How Not to Bid the Cloud

Prateek Sharma, David Irwin, Prashant Shenoy

UMassAmherst

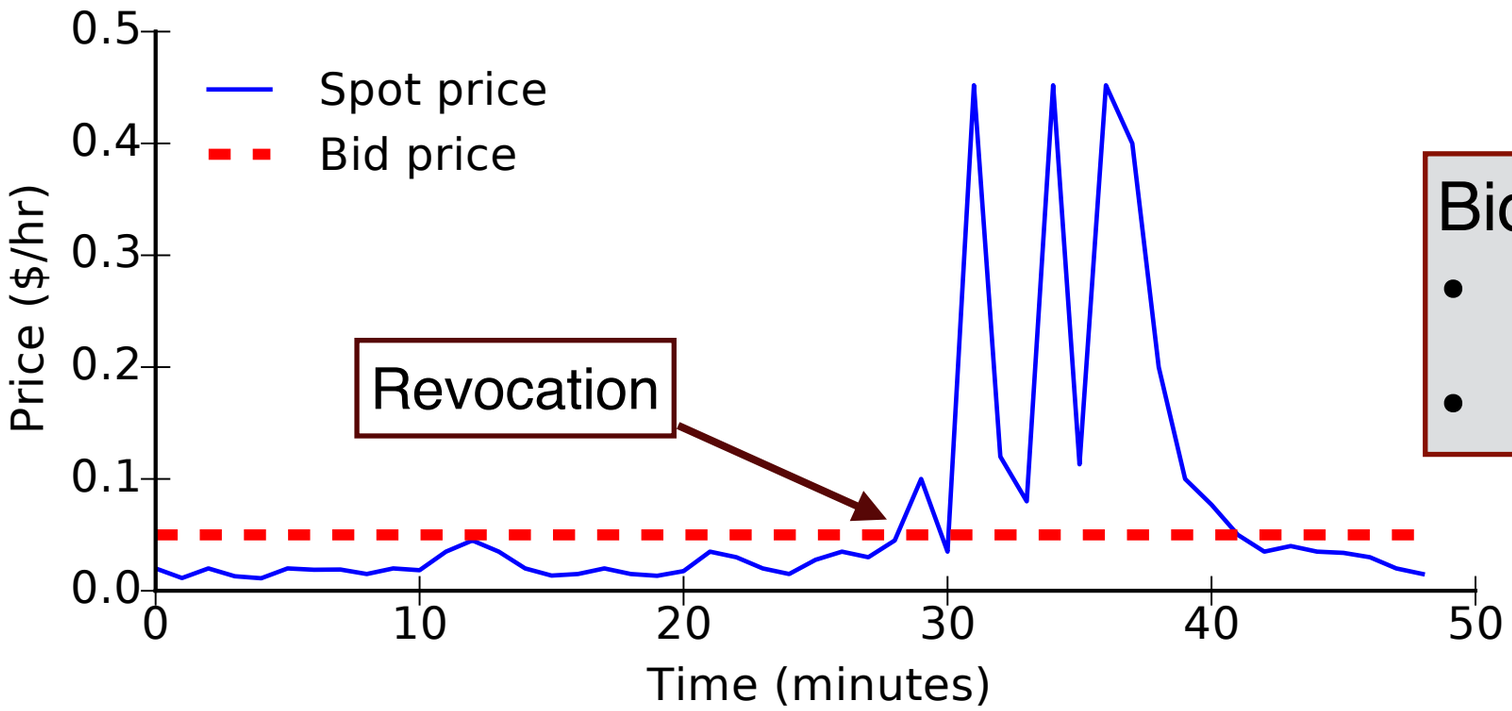
HotCloud 2016

Transient Servers in the Cloud

- Conventional cloud servers: on-demand servers
  - Fixed price, user-controlled life-span
- Cloud operators sell surplus capacity as low-cost transient servers:
  - Can be revoked at any time
  - **EC2 Spot instances**, GCE Pre-emptible VMs
- 70-90% lower costs ⇒ attractive for batch and delay tolerant applications

Bidding in EC2 Spot Markets

- Spot prices set by continuous second-price auction
- Users place a bid representing their maximum hourly price
- Spot price rises above bid ⇒ Server revoked after 2 minute warning



Bidding tradeoffs:
- High bid ⇒ high availability
- High bid ⇒ high cost

- Bidding strategies important to optimize cost, availability
- Zheng et.al. [SigComm '15], Zafer et.al., Tang et.al. [Cloud '12]
- **What is the impact of bidding on availability and cost?**

