LABORATÓRIO MÍDIACOM LabGen

Universidade Federal Fluminense

Towards an Autonomic and Decentralized Bandwidth Provisioning on a Multitenant Datacenter

Diogo Menezes Ferrazani Mattos

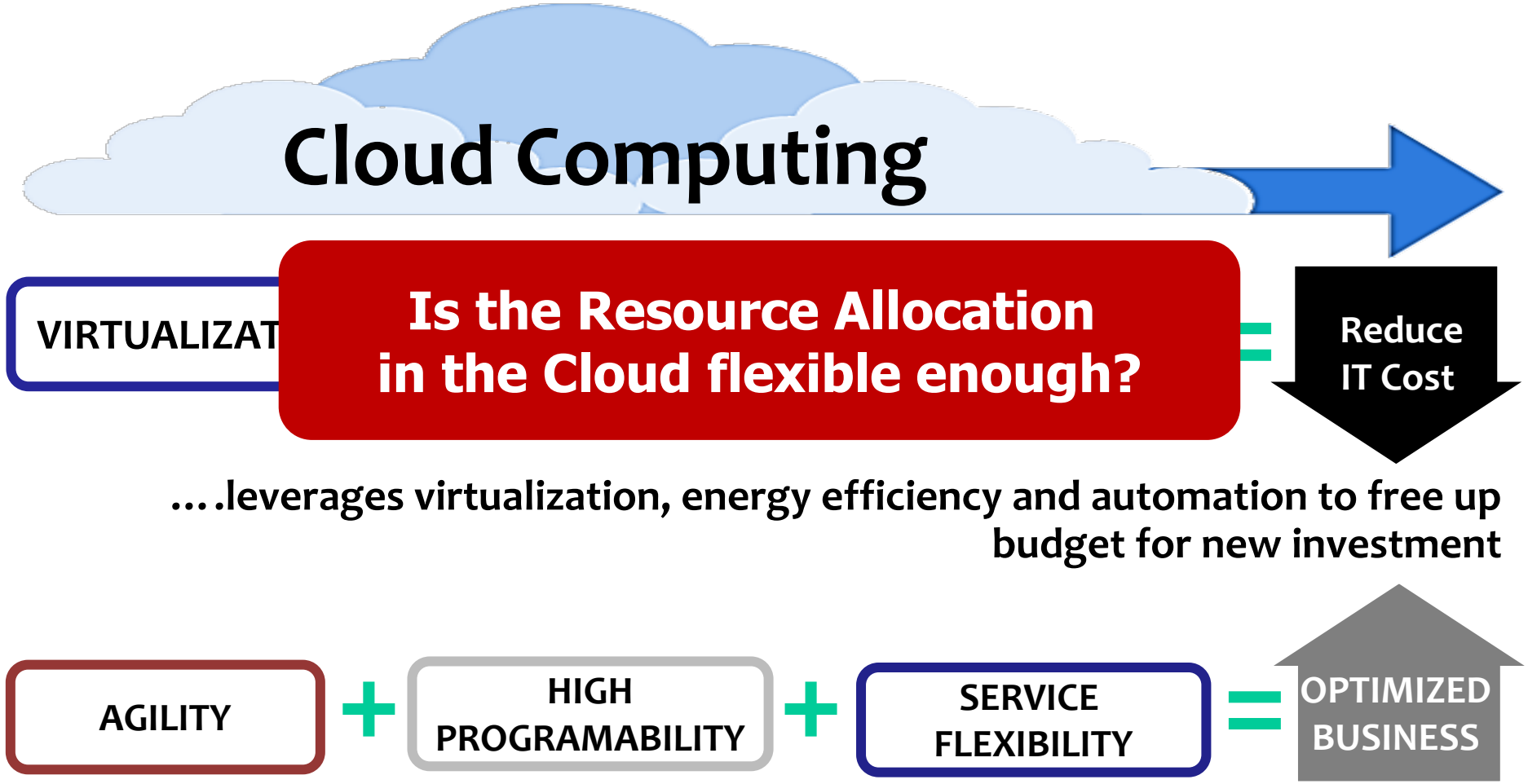
Thanks to Reiner H. Santos Filho, Tadeu N. Ferreira, Dianne S. V. Medeiros

