Analysis of System Performance

IN2072

Chapter 4 – Analysis of Markov Systems (Part 1/2)

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

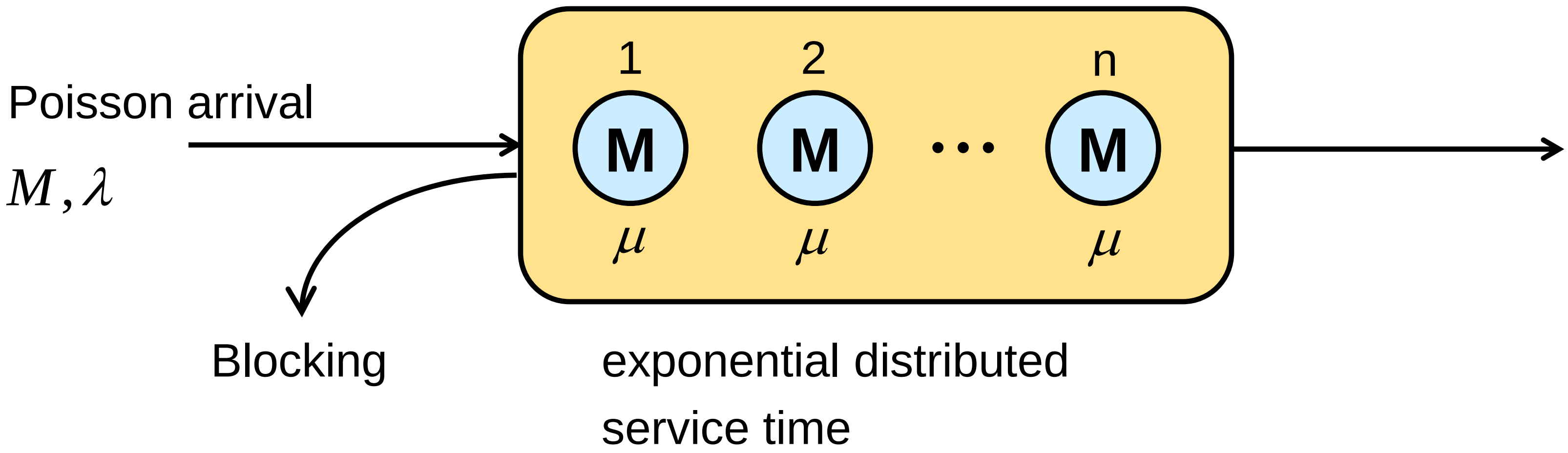
Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Markov Systems

- Content:
  - (Memory-less) systems
    - Poisson arrival process
    - Exponential distributed service times
  - Loss system M/M/n-s (**Infinite** number of sources)
    - Erlang-B equation
    - State probabilities
    - Blocking probability
    - Multiplexing gain
    - Dimensioning of systems

M / M / n – Loss system

- Model and parameter description:



- Model and parameter description:
  - M / M / n – 0 (No waiting slots!)
  - Arrival process is a Poisson process with an exponential distributed inter-arrival time A
  - Service time B is also exponential distributed
  - Jobs that arrive at a point in time when all service units are busy, are blocked and do not affect the future development of the system.

M / M / n – Loss system

- **Arrival process:**

  Arrival rate λ

  Average number of arriving jobs per time unit.