Talk outline

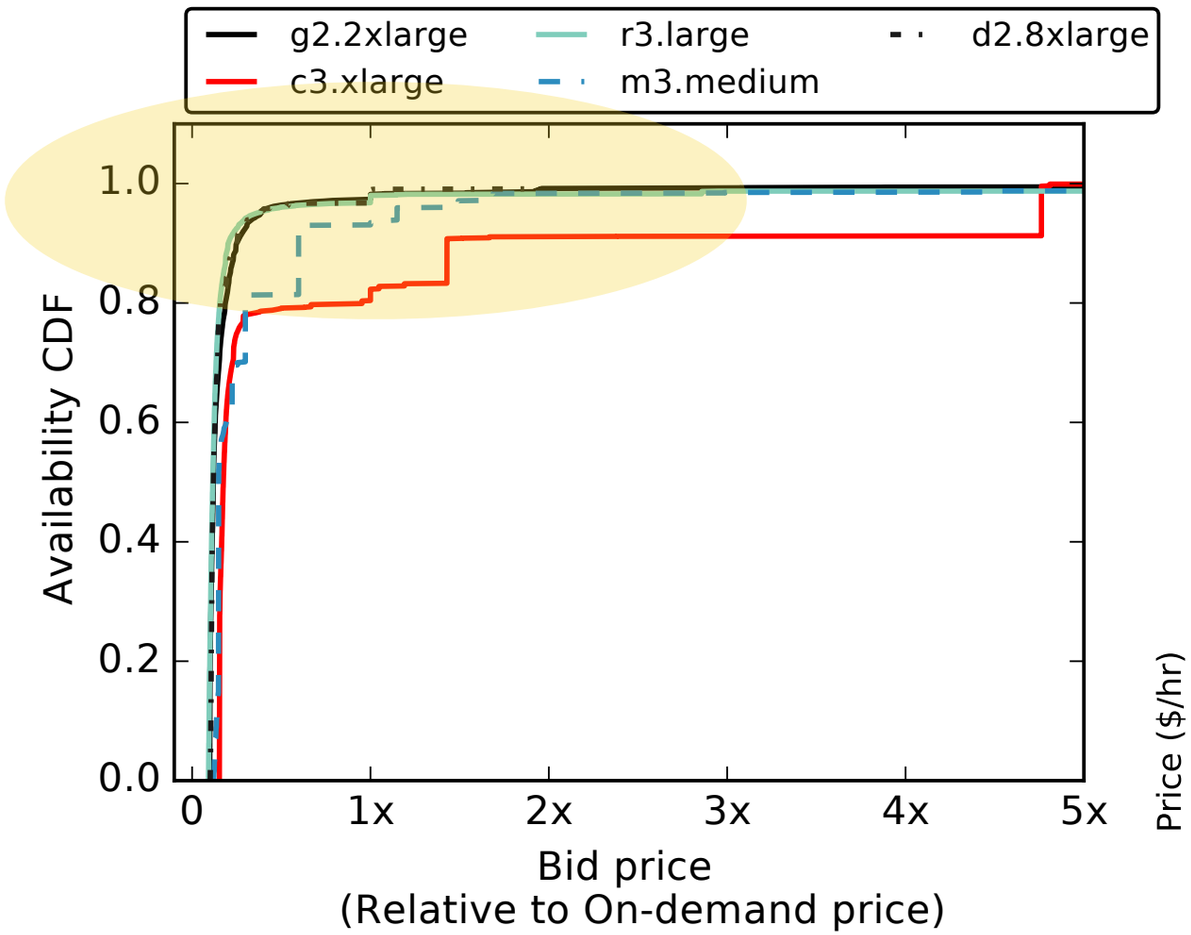
- Motivation: spot markets and bidding
- Comprehensive empirical analysis of effect of bidding
- Beyond bidding

Methodology

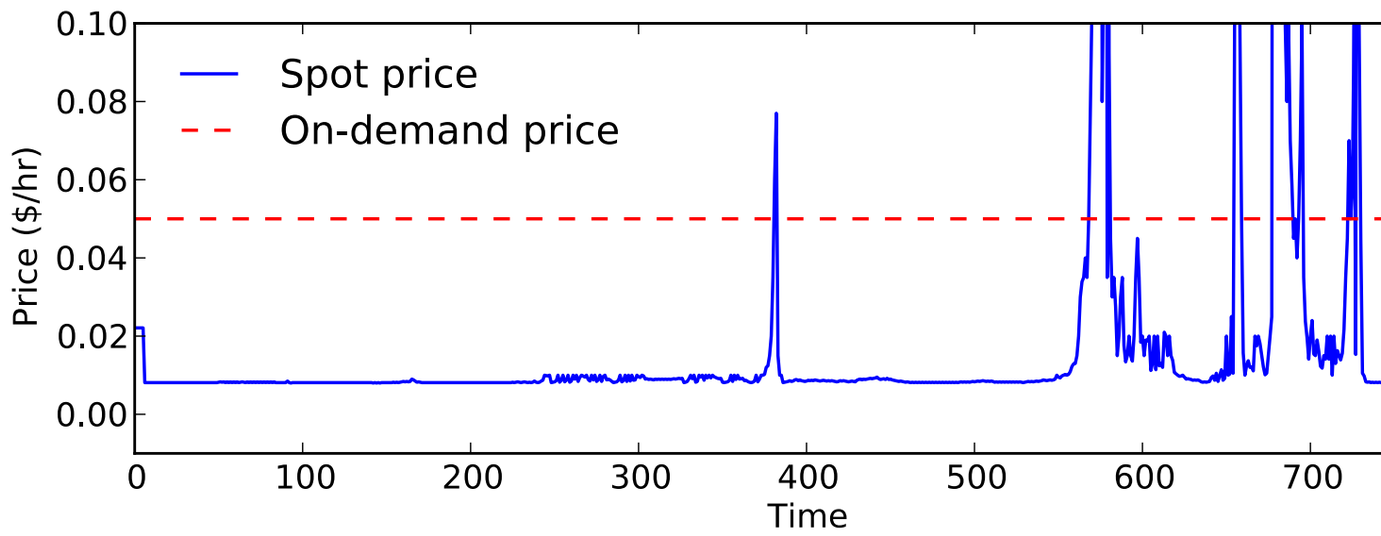
- Spot price traces published by Amazon
- Use spot price traces from March-October 2015
- 1500 markets : 8 geographic region, ~2 availability zones, 15 server types, 3 operating systems
  - Prior work is restricted to developing bidding strategies for a few (~10) markets
- Metrics: **Availability, Cost, Mean time between revocations**