Departamento de Engenharia de Telecomunicações – TET/TCE/UFF

Instituto de Computação – IC/UFF

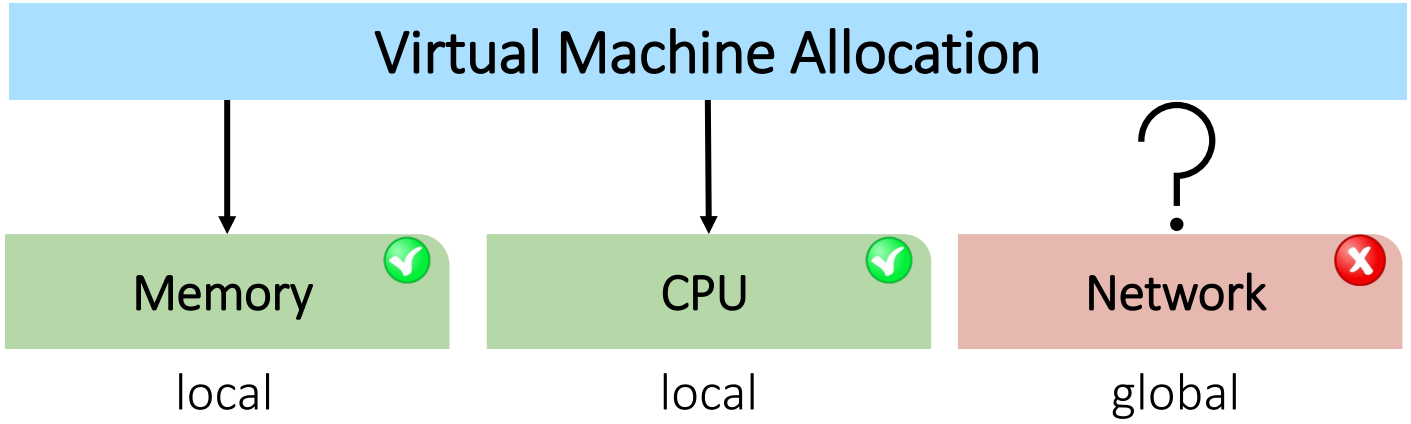
Universidade Federal Fluminense

Motivation

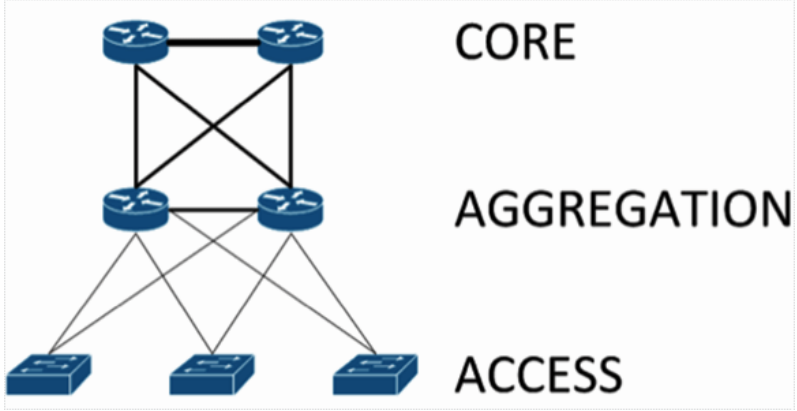

Cloud Computing
VIRTUALIZAT
Is the Resource Allocation in the Cloud flexible enough?
=
Reduce IT Cost
….leverages virtualization, energy efficiency and automation to free up budget for new investment
AGILITY
+
HIGH PROGRAMABILITY
+
SERVICE FLEXIBILITY
=
OPTIMIZED BUSINESS
… allowing you to optimize new investments for direct business benefits

Motivation

- Infrastructure as a Service (IaaS) providers offer on-demand resource provisioning



Network is a distributed resource

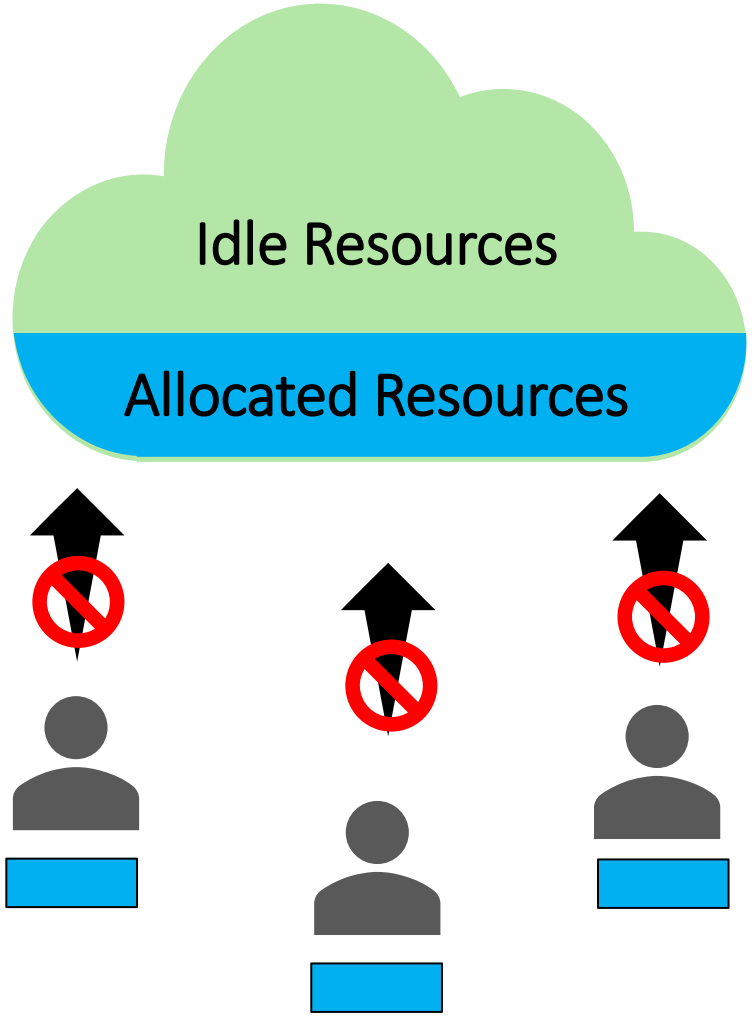


Motivation

Unallocated capacity

- Idle Resources
- Revenue Loss

Overprovisioning is the usual network allocation approach!



