

IAAS CLOUDS

- Infrastructure as a Service (IaaS)
- Examples : Amazon EC2, Google Compute Engine, Rackspace
- IaaS rents out physical or virtual computing resources



Cloud servers have different costs and availability tradeoffs:

On-demand Servers:

- Fixed price per unit time
- Non-revocable

Spot Servers:

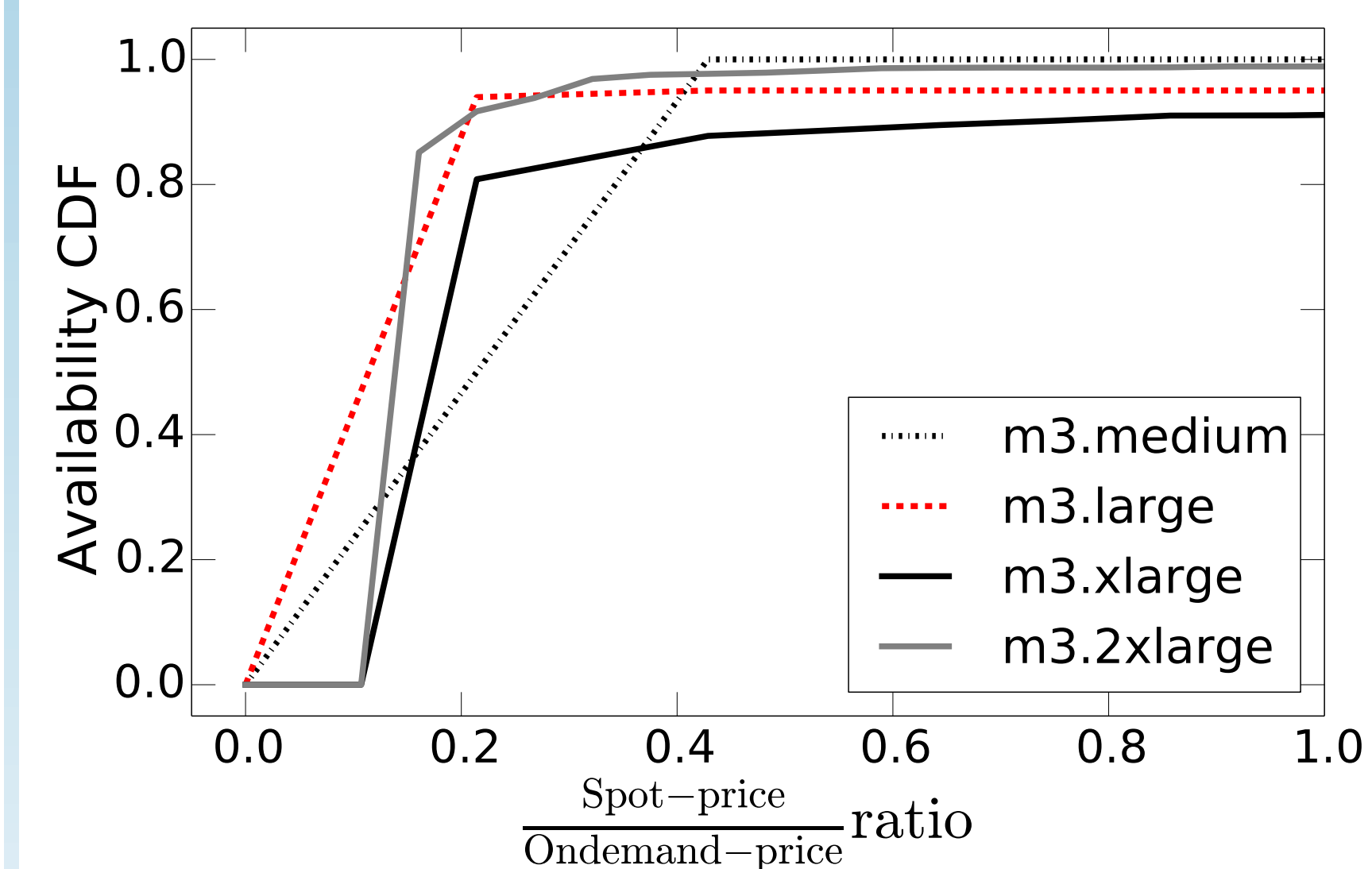
- Variable prices based on market conditions
- Revocable \implies lower availability
- Prices tend to be lower than On-demand servers
- Allow cloud provider to sell surplus capacity

