LCCC activities in Cloud Control

MARIA KIHL

Maria Kihl

- PhD in Teletraffic Systems, 1999
- Associate Professor at Dept. of Electrical and Information Technology since 2004.
- Internet related research projects, often with industry collaboration.
- Joint work with Dept. of Automatic Control since 2002.

Main research interests: performance modelling, analysis and control of internetworked systems.

LUND UNIVERSITY

Implementation in Networked and Embedded Systems

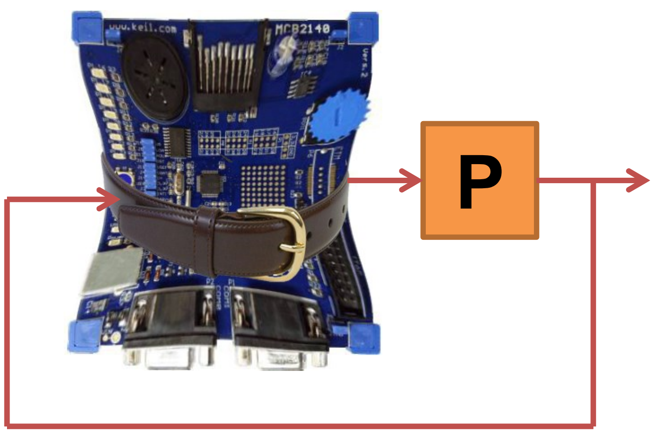
Maria Kihl
Karl Erik Årzén
Anders Robertsson
Bo Bernhardsson
Johan Eker (LU, Ericsson AB)
Martina Maggio
Anton Cervin
Alessandro Papadopoulos
Manfred Dellkrantz (PhD student)
Jonas Dürango (PhD student)
William Tärneberg (PhD student)
Yang Xu (PhD student)
Mikael Lindberg (PhD stud)
Payam Amani (PhD stud)

Background

- Application complexity and uncertainty increases for both embedded system and server infrastructures.
- Strict demands on both resource efficiency (e.g. power) and service performance.

