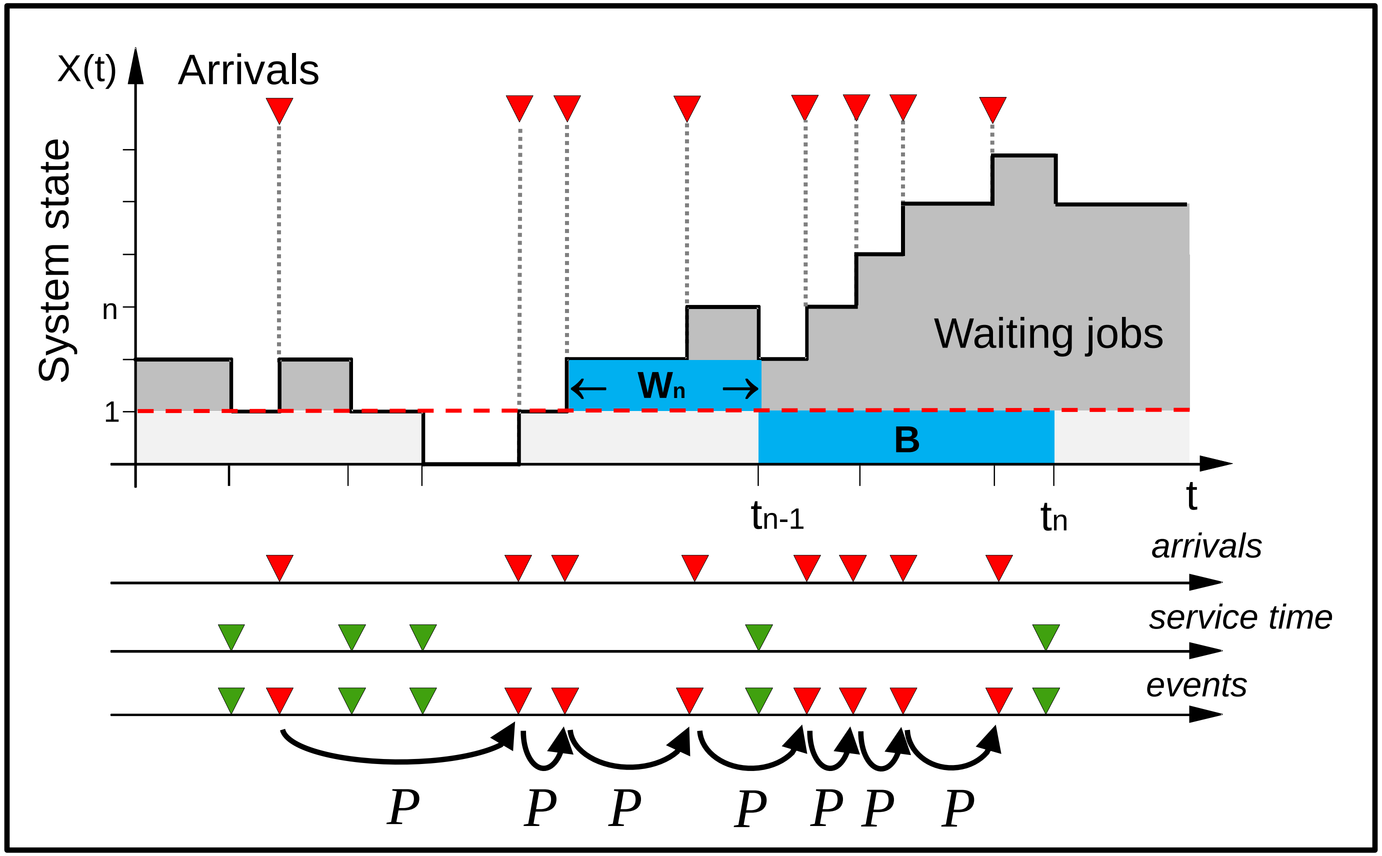
  Average number of service completions.
  (assuming the service unit only has two states – idle or busy)

$$B(t) = P(B \le t) = 1 - e^{-\mu t}, \qquad E[B] = \frac{1}{\mu}$$

- **System:**
  - Waiting system
  - Waiting queue with unlimited capacity
  - Queuing strategy – First In First Out (FIFO)

GI / M / 1 – Waiting system

- State space:
  - Random variable $X(t)$ describes the number of (waiting and currently served) jobs in the system.
  - State process is state discrete and time continuous stochastic process.



