Performance of Networked Systems

Lecture 2: Introduction to performance modeling

Overview of today's lecture

1. Wrap up of previous lecture
2. Brief history of time: emergence of telephony
3. Exponential distribution
4. Poisson processes and properties
5. Markov chains, equilibrium distributions
6. Erlang-B model
7. Reliability design

Background reading material for this course:
O.J. Boxma, Stochastic Performance Modelling, chapters 1, 2 and 3

The Dutch Soccer Frustrations

2010: Arjen Robben and Casillas' toe...

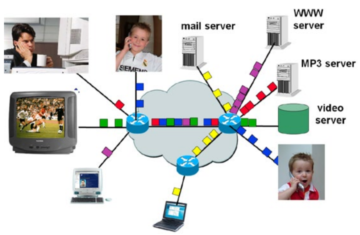
1974: 1-2 loss against West-Germany

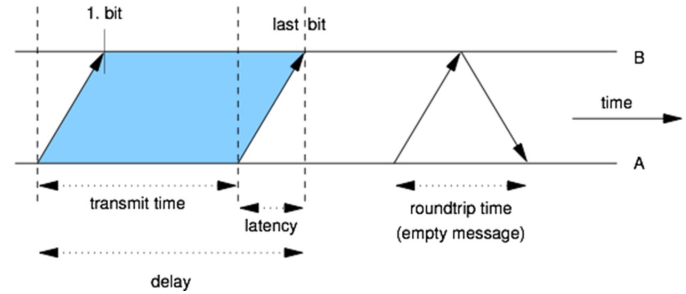
1978: Rob Rensenbrink hits the pole...

1988: European champion!

https://youtu.be/q3af4QIh1oc

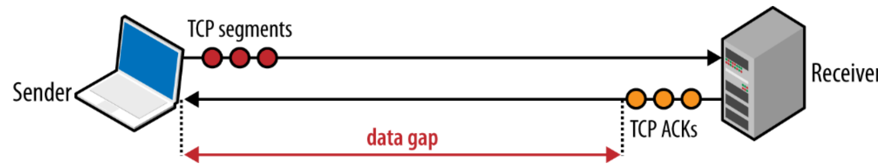
Wrap up of Last Lecture

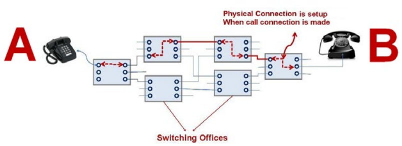


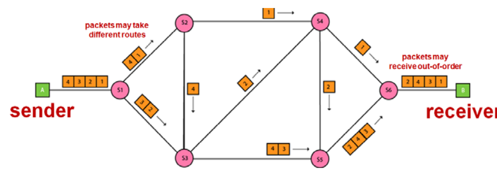


20% packet loss

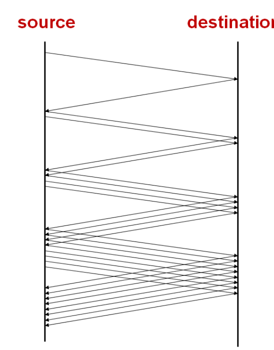
original

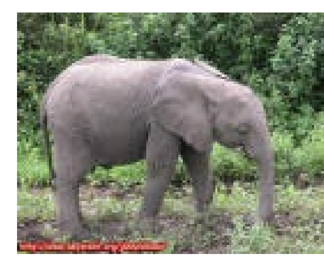




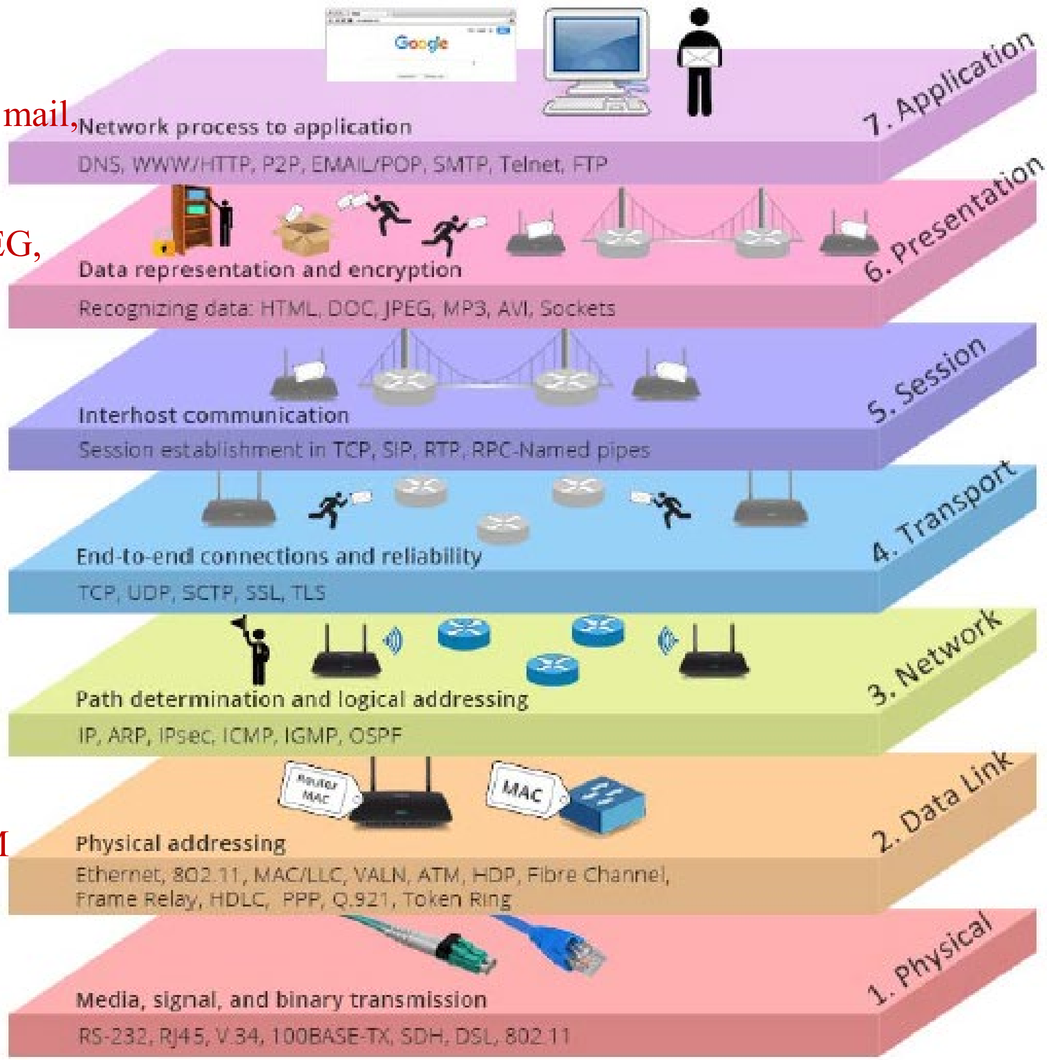


1. Examples of performance degradation
2. Performance-critical systems
3. Course overview
4. Basics and networking principles
5. Bandwidth versus latency





The OSI Reference Model



HTTP, FTP, mail, WWW	7. Application	supports application and end-user processes
HTML, JPEG, MPEG	6. Presentation	transforms data into the form that the application layer can accept
NFS, SIP	5. Session	establishes, manages and terminates connections between applications
TCP, UDP	4. Transport	flow control, data transfer between end systems (hosts)
IP	3. Network	switching, routing between nodes
802.11, ATM	2. Data Link	medium access control synchronization, error checking
RS232	1. Physical	hardware, cables

The OSI Reference Model

7 Layers

HTTP, FTP, browsers

JPEG, MPEG, HTML

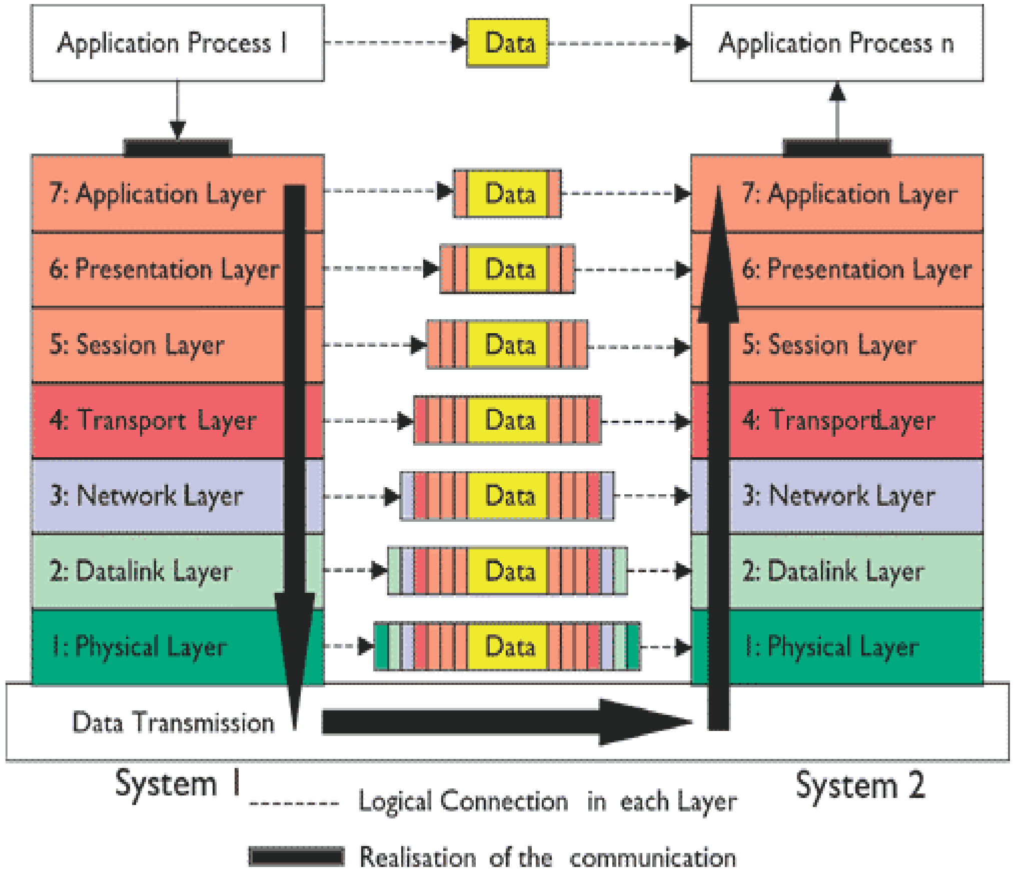
SIP, NFS

TCP, UDP

IP

802.11, ATM

RS232



supports application and end-user processes

transform data into the form that the application layer can accept

establishes, manages and terminates connections between application

flow control, data transfer between end systems (hosts)