PROBLEM STATEMENT

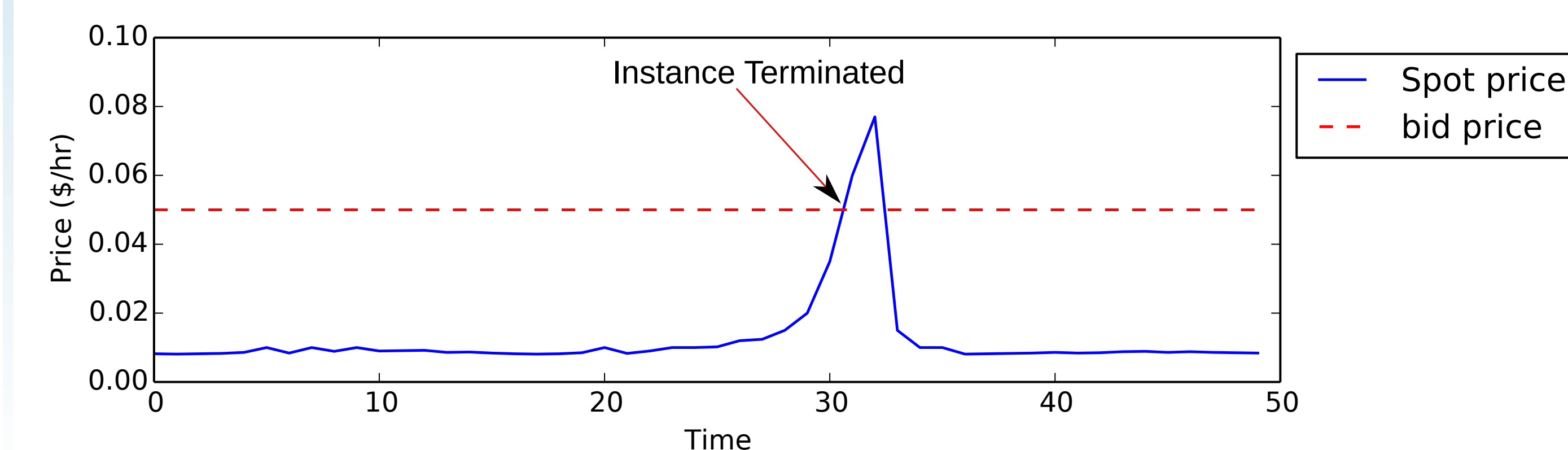
Run interactive applications on mix of Spot & On-demand servers

SPOT INSTANCES

- 2nd price auction determines spot price
- (Spot price > Bid) \implies termination
- *Small* termination warning (~2 minutes)

