IK1611, Spring term 2016
Dimensioning of Communication Systems

W E L C O M E !

Lena Wosinska, Matteo Fiorani
and Jiajia Chen
KTH/ICT/COS/ONLab

IK1611
Dimensioning of Communication Systems

LECTURE 1

COURSE INTRODUCTION

Lena Wosinska
KTH/ICT/COS/ONLab

Course Introduction

- General information
  - Prerequisites and aim of the course
  - The course at a glance
  - Course material
- Introduction to queuing systems
  - Examples
  - A general queuing system
  - System parameters
  - Performance measures
- Some basic terminology
- Summary

Prerequisites and aim of the course

Prerequisites:
Basics in communication systems and networks

Aim: after the course students shall be able to:

- define the basic queuing models for different communication systems
- dimension the systems in terms of capacity, delay and throughput.

The course at a glance

- Lectures, exercises, seminars, self study, etc. (TBD)
- Project
  - Work in the project group to solve the problem defined in the project
- Examination
  - Exam (written or oral, TBD) and the project work
  - Exam on Saturday 18/03, 08:00 – 13:00
  - The final grade based on results of both the written exam and the project

Lectures/seminars

	Topic	Reading
L1	Introduction to queuing systems, basic terms, course overview, introduction to the project	Chapters 1 and 2,
L2	Probability theory, basic principles and random variables, Z- and L-transforms	Chapter 7
L3	Poisson process and Markov chains in continuous time	Chapter 8
L4	Queuing systems. Basic formulas	Chapter 3
L5	M/M/1 system, unlimited queue	Chapter 4.1
L6	M/M/1 system, limited queue and finite client population	Chapters 4.2 and 4.3
L7	Project	Project instruction
L8	M/M/m	Chapter 4.4
L9	M/M/m system, limited queue	
L10	M/M/m/m loss systems	Chapters 4.5,
L11	M/G/1 system	Chapter 5
L12	Queuing networks	Chapters 8
L13	Course summary	

Course material

- Text:
  - Maria Kihl, ”Queuing systems”
- If you like a book:
  - Ng CheeHock, “Queuing Modeling Fundamentals”, Wiley, 1998.
  - L. Kleinrock, “Queuing Systems, Volume 1, Theory”, Wiley, 1975
  - D. Gross, C. M. Harris, “Fundamentals of Queuing Theory”, Wiley, 1998

  Be aware, the notations might differ !
- Supplementary material:
  - Erlang tables
  - Formula sheet
  - Instructions for the project

Contacts

- Course responsible: Jiajia Chen (*jiajiac@kth.se*)
- Examiner: Lena Wosinska (*wosinska@kth.se*)
- Teaching assistant: Matteo Fiorani (*matteof@kth.se*)
- Address: Electrum, level 4, stairs B

Introduction to Queuing Systems

- Examples
- A general queuing system
- System parameters
- Performance measures

Ode to a Queue

"If you want to model networks
Or a complex data flow
A queue's the key to help you see
All the things you need to know."

Leonard Kleinrock,
Ode to a Queue

Queuing problems

- Resource sharing systems
  - Customers with sporadic needs share a common resource
  - Communication networks and services
  - Computer systems
- The momentary need exceeds the available resource
  - Some users get served, others wait or are turned away
  - Waiting customers form a *queue* (hence the name)
- Examples
  - **Execution of time-sharing processes in a computer**
  - **IP packets in a highly-loaded Internet router**
  - **A web server receiving many http requests**
  - Morning rush hour in the subways and highways
  - The lunch service at a local restaurant

Queuing theory

- Mathematical modeling of resource sharing systems
- A tool for dimensioning of
  - Networks for fixed and mobile telephony
  - Data communication, Voice over IP
  - Servers for network-based services

Description of a queuing system

- Stochastic processes
  - Arrivals
  - Service
- System parameters
  - Order of service
  - Buffer size
  - Number of servers
  - Number of service stages

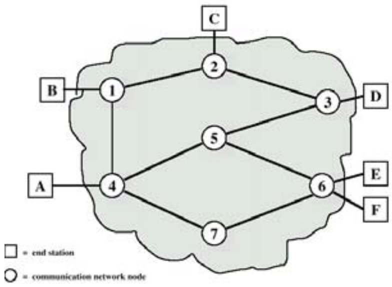
System dimensioning problems

- Given arrival intensity, traffic characteristic and requirements
  - Design a system that meets requirements on
    - Delay (waiting time and service time)
    - Loss probability
    - Number of customers in the system
    - Blocking probability

  etc
- Given system and requirements
  - Define the arrival process that fits the system and requirements
    - Arrival rate can't be too high
    - Arrival pattern should be appropriate

  etc

Switched Networks



- Switching nodes
  - not concerned with contents of data
  - purpose: provide switching facility
  - in general not fully connected
- End nodes
  - provides data to transfer
  - connected via switching nodes
- Links
  - physical connections between nodes