GI / M / 1 – Waiting system

- Embedded points:
  - State process becomes memory less at an arrival.
  - Embedded points of the markov chain are located directly before an arrival.
- Embedded Markov Chain:
  - The point in time of the n th embedded point corresponds to the n th service completion.
  - The sequence of the process states

$$\{X(t_0), X(t_1), \ldots, X(t_n), X(t_{n+1}), \ldots\}$$

at these points represent the embedded markov chain.

GI / M / 1 – Waiting system

- Analysis:
  - Introduce a random variable Γ which describes the number of served jobs within an inter-arrival interval.

$$\gamma(i) = P(\Gamma = i)$$

with $$\Gamma_{GF}(z) = \sum_{i=0}^{\infty} \gamma(i) \cdot z^i$$ Generation Function

$$\Rightarrow \quad E[\Gamma] = \left. \frac{d\Gamma_{GF}(z)}{dz} \right|_{z=1} = \mu \cdot E[A] = \frac{1}{\rho}$$

GI / M / 1 – Waiting system

- State transition:
  - State transition between consecutive embedded points t_n and t_{n+1}.

⇨ $$p_{ij} = P\{X(t_{n+1}) = j \mid X(t_n) = i\}$$

Case 1: System not empty (j ≠ 0) at time t_{n+1}.

- i jobs in the system right before the nth arrival.
- i+1 jobs are in the system the embedded point t_{n+1}.
- j jobs remain in the system at the next embedded point.

⇨ (i+1-j) jobs have to be served during the interval $[t_n ; t_{n+1}]$.

⇨ $$p_{ij} = \gamma(i+1-j), \qquad i = 0,1,\ldots, \qquad j = 1,2,\ldots,i+1$$

GI / M / 1 – Waiting system

Case 2: System is empty (j = 0) at time t_{n+1}

- At time t_n are i+1 jobs in the system.
- no jobs remain in the system at time t_{n+1}.

⇨ i+1 jobs have to be served during $[t_n; t_{n+1}]$.

⇨ $$p_{i0} = \sum_{k=i+1}^{\infty} \gamma(k) = 1 - \sum_{k=0}^{i} \gamma(k), \qquad i = 0,1,\ldots$$

⇨ $$P = \{p_{ij}\} = \begin{pmatrix} 1-\gamma(0) & \gamma(0) & 0 & 0 & \cdots \\ 1-\sum_{k=0}^{1}\gamma(k) & \gamma(1) & \gamma(0) & 0 & \cdots \\ 1-\sum_{k=0}^{2}\gamma(k) & \gamma(2) & \gamma(1) & \gamma(0) & \cdots \\ 1-\sum_{k=0}^{3}\gamma(k) & \gamma(3) & \gamma(2) & \gamma(1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

GI / M / 1 – Waiting system

- The state probabilities at the embedded points t_n can be described by a state probability vector.

⇨ $$\mathrm{X}_n = \{x(0,n), x(1,n), \ldots, x(j,n), \ldots \qquad j = 0,1,\ldots\}$$

⇨ $$x(j,n) = P\{X(t_n) = j\}$$

State transition probability matrix P which describes the relation between any consecutive state probability vectors X_n and X_{n+1} at embedded points t_n and t_{n+1} is given by

⇨ $$\mathrm{X}_{n+1} = \mathrm{X}_n \cdot P$$ Relation of consecutive state transition vectors.