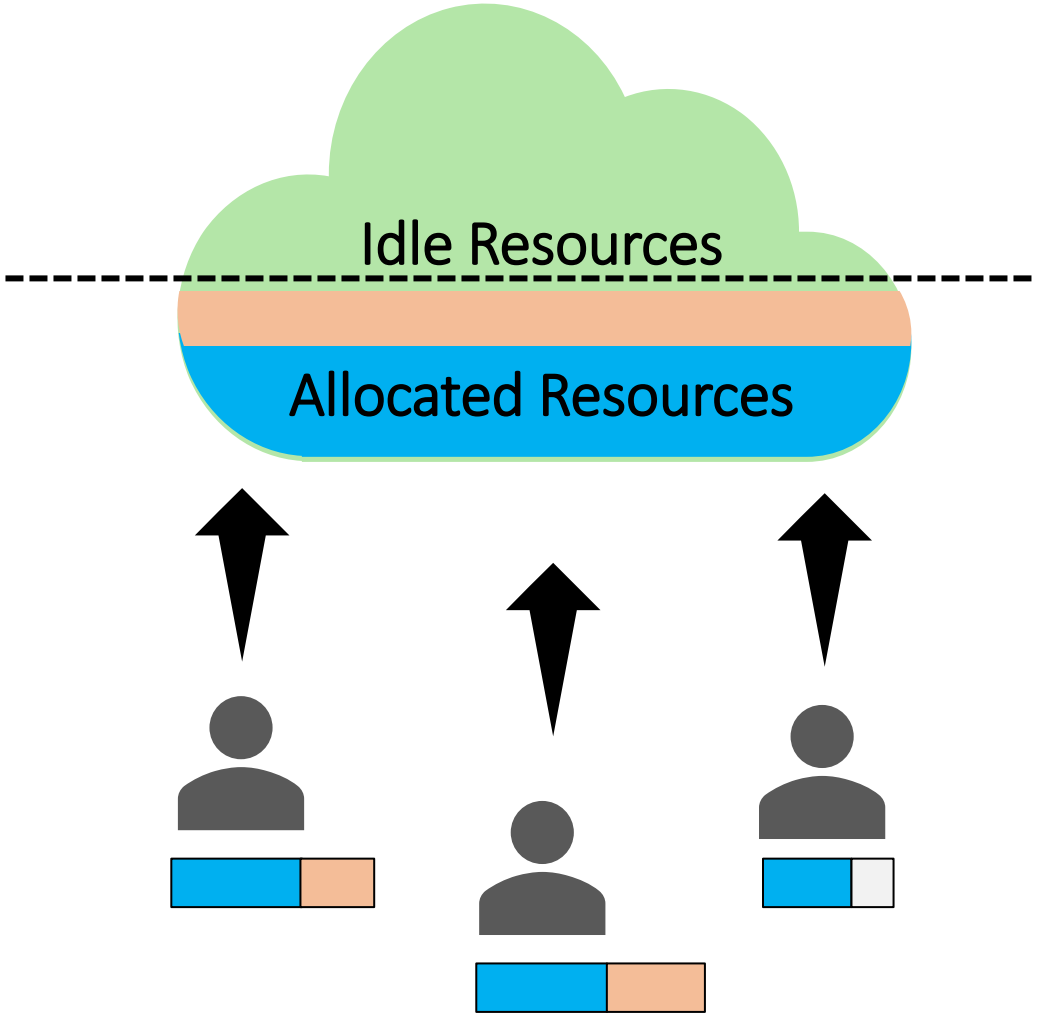
Motivation

Unallocated capacity

- Idle Resources
- Revenue Loss

↓

Redistribution of idle resources



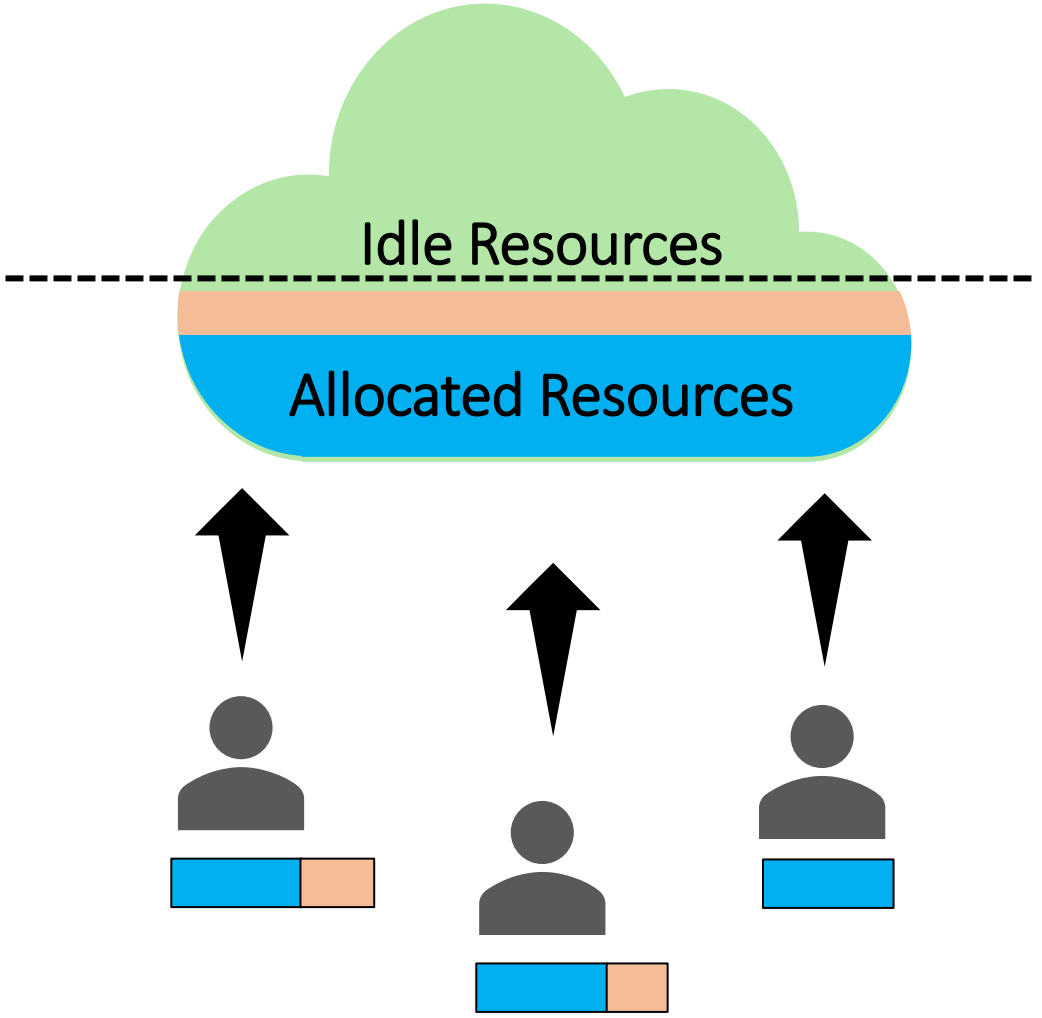
Motivation

Unallocated capacity

- Idle Resources
- Revenue Loss

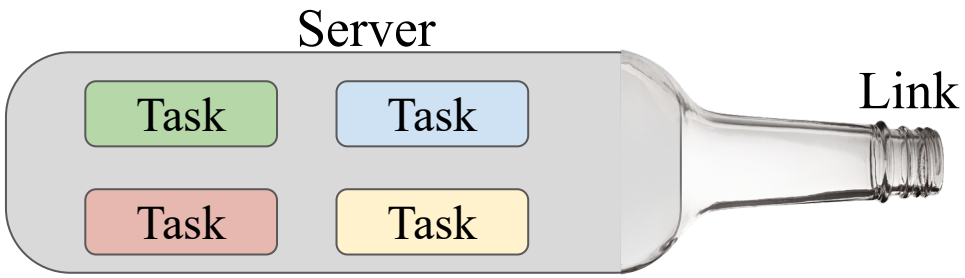
Redistribution of idle resources

Respect for service level agreements (SLAs)