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

- **Service process:**

  Service rate μ

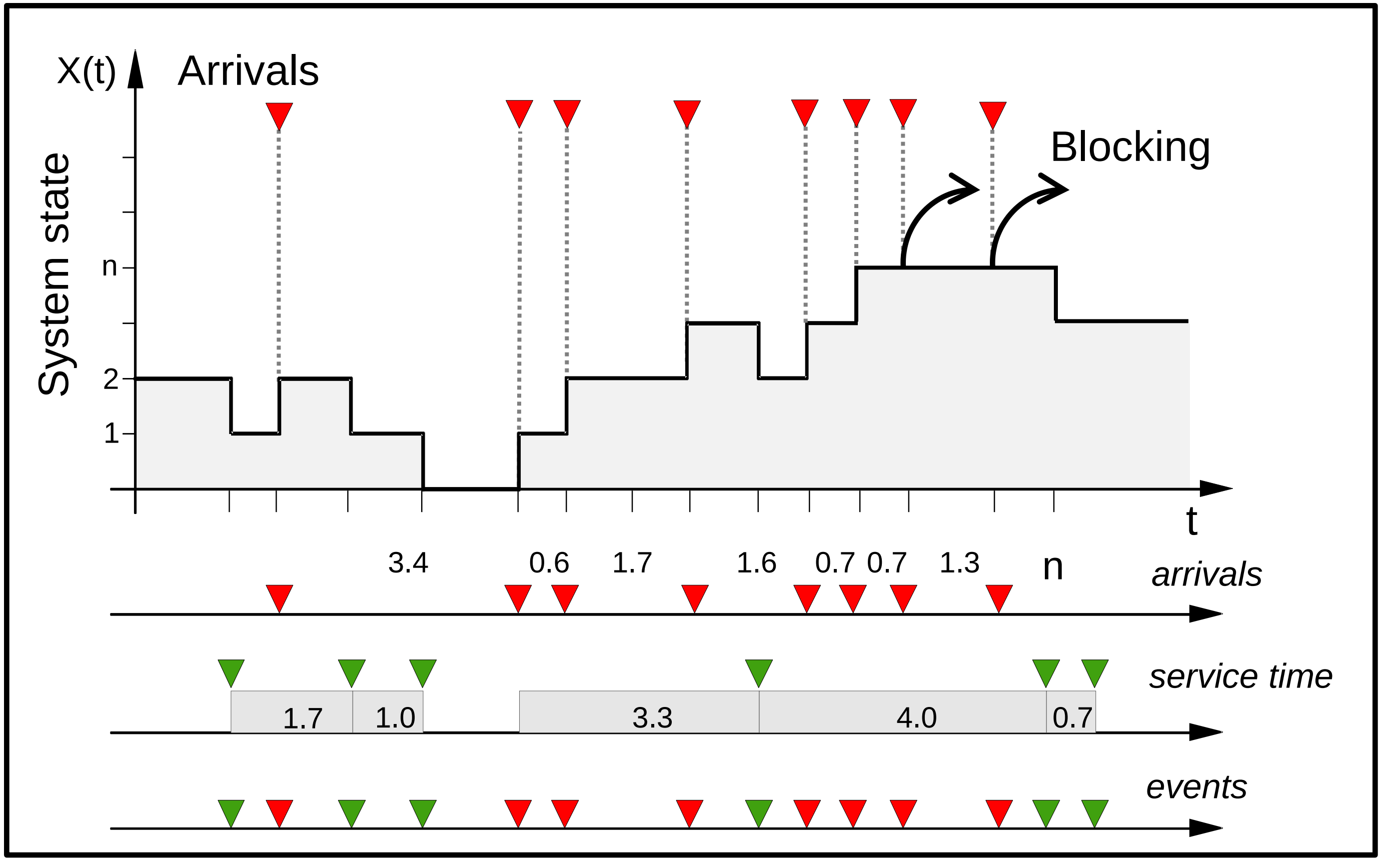
  Average number of service completions per time unit. (assuming a service unit with 100% utilization).

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

- **System:**
  - Loss system
  - No waiting queue
  - Blocked jobs do not affect the future development of the system.

State space and state probabilities

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



M / M / n – Loss system

- **Description:**
  - State $X(t)$ is incremented if a job can be served by an idle service unit .
  - State $X(t)$ is decremented if a service is completed.

Due to the memory-less characteristics of the arrival and the service process, the system is memory-less at any time of the process development.

- **Transient phase:**
  - The system starts in state $X(0)$ from which it develops through an instationary phase until it reaches a stationary state.
  - The state probabilities do not change any further as soon as the stationary state is reached.