A start vector X_0 is sufficient to calculate the future state probability vectors $\mathrm{X}_n, n = 1,2,\ldots$.

⇨ This method allows us to evaluate systems in overload or during transient phase which are typical issues in communication networks.

GI / M / 1 – Waiting system

- Stationary state equation:

  A system is called stable if it state probability vector does not further change. (c.f. Chapter 3)

$$\Rightarrow \quad \mathrm{X}_n = \mathrm{X}_{n+1} = \ldots = \mathrm{X}$$

$$\Rightarrow \quad \mathrm{X} = \{x(0), x(1), \ldots, x(j), \ldots\}$$

$$\Rightarrow \quad \mathrm{X} = \mathrm{X} \cdot P$$

Probability vector in stationary state.

$$\Rightarrow \quad x(0) = \sum_{i=0}^{\infty} x(i) \cdot \left(1 - \sum_{k=0}^{i} \gamma(k)\right) = \sum_{i=0}^{\infty} x(i) \sum_{k=i+1}^{\infty} \gamma(k)$$

$$\Rightarrow \quad x(j) = \sum_{i=j-1}^{\infty} x(i) \cdot \gamma(i+1-j) = \sum_{i=0}^{\infty} x(i+j-1) \cdot \gamma(i), \qquad j = 1,2,\ldots$$

Questions

- How can you analyze a GI/M/1 waiting system?
- What is the utilization of a GI/M/1 waiting system?
- What impact has the variation of the arrival process on the waiting time of a GI/M/1 waiting system?
- When does the system become memory-less?