Availability

- Availability : fraction of time for which spot price less than bid price



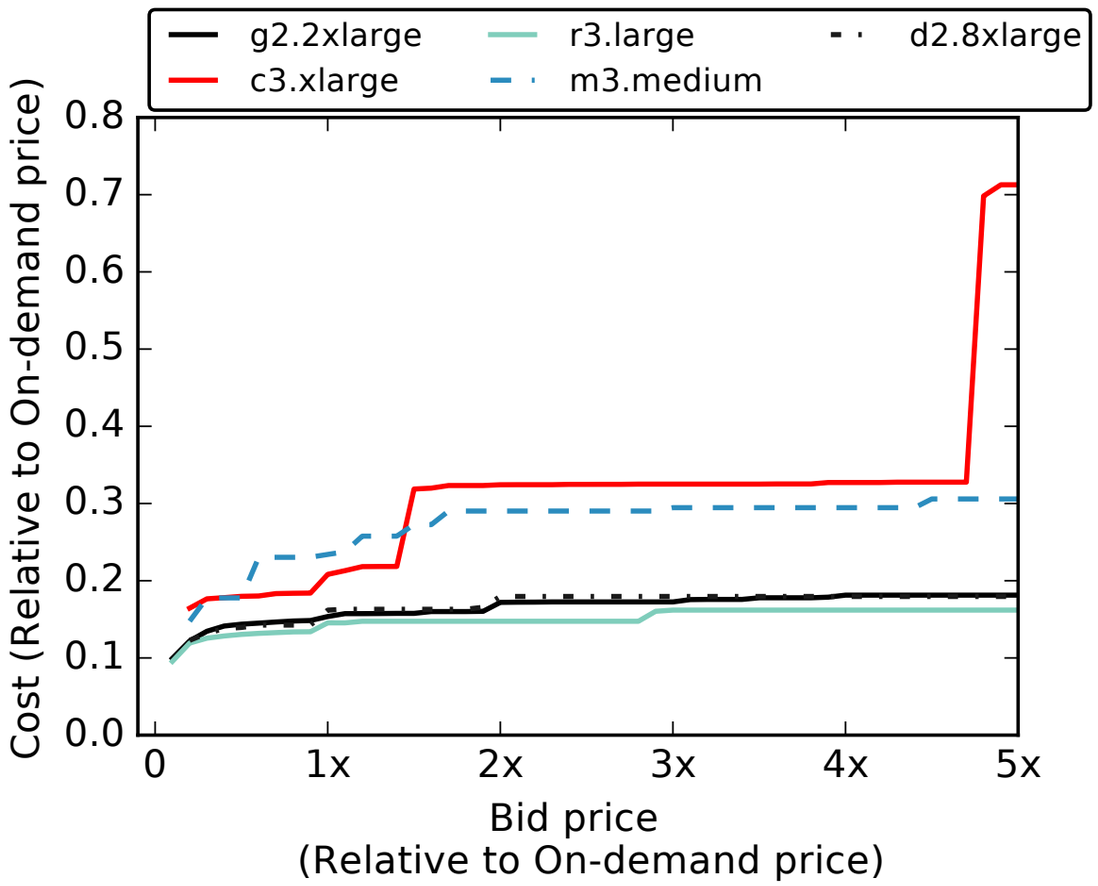
- Long tailed
- Availability is high (>90%) but does not reach 100%



- **Spot prices mostly low, with occasional large spikes**
- **High availability for wide range of bids**

Cost

- Cost of spot instances (relative to on-demand price) at different bid prices
- Costs determined by spot prices, not the bid-price itself

