

Index-based Trading in Cloud Spot Markets

Supreeth Shastri and David Irwin

UMassAmherst

Idle Cloud is Provider's Workshop



10-50%

typical utilization in large datacenters

[2013] The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines.

76 * 8

Hardware types

Contracts

IaaS is evolving into a marketplace

On-demand, Reserved (1 or 3), Spot, Spot-block, Burstable, Dedicated or Shared

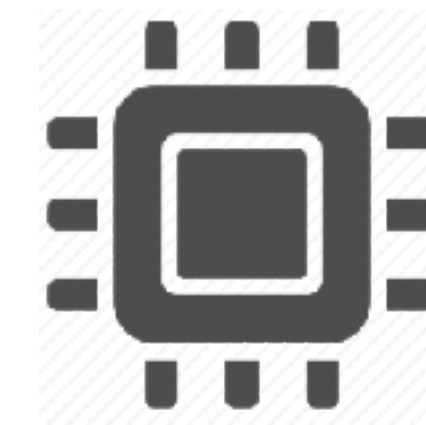
Selling Idle Cloud Capacity



Users bid in a 2nd price auction



EC2 continually evaluates supply-demand to price spot servers

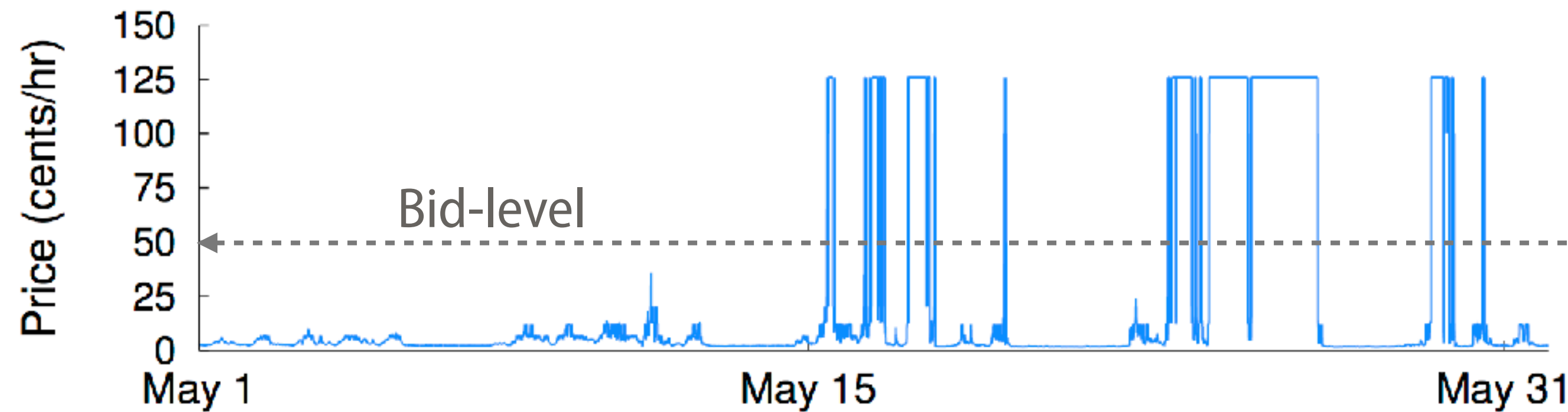


Allocate: bid price \geq spot price
Revoke: bid price $<$ spot price

“ On average, AWS customers are using more compute capacity on spot instances than across all of EC2 in 2012 ”