Challenges on Network Dynamic Provisioning

- Network resource sharing depends on several factors
  - Tasks allocated on the same server
  - Tenants communicating with the same server
  - Links shared by tenants
- Environment can be extremely dynamic
  - Variation in tenant traffic
  - Tenants entering and leaving
  - Modification in IaaS topology (links)



Application performance unpredictability

Proposal

- Improve the efficiency of network resource sharing in an IaaS with multiple tenants
  - Increase provider revenue
  - Ensure tenants' SLAs
  - Reduce idle network resources

Billing
Bandwidth
as well as
Traffic Volume

Development of an automated tool for provisioning network resources

Key Idea

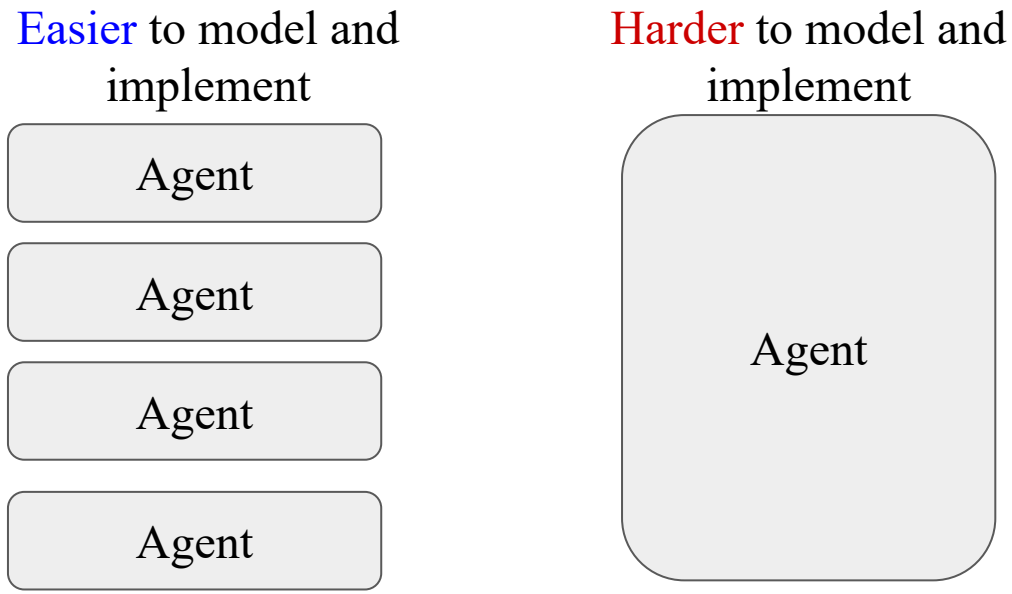
- Datacenter Network = Software Defined Networking + Machine Learning

Proposal

- Decision Agents
  - Machine Learning → Autonomic
  - Multiagent approach → Decentralized

Proposal

Fuzzy Inference
+
Reinforcement Learning



Decision Agent

Fuzzy Inference System

- Used as an universal approximator
- Flexible to adopt multi-agent approaches
- Easy to model

Reinforcement Learning

- Used in time varying and dynamic environments
- Multi-agent approaches are possible
- Modeled as a Markov Decision Process (MDP)