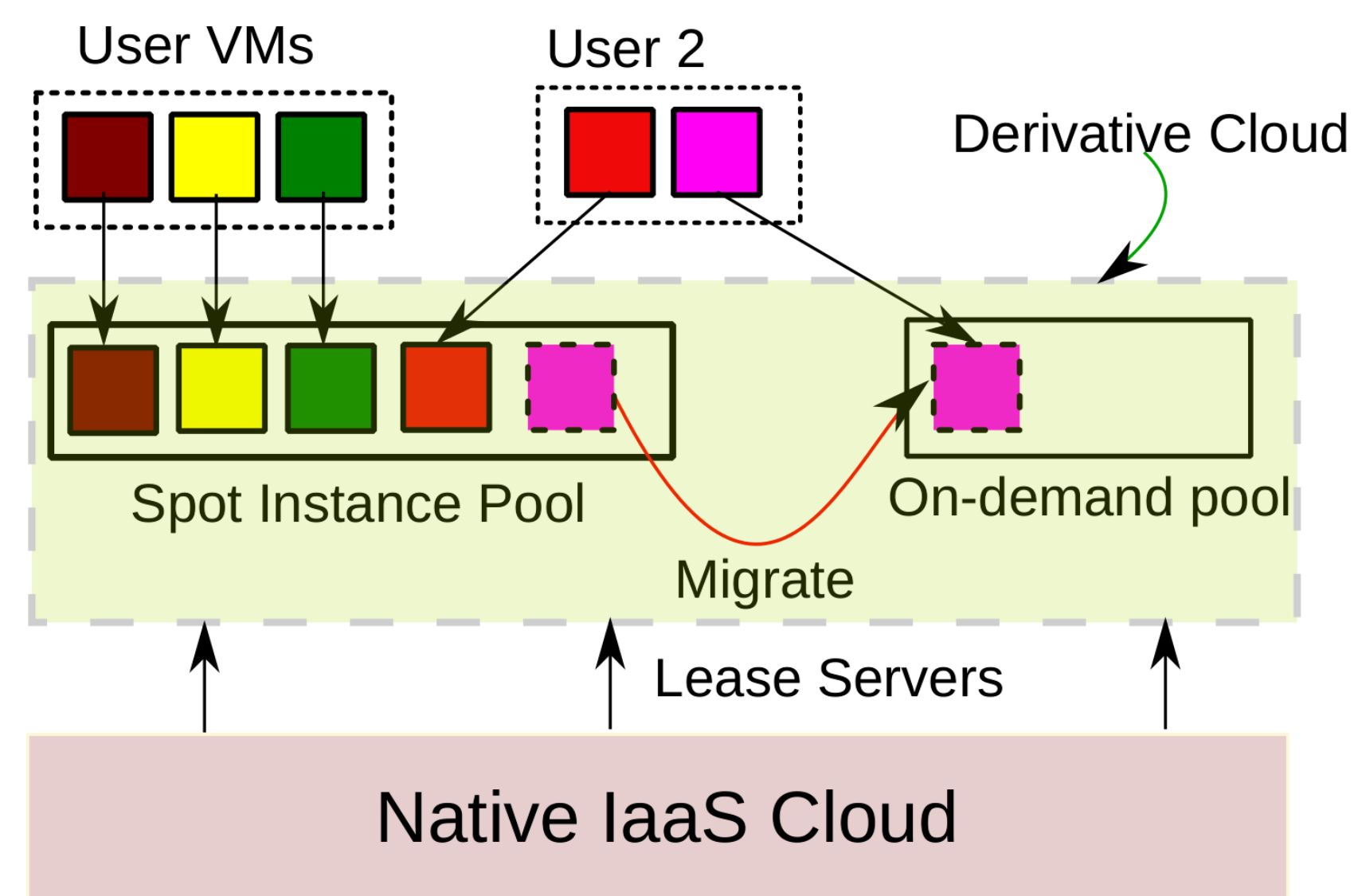


Spot instances are really cheap!



OUR SOLUTION: DERIVATIVE CLOUDS

- System to manage mix of spot and on-demand instances
- Intermediate layer between users and IaaS cloud provider
- Provide similar interface to users as provided by IaaS (Virtual Machines)
- Transparently migrate VMs between pools