switching, routing between nodes

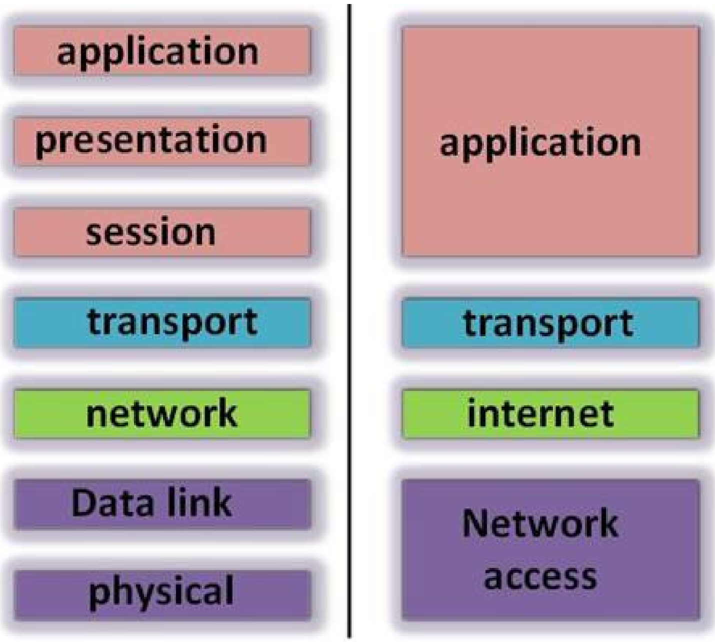
medium access control synchronization, error checking

hardware, cables

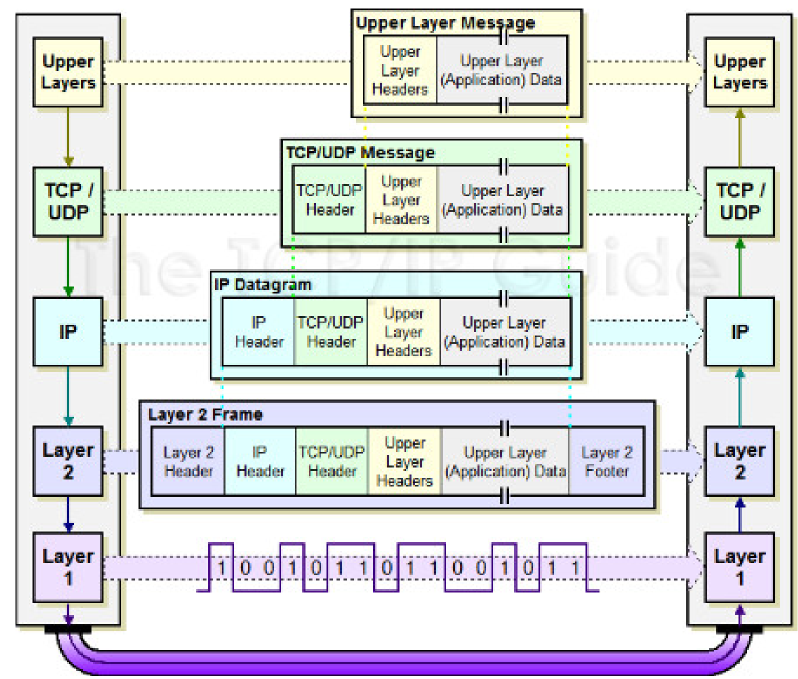
- OSI stack is a standard for how communication networks work
- Different layers of abstraction
- Encapsulation

OSI Model and TCP/IP Stack

OSI model **TCP/IP model**

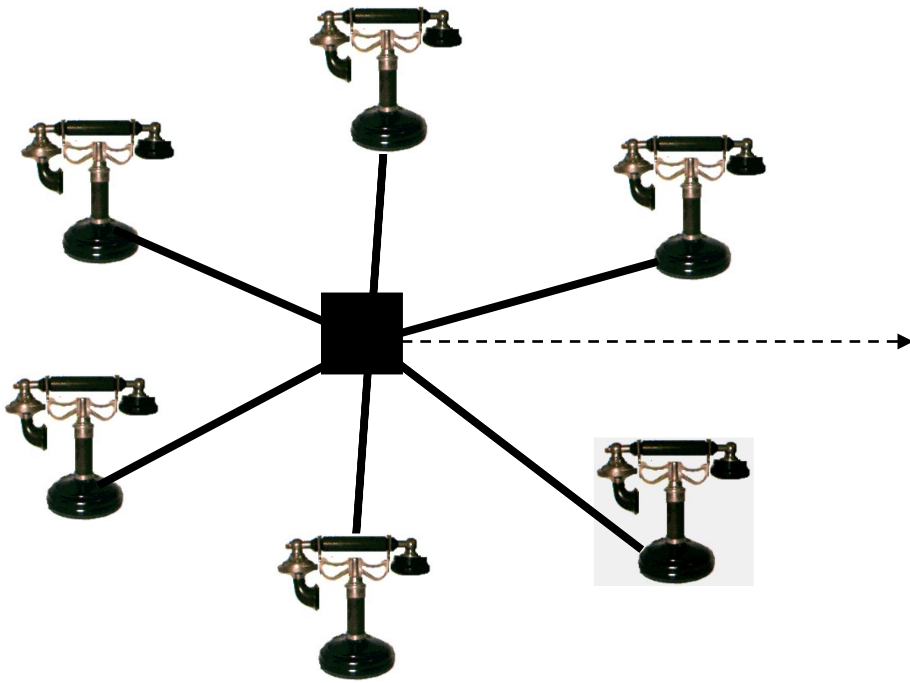


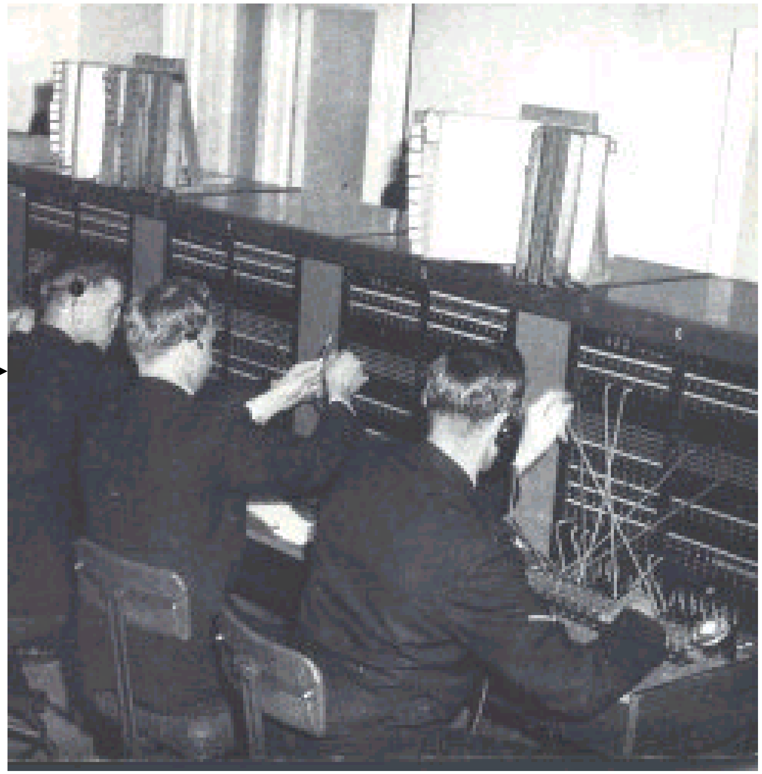
TCP/IP stack



- The **OSI model** is a **conceptual framework** using which the functioning of a network can be described (theoretical, for education purposes, to understand networking)
- The **TCP/IP model** is a 4-layer **practical model**, forms the basis for the Internet, built on specific protocols (TCP, UDP and IP), actually used for Internet communication
- The **TCP/IP stack** is a **concrete implementation** of the TCP/IP model

A Brief History of Time…



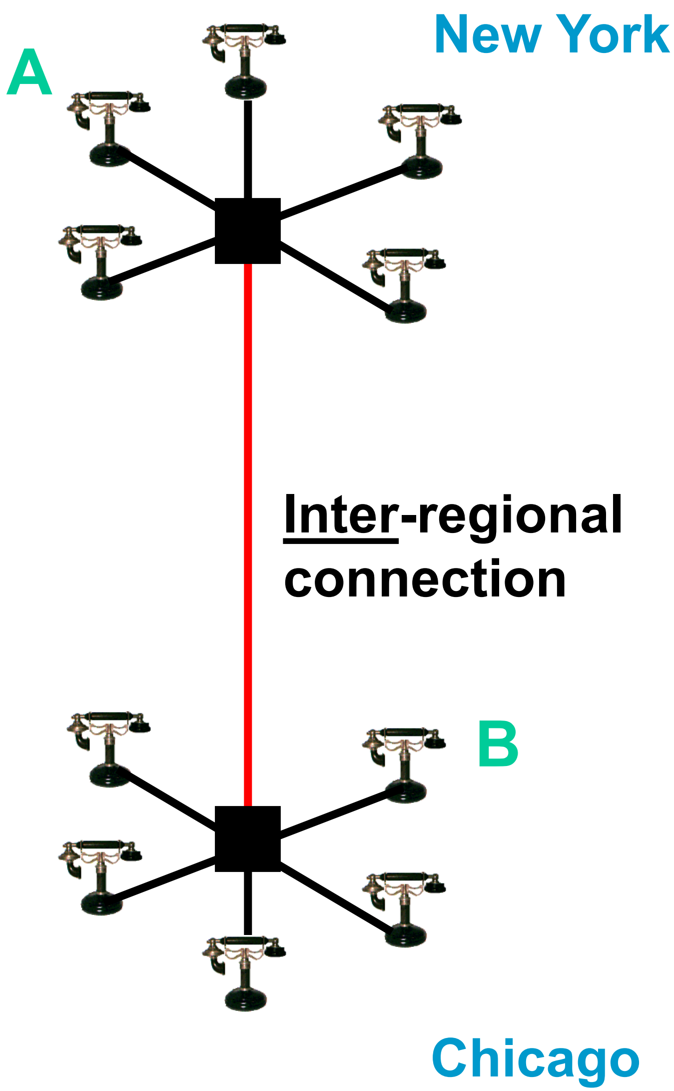


local network

switching center

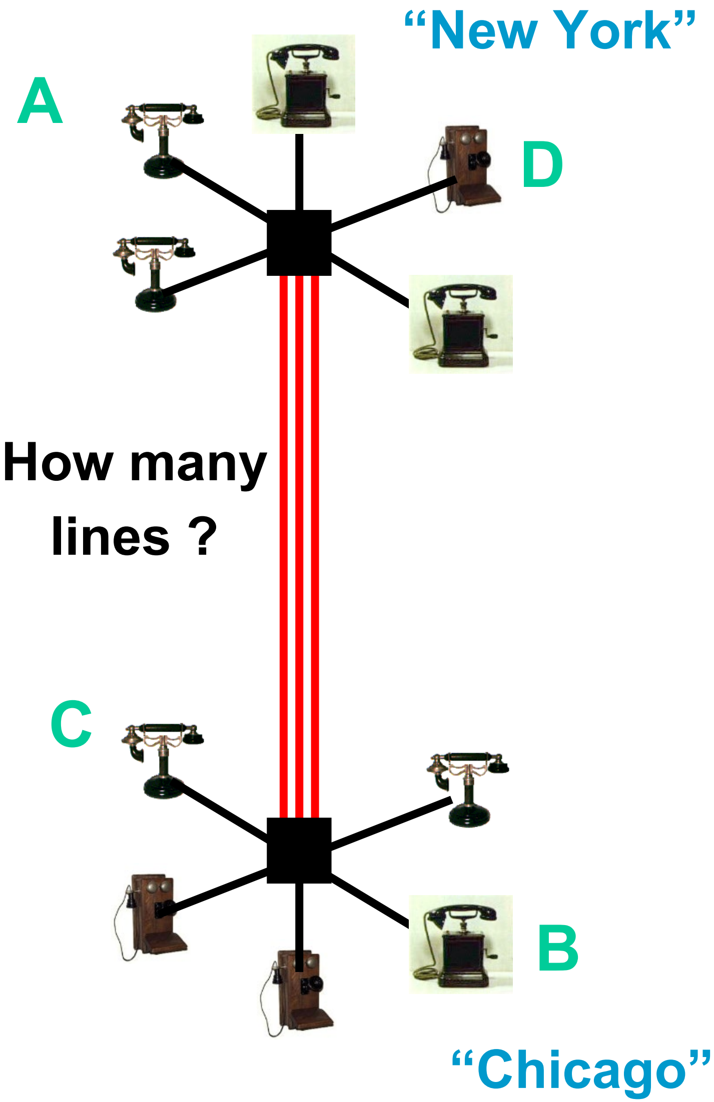
- The good-old days of Graham Bell…

Inter-Regional Calls...