Analysis of System Performance

IN2072

Chapter 5 – Analysis of

Non Markov Systems

M / GI[Θ,K] / 1 – S

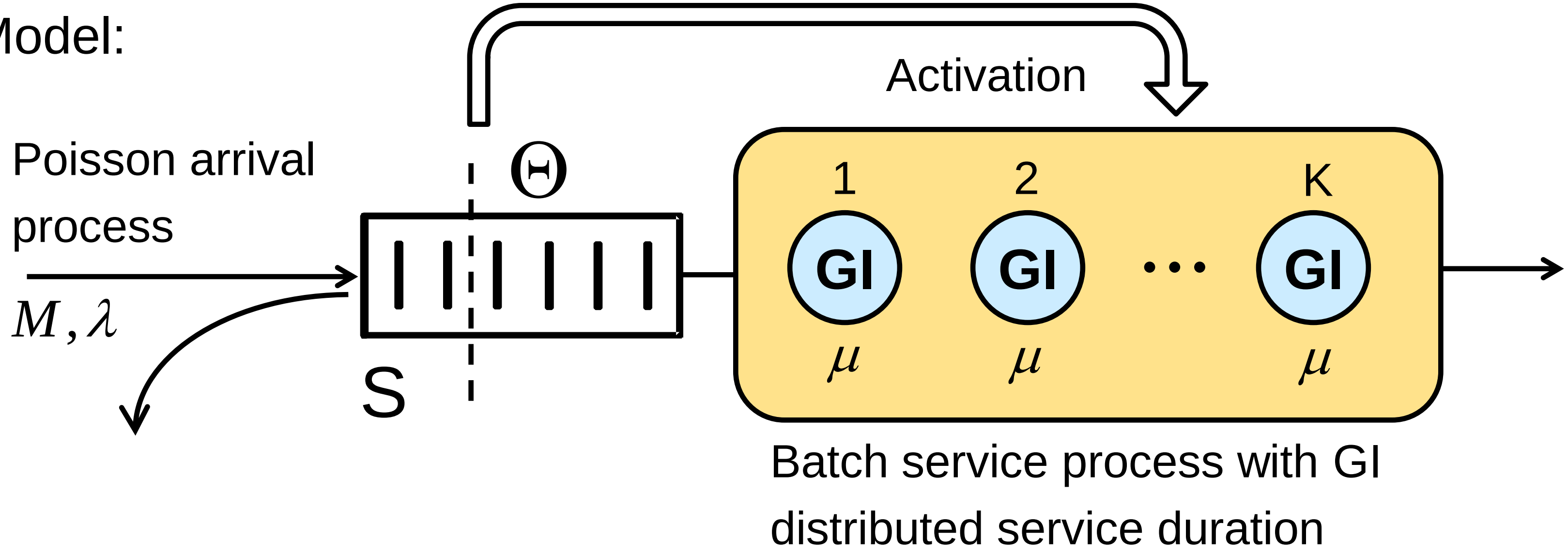
Batch Service with Start Threshold

Technische Universität München

M / GI[Θ,K] / 1 – S Loss system

Batch service system with start threshold:

Model:



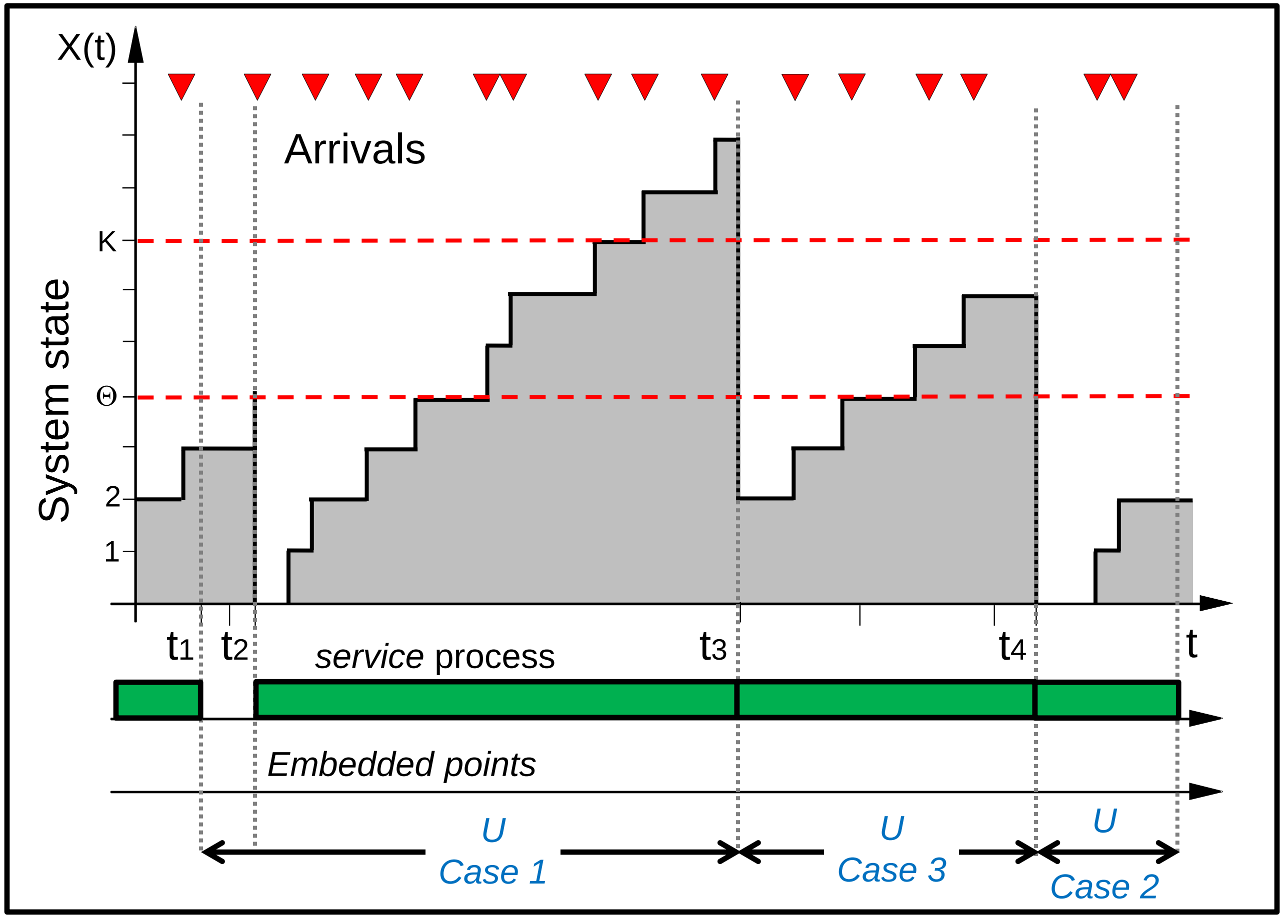
- Model and parameter description:
  - M / GI[Θ,K] / 1 – S – System with S waiting slots.
  - Arrival process is a Poisson process.
  - Service duration is GI distributed. Up to K jobs can be served in parallel.
  - Service unit is activated if Θ or more jobs are waiting.
  - All jobs of a batch experience exactly the same service time.
  - Arriving jobs have to wait until the current batch is served.