Simplicity

Lower bandwith oscillation during learning

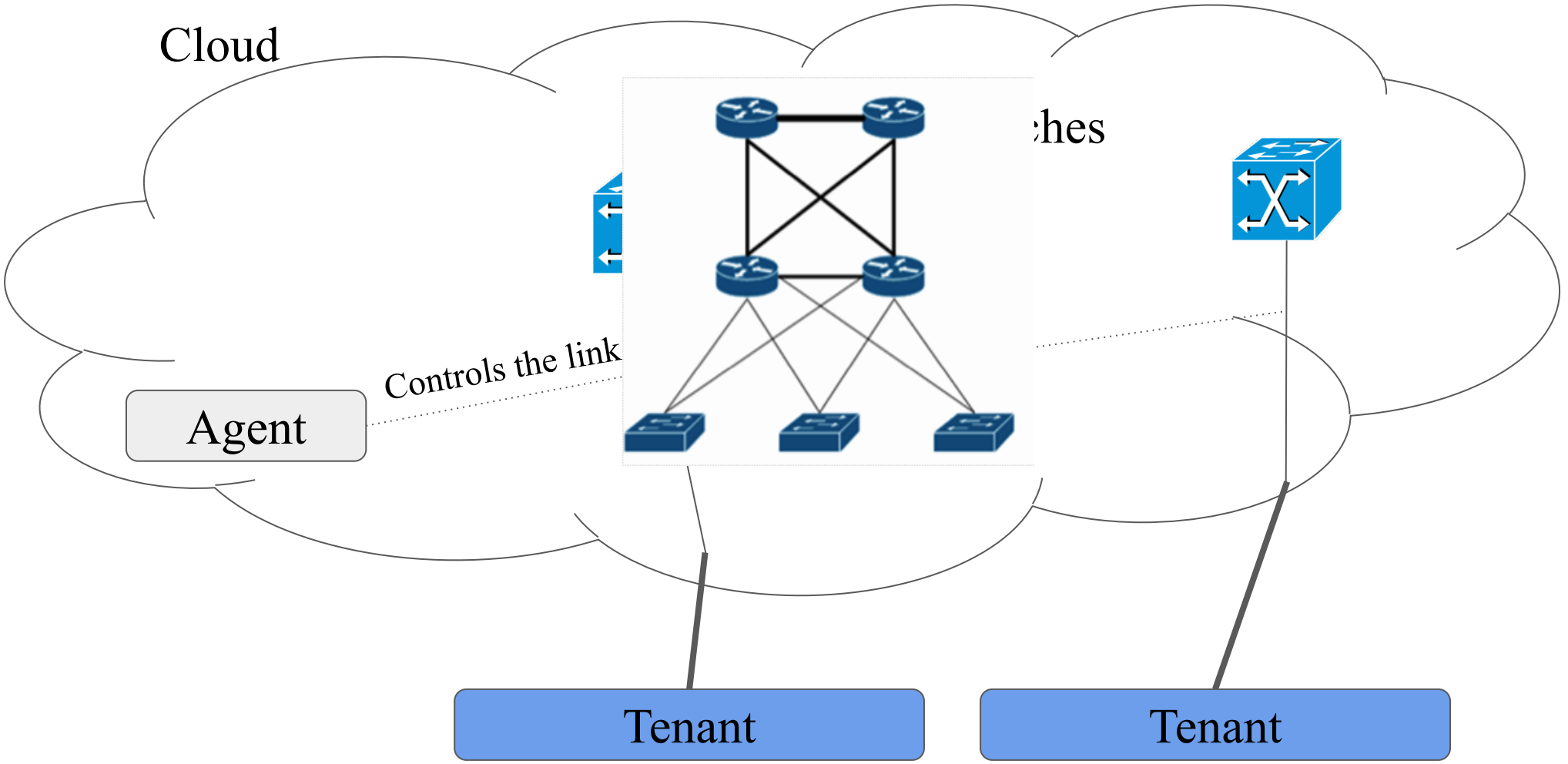
Model-free Algorithm

Ability to find the best policies

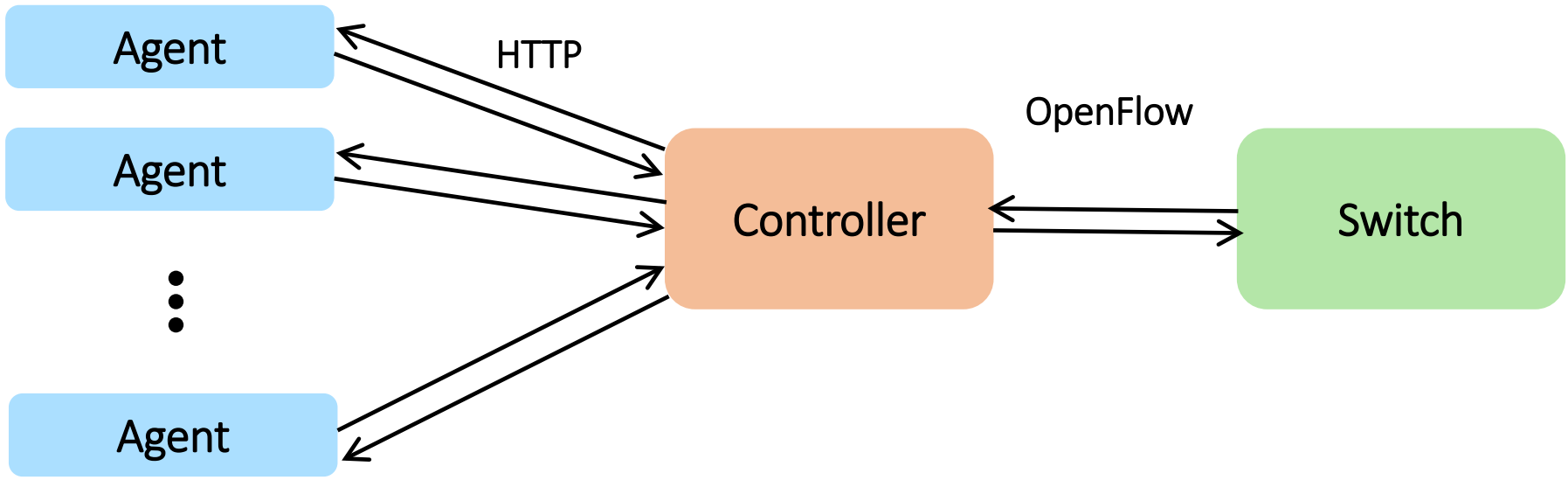
Discreet space of actions and states

Suitable for dynamic environments

Controlling the Cloud Network

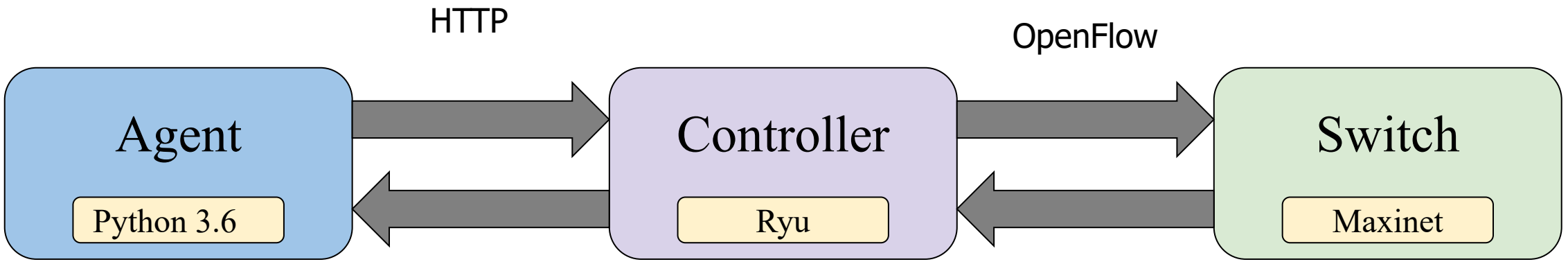


Controlling the Cloud Network



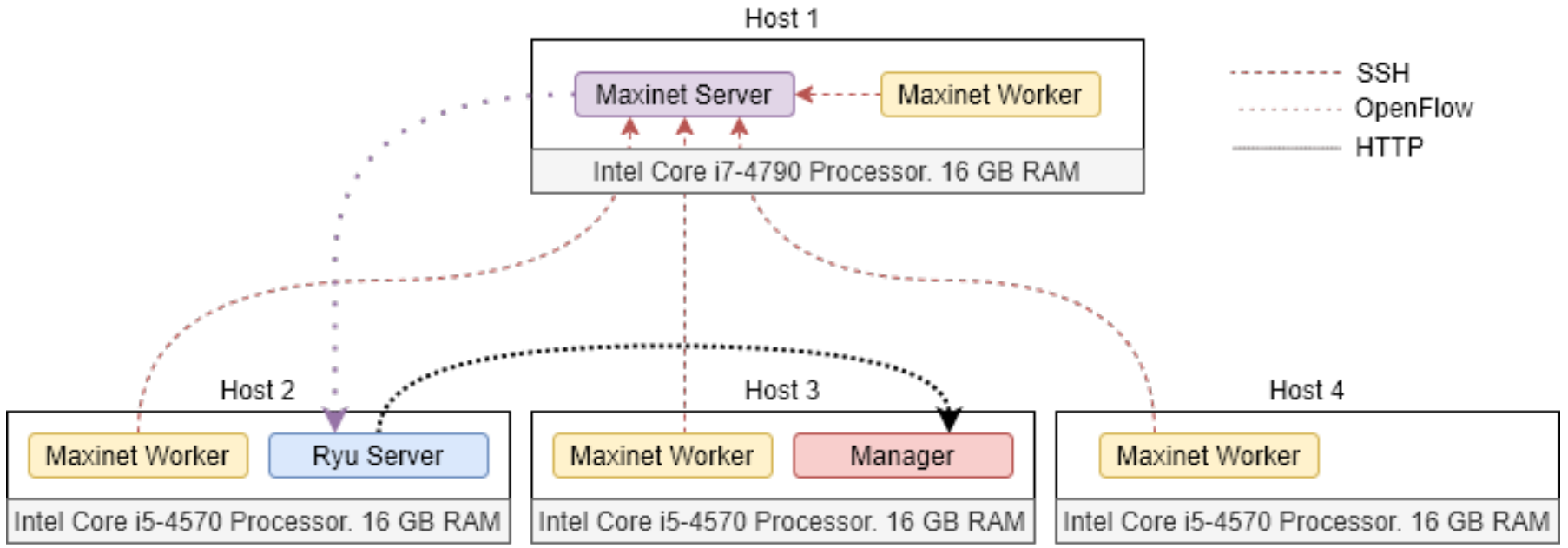