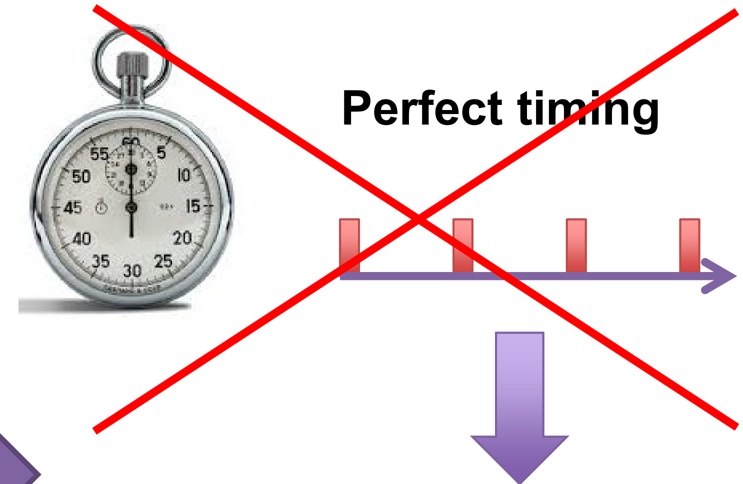
Dynamic resource management and control

Networked Embedded Control

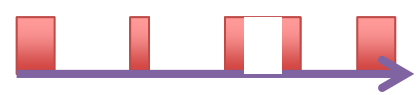


Embedded Control

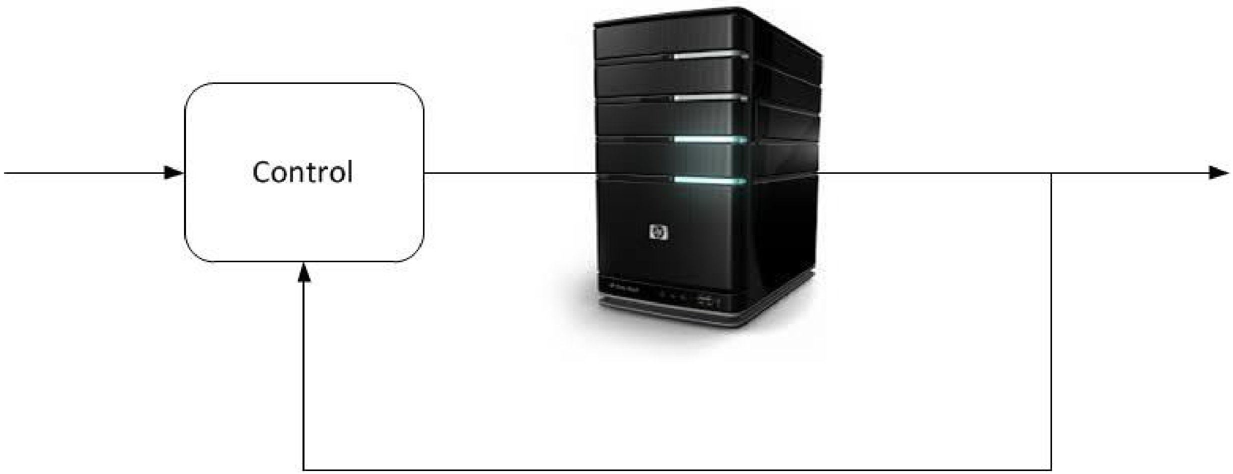
Implications for Control



Imperfect timing
- delays with jitter
- event-based control

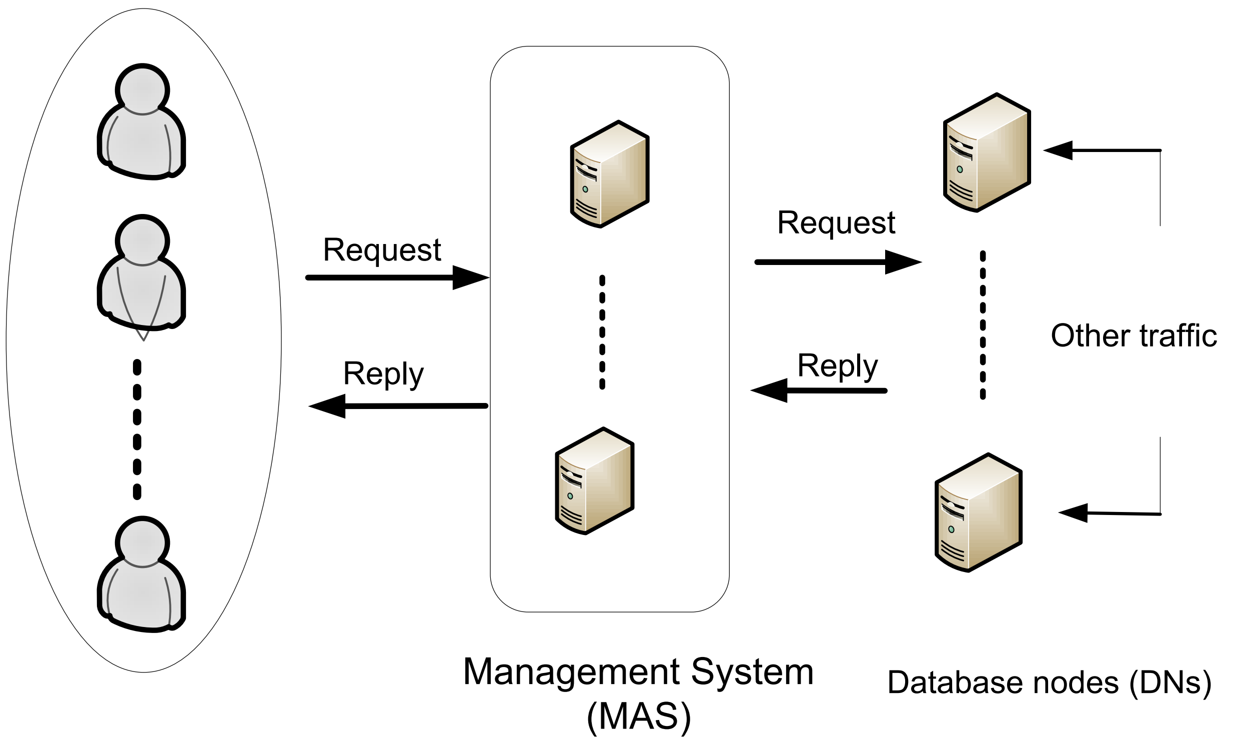


Resource optimization and control of server systems with latency constraints



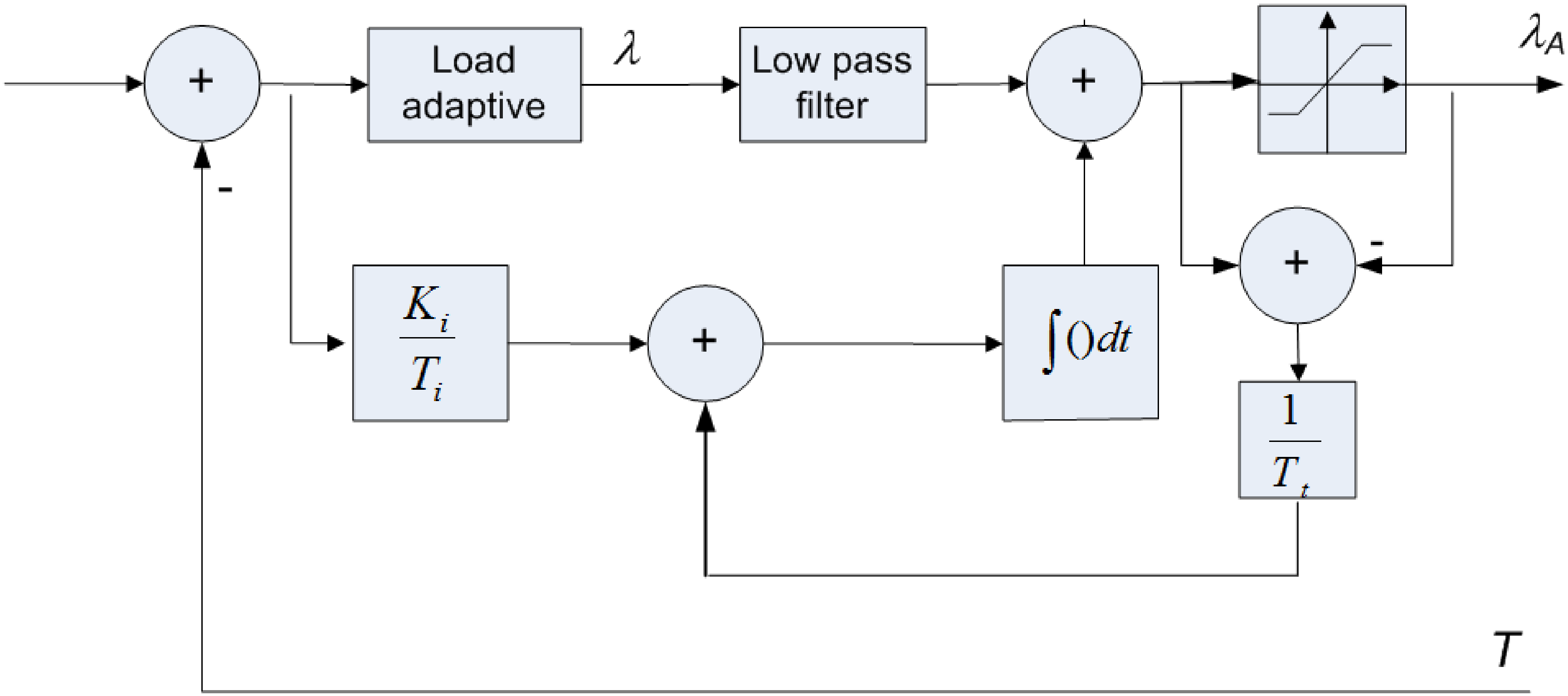
- Performance models
- Admission control
- Prediction based capacity optimization