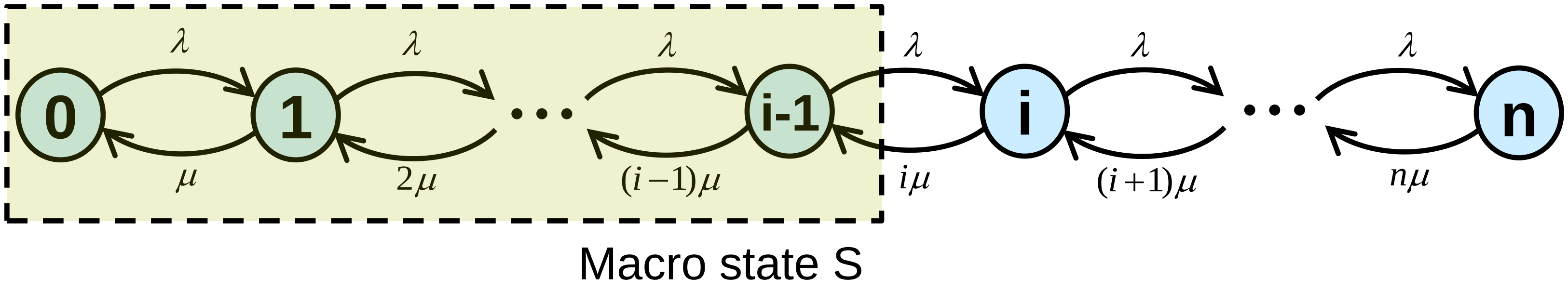
- State probabilities: $x(i) = P(X(t) = i) = P(X = i), \quad i = 0,1,\ldots,n$

- State probability vector: $X = \{x(0), x(1), \ldots, x(n)\}$

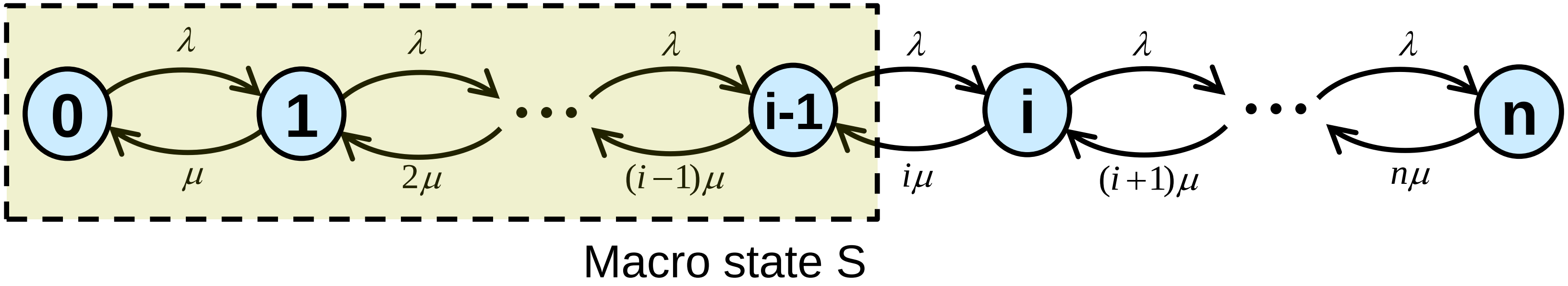
M / M / n – Loss system

- **Arrival event:**
  - According to the definition of a Poisson process the transition from $[X=i] \rightarrow [X=i+1]$ occurs with rate λ if the system is in state $x(i), \quad i=0,1,\ldots,n-1.$
  - Otherwise the system is in state $x(n)$ which results in the blocking of the arriving job.

- **Service completion event:**
  - If the system is in state $x(i)$, i jobs are in the system.
  - Thus, i service units are busy / i jobs are served.
  - The transition from $[X=i] \rightarrow [X=i-1]$ occurs with rate $i\mu, \quad i=1,\ldots,n$ if one of the currently served jobs has finished.