OUR SYSTEM: SPOTCHECK

SpotCheck : derivative cloud on spot and on-demand instances

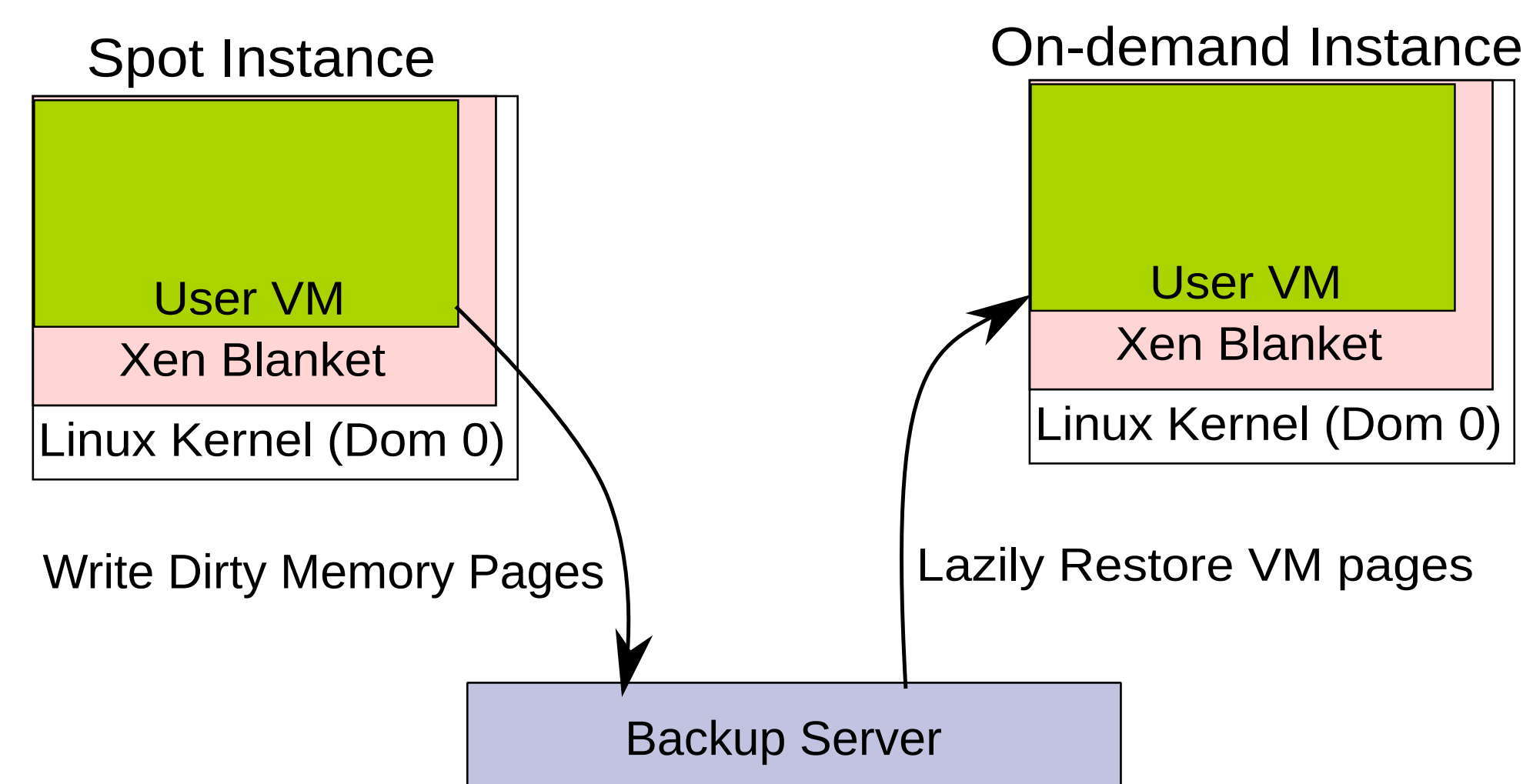
1. Ability to run interactive, disruption-intolerant applications
2. Not lose application state
3. Provide servers to customers at low cost
4. Not adversely impact application performance