<https://aws.amazon.com/10year/>

Spot Price Prediction



Characterized by **spikes** that are hard to predict

Accurate Prediction

- Reduces disruptive revocations
- Helps compare different servers

Research

SIGCOMM 2015
HotCloud 2016
HPC 2016
IC2E 2016
ICDCS 2016
ICPE 2017
SIGMETRICS 2017

Startups

SpotInst
Batchly
ClusterK

Accurately
Predicting Spot Prices is Difficult

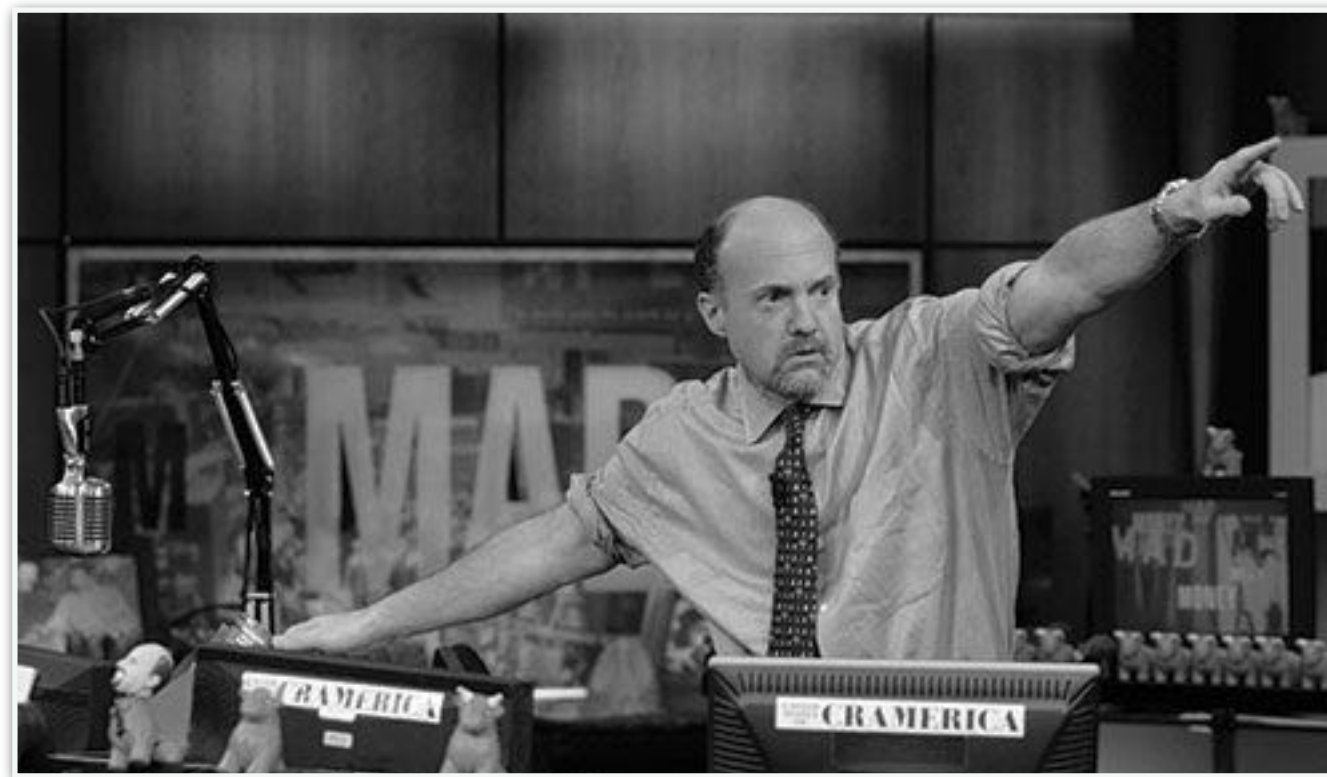
$$\begin{matrix} 68 & \cdot & 2 & \cdot & 2-5 & \cdot & 14 & \cdot & 2 & = & 7600^+ \\ \text{Hardware} & & \text{OS} & & \text{Zones} & & \text{Regions} & & \text{Time} & & \text{worldwide markets} \\ \text{config} & & \text{types} & & \text{(datacenters)} & & \text{(country, state)} & & \text{commitments} & & \end{matrix}$$

**One size fits all model
is unlikely**

**No visibility into
market internals**

**Limited correlation
with external variables**

Market-based Index *for CLOUD*



| U.S. FUTURES | | |
|----------------------------------|---|---------------|
| DOW JONES INDUS FUT 3/17 (D./H7) | | |
| 19,904.00 | ▲ | 60.00 (0.30%) |
| S&P 500 3/17 (ES/H7) | | |
| 2,280.75 | ▲ | 6.25 (0.27%) |
| NASDAQ 100 3/17 (NQ/H7) | | |
| 5,116.50 | ▲ | 20.20 (0.40%) |

Rather than focusing **exclusively** on predicting **individual servers**, cloud users should make decisions, in part, based on **broader market indices**