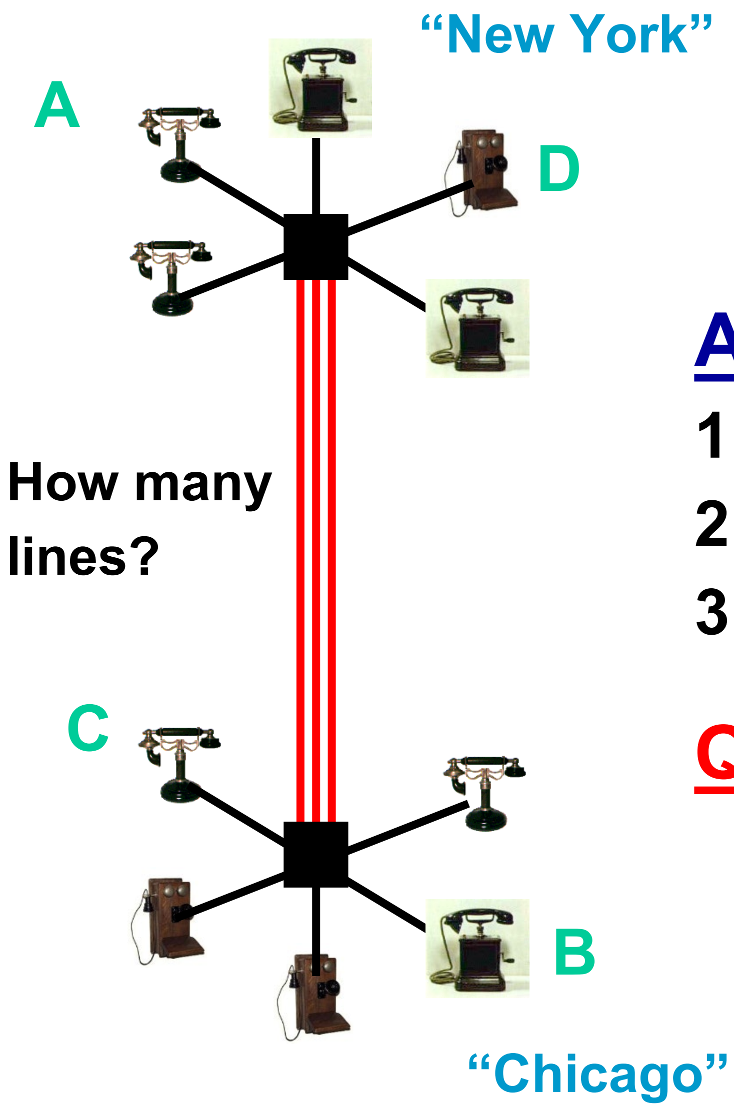
1892

The First Traffic Problems...



The Birth of Performance Analysis...

The single-rate model



Assumptions:

1. number of "calls" per minute = λ
2. average call duration = β
3. number of lines = N

Question: blocking probability ?

The "Erlang-B Formula"

Blocking probability =

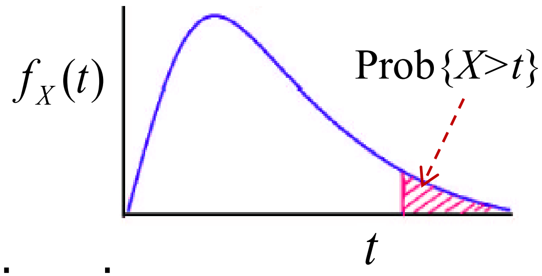
$$\frac{\dfrac{(\lambda\beta)^N}{N!}}{1+\dfrac{(\lambda\beta)^1}{1!}+\dfrac{(\lambda\beta)^2}{2!}+\ldots+\dfrac{(\lambda\beta)^N}{N!}}$$

"Erlang calculator"

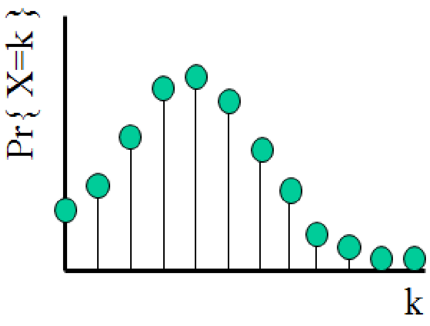
Agner Krarup Erlang
(1878-1929)

Probability Distributions

- **Continuous random variables**
  - **can have any real value (not restricted to integer)**
  - **typical examples:**
    - time between successive breakdown events
    - waiting time, job processing time, response time in RTS



- **Discrete random variables**
  - **can have only integer values**
  - **typical examples:**
    - number of job arrivals in given time interval
    - number of system failures in given time interval

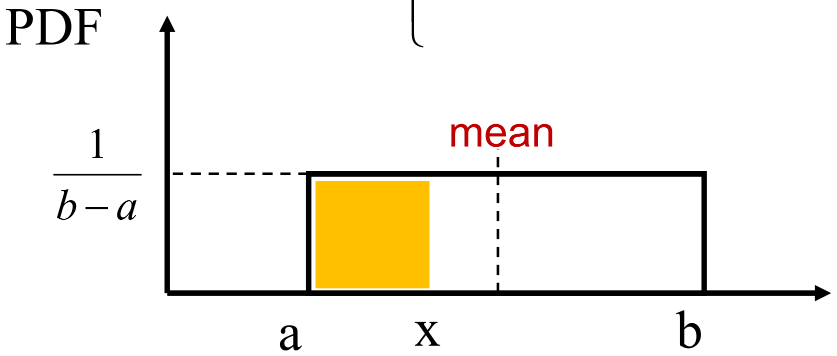


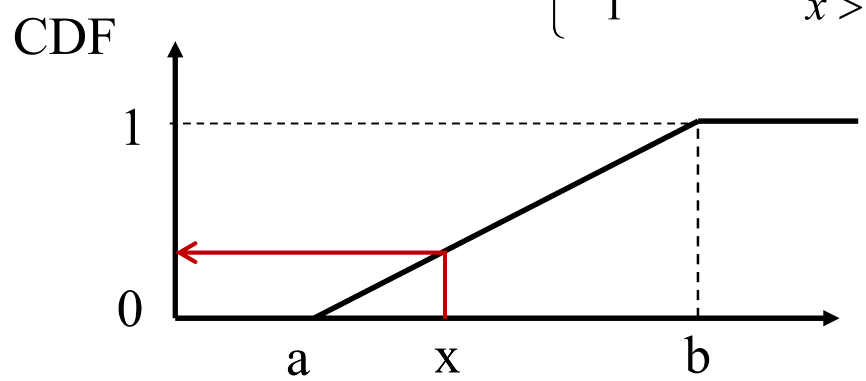
Uniform Distribution

‘Pick a random number between a and b’

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases}$$

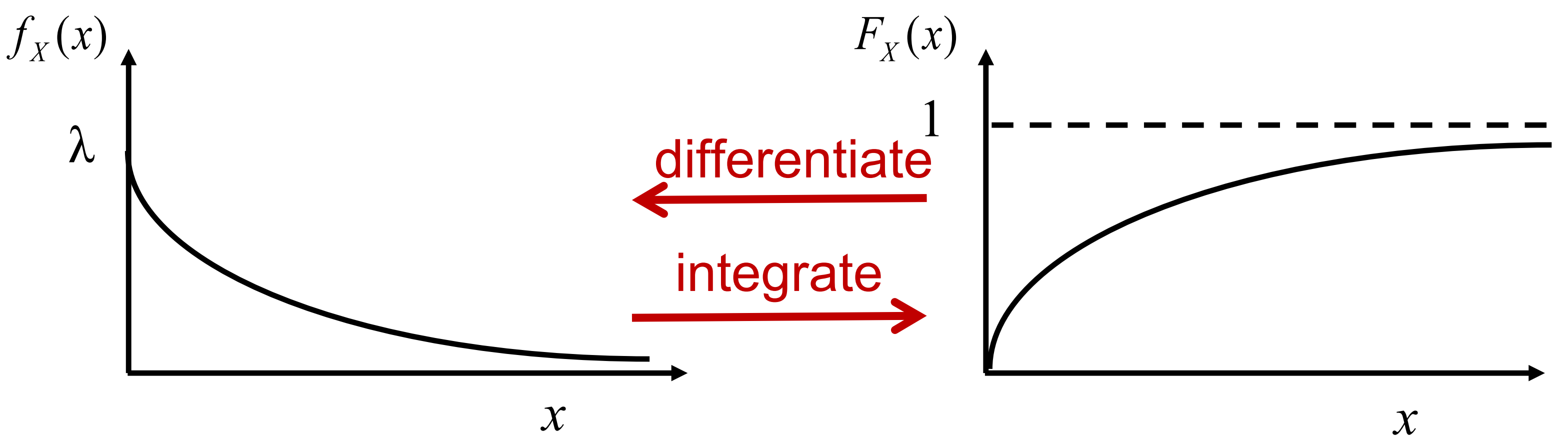




- **Interpretation:** “Pick random number between a and b”
- **Examples:**
  - waiting time at bus stop (when arriving at random time)
  - time until regular maintenance activity
- **Mean and variance:** $E[X] = \dfrac{a+b}{2}$ $\quad Var[X] = \dfrac{(b-a)^2}{12}$

Exponential Distribution

Probability Density Function (PDF) **Cumulative Density Function (CDF)**



- **PDF:** $f_X(x) = \lambda e^{-\lambda x} \; (x > 0)$
- **CDF:** $F_X(x) := \Pr\{X \le x\} = \int_{t=0}^{x} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \; (x > 0)$
- **Mean and variance**

$$E[X] := \int_{t=0}^{\infty} t f_X(t) dt = \lambda \int_{t=0}^{\infty} t e^{-\lambda t} dt = 1/\lambda; Var[X] = 1/\lambda^2$$