Switching

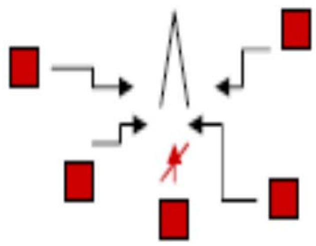
- Circuit switching
  - Exempel: traditional telephone network
  - Synchronous TDM (or WDM in optical networks)
  - Suitable for interactive services due to low and constant end-to-end delay
- Packet switching
  - Asynchronous (deterministic and statistic) multiplexing
  - Connection oriented (virtual circuits)
    - Exempel: ATM
  - Connection less (datagram)
    - Exempel: Internet

Circuit switching

- Traffic is concentrated to obtain efficiency of resource utilization
- Network resources are lower than a sum of all capacity offered to the customers, i.e. less time slots or wavelengths than a sum of all possible connections in the network would require
- Network/system designed to not exceed a certain level of blocking probability
- Dimensioning problem:
  - The momentary need exceeds the available resource, i.e. no time slots or wavelength channels are available at the moment
  - Calls are blocked
  - Dimension the network/system to not exceed the certain level of blocking probability

Example 1

- Voice calls in a GSM cell
  - call blocked if all channels busy



- Performance
  - Utilization of the channels
  - Probability of blocking a call
- Depends on:
  - How many calls arrive
  - Length of a conversation

  } → **Service demand**

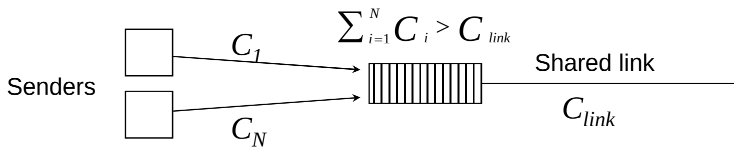
  - Cell capacity (number of channels) → **Server capacity**

Packet switching

- Network resources are dynamically shared
  - More efficient resource utilization due to statistical multiplexing
  - No idle time
- Services with variable bit rate (VBR) can be supported
- Better suited for data traffic
  - File transfer
- Buffers are needed at the nodes due to statistical multiplexing
- Less suitable for interactive services due to variable delay at the nodes

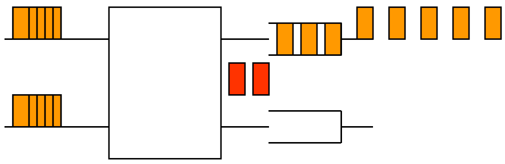
Statistical multiplexing

- Random access to the network resources
- Statistic multiplexing
  - Without Quality of Service QoS support (best effort)
  - With QoS support
  - Example: communication link



Congestion in PS Networks

- End-nodes transmit data independently from each other
- What happens if more packets arrive than it is possible to forward?

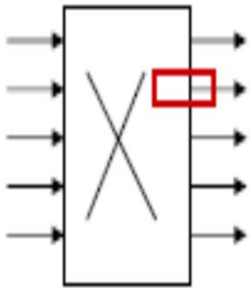


Congestion Scenario

- Output buffers become full
  - discard packets
  - sources retransmit packets
    - more messages in the network
    - more buffers saturated
  - delay increases
    - source times out
    - more retransmissions
  - capacity drops towards zero

Example 2

- Packet transmission at the output link of an IP router
  - packets wait for free output link