Market-based Index *for CLOUD*

Characterizing an individual server **i**

Price = P_i

Memory = M_i GB

Compute = C_i ECUs

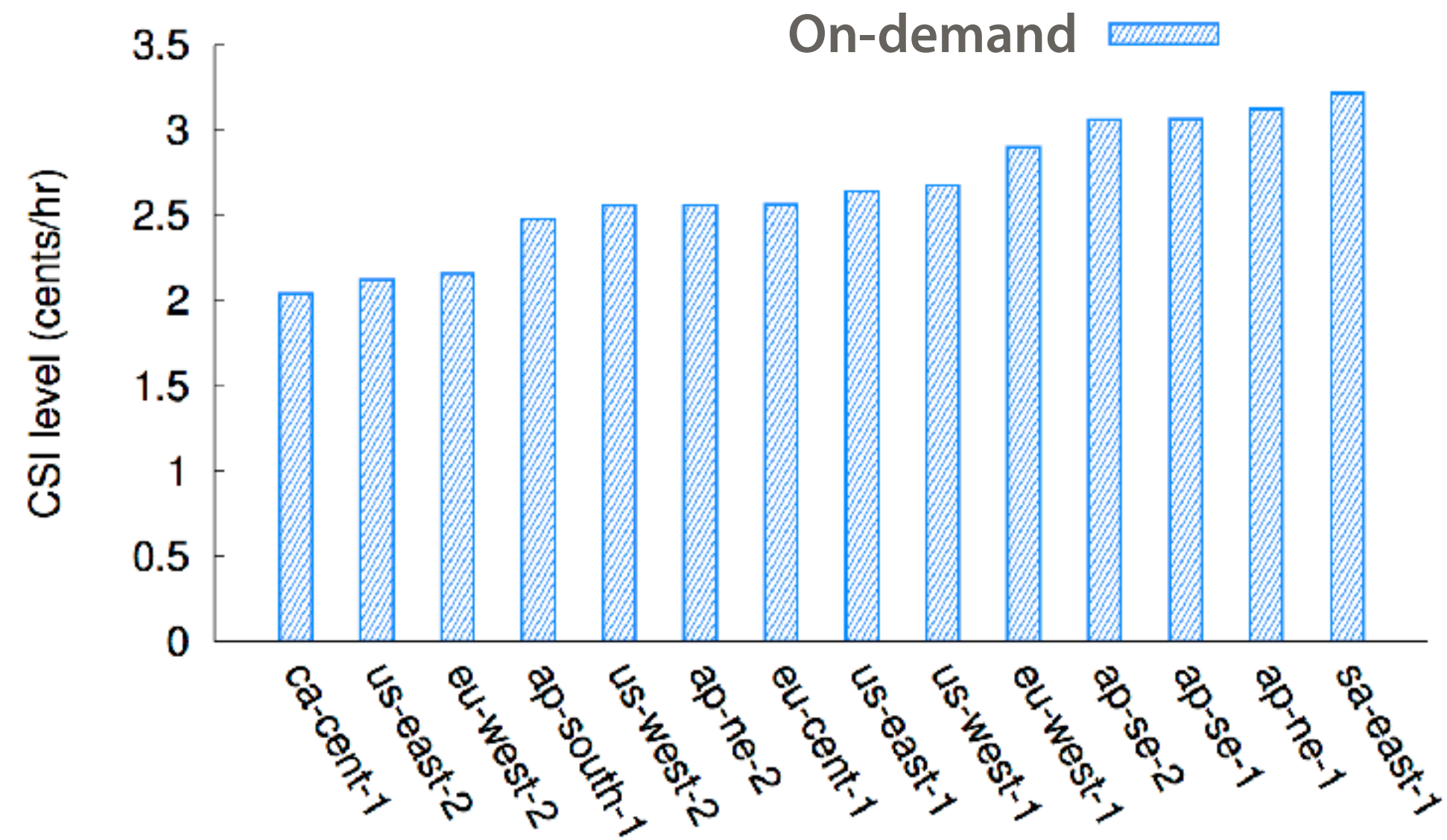
$$P_i^{\text{norm}} = \frac{P_i}{\sqrt{(C_i * M_i)}}$$