Properties of Exponential Distribution

Memoryless property: fundamental!

$$\Pr\{X > s + t \mid X > s\} = \Pr\{X > t\}$$

- **Example:**
  - X := lifetime of a component
  - Prob{ X > 30 | given X > 10 } = Prob{ X > 20 }
  - information about "past" no influence on future
  - "used is a good as new"

The exponential distribution occurs naturally in the modeling of 'completely random' events

Properties of Exponential Distribution

Minimum property: if X, Y independent exponential with parameters λ_1 and λ_2, then min{X,Y} exponential with parameter $\lambda_1 + \lambda_2$

- **Notation:**

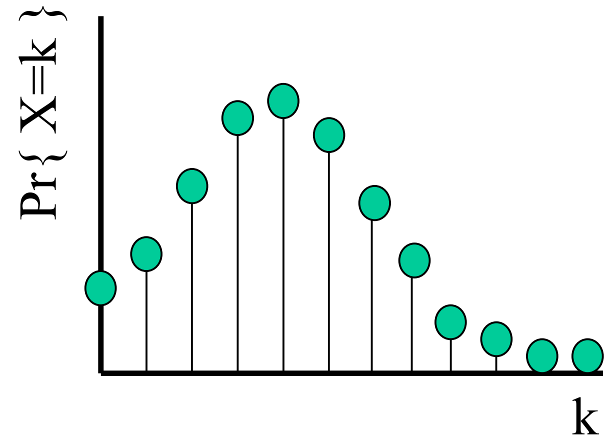
$$X \sim \exp(\lambda_1), Y \sim \exp(\lambda_2) \rightarrow \min\{X, Y\} \sim \exp(\lambda_1 + \lambda_2)$$

- **Interpretation:**
  - assume 10 items
  - each with exponential lifetime with mean L
  - T := time until first failure
  - then T exponential with mean L/10

Poisson Distribution

- X has a Poisson distribution with parameter λ if

$$\Pr\{X=k\}=e^{-\lambda}\lambda^{k}/k!\,(k=0,1,\ldots)$$



- mean

$$E[X]=\sum_{k=1}^{\infty}k\Pr\{X=k\}=e^{-\lambda}\lambda\sum_{k=1}^{\infty}\lambda^{k-1}/(k-1)!=\lambda$$

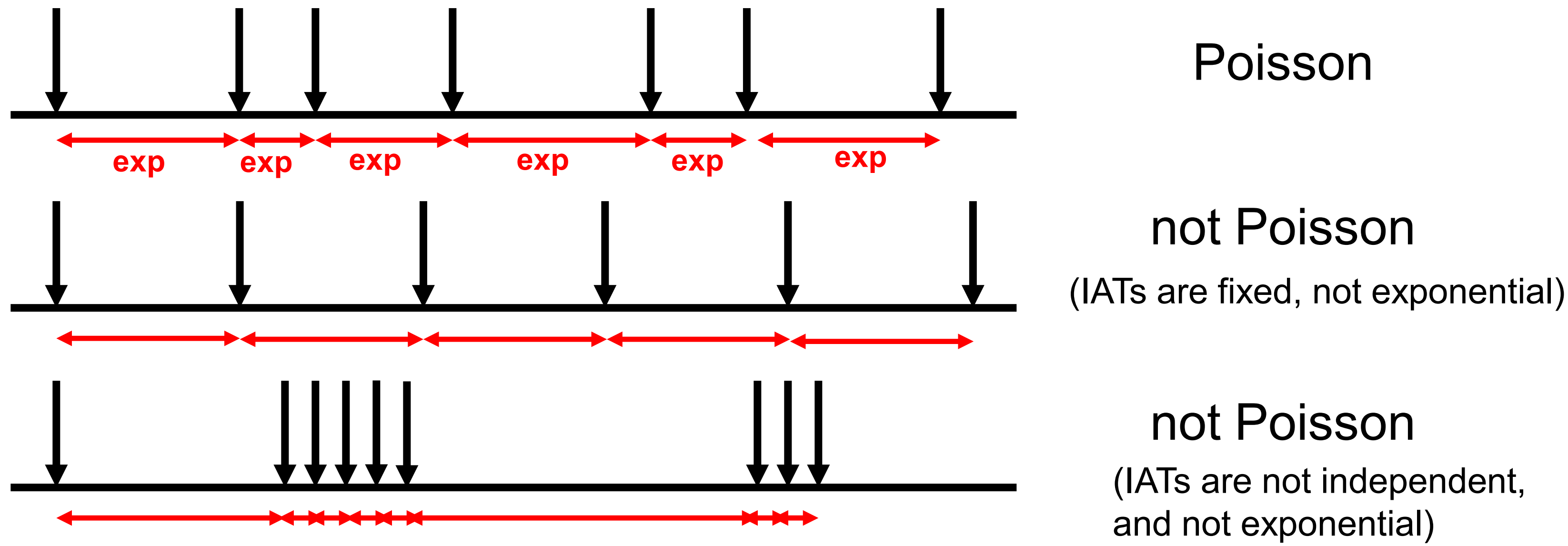
- variance

$$E[X(X-1)]=\sum_{k=2}^{\infty}k(k-1)\Pr\{X=k\}=e^{-\lambda}\lambda^{2}\sum_{k=2}^{\infty}\lambda^{k-2}/(k-2)!=\lambda^{2}$$

$$Var[X]=E[X^{2}]-\left(E[X]\right)^{2}=\lambda^{2}+\lambda-\lambda^{2}=\lambda$$

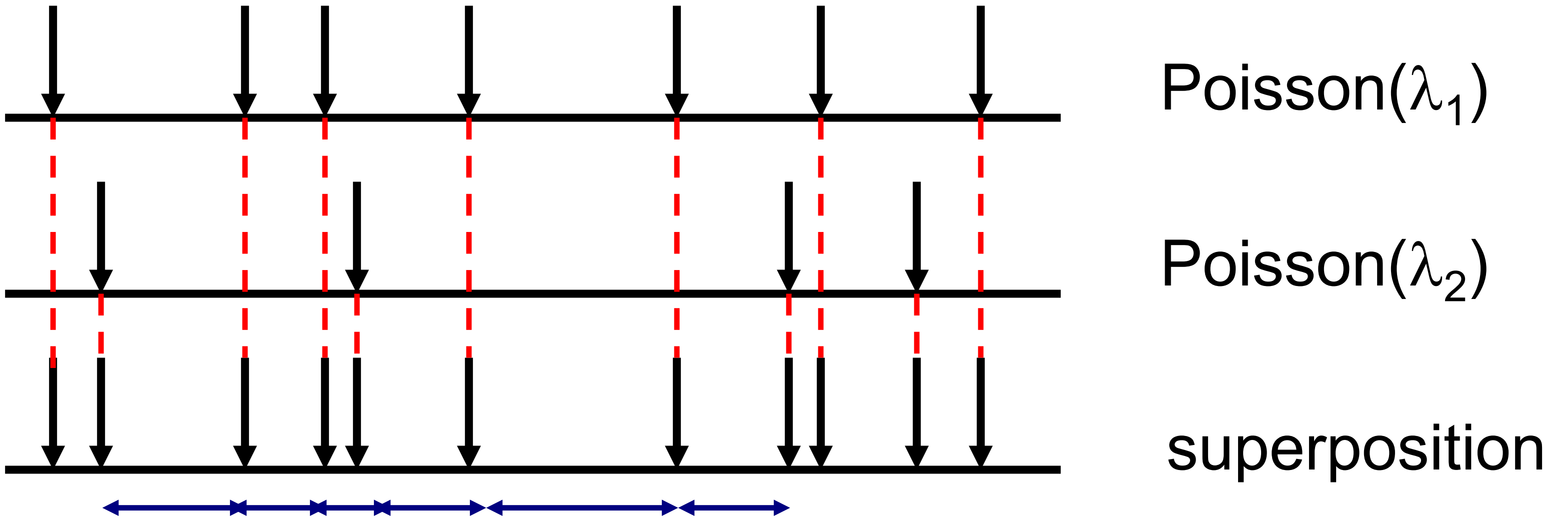
- occurs frequently in stochastic processes

Poisson Process



- Poisson process is a stochastic process (with rate λ)
- Inter-arrival (IAT's) times should satisfy following **two requirements**
    1. mutually independent
    2. exponentially distributed (each with the same mean $1/\lambda$)
- Poisson process models "completely random" events
- natural way to model human-initiated events
    - occurrence of rare events (for example in real-time systems)
    - customers arriving at a post office
    - initiation of telephone calls

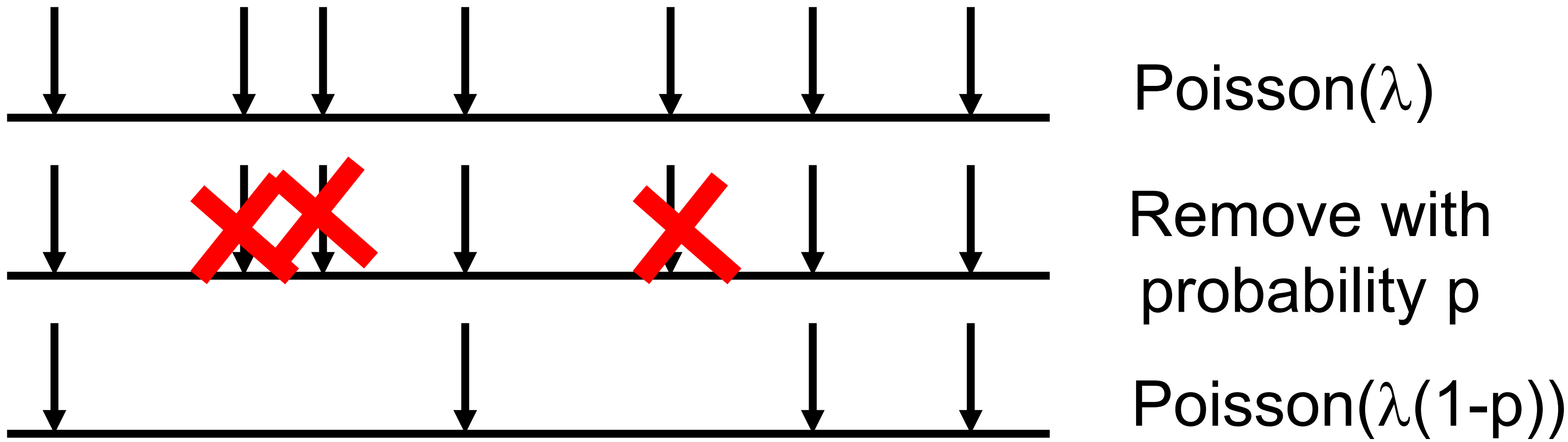
Poisson Process Properties:
Superposition Property