Run on spot when possible, move to on-demand when evicted

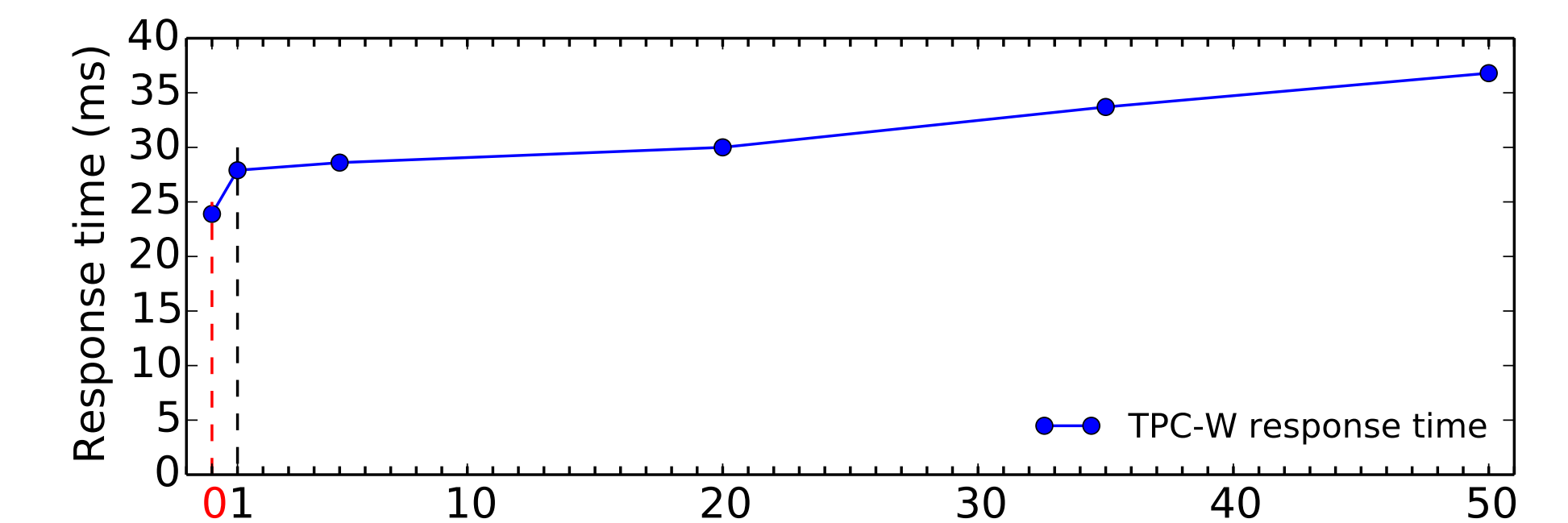
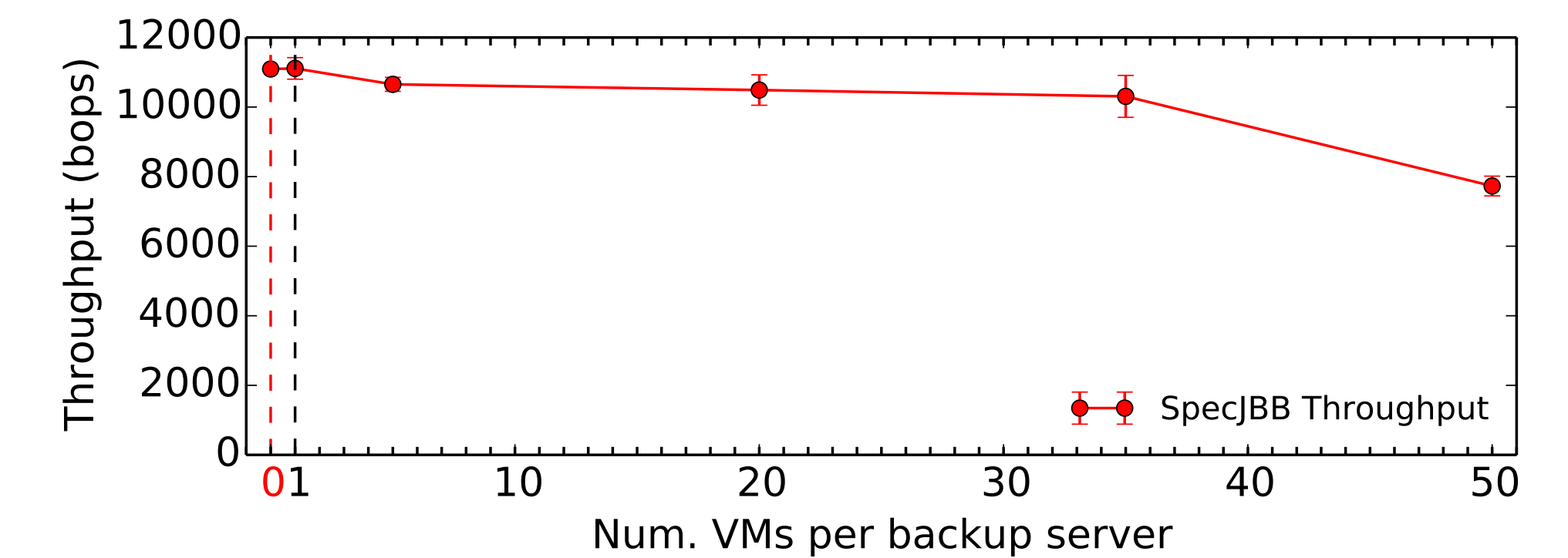
BOUNDED TIME VM MIGRATION