Example: Mobile Service Support Systems developed by Ericsson AB

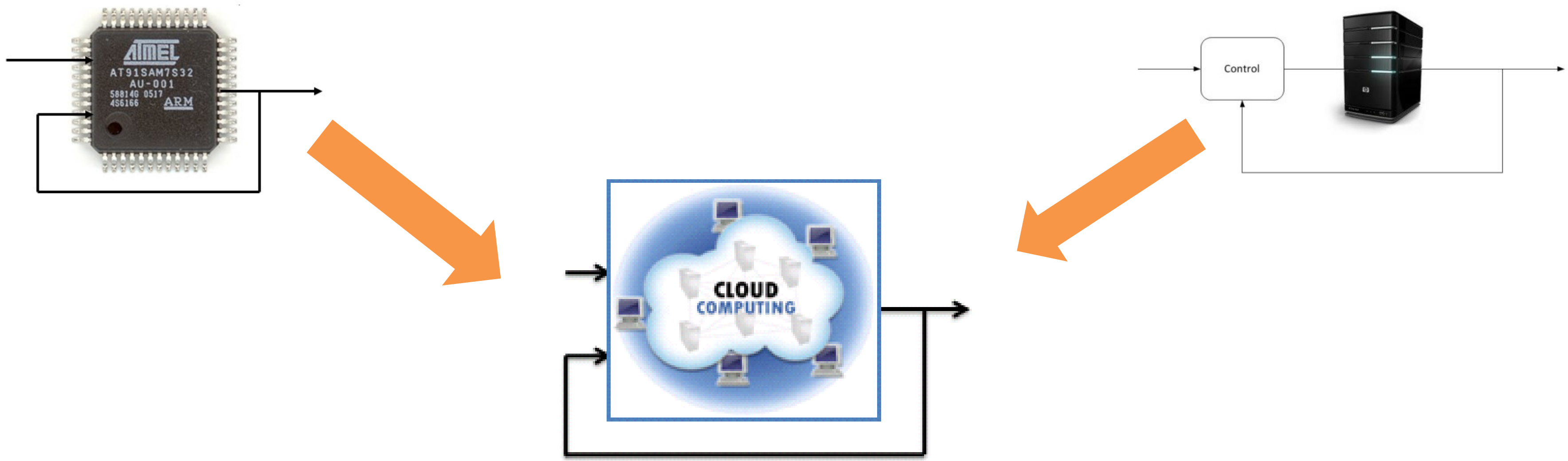


Load control of database nodes with unknown high priority background traffic.