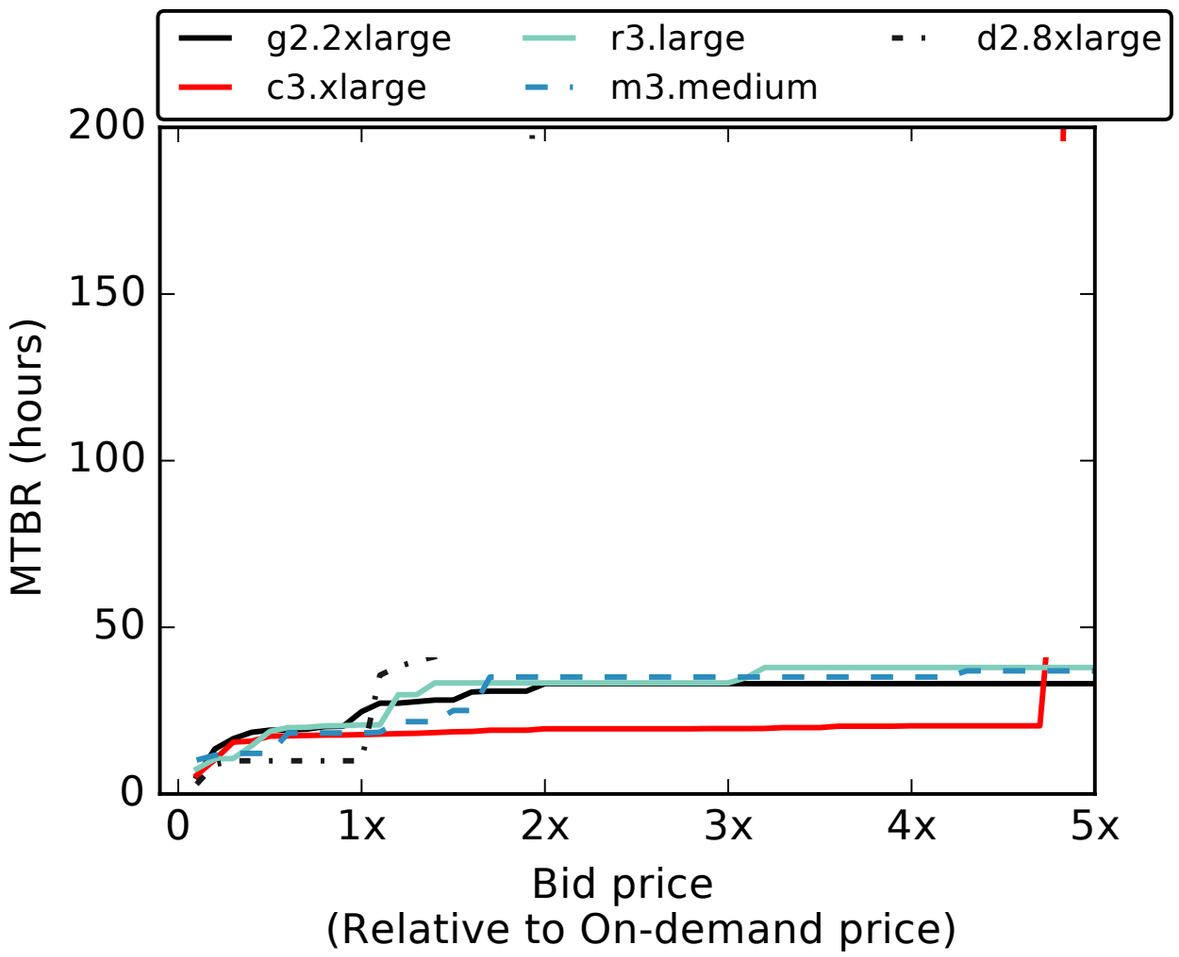


- Long tailed
- Cost of spot instances is low for wide range of bid prices

- **No cost penalty for high bid prices**
- **Cost not particularly sensitive to bidding**

Mean Time Between Revocations

- Mean time between revocations : how long applications can run uninterrupted
- MTBR≠Availability : Short, frequent spikes cause low MTBR