M / GI[Θ,K] / 1 – S Loss system

- Model and parameter description:
  - System has S waiting slots.
  - Jobs are blocked if S jobs are waiting in the queue.
  - At the end of a batch service, new jobs are loaded into the servers if at least Θ jobs are waiting.
  - Up to K jobs are loaded into the system after a batch service.
  - If less than Θ jobs are waiting in the queue at the end of a batch service, the servers remain idle until Θ jobs are in the queue and a new batch service starts.

M / GI[Θ,K] / 1 – S Loss system

- State space:
  - Random variable $X(t)$ describes the number of waiting the system.
  - State process is state discrete and time continuous stochastic process.



M / GI[Θ,K] / 1 – S Loss system

- Embedded points:
  - State process becomes memory less at the time of a service completion.
  - Embedded points of the markov chain are located directly after service completion.

- Embedded Markov Chain:
  - The point in time of the n th embedded point corresponds to the n th (batch) service completion.
  - The sequence of the process states

    $$\{X(t_0), X(t_1), \ldots, X(t_n), X(t_{n+1}), \ldots\}$$

    at these points represent the embedded markov chain.

M / GI[Θ,K] / 1 – S Loss system

- Analysis:
  - Introduce a random variable Γ which describes the number of arrivals during a service duration.

$$\gamma(i) = P(\Gamma = i)$$

The state probabilities at the embedded points t_n and t_{n+1} can be described by a state probability vector.

$$\Rightarrow \quad p_{ij} = P\{X(t_{n+1}) = j \mid X(t_n) = i\}$$

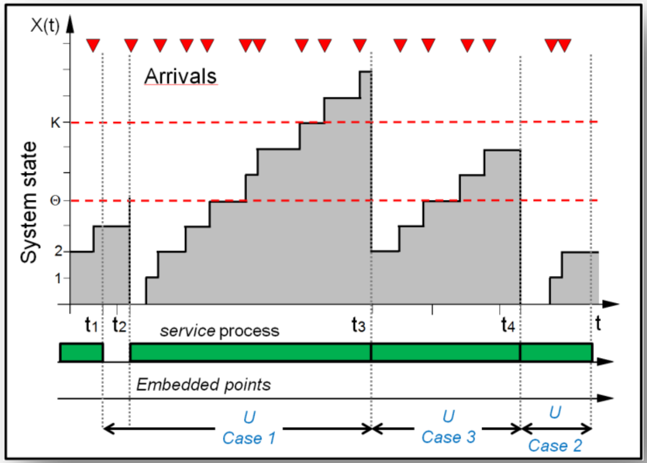
M / GI[Θ,K] / 1 – S Loss system

- State transition:
  - State transition between consecutive embedded points t_n and t_{n+1}.

$$\Rightarrow \quad p_{ij} = P\{X(t_{n+1}) = j \mid X(t_n) = i\}$$

Case 1: $i < \Theta$ Less than Θ jobs in the queue at time t_1.