Superposition Property: superposition of two independent Poisson processes with rate λ_1 and λ_2 is a Poisson process with rate $\lambda_1 + \lambda_2$

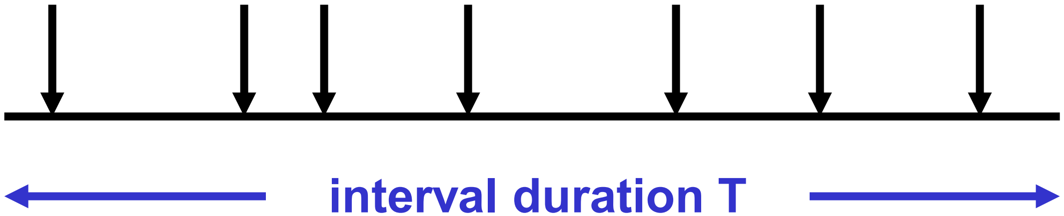
Poisson Process Properties:

Thinning-out Property



Thinning-out Property: if Poisson arrivals are "removed" with some fixed probability p, then resulting process is Poisson process with rate $\lambda(1-p)$

Poisson: Process and Distribution



Poisson process
with rate λ

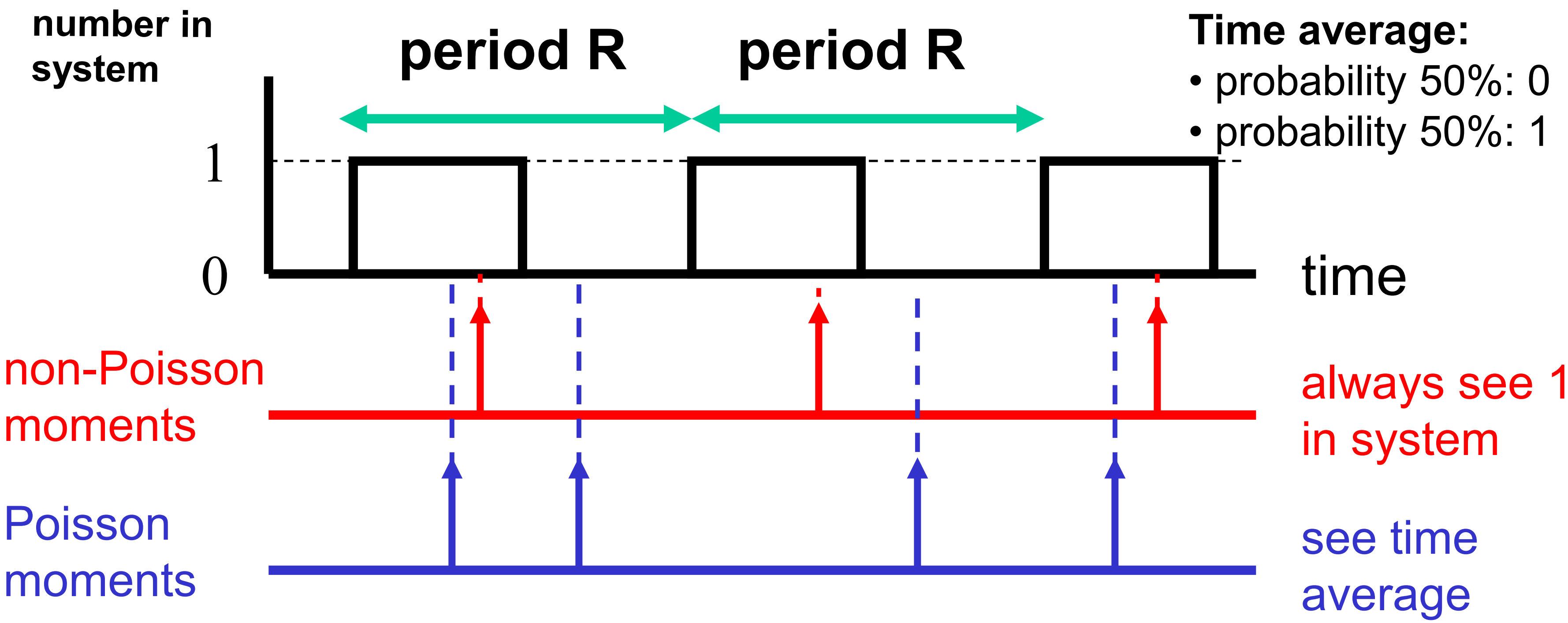
<u>If</u> events occur according to Poisson <u>process</u> with rate λ
<u>and</u> N:= number of events in interval of length T
<u>then</u> N has Poisson <u>distribution</u> with parameter λT:

$$\Pr\{N = k\} = e^{-\lambda T}\frac{(\lambda T)^k}{k!}(k = 0,1,\ldots)$$

Mean and variance:

$$E[N] = \lambda T \qquad Var[N] = \lambda T$$

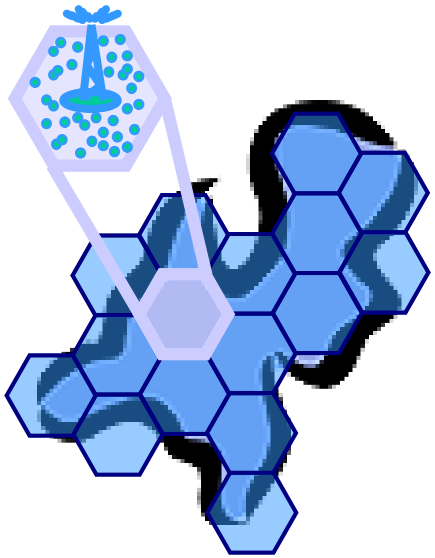
Poisson Arrivals See Time Averages

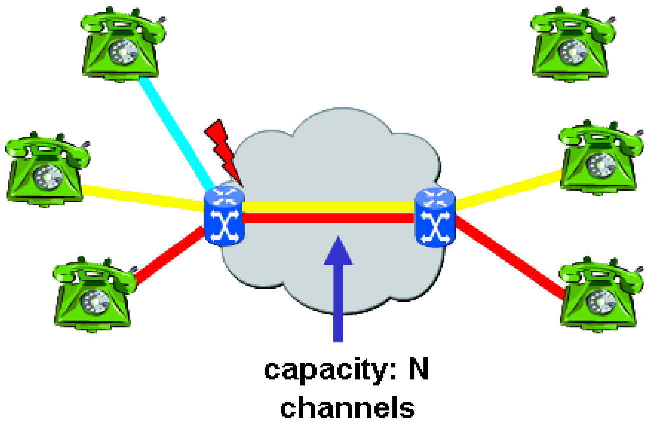


Time average:
- probability 50%: 0
- probability 50%: 1

- **time average distribution: what there <u>is</u>**
- **not always same as what an <u>arriving</u> customer <u>sees</u>**
- **PASTA: for Poisson arrivals, arriving customers see time average**

Planning Problem

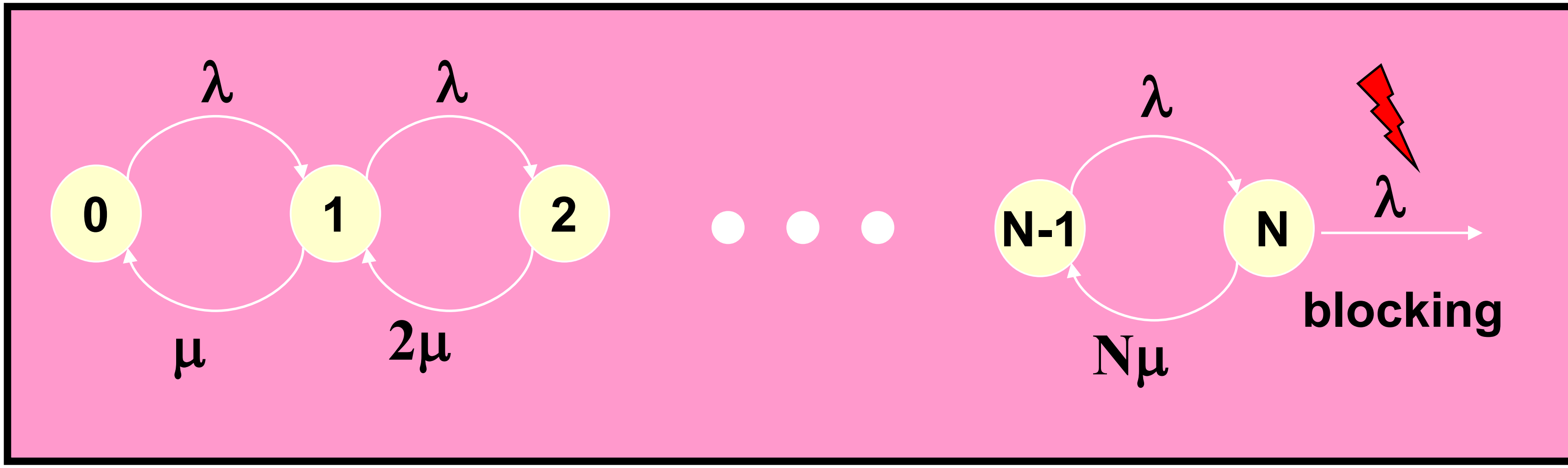




- **<u>Given</u>:** 1. maximum blocking probability $\alpha \in (0;1]$
  2. Poisson arrival process with rate λ
  3. mean call holding time β (exponential)

- **<u>Question</u>:** "Determine smallest N such that Prob { call blocked } $\leq \alpha$?" (α is target quality)

Continuous-Time Markov Chain



- $N(t)$:= # busy lines at time t ($t > 0$)
- continuous-time Markov chain $\{ N(t), t > 0 \}$
- state space $S := \{0,1,\ldots,N\}$
- birth-and-death process
- 'transition rates':

$$\text{Prob}\ \{ i \rightarrow i + 1 \text{ in } (t; t + \Delta t] \} = \lambda\Delta t + o(\Delta t), \text{ for } \Delta t \downarrow 0$$