Example: Load adaptive controller for mobile service support systems

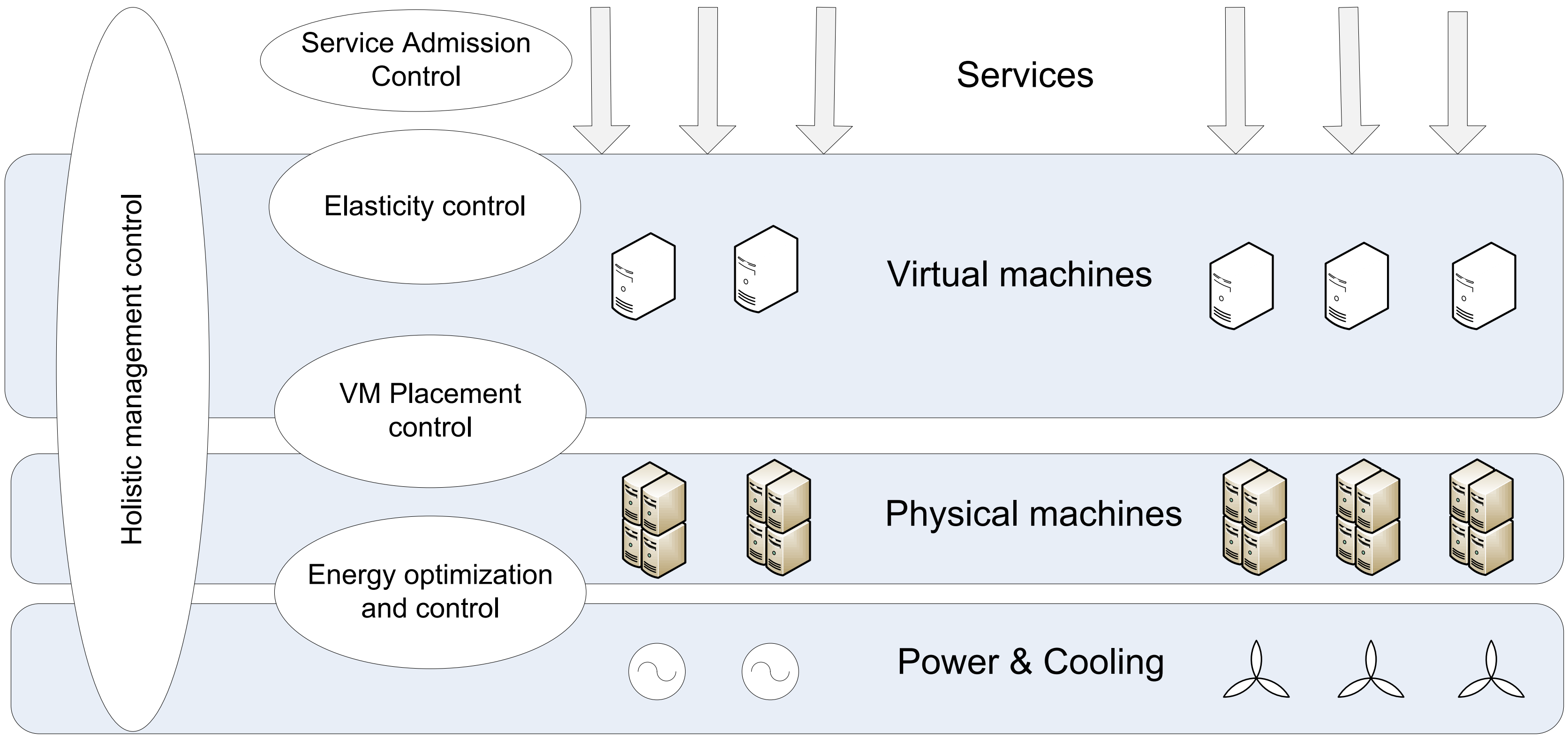


Cloud Control



- Joint work using our competences in both embedded control and server systems
- A control theoretic approach to a range of cloud management problems.

Challenges for Cloud Control



Some ongoing work

- Brownout
  - Martina Maggio, Alessandro Papadopoulos, Jonas Dürango, Manfred Dellkrantz
- VM startup time compensation
  - Manfred Dellkrantz
- Omnipresent clouds
  - William Tärneberg
- Modeling and autoscaling
  - Jonas Dürango

Brownout
(Martina Maggio, Alessandro Papadopoulos,
Jonas Dürango, Manfred Dellkrantz)

- Cloud applications need to cope with many unexpected events:
  - Flash crowds
  - Hardware failures
  - Unexpected performance degradations

Objective: Applications that **withstand variations** and have **similar behaviors** in **similar conditions**

Motivation: The need for predictability

- Tools like **cloud operating systems** and **virtualization** greatly simplify the development, management and deployment of software applications

A service provider should just use **the cloud**

A service provider should not need to think about **where** the application **will be executed**

how many resources it will consume

how to distribute it

Application performance challenge

However, in reality, another type of problems arise when an application is deployed in the cloud...

Applications do not behave always in the same way

Due to a varying amount of physical resources assigned to an application

Use resource allocation mechanisms inspired from embedded systems .

VM startup time compensation
(Manfred Dellkrantz)

Virtual machines take time to start up.

Controller saying, “Give me *m* VMs!” will have to wait for control signal to have effect.