- Long tailed

- **Revocations are unavoidable if prices spike too high**

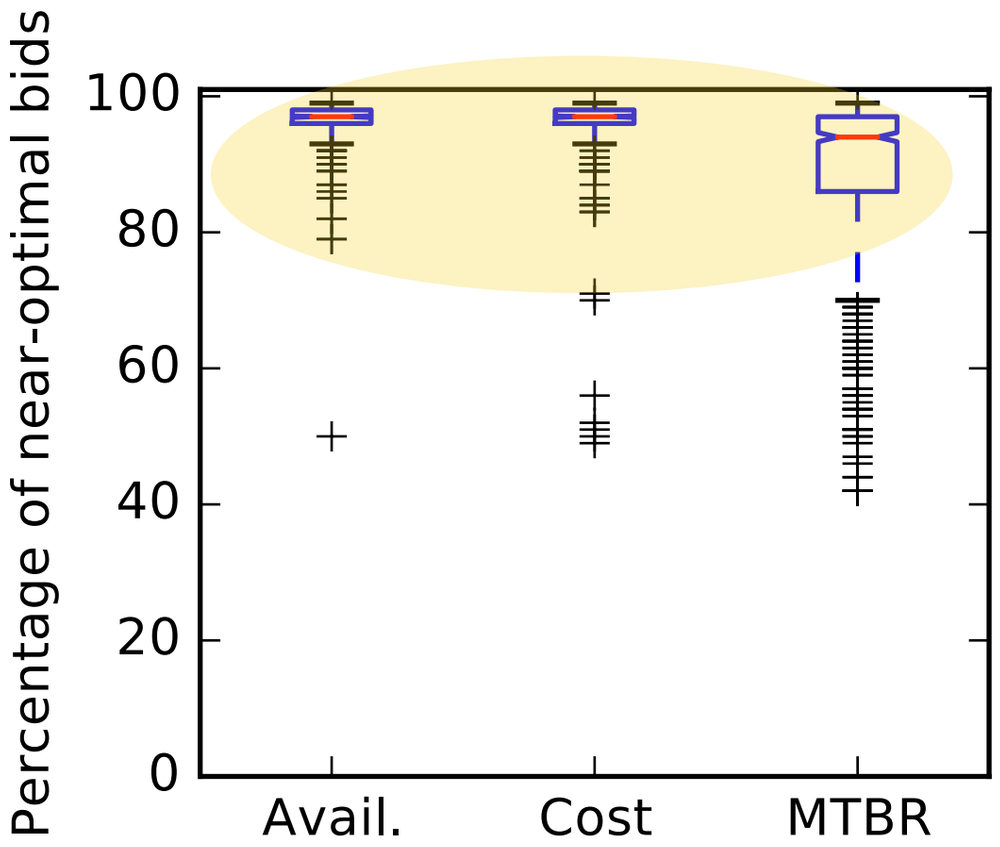
Impact of Bidding

- Availability, Cost, MTBR not particularly sensitive to bidding
- Low-cost, highly available spot servers for wide range of bids

Do we need sophisticated bidding strategies?

Analyzing 1500 markets

- Percentage of all bids that yield availability, cost, MTBR that are 10% within the optimal

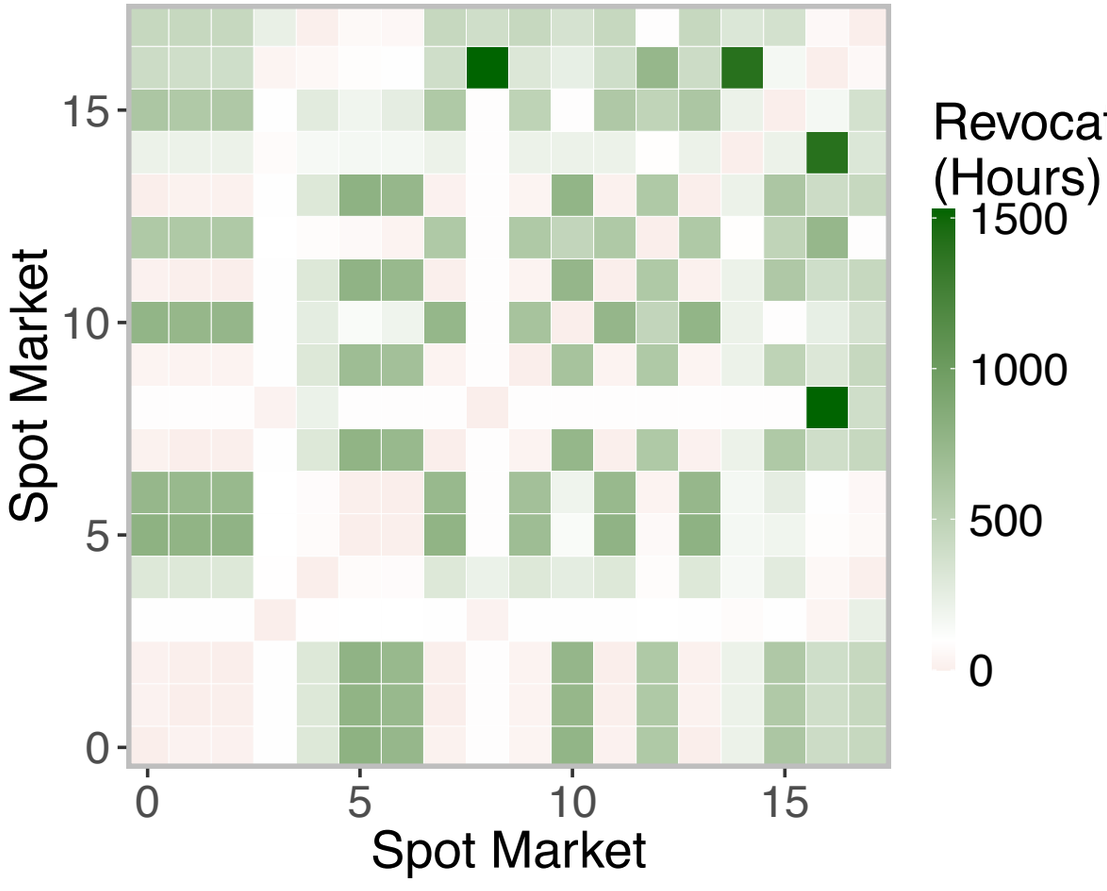


- 90% of all bids yield availability, cost, and MTBR that are “near optimal”
- Vast majority of 1500 markets have long tails

- **In the current spot markets, bidding has negligible impact**
- **Different bidding strategies yield same practical end-result**

Beyond Bidding

- **Look beyond bidding and focus on systems problems**
- Simple strategy: Bid the on-demand price, migrate when revoked
  - Requires efficient migration and checkpointing
- Avoid simultaneous revocations by using multiple markets
  - Revocation gap: time difference between revocations in two markets



- Many markets have large revocation gaps (>24 hours)
- “Independent” failures
- Distribute applications, migrate to uncorrelated markets

Conclusion

- Spot instances : auction based pricing
- Empirically study effect of bidding on cost, availability, and failure-rates
- Large range of bids have same effect ⇒ bidding is not crucial
- Sophisticated bidding strategies do not outperform simple ones
- Simple bidding strategies and using mutually uncorrelated markets : easier and practical alternative
- Beyond bidding: fault-tolerance and market selection