M / M / n – Loss system



Macro state S consists of micro states $\{X = 0,1,\ldots,i-1\}$.

$$\Rightarrow \quad \lambda \cdot x(i-1) = i \cdot \mu \cdot x(i), \qquad i = 1,2,\ldots,n$$

$$\Rightarrow \quad \sum_{i=0}^{n} x(i) = 1$$

M / M / n – Loss system

- Load $A = \frac{\lambda}{\mu}$

⇨ $\lambda \cdot x(0) = 1 \cdot \mu \cdot x(1)$ $\qquad x(1) = x(0) \cdot \frac{A}{1}$ $\qquad = x(0) \cdot \frac{A^1}{1!}$

⇨ $\lambda \cdot x(1) = 2 \cdot \mu \cdot x(2)$ $\qquad x(2) = x(1) \cdot \frac{A}{2}$ $\qquad = x(0) \cdot \frac{A^2}{2!}$

⇨ $\lambda \cdot x(2) = 3 \cdot \mu \cdot x(3)$ $\qquad x(3) = x(2) \cdot \frac{A}{3}$ $\qquad = x(0) \cdot \frac{A^3}{3!}$

➡ $x(i) = x(0) \cdot \frac{A^i}{i!}, \qquad i = 0,1,2,\ldots,N$

➡ $x(i) = \frac{\frac{A^i}{i!}}{\sum_{k=0}^{n} \frac{A^k}{k!}}, \qquad i = 0,1,2,\ldots,N$