↓

Dead time

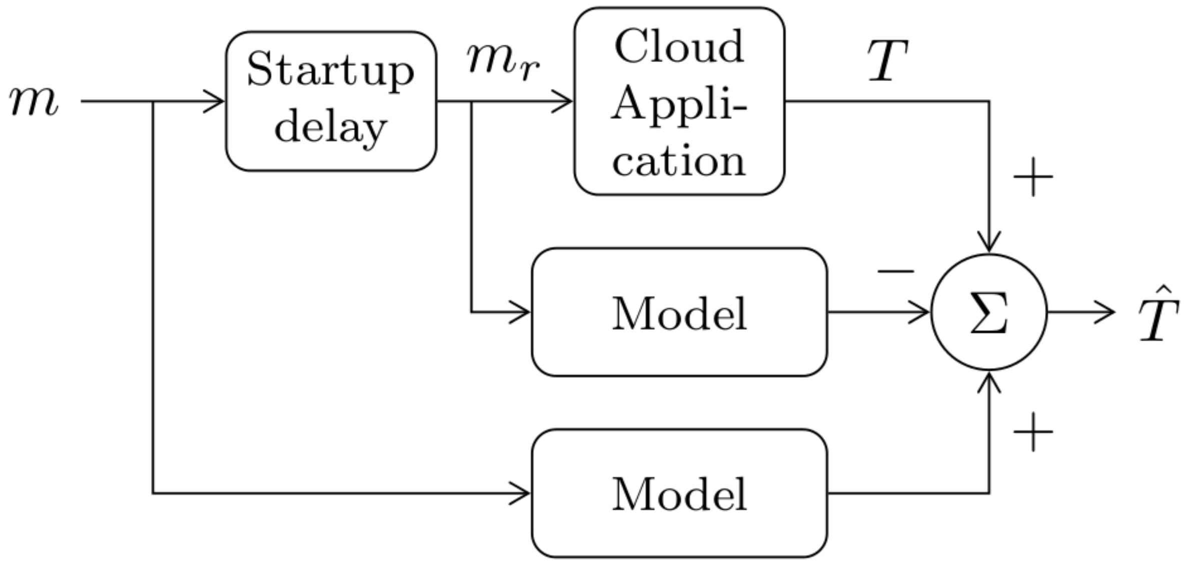
VM Startup Time Compensation

- VM startup time (“dead time” in control lingo) is a major challenge.
  - Controller reacts several times before its first action has an effect
  - Dead time-unaware controller gives overshoot (unnecessary costs)

Dead time compensation

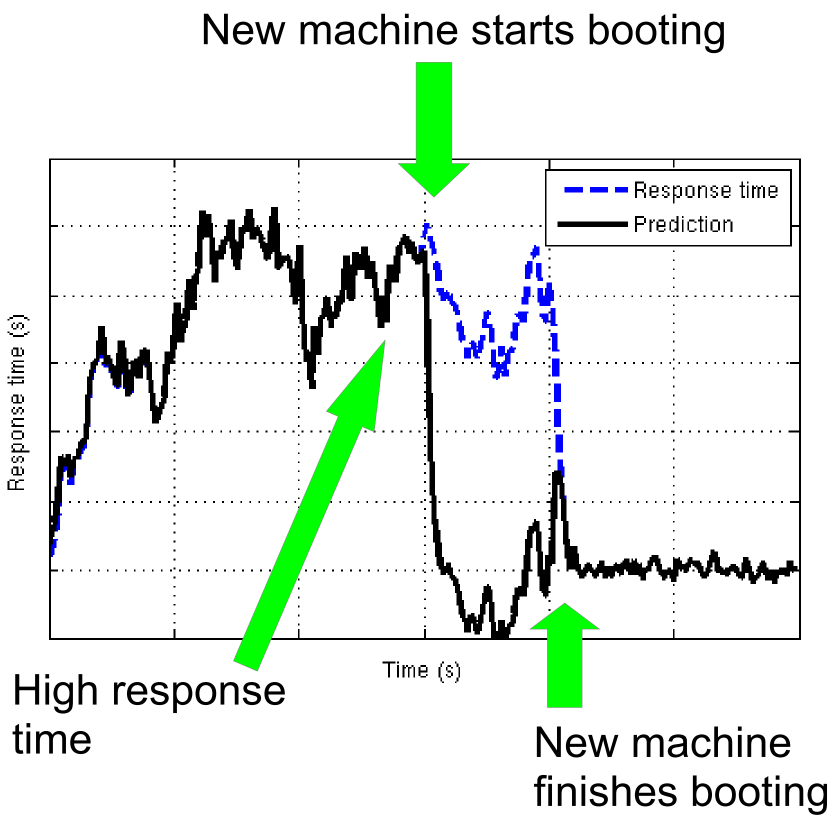
- Dead time compensation [Smith, 1957] calculates what the output would have been without dead time.
- Allows you to do faster control, still maintaining stability and avoiding oscillations.
- Requires a model!

Dead time compensator



- m – wanted # machines,
  m_r – actual # machines
- T – response time
- $\hat{T}$ – compensated response time