Thank You

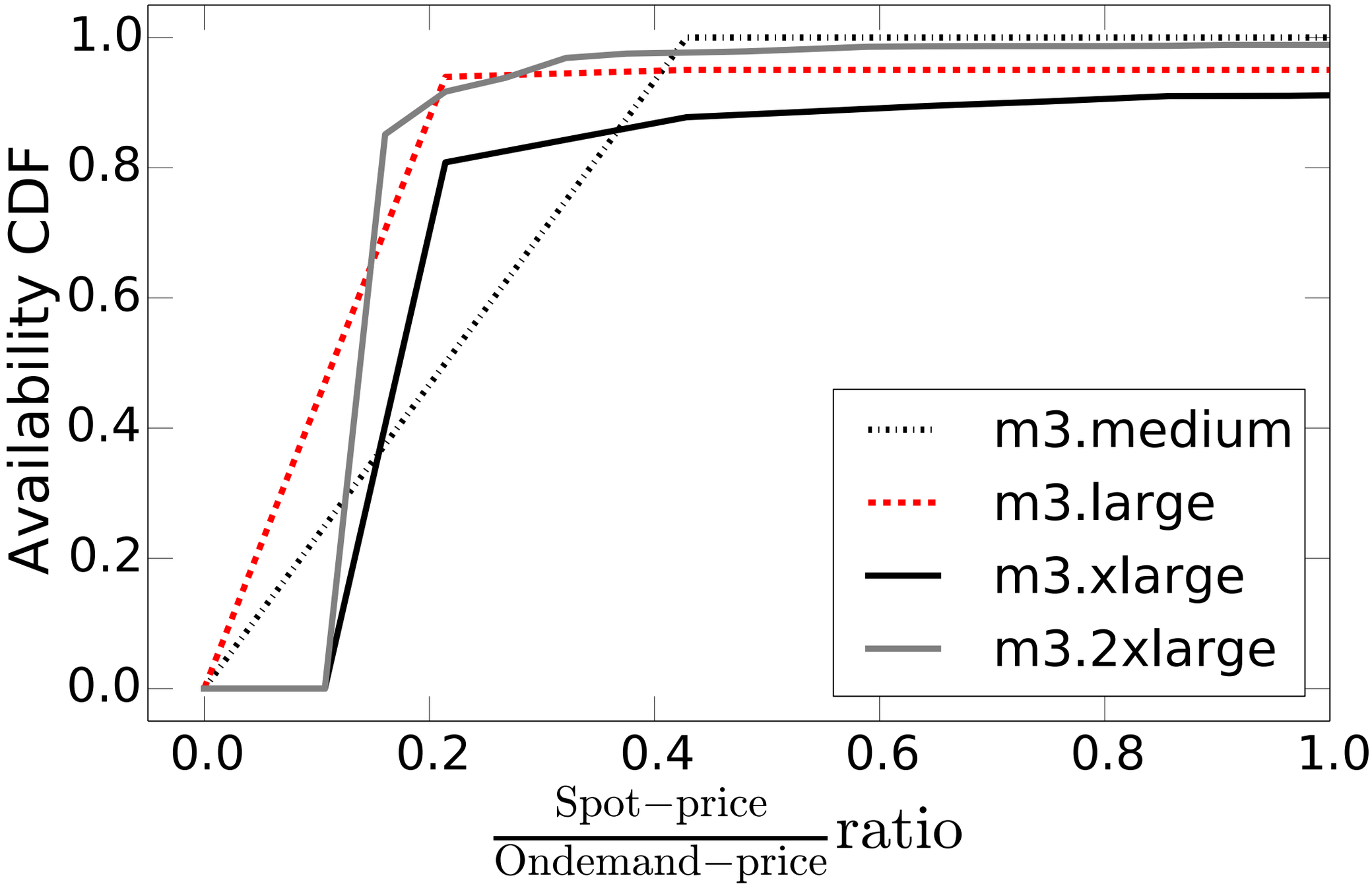
prateeks@cs.umass.edu

Why bidding strategies are not crucial

- Wide range of optimal bids
- Resources always available
- No penalty for high bids