Bounded time migration : VM migrates within specified time

1. VM dirty memory pages transferred to a backup server continuously, incrementally, and asynchronously
2. Backup server able to support multiple (~50) VMs
3. Memory checkpoint lazily restored from backup server if spot instance terminates — fetch page on first access



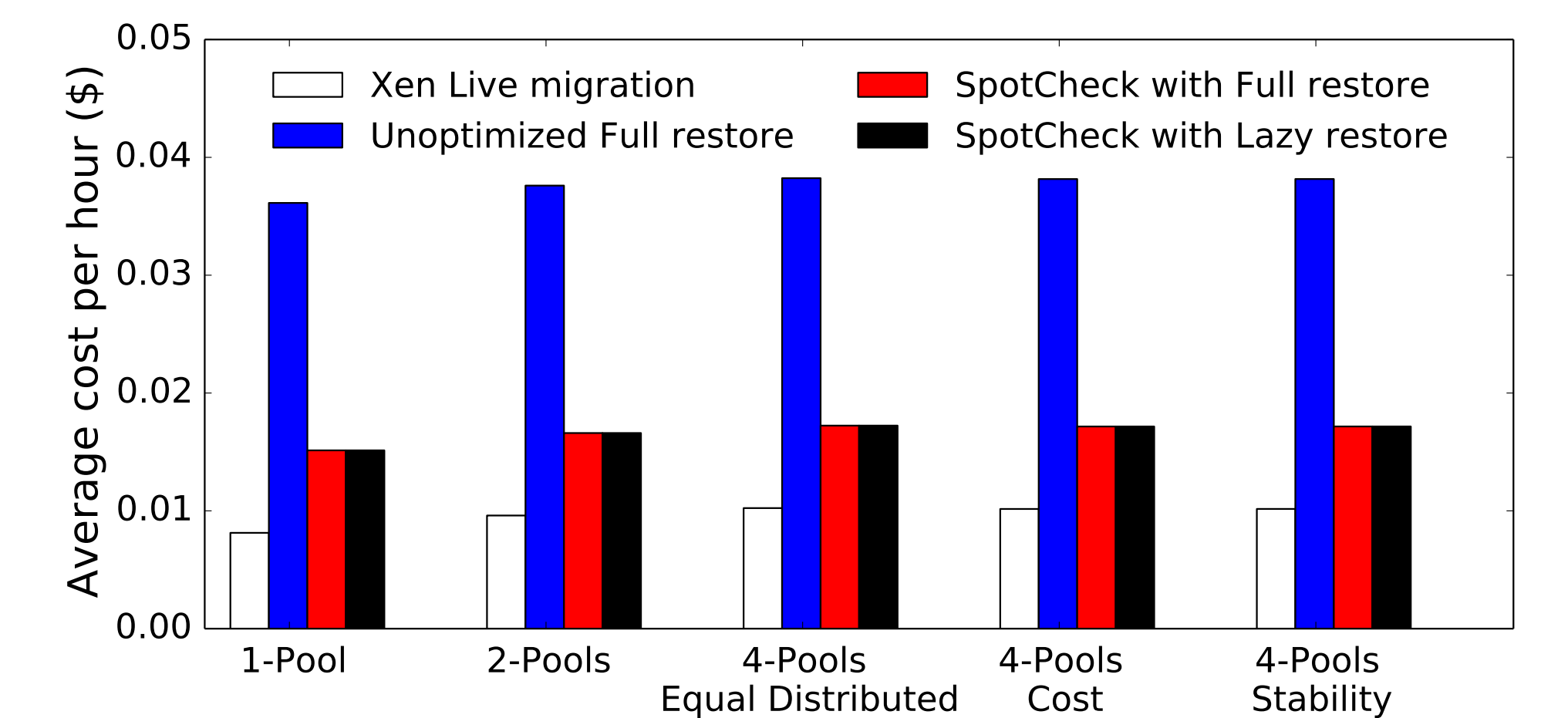
PERFORMANCE & SCALABILITY