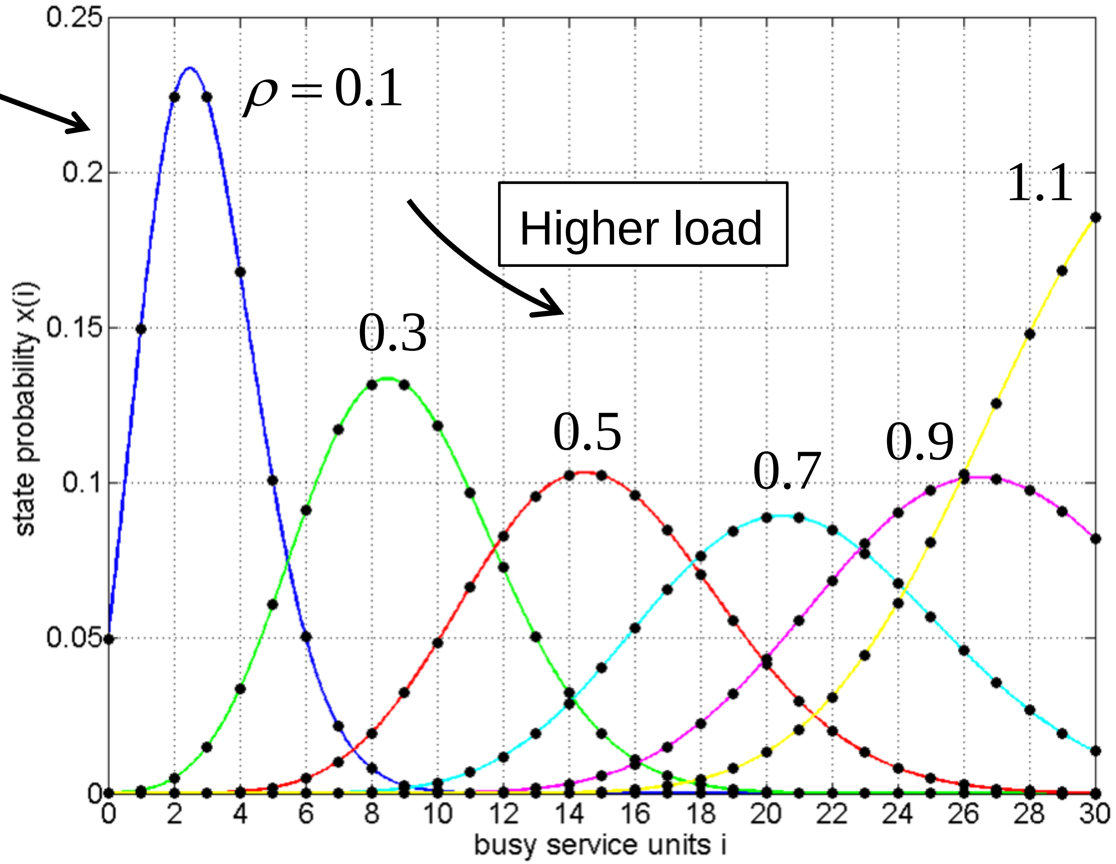
Erlang-B equation for loss systems

M / M / n – Loss system

- State probabilities of a **M / M / 30 – loss system** depending on the offered work load $\rho = \frac{A}{n} = \frac{\lambda}{n\mu}$

Probability that the system is idle

Blocking probability



M / M / n – Loss system

- The load $A = \frac{\lambda}{\mu}$ is often described with in the pseudo unit [Erl] Erlang.
- Higher load shifts the state probabilities to the right side.
- Probability x(n) represents the system blocking probability.
- Probability x(0) indicates an idle system.

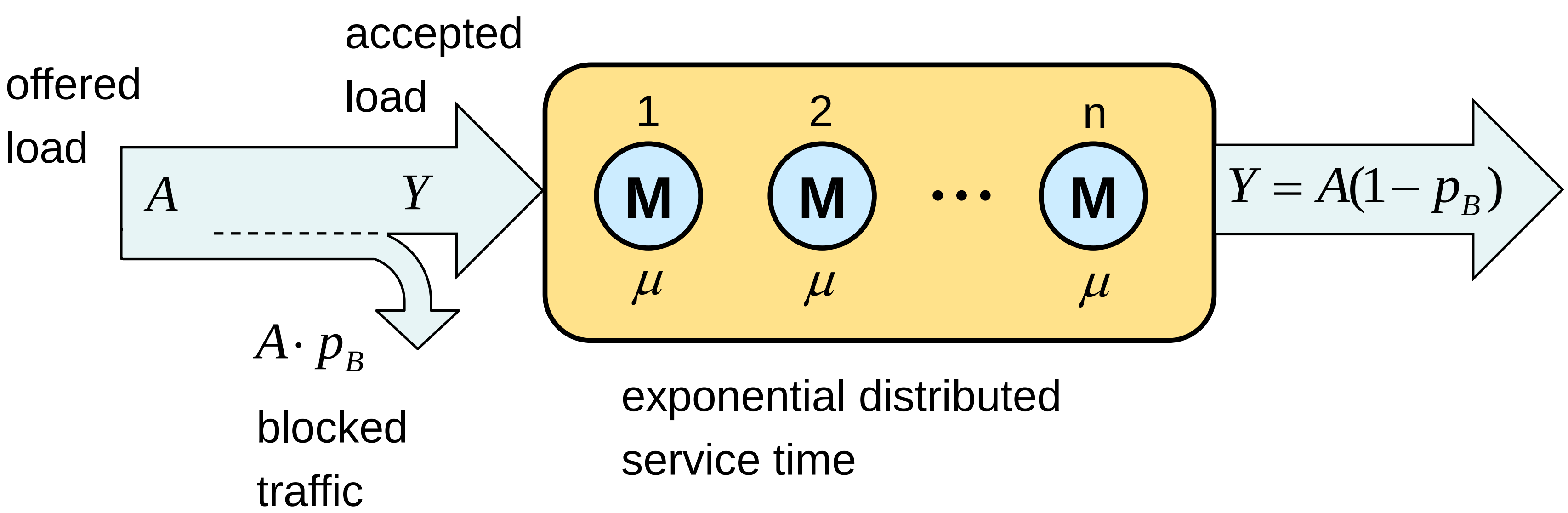
PASTA – Poisson arrivals see time averages

Due to the fact that the arrival process is a memory-less Poisson arrival process. The state probabilities $\{x(i), \quad i = 0,1,\ldots,n\}$ are also valid at the time of arrivals $\{x_A(i), \quad i = 0,1,\ldots,n\}$.