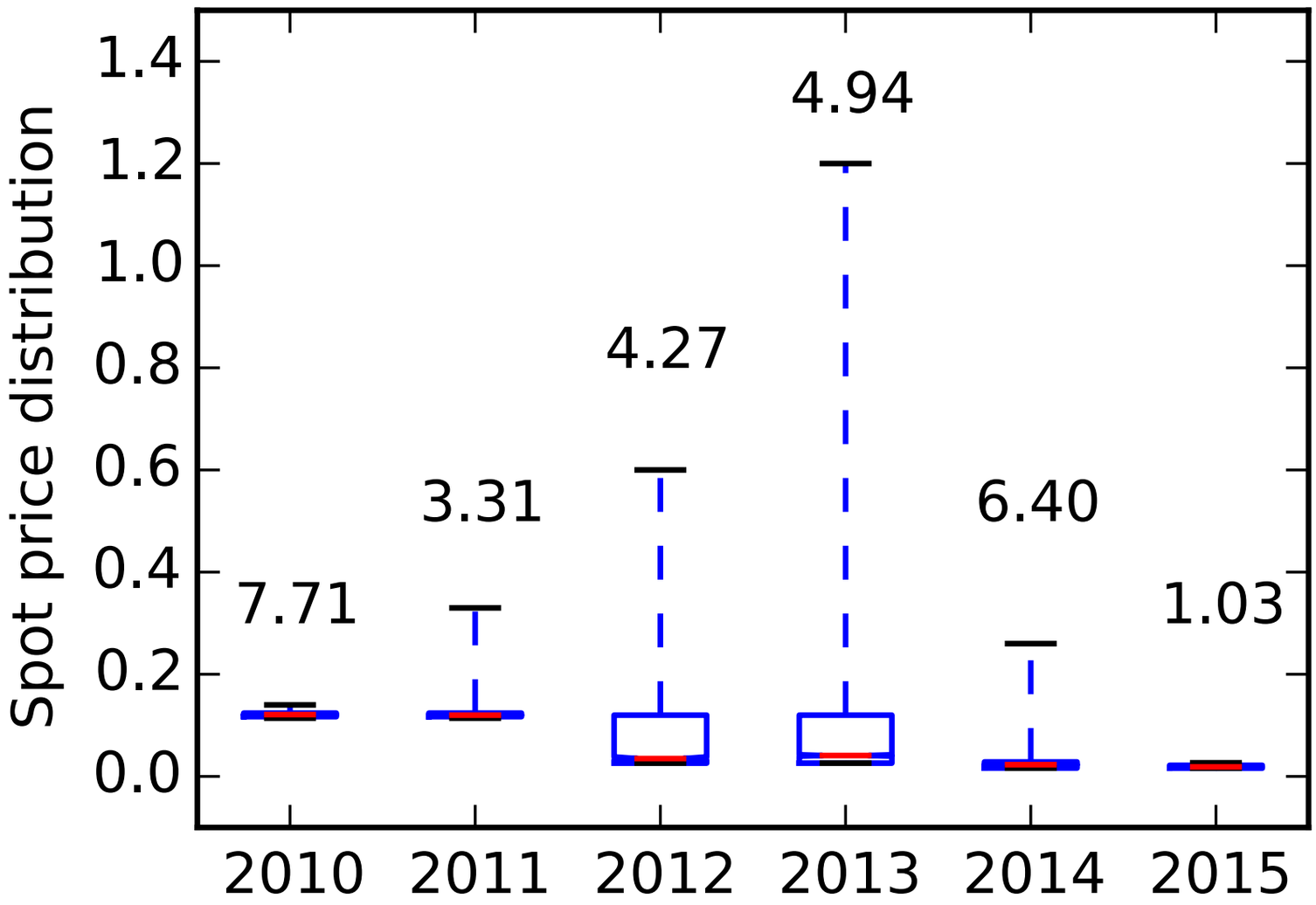
When will bidding be relevant?

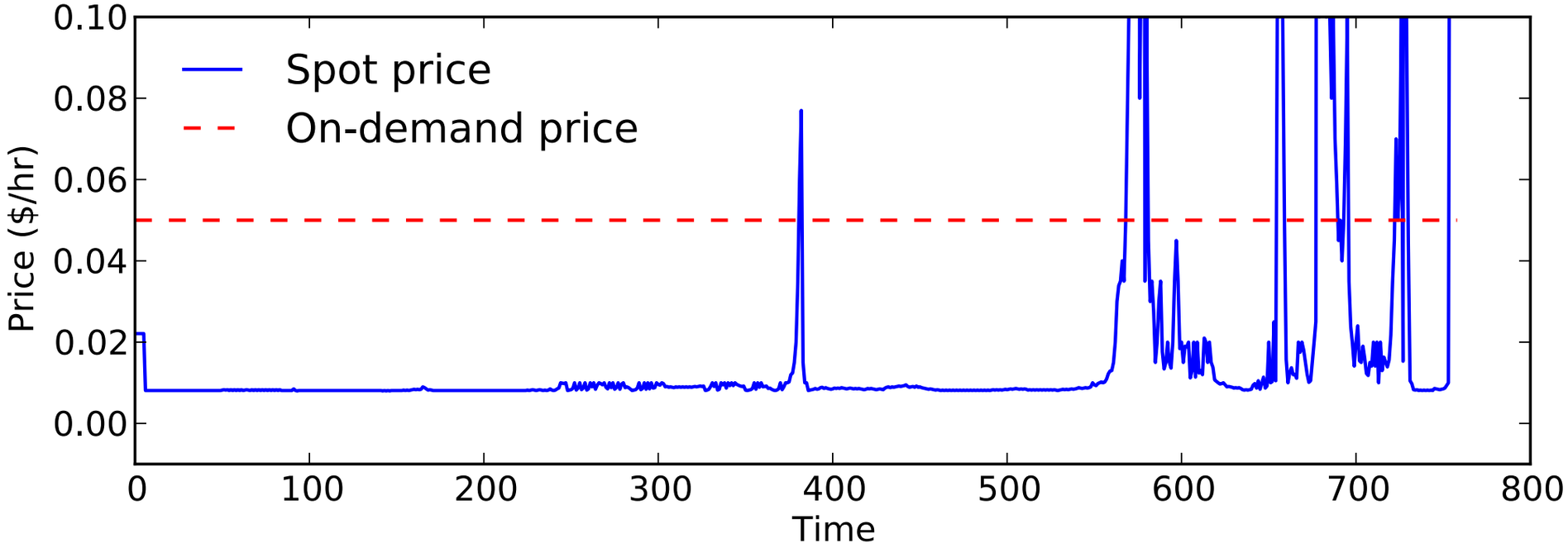
- Availability, cost CDFs not long tailed
- More penalty for bidding too high
- Higher market volatility
- Users and systems exploiting arbitraging opportunities
- Still need systems to handle the transiency gracefully