40 VMs can share one backup server

COST

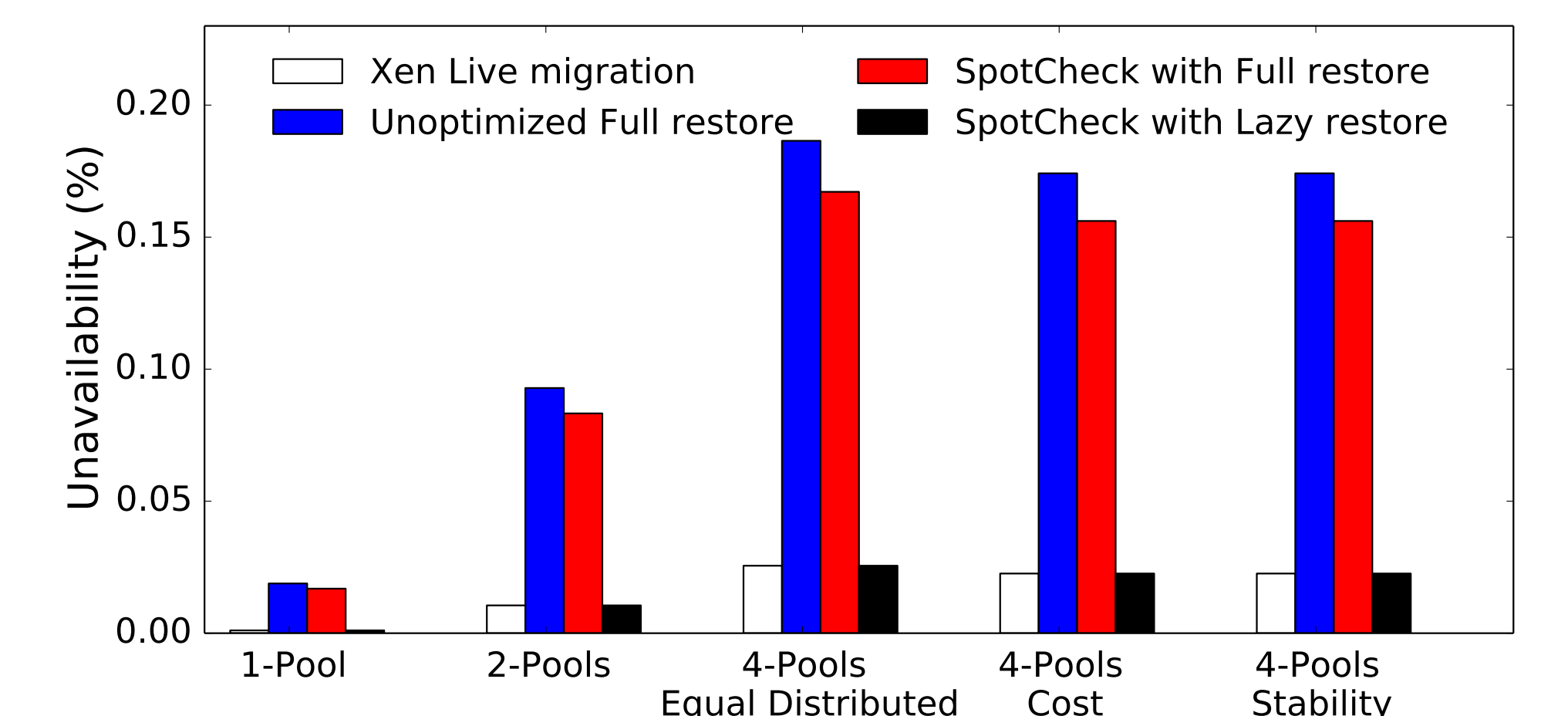
Expected Cost = 0.2 × On-Demand = \$ 0.014 / hour



Save 80% on your EC2 bill

AVAILABILITY

- Unavailability is due to migrations from spot to on-demand
- Small downtime (~20 seconds) during migration
- Due to latency of IaaS operations detaching & reattaching network & storage



99.9989% Availability