- Θ-i jobs have to arrive until the service process is started.
- This waiting period is Erlang (Θ-i) distributed $E_{\Theta - i}$.
- Service unit is activated as soon as Θ jobs are waiting.
- j jobs arrive during the service.
- Transition time U is given by:



$$\Rightarrow \quad U = E_{\Theta - i} + B$$

$$\Rightarrow \quad p_{ij} = \gamma(j), \qquad j = 0,1,\ldots,S-1$$

$$\Rightarrow \quad p_{iS} = \sum_{k=S}^{\infty} \gamma(k), \qquad j = S$$

M / GI[Θ,K] / 1 – S Loss system

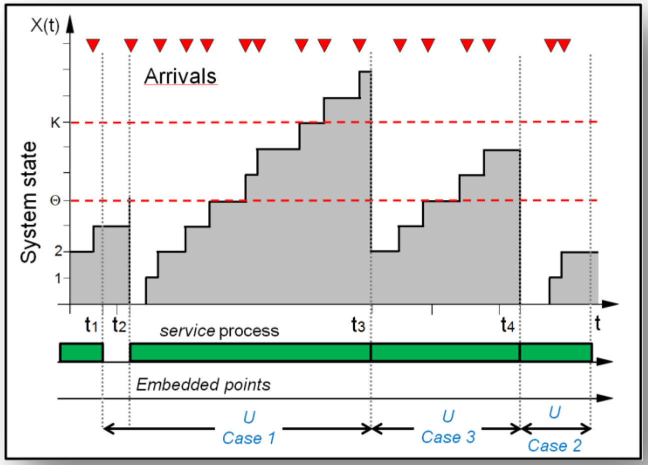
State transition:

- State transition between consecutive embedded points t_n and t_{n+1}.

$$\Rightarrow \quad p_{ij} = P\{X(t_{n+1}) = j \mid X(t_n) = i\}$$

Case 2: $\Theta \le i \le k$ Number of jobs in the queue higher than the threshold.

- Service process is started immediately and the queue is emptied.
- Transition time U is identical with the service duration.
- State transition probability is identical to that in case 1.
- j jobs arrive during the service.



$$\Rightarrow \quad U = B$$

$$\Rightarrow \quad p_{ij} = \gamma(j), \qquad j = 0,1,\ldots,S-1$$

$$\Rightarrow \quad p_{iS} = \sum_{k=S}^{\infty} \gamma(k), \qquad j = S$$

Case 3: $K \le i \le S$

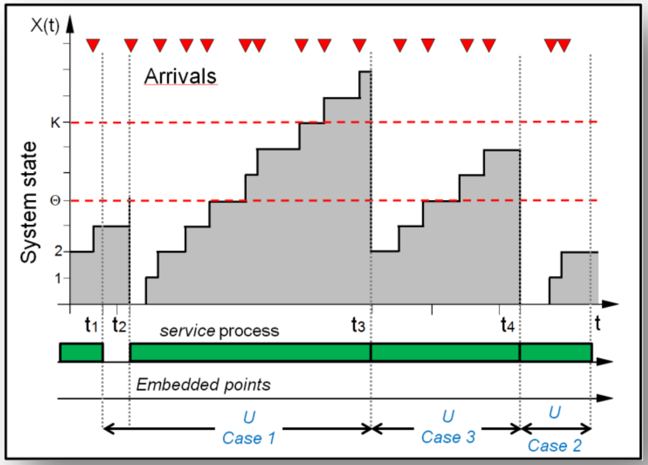
Number of jobs in the queue is higher than the number of servers.