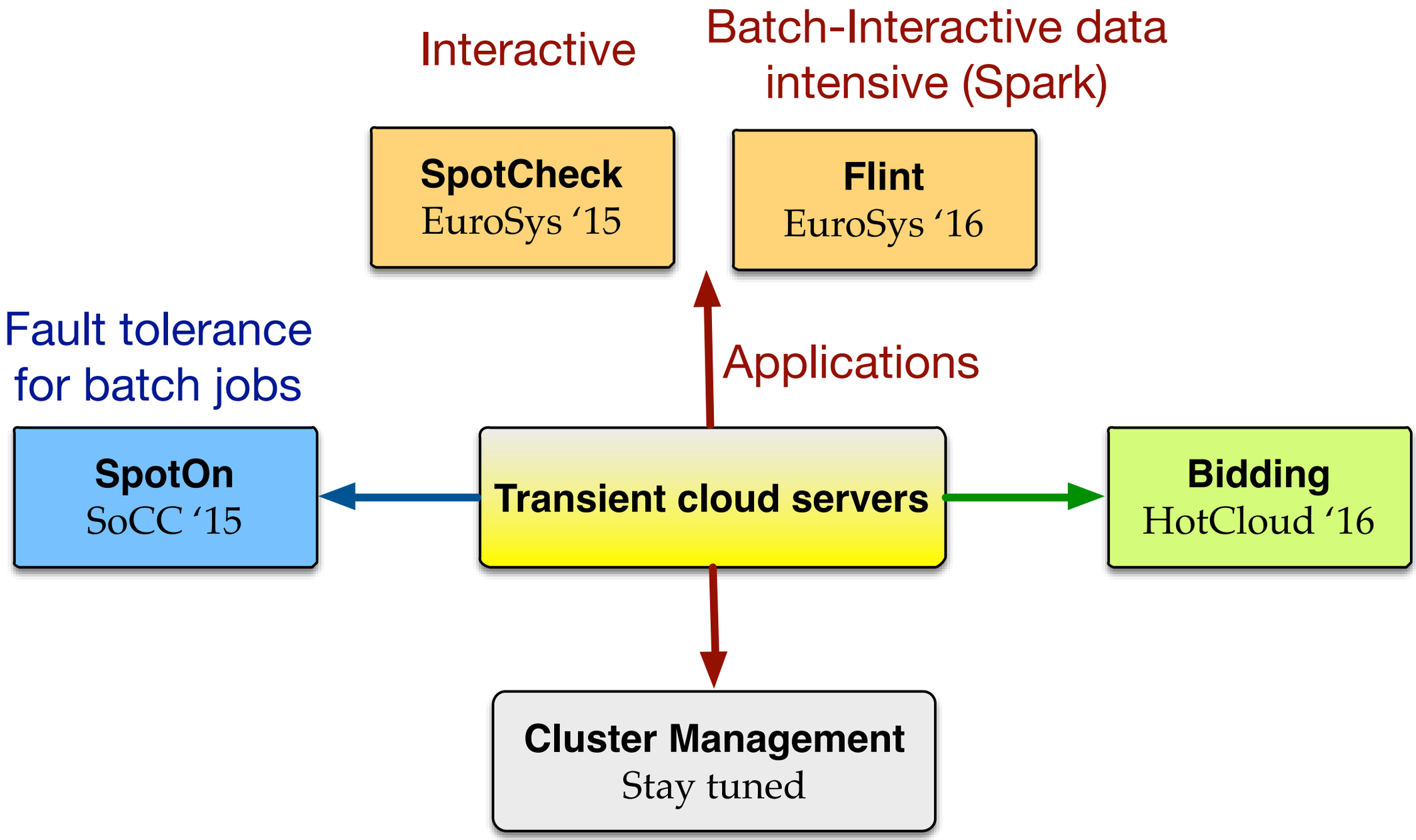

Availability CDF
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
m3.medium
m3.large
m3.xlarge
m3.2xlarge
Spot−price / Ondemand−price ratio

Spot market volatility over the years

- m1.large price range and skewness




Spot price
On-demand price
Price ($/hr)
0.10
0.08
0.06
0.04
0.02
0.00
0
100
200
300
400
500
600
700
800
Time


Interactive
Batch-Interactive data intensive (Spark)
SpotCheck
EuroSys '15
Flint
EuroSys '16
Applications
Fault tolerance for batch jobs
SpotOn
SoCC '15
Transient cloud servers
Bidding
HotCloud '16
Cluster Management
Stay tuned