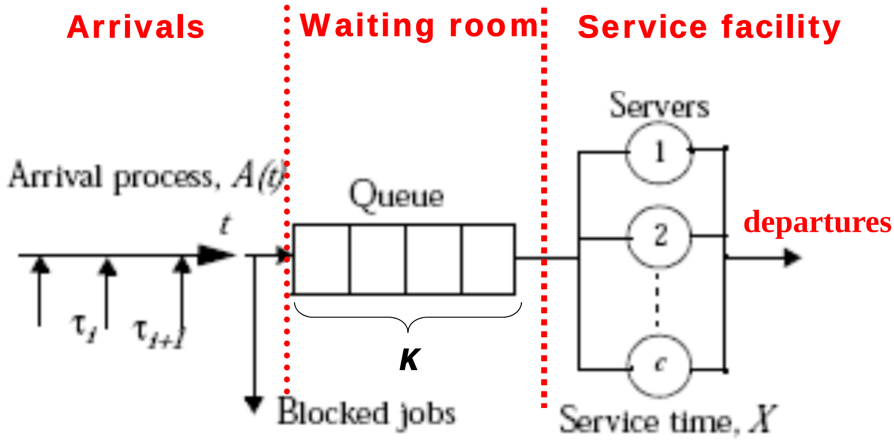
- Performance
  - Utilization of the output link
  - Waiting time in the buffer
- Depends on:
  - How many packets arrive
  - Packet size

  → Service demand
  - Link capacity →Server capacity

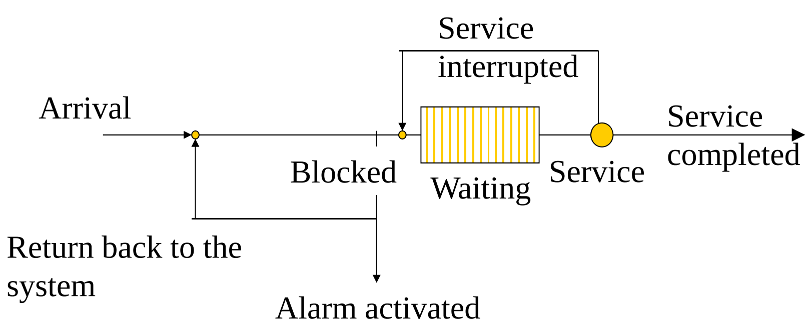
Content of the course

- Terminology, definitions and basic formulas
- Basics of probability theory and Markov chains.
- Modeling of communication systems in terms of delay, packet loss probability, system utilization etc.
- Open and closed queuing networks.
- Solving practical dimensioning problems for communication systems and networks.

Queuing system



Queuing system, cont.



Arrival process

- Interarrival time distribution
  - Time between arrivals is assumed to be independent from customer to customer
  - Stationary, the distribution does not change in time
- Do customers arrive in batches or one at a time ?
  - In any case arrivals are independent
- Is the customer impatient?
  - Might choose not to enter the queue if it is too long
  - Could leave the queue if waiting is exceedingly long
  - Might choose to switch to another queue if available

Wrights' Axioms of Queuing Theory

- It's always better to wait at the front of the line.
- There is no point in waiting at the end of the line.

But (note Wrights' paradox):

- If you don't wait at the end of the line, you'll never get to the front.
- Whichever line you are in, the others always move faster.

Service process

- Service time distribution
  - Assumed to be independent from customer to customer
  - Assumed to be independent of the arrival process
  - Stationary, the distribution does not change in time
- Are the customers served one at a time or in batches?
- Are the service times dependent of the number of customers in the queue?
  - State-dependent service times
  - Do not mix it with non-stationary (time-dependent) service distribution

Queuing discipline (order of service)

- Next in line
  - First come, first served, FCFS (a k a first in, first out, FIFO)
- Last in line (a stack)
  - Last come, first served, LCFS (a k a last in, first out, LIFO)
- Random order, independent of arrival time
  - Random service selection (RSS)
- Priority order
  - Preemptive
    - Service of customer might be interrupted by an arriving customer
      - Its service is resumed from the point of interruption
      - Its service restarts from the beginning (prior service time wasted)
  - Non-preemptive
    - Customer in service always completes before new arrival is served
    - New customer delayed by the other customer's residual (remaining) service time

Queuing system parameters

- Buffer capacity
  - Infinite so that any number of customers can wait
  - Finite
  - Equal to zero, i.e., no buffer (loss system)
- Number of servers
  - Work in parallel
  - Independent of one another
  - The same service rate for all of them
- Number of service stages
  - The service could consist of several subtasks (no pipelining)
- Queuing discipline (order of service)