⇒ $\{x_A(i) = x(i), \quad i = 0,1,\ldots,n\}$.

M / M / n – Loss system

- Traffic visualization of an M / M / n – loss system



- M / GI / n – loss system

It can be shown that the Erlang-B equation also holds for loss systems where the service time is NOT exponential distributed.

The proof is not discussed in this lecture since it can be found in:

Syski, R., *Introduction to Congestion Theory in Telephone Systems*, North-Holland, Amsterdam, 1985.

M / M / n – Loss system

- **Blocking probability:**

  Arriving jobs are blocked if all service units are busy.

$$\Rightarrow \quad p_B = x(n) = \frac{\frac{A^n}{n!}}{\sum_{k=0}^{n} \frac{A^k}{k!}}, \qquad i = 0,1,2,\ldots,N$$

- **Traffic load:**

  The traffic load Y represents the average number of busy service units within a system.

$$\Rightarrow \quad Y = \sum_{i=0}^{n} i \cdot x(i)$$

The traffic load is typically described in Erlang [Erl].

M / M / n – Loss system

- The traffic load Y can be derived following the Little-Theorem.

System:

- Average arrival rate of **accepted** jobs: $\lambda(1-p_B)$
- Average retention time in the system = average service time: $E[B]=\frac{1}{\mu}$
- Average number of jobs in the system is given by the traffic load Y.

$$\Rightarrow \quad Y=\lambda(1-p_B)E[B]=\lambda(1-p_B)\frac{1}{\mu}=A(1-p_B)$$

(!) Traffic load depending on the offered load and the blocking probability.