Proposal Evaluation

- Network Emulation



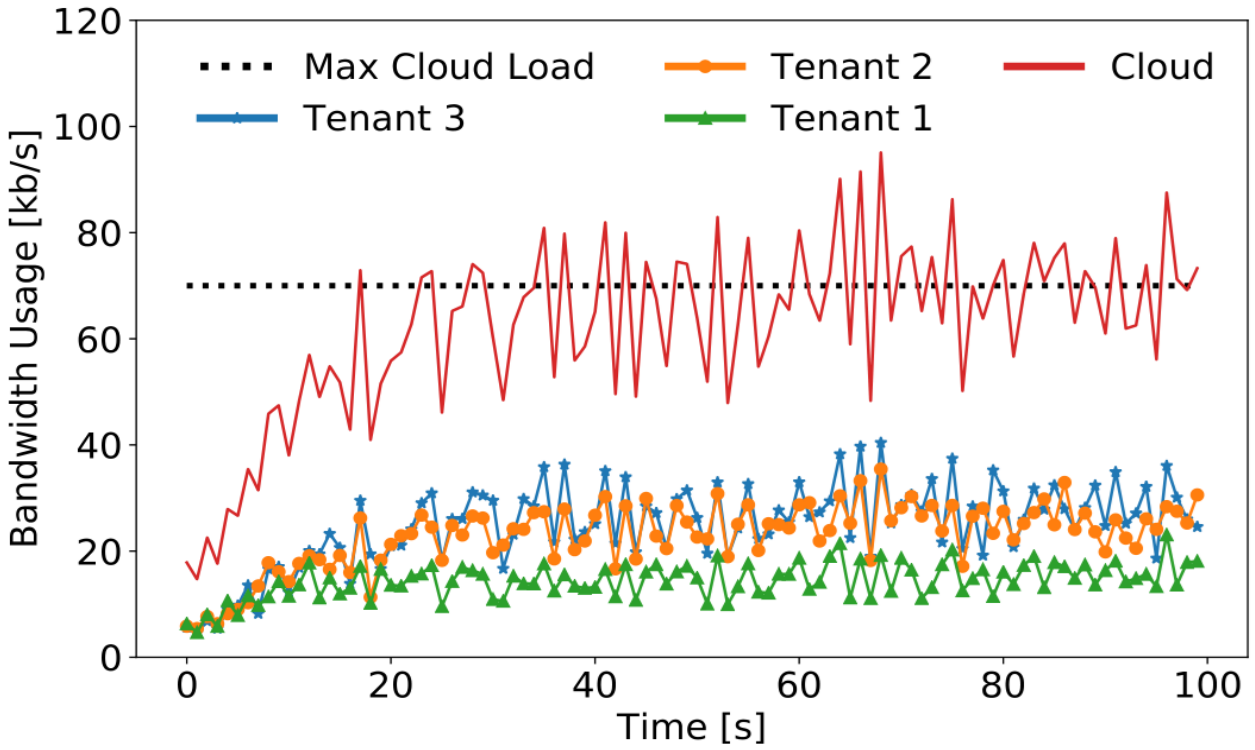
Emulation Environment

- Distributed Emulation with Maxinet
  - 4 workers



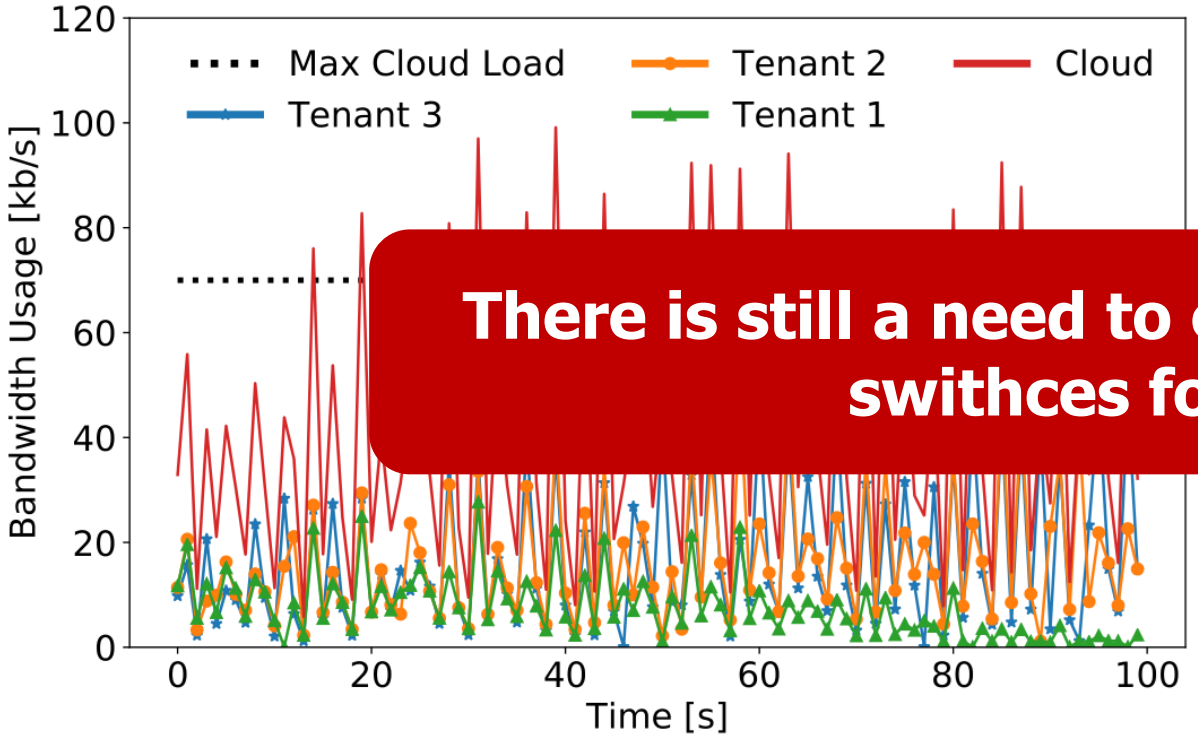
Emulation Results

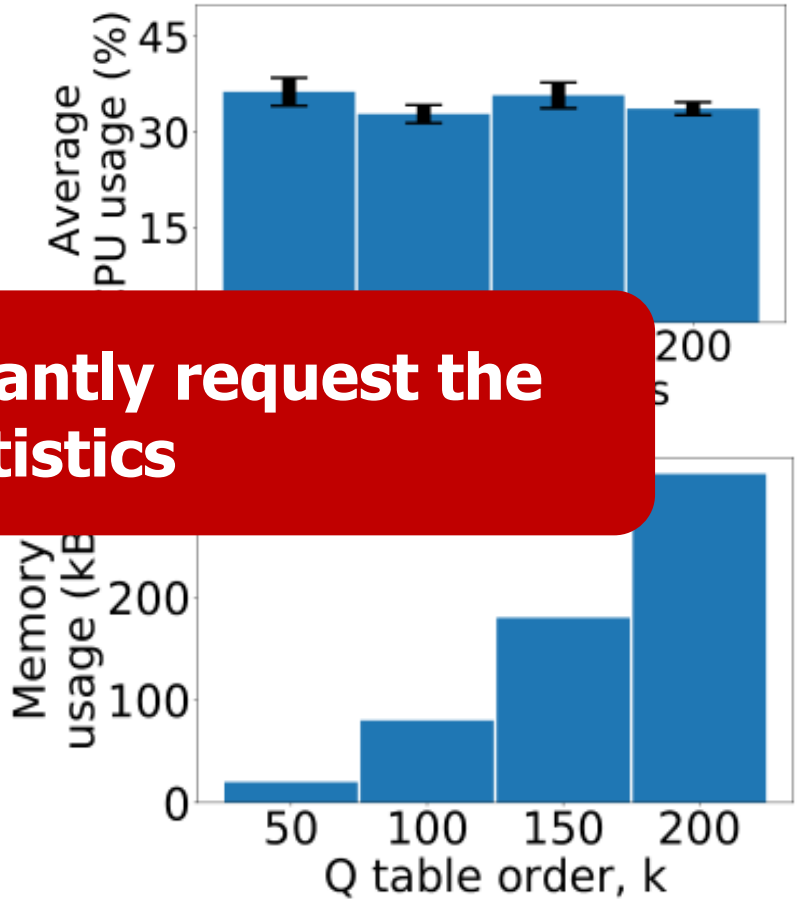
- Emulation: 3 clients, constant traffic, fuzzy inference system



Emulation Results

- Emulation: 3 clients, constant traffic, Q-Learning, 4 actions, 15 states





There is still a need to constantly request the swithces for statistics

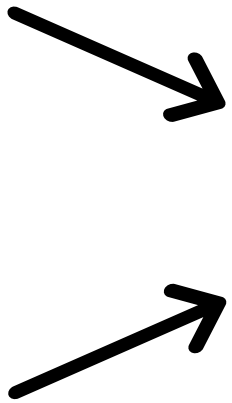
Emulation Results

Observable Case (OC)

Before the agent determines the action, the cloud load is measured

Partially Observable Case (POC)

Before the agent determines the action, cloud processing load is measured and mapped using bandwidth usage