- Service process is started immediately and K jobs are served.
- (i-k) jobs remain in the queue after the service is started.
- Transition time U is identical with the service duration.
- j jobs are in the queue after the batch service is completed.
- (j-i+K) jobs arrive during the service duration.

$$U = B$$

$$p_{ij} = \gamma(j - i + k), \qquad j = 0,1,\ldots,S-1$$

$$p_{iS} = \sum_{k=S-i+K}^{\infty} \gamma(k), \qquad j = S$$

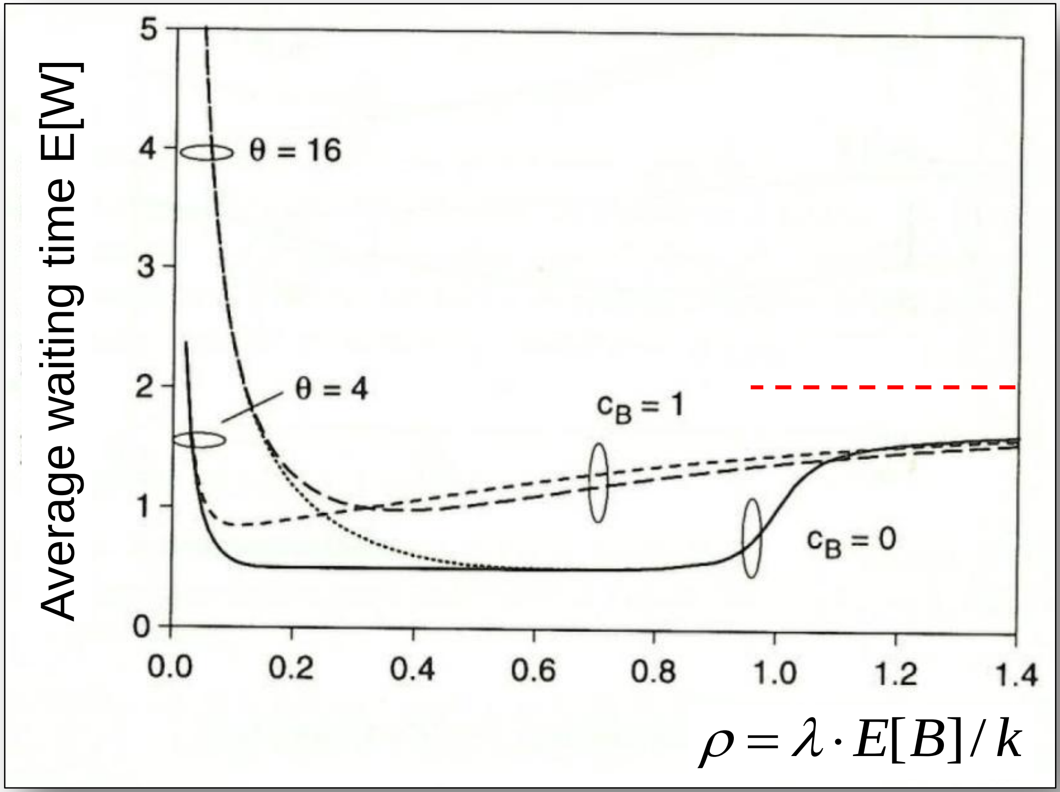


State transition matrix:

$$
\begin{array}{c}
 \\
P = \{p_{ij}\} =
\end{array}
\begin{array}{c}
\begin{array}{cccccc}
\mathbf{0} & \mathbf{1} & \mathbf{2} & \cdots & \mathbf{S\text{-}1} & \mathbf{S}
\end{array} \\
\begin{pmatrix}
\gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(S-1) & \sum_{k=S}^{\infty}\gamma(k) \\
\gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(S-1) & \sum_{k=S}^{\infty}\gamma(k) \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(S-1) & \sum_{k=S}^{\infty}\gamma(k) \\
0 & \gamma(0) & \gamma(1) & \cdots & \gamma(S-2) & \sum_{k=S-1}^{\infty}\gamma(k) \\
0 & 0 & \gamma(0) & \cdots & \gamma(S-3) & \sum_{k=S-2}^{\infty}\gamma(k) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \gamma(K-1) & \sum_{k=K}^{\infty}\gamma(k)
\end{pmatrix}
\end{array}
\begin{array}{c}
\mathbf{0} \\ \mathbf{1} \\ \vdots \\ \mathbf{K} \\ \mathbf{K+1} \\ \mathbf{K+2} \\ \vdots \\ \mathbf{S}
\end{array}
$$