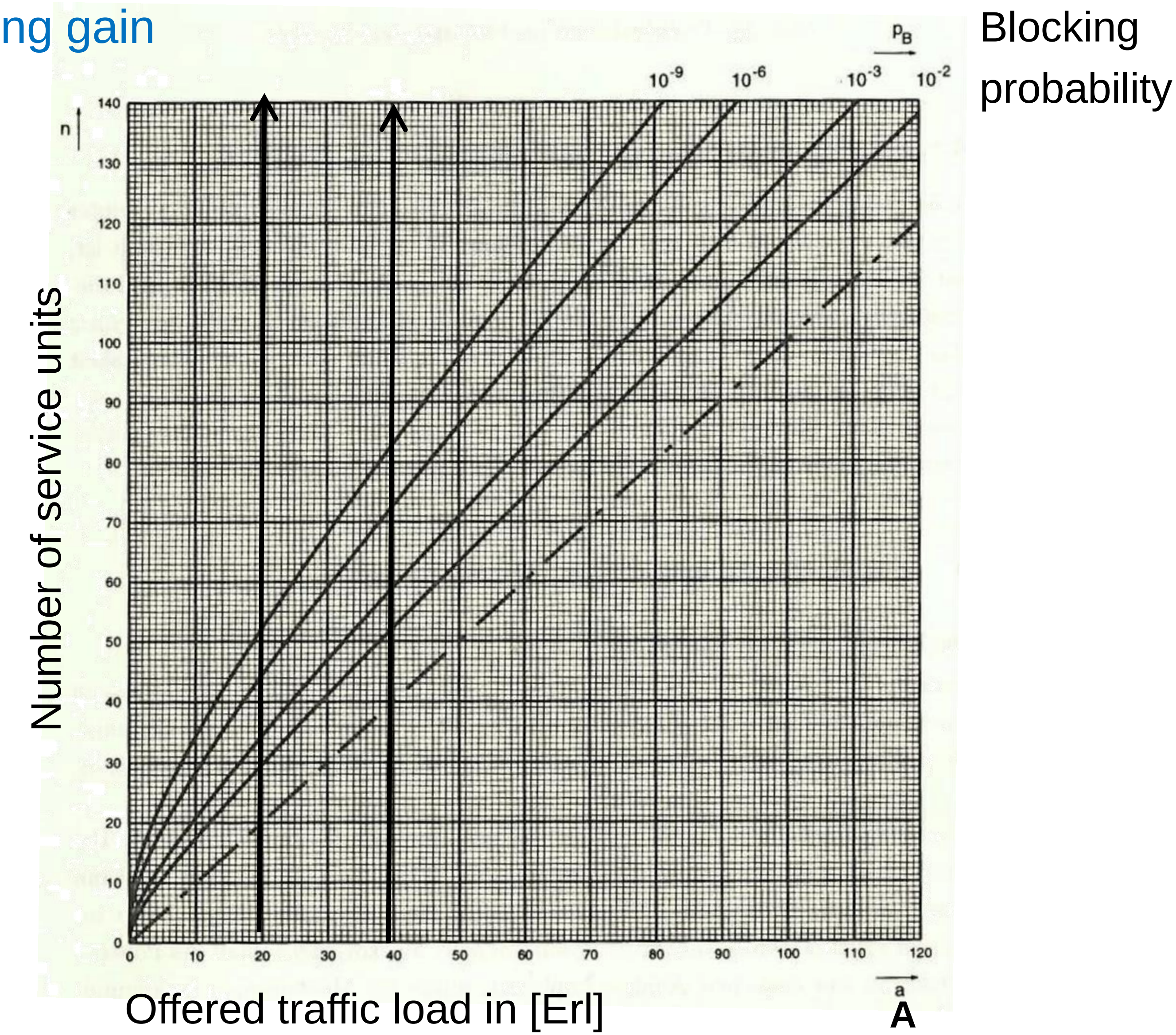
M / M / n – Loss system

- **Multiplexing in circuit-switched networks:**
  - Dimensioning of circuit-switched networks
  - System:
    - Arrival process:
      - An arrival results in a busy link / service unit.
      - Arrivals are blocked if all links / service units are busy.
      - The arrivals are generated by a large group of participants which allows as to assume that the arrival process is a Poisson process.
    - Service process:
      - An accepted arrival results in a busy link / service unit.
  - Question:

    How many links are required for a given traffic load such that the blocking probability remains below a certain threshold?

M / M / n – Loss system

Multiplexing gain



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

M / M / n – Loss system

- Assumption:
  - Offered traffic load is known in advance.

- Example:
  - Offered traffic load $\lambda = 30$ [calls per minute]
  - Average service/call duration $E[B] = 90[Seconds]$

$$\Rightarrow \quad A = \lambda \cdot E[B] = \frac{30}{60s} 90s = 45[Erl]$$

- Target blocking probability $p_B = 10^{-2} \quad \rightarrow \quad n = 58$
- Target blocking probability $p_B = 10^{-3} \quad \rightarrow \quad n = 65$
- Target blocking probability $p_B = 10^{-6} \quad \rightarrow \quad n = 80$

The Erlang-B equation cannot be solved for variable n. Therefore, Pre-calculated tables for different values of A and p_B are used.

M / M / n – Loss system

- Multiplexing gain:
  - In wide-area network, links are usually aggregated in order to benefit from the multiplexing gain.
  - Larger links can be used more efficiently since load variations of single users are ‚compensated' by the large user group.