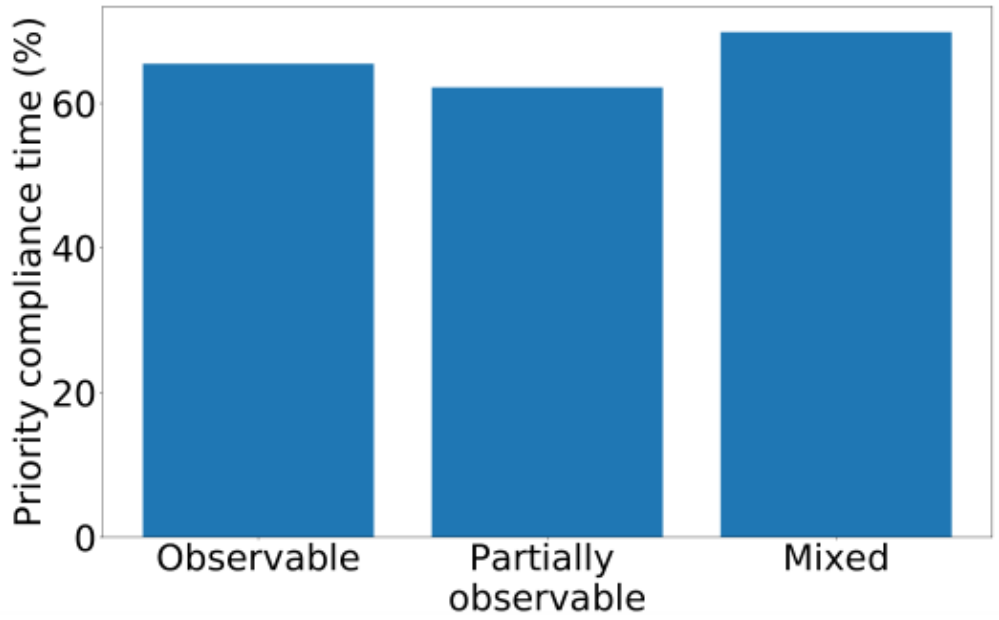


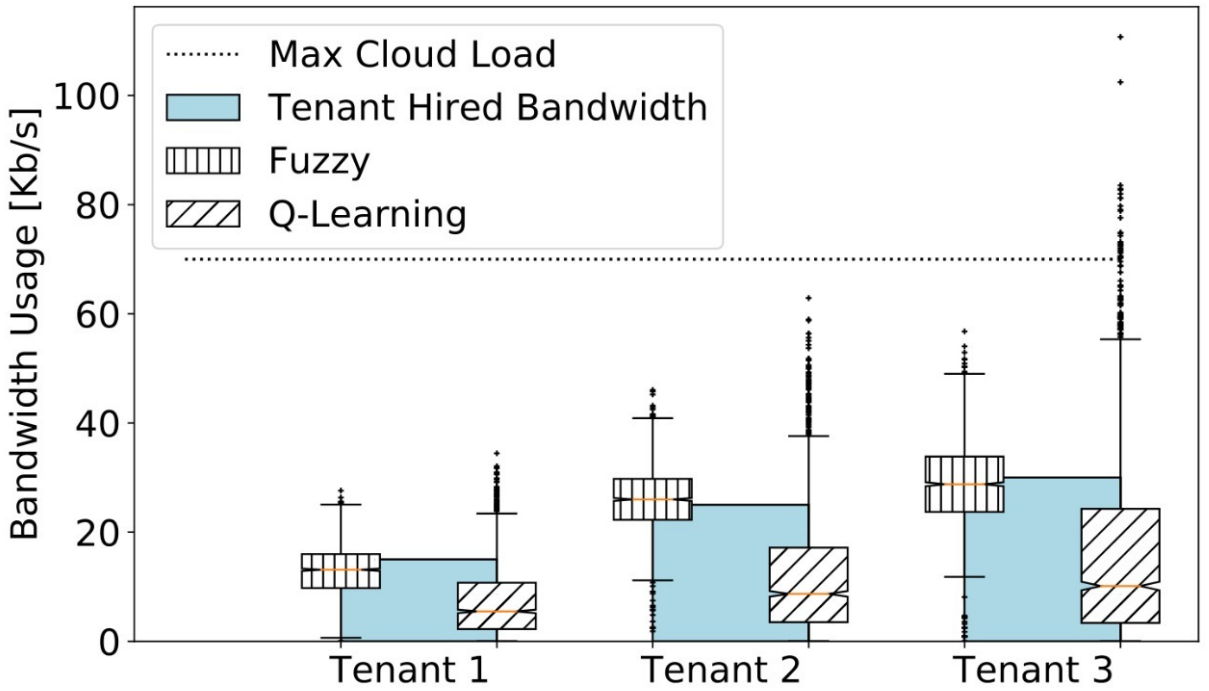
Mixed Case

Alternation between the two cases

Emulation Results

- Fuzzy Inference System vs. Q-Learning





Emulation Results

- Fitness for the Fuzzy Inference System

Interval	Tenant 1	Tenant 2	Tenant 3	Server
1	85.49%	96.40%	95.67%	96.32%
2	87.49%	96.00%	97.74%	97.35%
5	83.90%	93.87%	98.05%	99.57%

$$fitness = 1 - \left| 1 - \frac{\text{Measured Throughput}}{\text{Expected Bandwidth}} \right|.$$

Conclusion

- Proposed approach includes use of diffuse inference and reinforcement learning for bandwidth provisioning

- Technical feasibility analysis in progress
  - Initial tests performed in an emulated environment
    - Initial results obtained indicate feasibility of the proposed approach

- On going project
  - Develop a bandwidth allocation tool for datacenter networking
  - https://rlp.labgen.uff.br (Available in Brazilian Portuguese)

Funding

Future Work

- Generalize the mechanism for a bigger number of tenants and realistic traffic
- Compare other methods to solve the same scenario
- Get access to a real dataset for testing and validation

Publications

- SANTOS FILHO, R. H. ; FERREIRA, T. N. ; DIOGO MATTOS, M. F. ; MEDEIROS, DIANNE S. V. . *A Rapid Fuzzy Controller for Decentralized Bandwidth Provisioning on a Multitenant Data Center*. In: 11th International Conference on Network of the Future (NoF 2020), 2020, Bordeaux, França. Proceedings of NoF 2020, 2020.
- SANTOS FILHO, R. H. ; FERREIRA, T. N. ; MATTOS, D. M. F. ; MEDEIROS, DIANNE S. V. . *A Lightweight Reinforcement-Learning-based Mechanism for Bandwidth Provisioning on Multitenant Data Center*. In: 27TH INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING, 2020, Niterói/RJ, Brasil. Proceedings of IWSSIP 2020, 2020.

LABORATÓRIO MÍDIACOM LabGen

Universidade Federal Fluminense

Towards an Autonomic and Decentralized Bandwidth Provisioning on a Multitenant Datacenter

Diogo Menezes Ferrazani Mattos

Thanks to Reiner H. Santos Filho, Tadeu N. Ferreira, Dianne S. V. Medeiros

Departamento de Engenharia de Telecomunicações – TET/TCE/UFF

Instituto de Computação – IC/UFF

Universidade Federal Fluminense