Distributed Computer Systems Lab
http://disco.informatik.uni-kl.de

Technische Universität Kaiserslautern

Performance Modelling of Distributed Systems

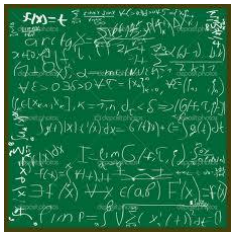
1. Introduction & Motivation

Prof. Dr.-Ing. Jens B. Schmitt
(jschmitt@cs.uni-kl.de)

Performance Makes the Difference

- Performance you feel with every interaction
  - Who likes a slow network, database, server?
  - For example, (in)security you only feel if it actually happens
- Performance needs to be
  - measured
  - controlled

- Oftentimes, it needs to be predicted
  - Performance simulation
  - **Performance analysis**



- Performance analysis is often the simplest and provides deep insights in what is going wrong and why (→ control)
- Performance models resulting from analysis can or even should drive system development/evolution
- Performance analysis of *distributed* systems is challenging

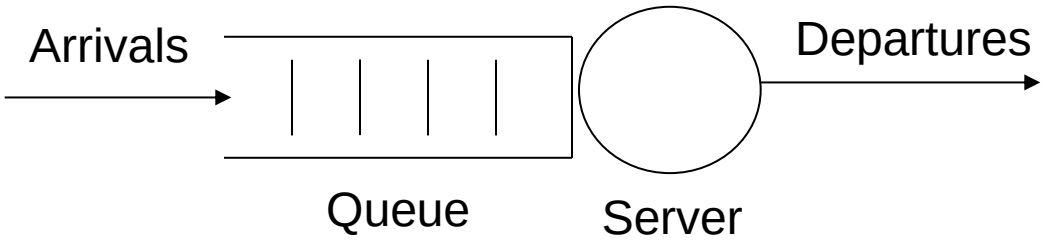
Performance Analysis of Distributed Systems

- Classical method: Queueing Theory (QT)
  - Huge success: telephone network
  - Poisson arrivals: M/M/1, etc.
  - Product-form networks rely on Poisson assumption
  - Mainly for **average-case** analysis

- Lately: Deterministic Network Calculus (DNC)
  - Get rid of stochastic assumptions
  - Work out **worst-case** behaviour
  - Elegant network analysis without many assumptions
  - Yet, does not capture statistical multiplexing
- Most recently: Stochastic Network Calculus (SNC)
  - Falls in the middle between QT and DNC: **probabilistic worst-case**
  - Captures statistical multiplexing
  - Promises elegant network analysis

A Look Ahead: Motivating Example

Goal: Compute delay for a single flow at a single server



- QT approach: knowledge about arrival and service distributions
- DNC approach: deterministic bounds on arrivals and service
- SNC approach: probabilistic bounds on arrivals and service

DNC at Work

- Assumption: discrete time (slots) $t = 1, 2, \ldots$
- Cumulative arrivals up to time t: $A(t)$
- Server provides constant rate C
- Arrivals are constrained by so-called (σ, ρ) envelope

$$\forall s \leq t : A(t) - A(s) \leq \sigma + \rho(t - s)$$

- What is a (deterministic) bound on the delay?

SNC at Work

- Now, we further assume that the arrivals in a time slot are independently uniformly distributed over $[0, 1]$, i.e. $U(0, 1)$
- This means, arrivals are still (σ, ρ)-constrained with $\sigma = 0, \rho = 1$
- This, in turn means we need a server with rate $C = 1$ at least
- Hmm, couldn‘t we do better if we were less demanding?
- Could we benefit from several flows being multiplexed?

Conclusion and Outlook

- Performance is important and thus needs to be controlled
- Throughout the course, we will see

  SNC provides a general and uniform framework to derive performance bounds in distributed systems
- SNC provides systematic ways of modelling arrival and service processes
  - Several flavours exist (MGF vs. Tail Bounds)
- SNC enables an elegant end-to-end performance analysis
- The strength of SNC is its practical relevance as well as its volatility in dealing with fundamentally hard queueing problems
- Let the games begin!