M / GI[Θ,K] / 1 – S Loss system

- Average waiting time:
  - Start threshold
  - Traffic load
  - Variance of the service process
- Characteristics:
  - High waiting time for systems with low traffic load and high start threshold.
  - Start threshold becomes dominating for systems with low traffic load since it takes a long time until the start threshold is reached.

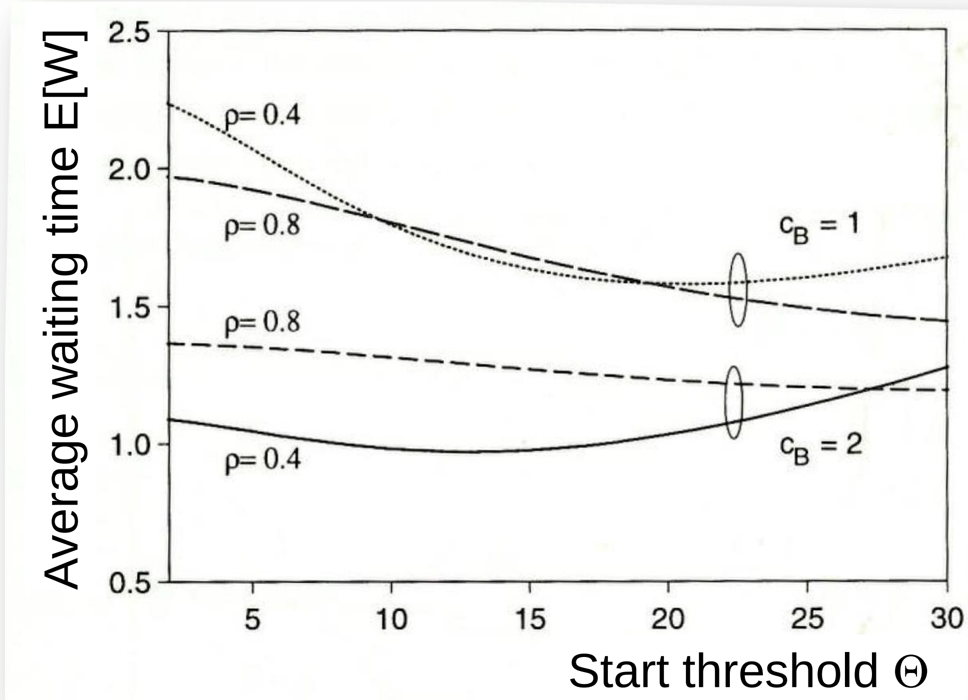


M / GI[Θ,K] / 1 – S with K=32 service units and S=64 waiting slots

M / GI[Θ,K] / 1 – S Loss system

- Average waiting time:
  - Start threshold
  - Traffic load
  - Variance of the service process

- Characteristics:
  - Impact of variance of the service process decreases with higher start thresholds.
  - Optimum in terms of average waiting time is highly parameter sensitive.



M / GI[Θ,K] / 1 – S with K=32 service units and S=64 waiting slots

Questions

- Describe the M/GI[Θ,K]/1-S loss system and its parameter.
- What impact has the start threshold on the waiting duration?
- How can you analyze the M/GI[Θ,K]/1-S loss system?
- Which state changes are possible between to consecutive embedded points?