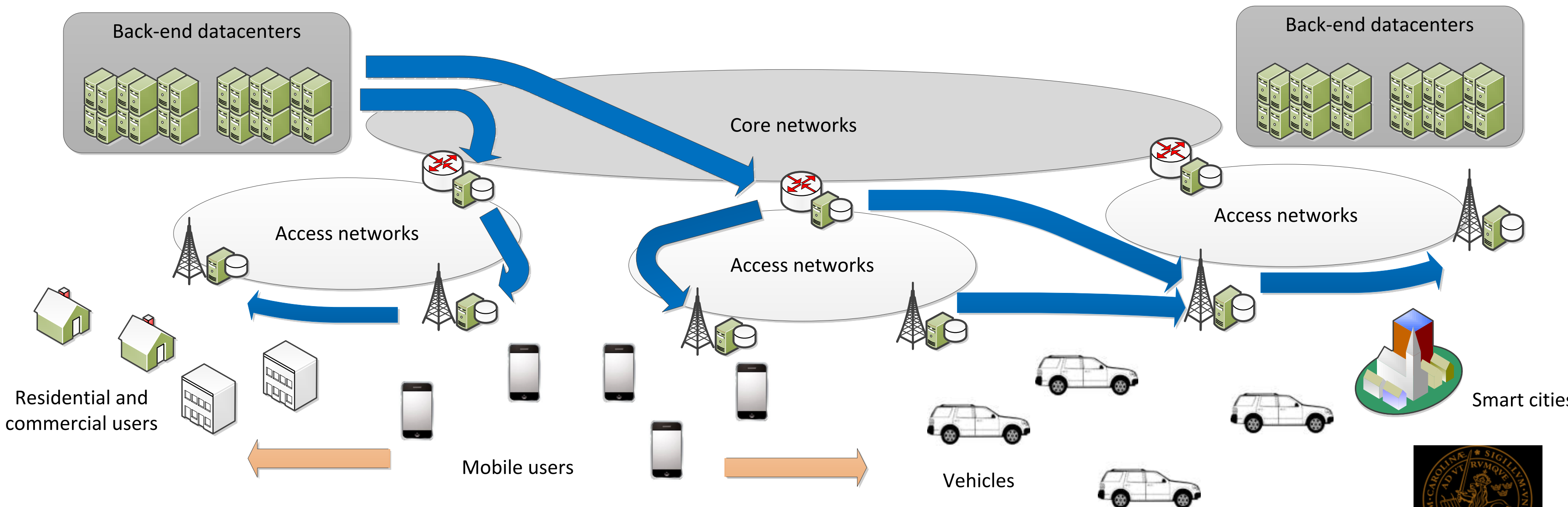
Proof-of-concept



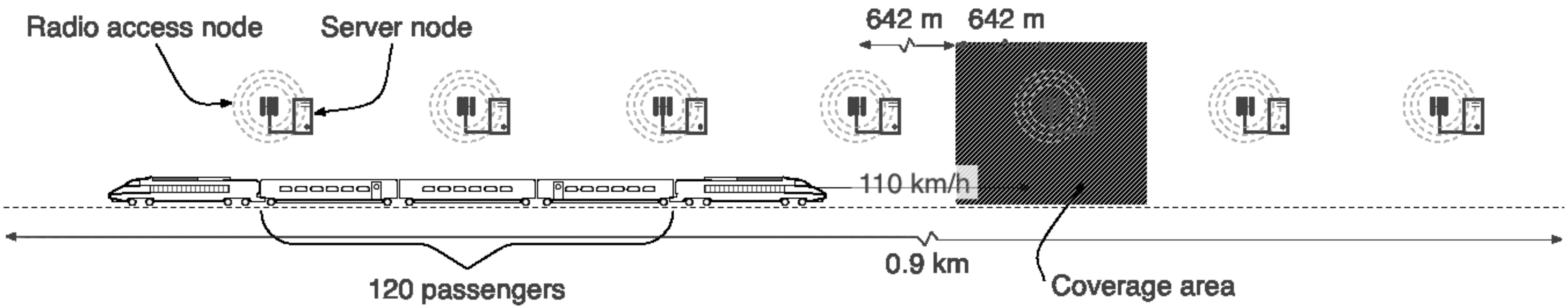
- Open-loop proof-of-concept
- Cloud infrastructure modelled as queuing system.
- Nonlinear continuous flow model [Tipper, 1990]
- Step up in number of VMs (*m*)

Omnipresent clouds
(William Tärneberg)

In the omnipresent cloud paradigm, cloud compute resources are migrated or located to the capillaries of the mobile networks.

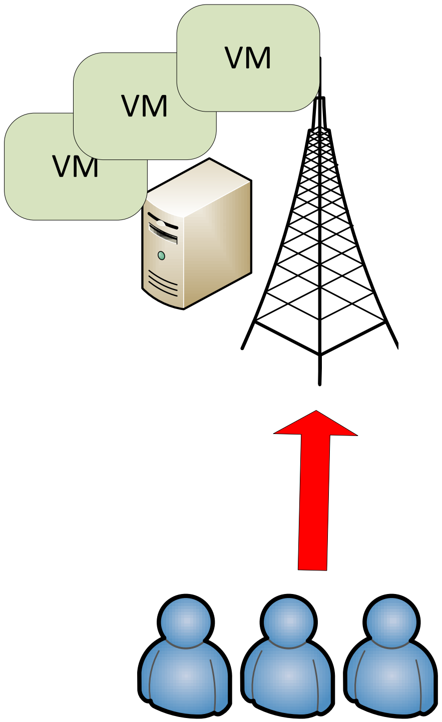


Scenario



- Cellular network corresponding to 4G/LTE.
- First mobility model corresponding to a train.
  - Easily extended to a highway with cars.
- Users request a web-like cloud service according to a stochastic process.
- VMs can be migrated between cells with a certain delay.