Service demand

Stochastic processes

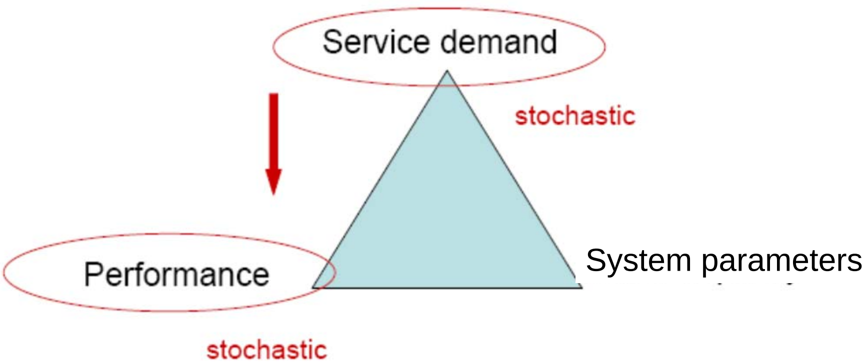
- Arrival process: How do the customers arrive to the system
- Service process: How much service does a customer demand
  - Customer:
    - IP packet
    - Phone call

System performance measures

- Stationary measures
  - How does the system behave on the long run?
  - Average measures (often considered in this course)
- Average number of customers in the system
  - Average number of customers waiting in the queue
  - Average number of customers in the server
- Average system time (response time)
  - Average waiting time
  - Average service time
- Probability of blocking (blocked customers / all arrivals)
- Utilization of the server (time the server is occupied / entire considered time)

Performance of queuing systems

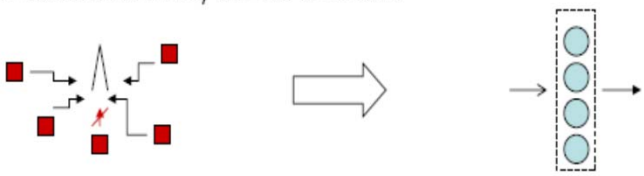
- The triangular relationship in queuing



- Tradeoffs in queuing system performance
  - Efficient use of the common resource (or resources)
  - Risk of turning customers away or letting them wait

Example 1

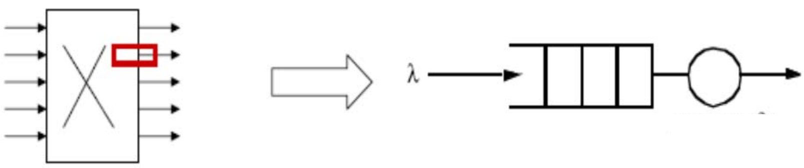
- Voice calls in a GSM cell
  - channels for parallel calls, each call occupies a channel
  - if all channels are busy the call is blocked



- Arrival process: calls attempts in the GSM cell
- Service process: the phone call (service time = length of the phone call)
- Number of servers: number of parallel chanells
- Number of service stages: 1
- Buffer capacity: no buffer

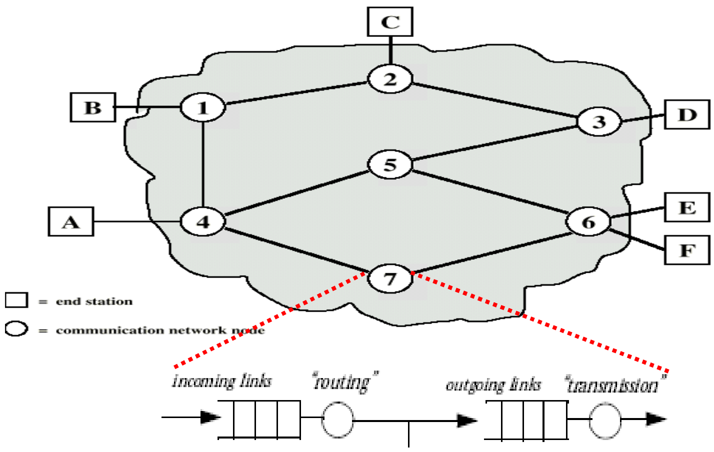
Example 2

- Packet transmission at the output link of a large IP router



- Arrival process: IP packet multiplexed at the output buffer
- Service process: transmission of one IP packet (service time = packet length / link transmission rate)
- Number of servers: 1
- Number of service stages: 1
- Buffer capacity: max. number of packets IP

A Network of Queues



How to solve dimensioning problems?

- Analytical solution (queuing theory) for tractable systems
  - Develop a mathematical model of the system
  - The model should describe the system as accurate as possible
  - Based on your model you can be able to obtain the performance measures or dimension the system according to the requirements.
- Computer simulations for more complex systems

Project work

- To be defined

Summary

- **Today:**
  - General information about the couse
  - Introduction to queuing systems
- **Next lecture:**
  - Probability theory and transforms