Autonomous Resource Management for Robust, Efficient, and High-Performance & Cloud Computing

Erik Elmroth
Umeå University
elmroth@cs.umu.se

www.cloudresearch.org

Senior researchers

Erik Elmroth, Professor
Johan Tordsson, Docent
P-O Odelberg, Researcher

Post docs, Researchers

Ahmed Ab-Elin, PhD
Christian Klein, PhD
Danuel Bayuh Lakelo, PhD
(Vacant)

PhD students

Amardeep Mehta
Jakub Krzywda
Olumuyiwa Okunmoye
(Lars Larsen, PhD)
Chinh Nguyen
Abel Souza
Koiten Selome Tesfahun
Gonzalo Rodrigo

Others

Lennart Edblom, Senior lecturer
Peter Garthäll, Sys. Dev, PhLic
Tomas Forsman, Systems expert

Research assistants

Simon Kullberg
Emil Henrikund

www.cloudresearch.org
High-level objectives

- Autonomous resource management algorithms and systems for optimizing
  - Performance (throughput, responsiveness, availability, reliability, etc)
  - Utilization
  - Usability
  - Energy efficiency
- Infrastructure scope from individual servers to wide-spread distributed systems including datacenters and telecom networks

Scheduling for exascale systems

- Learning/predicting future exascale workloads
  - Incl. the convergence of cloud and HPC
- Scheduling algorithms and systems for exascale workloads and systems
- Workflow scheduling for performance and resource utilization
- Tight collaboration with Lawrence Berkeley Lab
- To be presented at the eSSENCE Academy next week
Resource management challenge

- Robustness & performance
- Cost- & energy efficiency

Approach

- Autonomic resource management based on control, analytics, and optimization

How much and what type of resources to allocate and when and where to deploy them?
Interdisciplinary collaborations

Images from http://artsandhumanities.pressible.org

Bottlenecks and Performance Anomalies

1. Controlling end-user performance
2. Managing extreme amounts of incoming data
Design and Management of Software-Defined Infrastructures

- Future computers will be designed in software, not in hardware
- Massive scale disaggregated hardware
- Dynamic definition (and redefinition) of virtual system
  - Arbitrarily large “imbalance” between virtual systems’ CPU-memory-network
  - Less constraints in resource management optimization
- Partial rack-scale SDIs are (soon) here
- Disaggregated datacenters in planning
Academic & Industrial collaboration

- Deep collaboration with LU (LCCC & EIT)
  - 20 msek VR framework grant
  - 790 man-days exchange time UmU-LU
    - 253 visits of length 1-17 days by 44 different individuals (23 from UmU and 21 from LU)
- Several UU collaborations
  - Holmgren, Hellander, Kaxiras, Sagonas
- Joint papers during the last year:
  - Google, IBM, RedHat, Ericsson, Intel, SAP
  - UU, LU, KTH, Lawrence Berkeley Lab, TU Vienna, Imperial College, Dublin City Univ., Univ. Castilla-La Mancha, Queens Univ. Belfast, ...
- Seven EU FP7 / H2020 projects

Cloud Control Workshops

- Foster multi-disciplinary research on Cloud Management, leveraging
  - autonomic computing, control theory, distributed systems, energy management, machine learning, mathematical statistics, performance modeling, systems management, etc.
- Flexible format
  - Often mainly discussion-oriented
- 9 workshops in 3.5 years
- (Surprisingly) easy to attract participants
  - Key people from, e.g., Google, VMware, Intel, Ericsson, Netflix, IBM and academia world-wide

cloudresearch.org/workshops