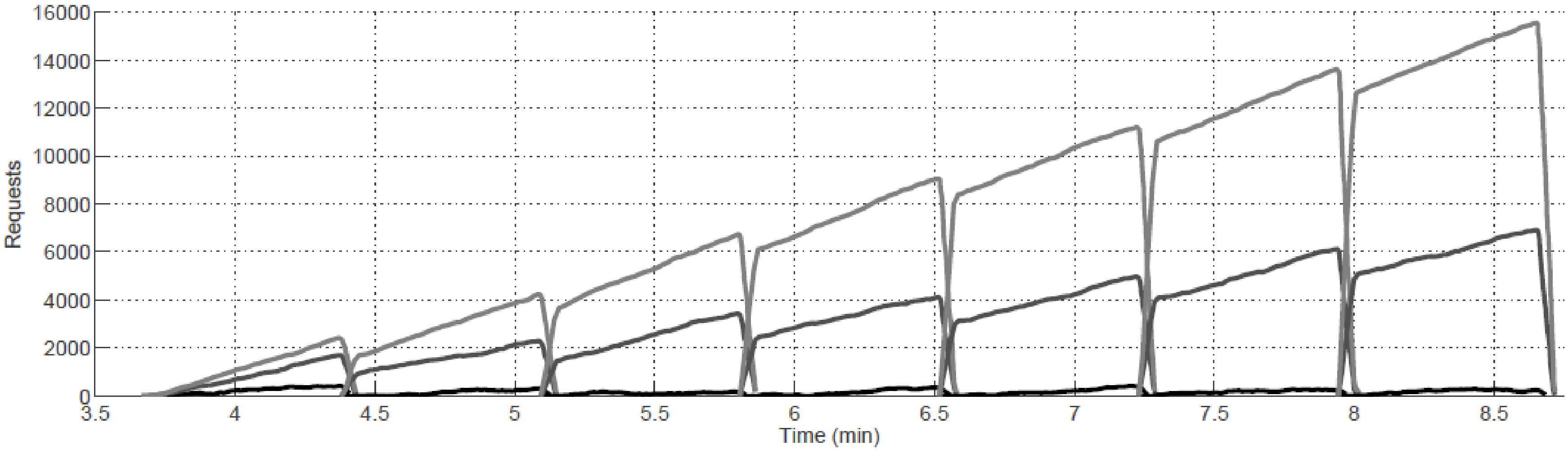
Cloud capacity moves with users



Cloud control challenges

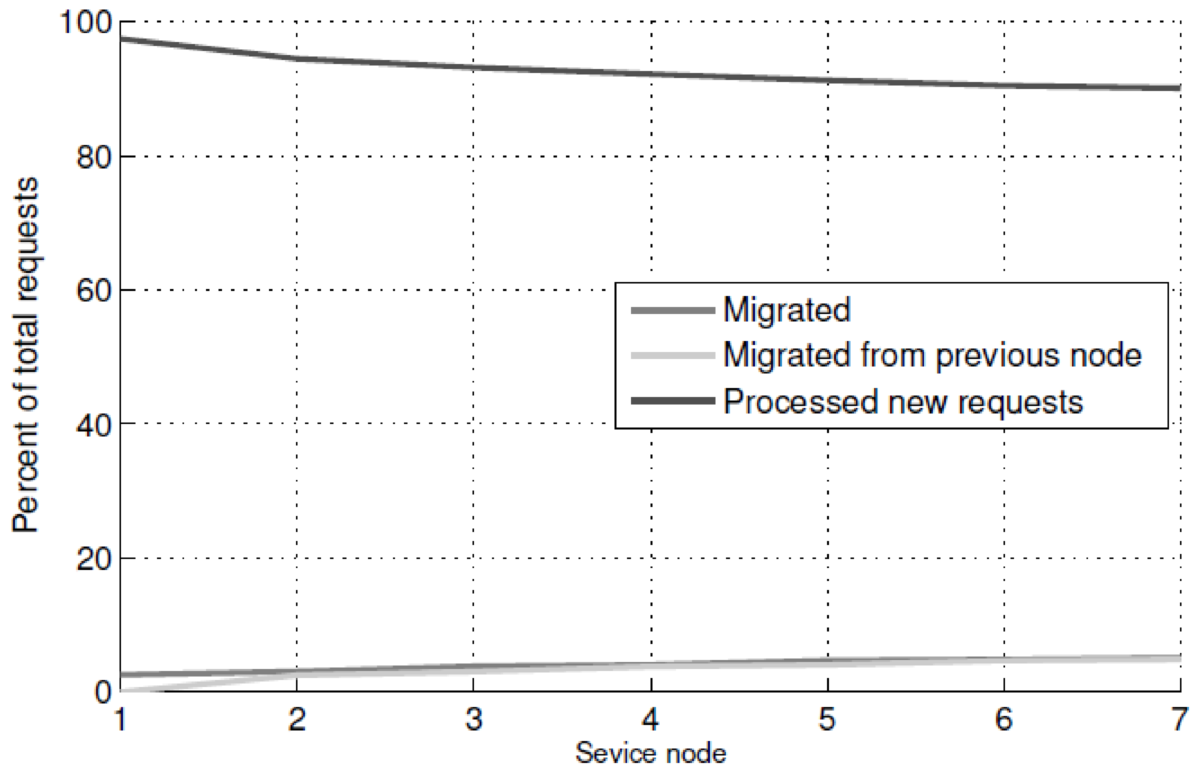
- Lack of good system models
  - Architecture models
  - Service models
  - Mobility models
  - Workload models
- Workload and Capacity demands in both time and space.
- Auto-scaling: When and Where?
- VM migration: Needs to be performed fast enough dependent on the mobility of the users.
- Heterogeneous nodes must be taken into account.

