Availability Knob: Flexible User-Defined Availability in the Cloud

Mohammad Shahrad and David Wentzlaff
Princeton University

amazon web services EC2 | Microsoft Azure

They all provide fixed 99.95% availability!

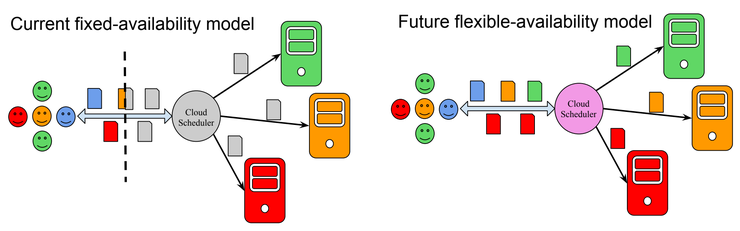
What doesn't seem right?

Cloud Customers:
- Various downtime demands
- Different willingness to pay

Cloud infrastructure:
- Heterogeneous HW/SW
- Reliability costs dollars

The Availability Knob (AK) provides clients with flexible availability.

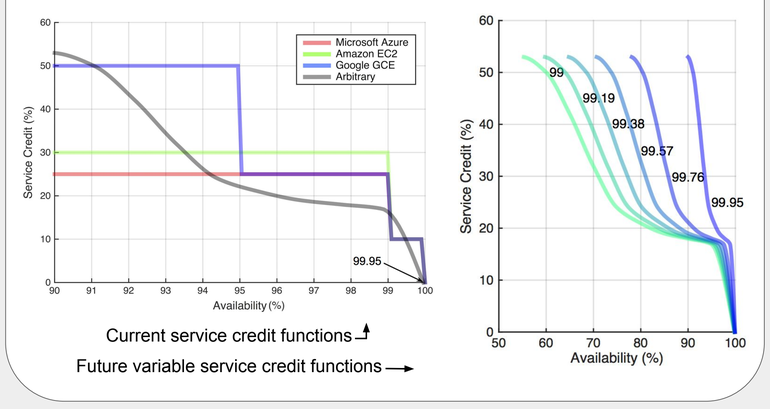
What needs to be changed?



1. Service Level Agreements (SLAs)
2. Cloud Management

SLA for Flexible Availability

1. Desired availability & measurement period
   E.g: 99.98 % & 7-day period
2. Price scale for desired availability
   E.g: 1.12 price scale for 99.98%
   (12% price increase compared to default availability)
3. Variable service credit (penalty) functions



Availability Knob Scheduler

Any machine unavailability is recorded in a failure DB.

Scheduling performed considering:
- Expected time-to-next-failure for available machines
- VM size and expected downtime in case of failure
- User's delivered service

✚ New features enabled by extra knowledge on user demand:

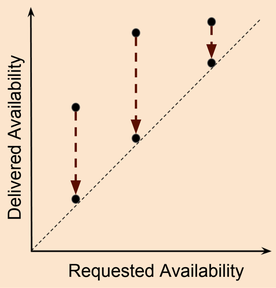
Benign VM Migration (BVM)

Virtual machines (VMs) can be over-served due to:
- Low failure rate
- Initial assignment to high-reliability resources

BVM: Periodic migration of over-served VMs to available cheaper resources

Deliberate Downtimes (DDT)

- Deliberately fail VMs near the end of period.
- Main motivation: building market incentives.



Economics of Availability Knob

Pricing for Incentive Compatibility

Providers can:
- Neglect meeting service level objectives (SLOs)

Clients can:
- Run buggy VM
- Intentionally cause downtime

We use game theory to make sure both parties have $ incentives to act truthfully.

How does AK make money?

Supply Chain Flexibility

More Efficient Resource Utilization

Compensating Risks by Variable Profit Margins

Deployment of AK

1. No hardware change is required
   - Low technology adoption cost
2. Existing fixed availability is a subset of AK
   - AK can continue to serve current customers
3. AK can be used under the hood automatically