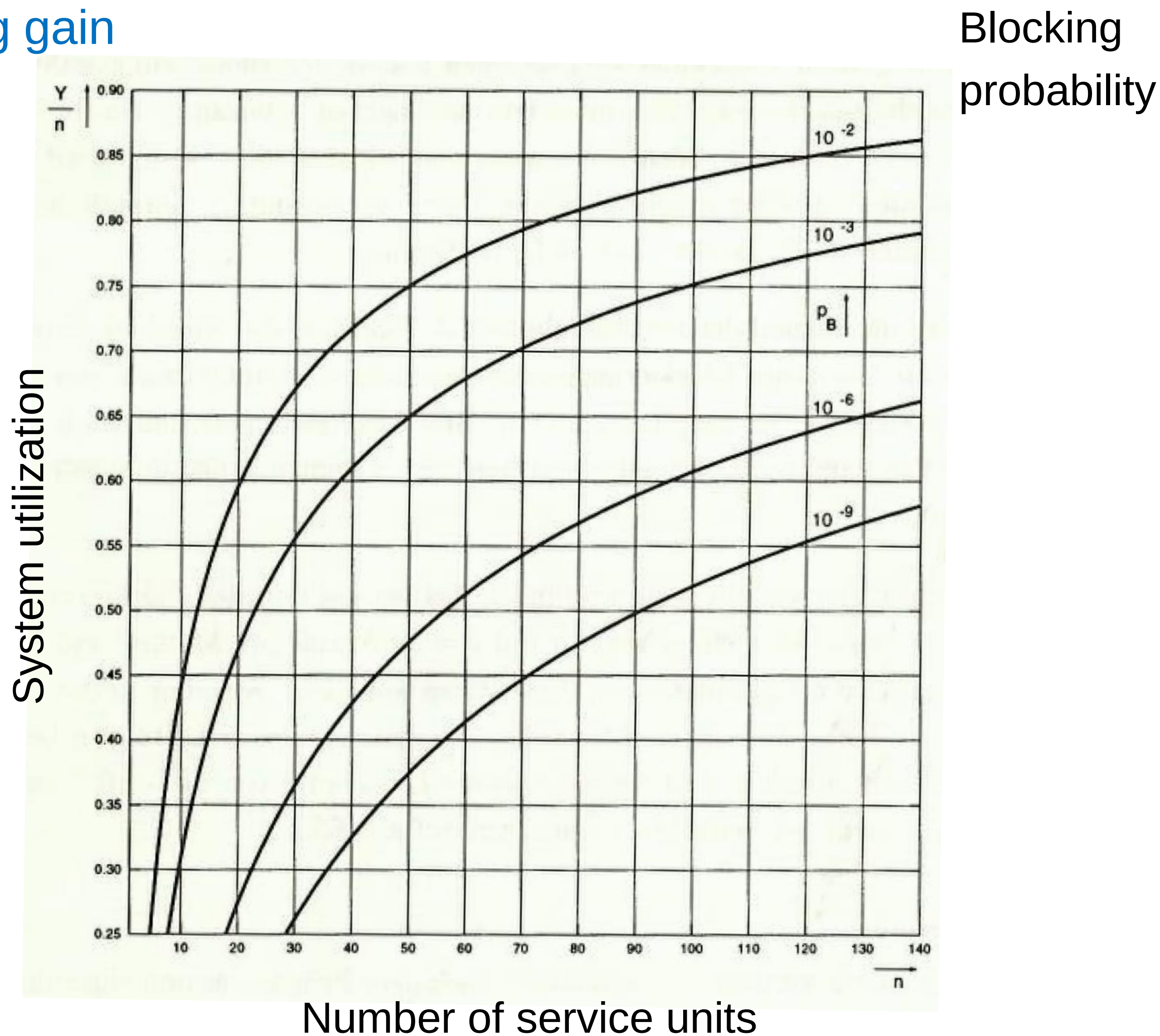
This correlation can be explained and calculated by the Erlang-B equation.

Characteristics:

- The utilization of each link / service unit increases the more links are aggregated while the blocking probability remains on the same level.
- The slope of factor Y/n corresponds to the multiplexing gain.

M / M / n – Loss system

Multiplexing gain



M / M / n – Loss system

- Multiplexing gain:

  ⇨ The multiplexing gain converges for large values of n.

  ⇨ Economical aspects have to be taken into account! A high capacity link may cost much more than two or three low capacity links.

  ⇨ System design is always a trade-off between efficient use of available resources and their costs!

Questions

- How would you analyze a M/M/n – Loss system?
- What is multiplexing gain and how does it work?
- Can you describe the system by a birth-death process?
- Would you prefer a system with a single fast serving unit over a system with many slow serving units?
- What does PASTA mean?
- What is the traffic load in a M/M/n-Loss system?