Introducing Service-level Awareness in Clouds
Cristian Klein1, Martina Maggio2, Karl-Erik Årzén2, Francisco Hernández-Rodriguez1
1Umeå University, Sweden, 2Lund University, Sweden

cristian.klein@cs.umu.se, martina.maggio@control.lth.se, karlerik@control.lth.se, francisco@cs.umu.se

https://github.com/cristiklein/cloudish

Problem: unexpected events
- flash-crowds
- hardware failures

Idea: reduce service-level to avoid data center overload
Downgrade user experience to reduce capacity requirements and serve more users

Our Proposal

Results

Conclusions
- clouds can more robustly withstand capacity shortages
  - flash-crowds
  - hardware failures
  - applications can reduce their service level (turn off optional features)
  - infrastructure can rebalance resources among hosted applications

Perspectives
- combining with other mechanism
  - migration
  - horizontal scaling
  - overbooking
  - devising a billing model