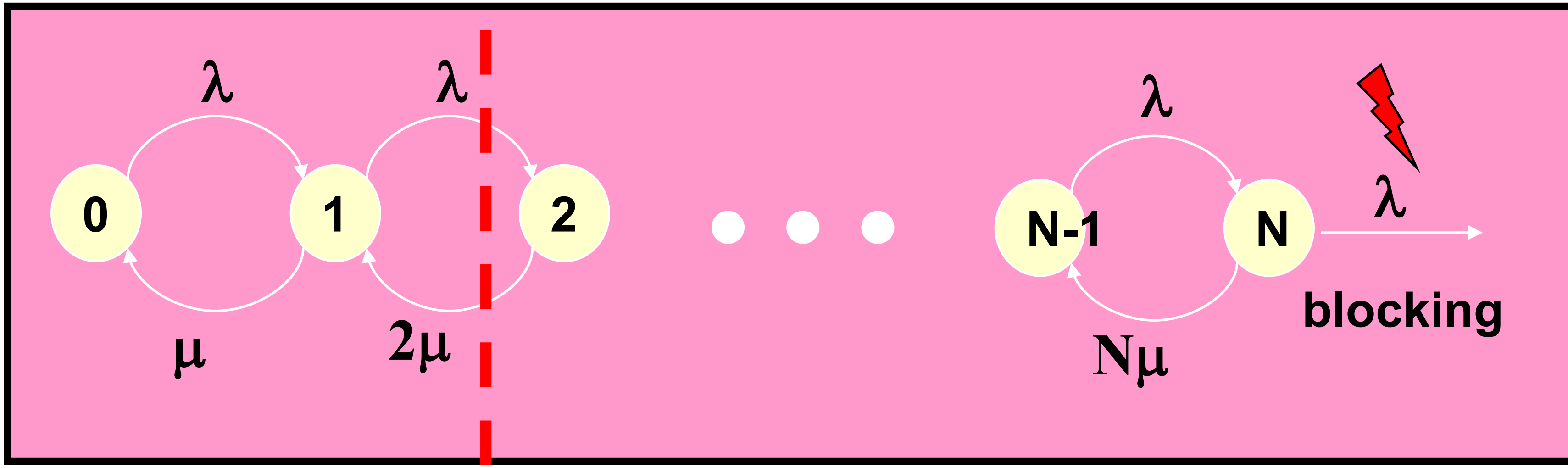
$$\text{Prob}\ \{ i \rightarrow i - 1 \text{ in } (t; t + \Delta t] \} = i\square\ t + o(\Delta t), \text{ for } \Delta t \downarrow 0$$

- interpretation of **transition rate**: reciproke of average time

$\beta=1/\mu$

β = mean service time
μ = service rate

Continuous-Time Markov Chain



- **stationary distribution**

  $\pi_k := \lim_{t \to \infty}$ Prob{ N(t) = k }   (k = 0, 1,…, N)

- **balance equations ("rate in = rate out")**

  1. $\lambda\, \pi_{k-1} = k\mu\, \pi_k$ (k=1,2,…,N)
  2. $\Sigma\, \pi_k = 1$ (normalisation)
  3. $\pi_0 \Rightarrow \pi_1 \Rightarrow \ldots \Rightarrow \pi_N$, and then normalize

Erlang Blocking Formula

Solution of Markov chain:

$$\pi_k = \frac{(\lambda\beta)^k / k!}{\sum_{i=0}^{N} (\lambda\beta)^i / i!} \quad (k = 0, 1, \ldots, N)$$

Agner Krarup Erlang
(1878-1929)

- **Use Poisson Arrivals See Time Averages (PASTA)**
- **Blocking probability**

Prob { call blocked } = $\pi_N = \frac{(\lambda\beta)^N / N!}{\sum_{i=0}^{N} (\lambda\beta)^i / i!}$

Insensitivity Property: formula also valid for non-exponential call holding times

The "Erlang-B Formula"

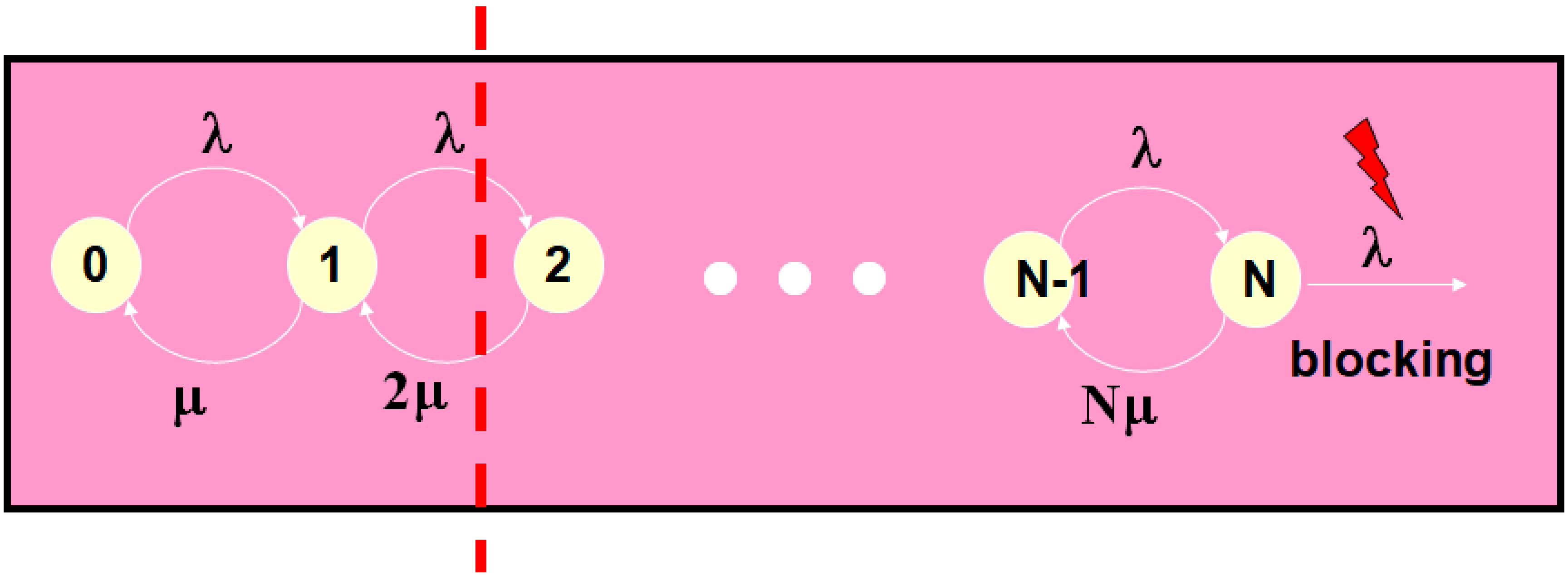
Blocking probability =

$$\frac{\dfrac{(\lambda\beta)^N}{N!}}{1+\dfrac{(\lambda\beta)^1}{1!}+\dfrac{(\lambda\beta)^2}{2!}+\ldots+\dfrac{(\lambda\beta)^N}{N!}}$$

"Erlang calculator"

Agner Krarup Erlang
(1878-1929)

Recipe for Solving Blocking Probabilities



Step 1: Define Markov Chain to describe the system

Step 2: Write down balance equations

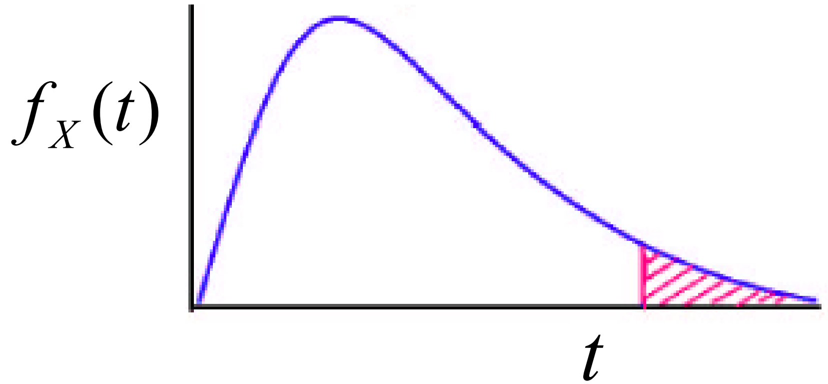
Step 3: Determine steady-state distribution $\underline{\pi} = (\pi_0, \ldots, \pi_N)$

Step 4: Use PASTA

Step 5: Express blocking probabilities in terms of $\underline{\pi}$

Design for Reliability

Probability density function of X



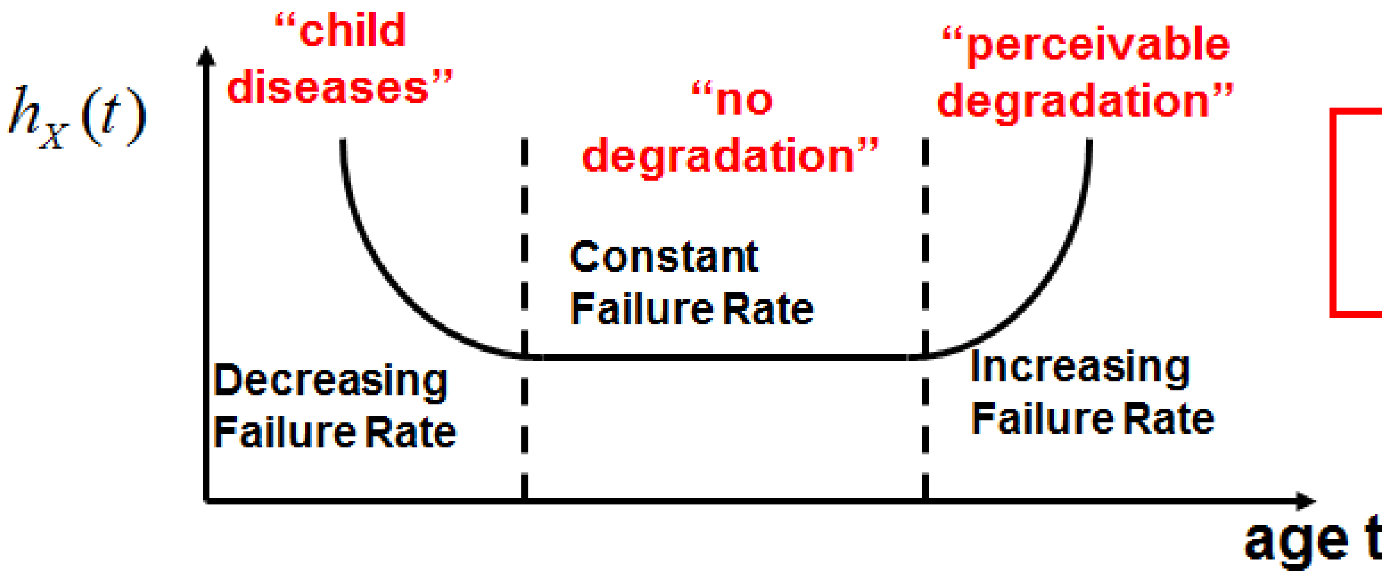
<u>Reliability</u> of a component (or system) is its ability to function correctly over a specified period of time

X := lifetime of a system (also called time-to-failure)

R(t) := Pr { S is fully functioning in [0,t] } (t≥0)

