Availability Knob: Flexible User-Defined Availability in the Cloud

Mohammad Shahrad and David Wentzlaff
Princeton University

They all provide fixed 99.95% availability!

What doesn't seem right?

<table>
<thead>
<tr>
<th>Cloud Customers:</th>
<th>Cloud Infrastructure:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Various downtime demands</td>
<td>- Heterogeneous HW/SW</td>
</tr>
<tr>
<td>- Different willingness to pay</td>
<td>- Reliability costs dollars</td>
</tr>
</tbody>
</table>

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

What needs to be changed?

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

Economics of Availability Knob

Pricing for Incentive Compatibility

<table>
<thead>
<tr>
<th>Providers can:</th>
<th>Clients can:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Neglect meeting service level objectives (SLOs)</td>
<td>- Run buggy VM</td>
</tr>
<tr>
<td>- Intentionally cause downtime</td>
<td></td>
</tr>
</tbody>
</table>

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

Some Results

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

Applying DDT shapes delivered availability.

AK Evaluation

Stochastic simulator written in MATLAB

Prototype implementation in OpenStack

AKSim: Stochastic Cloud Simulator

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

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.

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

Some Results

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

Applying DDT shapes delivered availability.

AK Evaluation

Stochastic simulator written in MATLAB

Prototype implementation in OpenStack

AKSim: Stochastic Cloud Simulator

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

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.

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

Some Results

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

Applying DDT shapes delivered availability.

AK Evaluation

Stochastic simulator written in MATLAB

Prototype implementation in OpenStack

AKSim: Stochastic Cloud Simulator

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