AK Evaluation

Stochastic simulator written in MATLAB

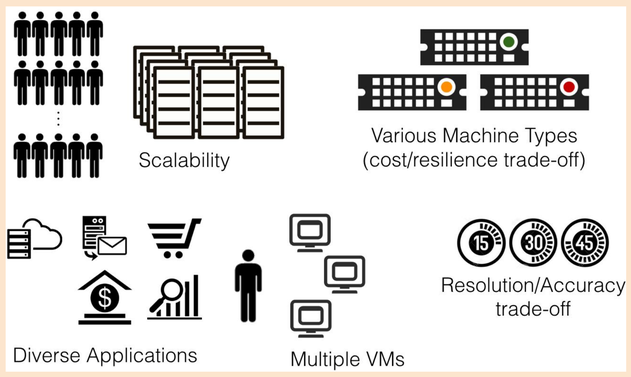
MATLAB

Prototype implementation in OpenStack

openstack CLOUD SOFTWARE

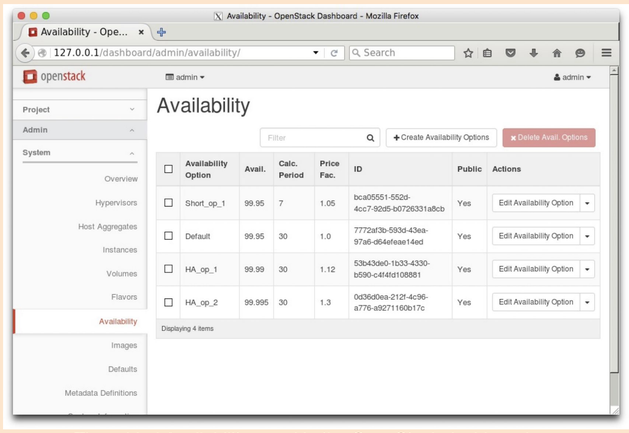
AKSim: Stochastic Cloud Simulator

For our study analyzing large scale systems over long time periods is required.



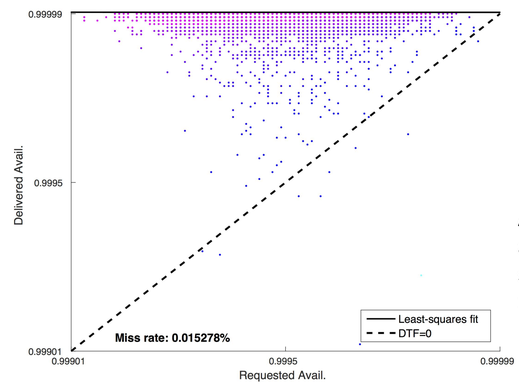
AK OpenStack Prototype

We modified/added some APIs, DBs, and the scheduler.



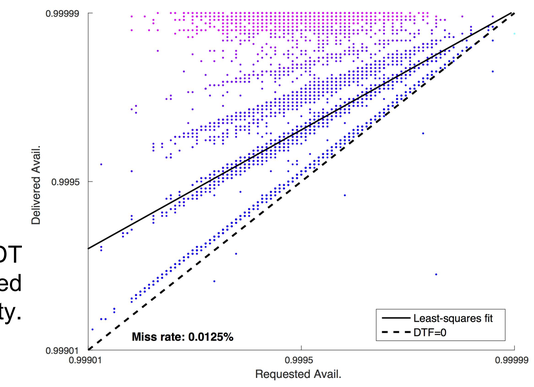
The added Availability panel in the OpenStack dashboard.

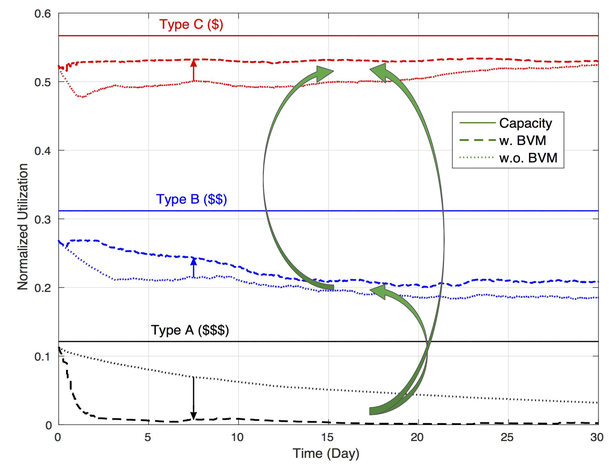
Some Results



AK's availability-aware scheduler meets >99.98% of SLOs.

Applying DDT shapes delivered availability.





BVM offloads VMs to cheaper less reliable resources.

$[Type_A, Type_B, Type_C]$	Cost Redu. (%)	Δ Miss Rate (%)
[50,150,800]	5.8	0.60
[50,250,700]	5.3	0.58
[100,300,600]	6.6	0.69
[150,350,500]	5.3	0.65
[200,400,400]	3.6	0.63
[300,400,300]	2.5	0.76
[450,100,450]	2.4	0.68

Dependence of BVM cost reduction and induced SLO misses on machine type blend.

Summary of results:

- **~10% Cost Reduction**
- **Up to 20% Profit Margin Increase**
- **Improved User Satisfaction**

Final Insights

- Traditional mindset to provide higher and higher availability is not necessarily correct, scalable, or economically efficient.
- Providing supply chain flexibility can save and earn money.
- IaaS providers can deploy Availability Knob to:
  - Reduce costs
  - Create more revenues
  - Improve user satisfaction
- Future work includes:
  - Automatic user demand extraction for AK
  - Extending AK to PaaS and federated cloud

Download Paper

SOCC ACM Symposium on Cloud Computing
October 5-7, 2016, Santa Clara, CA

Mohammad Shahrad
Email: mshahrad@princeton.edu
Webpage: http://www.princeton.edu/~mshahrad/

My Website