Usually we assume that system works at time t=0: $R(0)=1$

Bathtub Curve



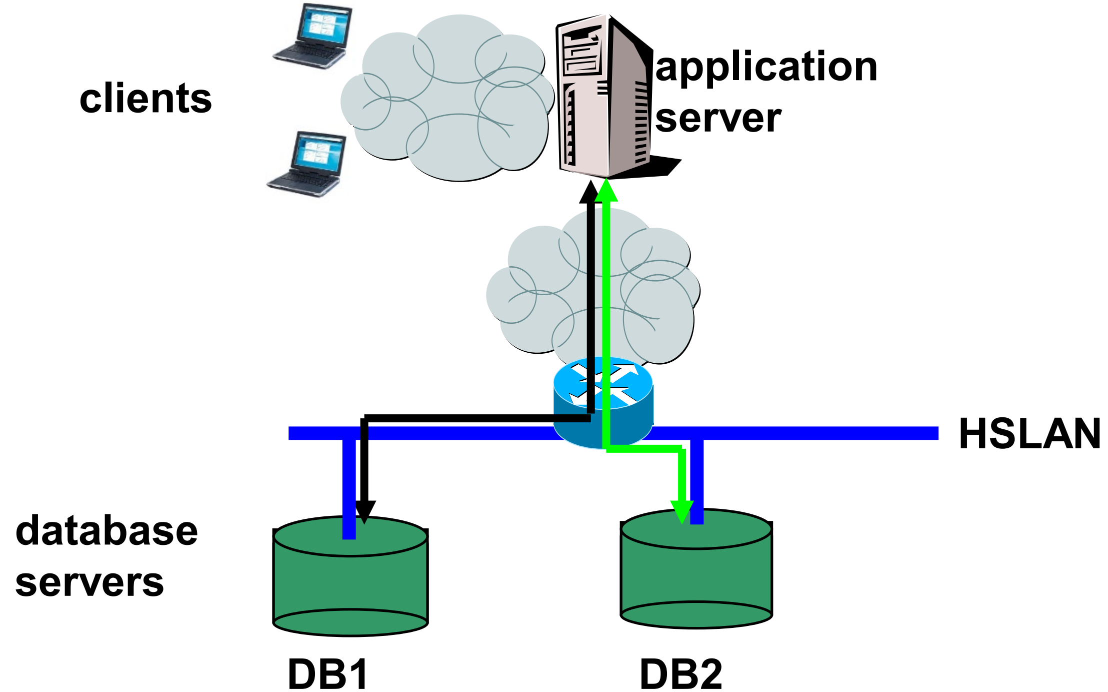
"Bathtub-curve"

Hazard rate of failure (HRF):

$$h_X(t) = \frac{\Pr\{\text{failure occurs in } [t, t + \Delta t], \text{ given component has age } t\}}{\Delta t}$$

Interpretation: HRF of a component X is an indication of the "likelihood" of failure at age t

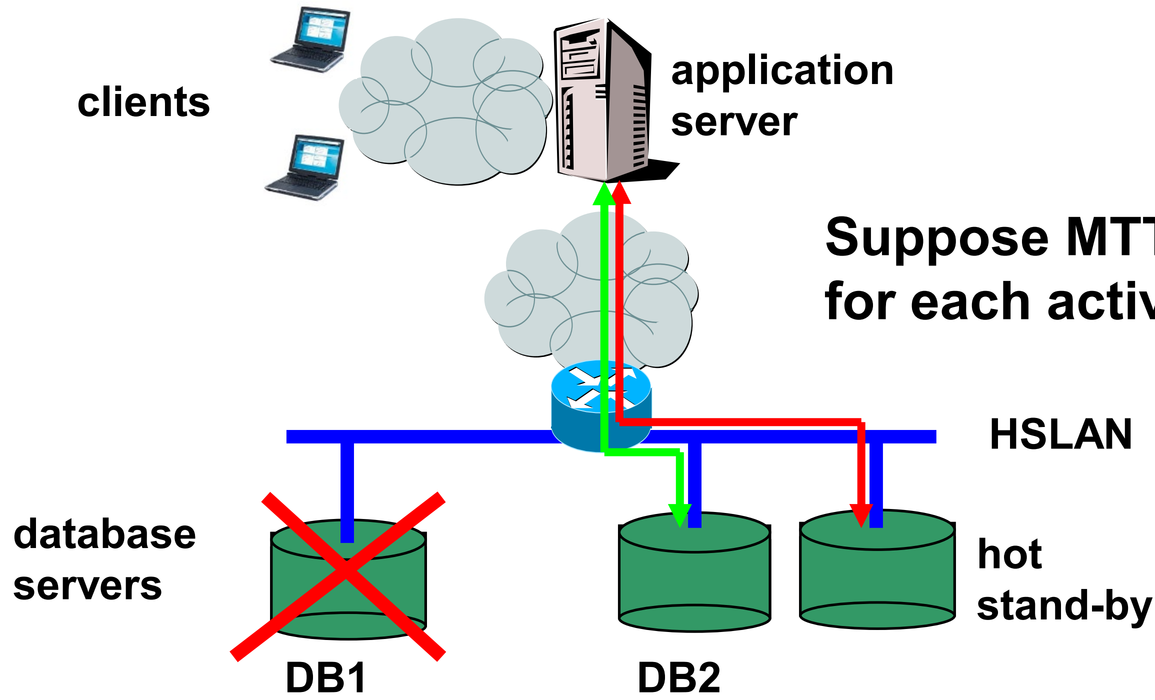
DBMS Reliability Design



Suppose Mean Time to Failure (MTTF) = 1 year for each DB server

- overall MTTF = 6 months
- probability TTF > 6 months = 37%
- probability TTF > 12 months = 14%
- probability TTF > 60 months = 0.0545%

DBMS Reliability Design



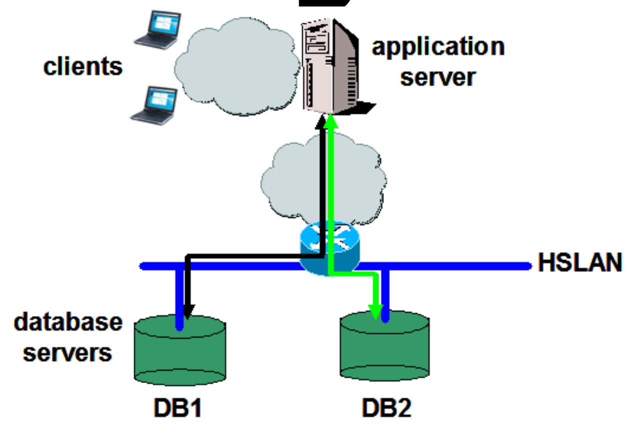
Suppose MTTF = 1 year for each active DB server

- overall MTTF = 6 + 6 = 12 months
- probability TTF > 6 months = 74% (roughly 2 times case without standby)
- probability TTF > 12 months = 41% (about 3 times case without standby)
- probability TTF > 60 months = 0.05% (> 10 times case without standby)

Great impact of "hot standby" … but how to calculate these figures?

DBMS Reliability Design

Suppose: MTTF = 12 months for each active server



Then

1. TTF_1 and TTF_2 are independent and exponentially distributed with mean 12 months (rate $\lambda_1 = \lambda_2 = 1/12$ per month)

2. Total TTF = $\min\{TTF_1, TTF_2\}$ exponential with rate $\lambda = \lambda_1 + \lambda_2 = 1/6$ per month

number of failures in time interval [0;t] has a Poisson distribution with mean λt

$$\Pr\{0 \text{ failures}\} = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

Reliability function:

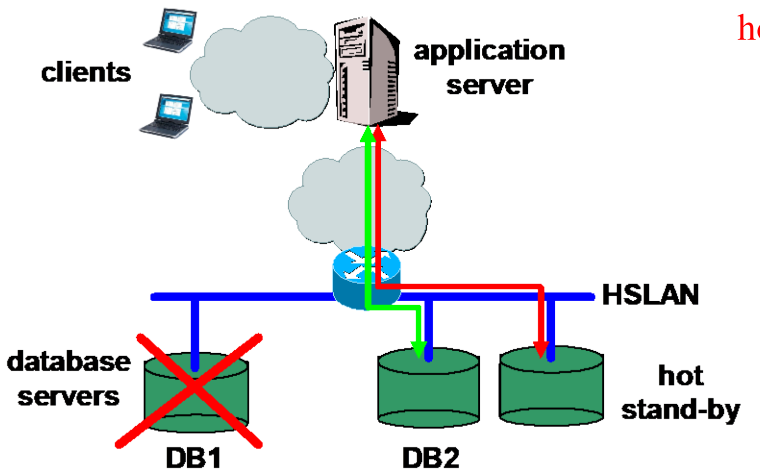
$$R(t) = \text{Prob}\{TTF > t \text{ months}\} = \text{Prob}\{0 \text{ failures in } [0,t]\} = e^{-\lambda t} = e^{-(1/6)t}$$

$$\text{Prob}\{TTF > 6 \text{ months}\} = e^{-(1/6)\times 6} = 1/e \approx 37\%$$

$$\text{Prob}\{TTF > 12 \text{ months}\} = e^{-(1/6)\times 12} = 1/e^2 \approx 14\%$$

$$\text{Prob}\{TTF > 60 \text{ months}\} = e^{-(1/6)\times 60} = 1/e^{10} \approx 0.00454\%$$

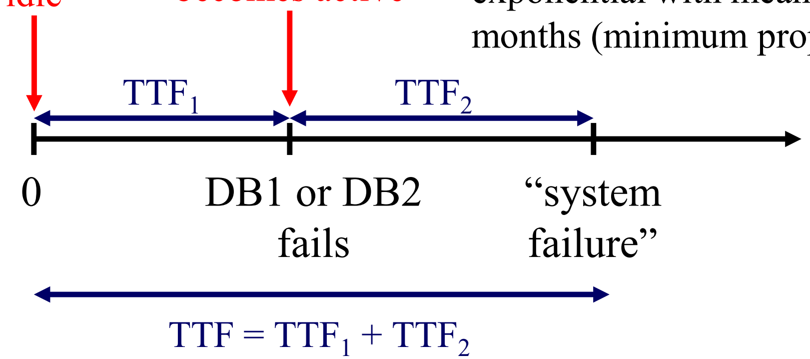
DBMS Reliability Design



hot standby idle

hot standby becomes active

Then both TTF_1 and TTF_2 exponential with mean $1/\lambda = 6$ months (minimum property)



number of failures in time interval [0;t] has a Poisson distribution with mean λt

$$\Pr\{0 \text{ failures}\} = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t} \qquad \Pr\{1 \text{ failure}\} = e^{-\lambda t}\frac{(\lambda t)^1}{1!} = \lambda t e^{-\lambda t}$$

Suppose: MTTF = 12 months for each active DB

Reliability function:

$$R(t) = \text{Prob}\{TTF > t\} = \text{Prob}\{0 \text{ or } 1 \text{ failure before time t}\} = e^{-\lambda t}(1+\lambda t)$$

$$\text{Prob}\{TTF > 6 \text{ months}\} = e^{-(1/6)\times 6}(1+6/6) = 2/e \approx 74\%$$

$$\text{Prob}\{TTF > 12 \text{ months}\} = e^{-(1/6)\times 12}(1+12/6) = 3/e^2 \approx 41\%$$

$$\text{Prob}\{TTF > 60 \text{ months}\} = e^{-(1/6)\times 60}(1+60/6) = 11/e^{10} \approx 0.05\%$$

Wrap Up

Done today:

1. Revisit ISO reference model
2. Emergence of telephony
3. Exponential distribution
4. Poisson processes and properties
5. Markov chains, equilibrium distributions
6. Erlang-B model
7. Reliability design

Background reading material for this course:

O.J. Boxma, Stochastic Performance Modelling, chapters 1, 2 and 3

Appendix

A World of Uncertainties

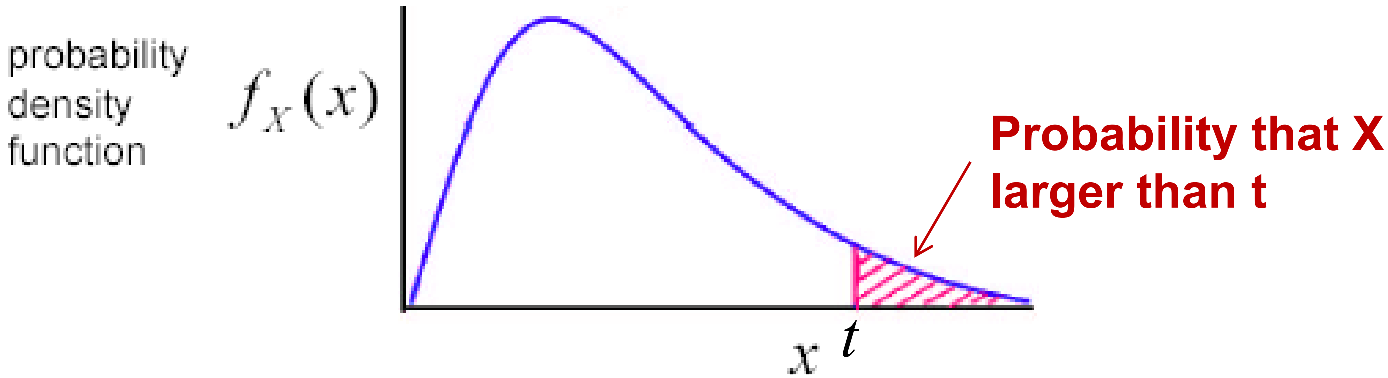
- **Random variables are anywhere…**
  - duration of an Internet session
  - time between successive financial transactions
  - CPU, memory and I/O utilization
  - time between system breakdowns
  - load of network connections
  - database access frequency distribution
  - file download times, job processing times
  - size of an E-mail
  - …

Probability Distributions

- <u>Continuous</u> random variables
  - can have any real value (not restricted to integer)
  - typical examples:
    - time between successive breakdown events
    - waiting time, job processing time, response time in RTS

- <u>Discrete</u> random variables
  - can have only integer values
  - typical examples:
    - number of job arrivals in given time interval
    - number of system failures in given time interval

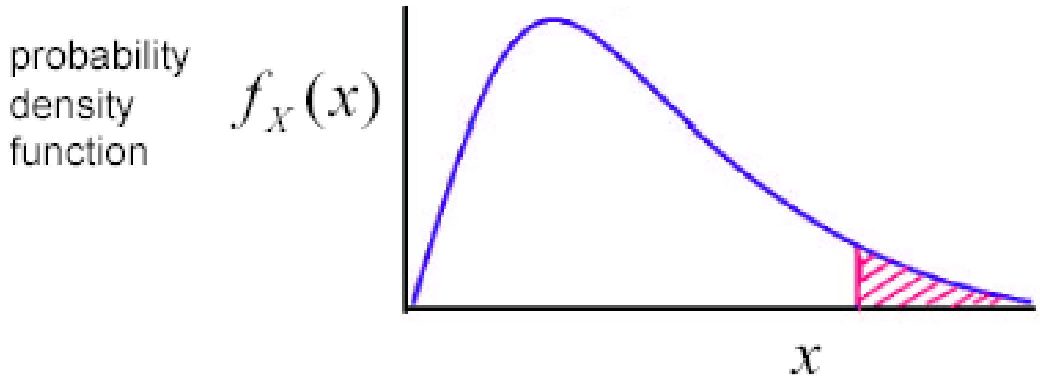
Continuous Distributions



- **Random variable *X* (continuous, non-negative)**
- **Probability density function (PDF):** $f_X(x), x \geq 0$
- **Cumulative distribution function (CDF):** $F_X(x) := \int_{t=0}^{x} f_X(t)dt$
- **Mean (1st moment):** $\mu := E[X] := \int_{t=0}^{\infty} t f_X(t)dt$
- **Interpretation:** $\frac{X_1 + \cdots + X_N}{N} \xrightarrow{N\to\infty} E[X]$

'Law of Large Numbers'

Measures of Variability



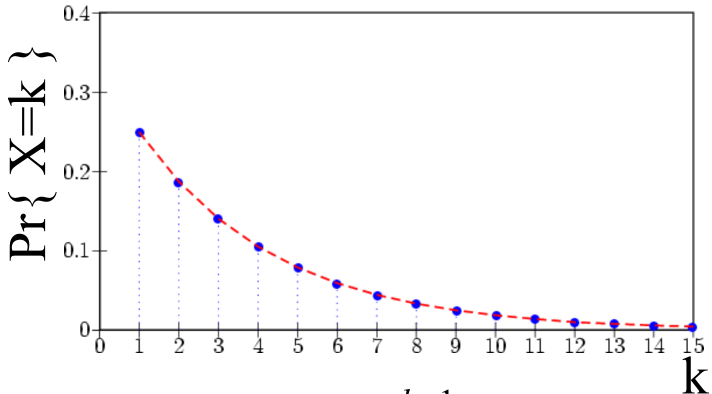
- 2nd moment, variance and standard deviation

$$E[X^2] := \int_{t=0}^{\infty} t^2 f_X(t)dt$$

$$\sigma^2 := Var[X] := E[X^2] - (E[X])^2 \quad SD[X] := \sqrt{Var[X]}$$

- Squared coefficient of variation (SCV) $c_X^2 := \dfrac{Var[X]}{(E[X])^2}$
- Variance, standard deviation and SCV are measures for the <u>variability</u> of X (but have different units!)
- k-th moment $E[X^k] := \int_{t=0}^{\infty} t^k f_X(t)dt, k = 1,2,...$

Geometric Distribution



First k-1 times failure, and then success

- Distribution: $\Pr\{X = k\} = p(1-p)^{k-1}, k = 1,2,\ldots$
- **Interpretation:**
  - flip a coin, success with probability p
  - X := number of attempts until first success (inclusive)

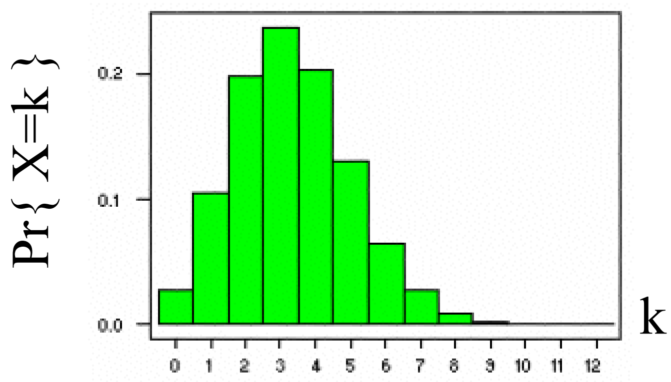
$$E[X] := \frac{1-p}{p} \qquad Var[X] := \frac{1-p}{p^2}$$

  - $\Pr\{X > t\} = \Pr\{\text{at least t times failure}\} = (1-p)^t$

- **Memoryless property:**

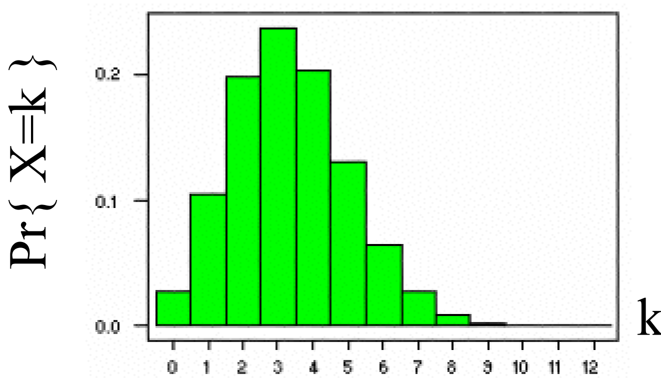
$$\Pr\{X > s+t \mid X > s\} = \frac{(1-p)^{s+t}}{(1-p)^s} = (1-p)^t = \Pr\{X > t\}, s,t = 1,2,\ldots$$

Binomial Distribution



- Let $X := X_1 + \cdots + X_n$ with $X_i = \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$ independent
- Distribution: $\Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, k = 0,1,\ldots,n$ → **see next slide**
- Notation: $X := X_1 + \cdots + X_n \sim bin(n,p)$
- **Interpretation:**
  - throw n times with coin, head with probability p
  - X := number of successes, then X ~ bin(n,p)
  - $E[X] := np \qquad Var[X] := np(1-p)$

Binomial Distribution



$$\Pr\{X=k\} = \binom{n}{k} p^k (1-p)^{n-k}, k = 0,1,\ldots,n$$

number of combinations with k success out of n

k times success

n-k times failure

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \text{ with } k! = k(k-1)(k-2)\cdots 2\cdot 1$$

Examples:

$1!=1,\ 2!=2,$

$3!=6,\ 4!=24$

Pascal's Triangle to calculate binomial coefficients

$$1$$

n=1 ⟶ $1 \quad 1$

n=2 ⟶ $1 \quad 2 \quad 1$

n=3 ⟶ $1 \quad 3 \quad 3 \quad 1$

n=4 ⟶ $1 \quad 4 \quad 6 \quad 4 \quad 1$

$$\binom{0}{0}$$

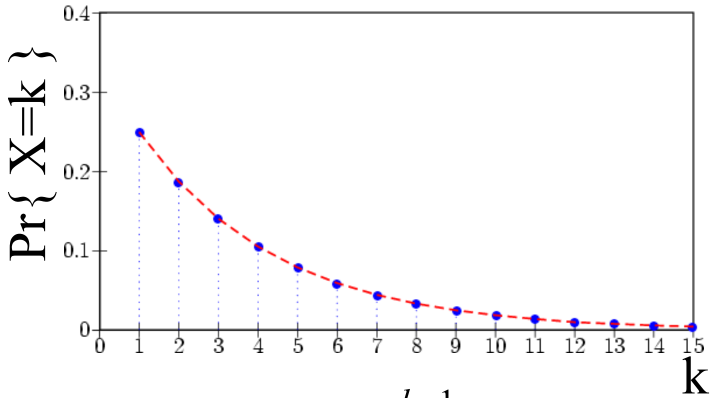
n=1 ⟶ $\binom{1}{0} \quad \binom{1}{1}$

n=2 ⟶ $\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$

n=3 ⟶ $\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$

n=4 ⟶ $\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$

Geometric Distribution



First k-1 times failure, and then success

- Distribution: $\Pr\{X = k\} = p(1-p)^{k-1}, k = 1,2,\ldots$
- **Interpretation:**
  - flip a coin, success with probability p
  - X := number of attempts until first success (inclusive)

$$E[X] := \frac{1-p}{p} \qquad Var[X] := \frac{1-p}{p^2}$$

  - $\Pr\{X > t\} = \Pr\{\text{at least t times failure}\} = (1-p)^t$

- **Memoryless property:**

$$\Pr\{X > s+t \mid X > s\} = \frac{(1-p)^{s+t}}{(1-p)^s} = (1-p)^t = \Pr\{X > t\}, s,t = 1,2,\ldots$$