Load displacement example



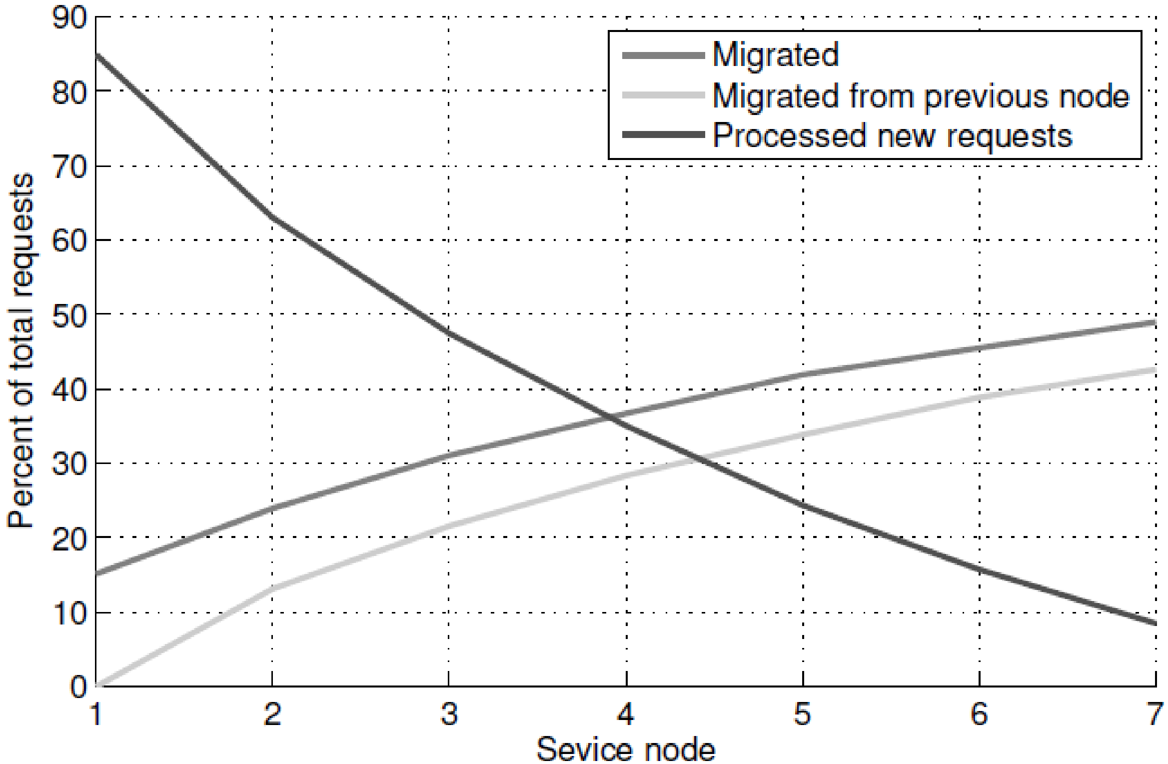
- The workload willl move between cloud servers, dependent on the mobility of the users.
- An underprovisioned system will spread in both time and space.

Performance during normal load



During normal load, the system migrates VMs according to the mobility and handles requests with low response times.

Performance during overload

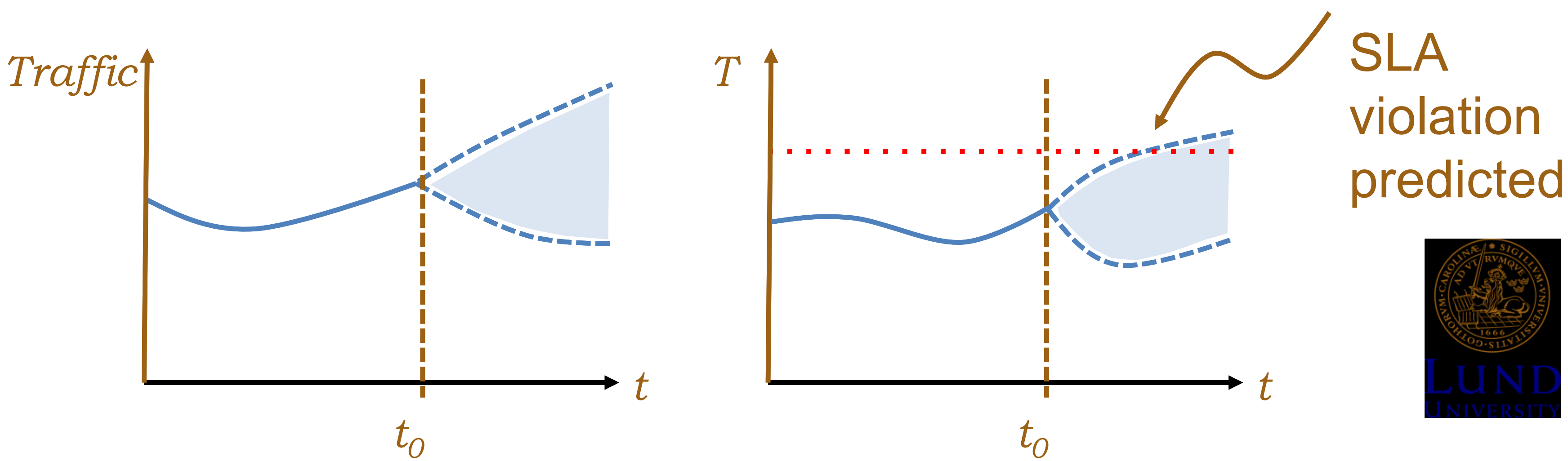


The system needs to cope with the mobility of users, or it will become overloaded. One major challenge is the delay for VM migration.

Modeling and autoscaling
(Jonas Dürango)

Computer system models suitable for control are usually tricky to find.

Understanding of dominant system dynamics is vital for decision making (Scale up? Down?)



Modeling and autoscaling

Use system identification to find models for different purposes:

- Simulation (model-based feedback control)
  - Better understanding of system dynamics -> allows better tuning of controllers.
- Prediction (proactive control)
  - Possibility to anticipate potential SLA violations

Summary

- Within LCCC, we have established a cross-disciplinary research environment with competence in control theory, real-time systems, and internetworked systems.
- Ongoing work on different aspects of Cloud Control.

Bring theory to practice