Characterizing a group of servers

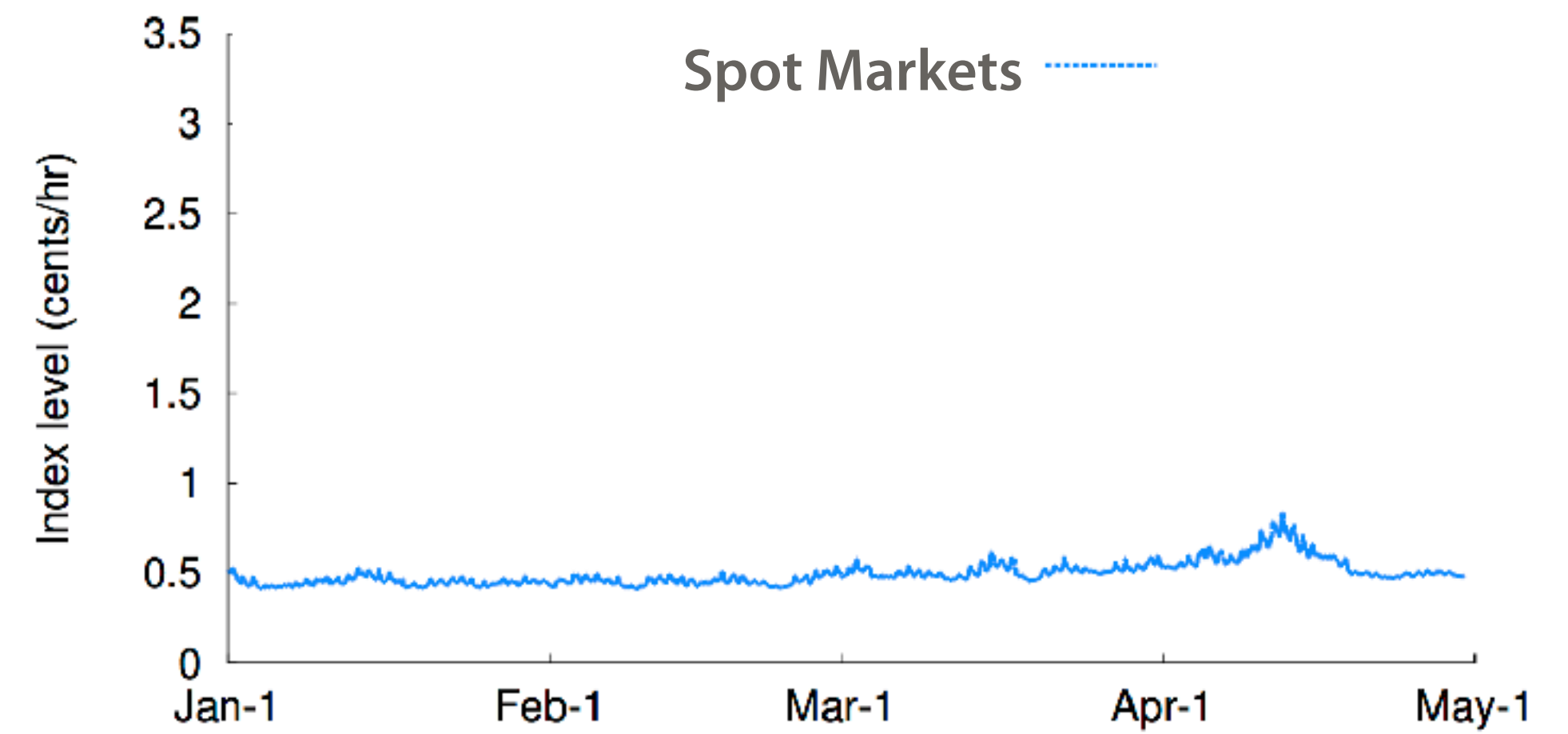
Average of normalized prices

$$\text{Index-level} = \frac{\sum_{i=1}^N P_i^{\text{norm}}}{N}$$

Market Indices at Global Level

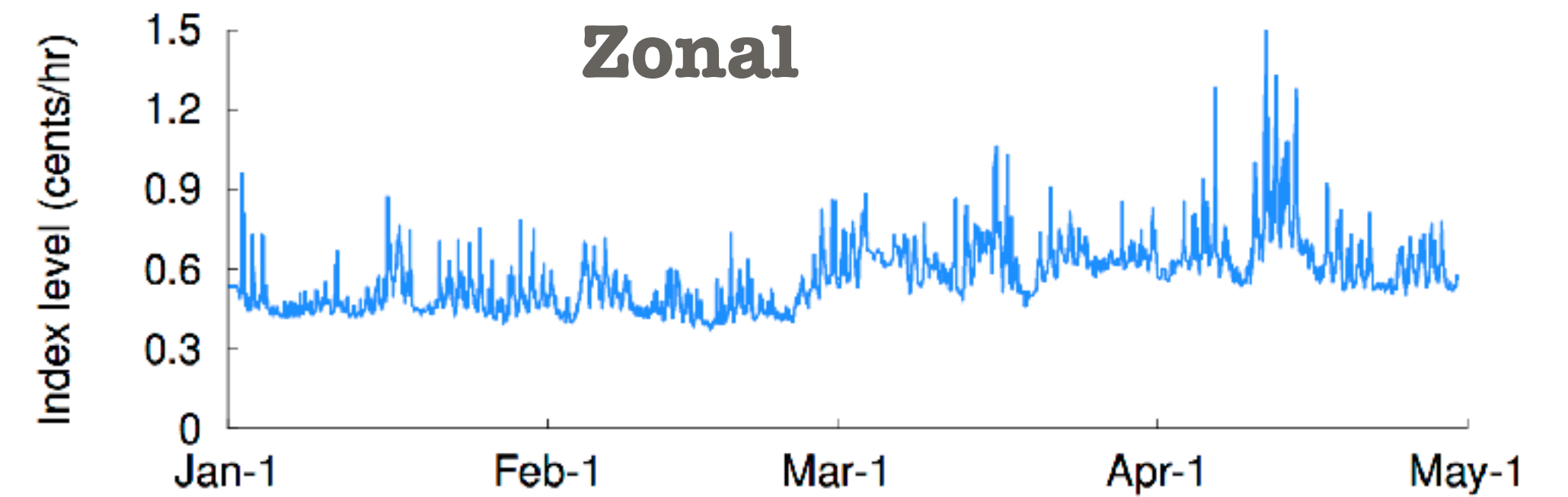
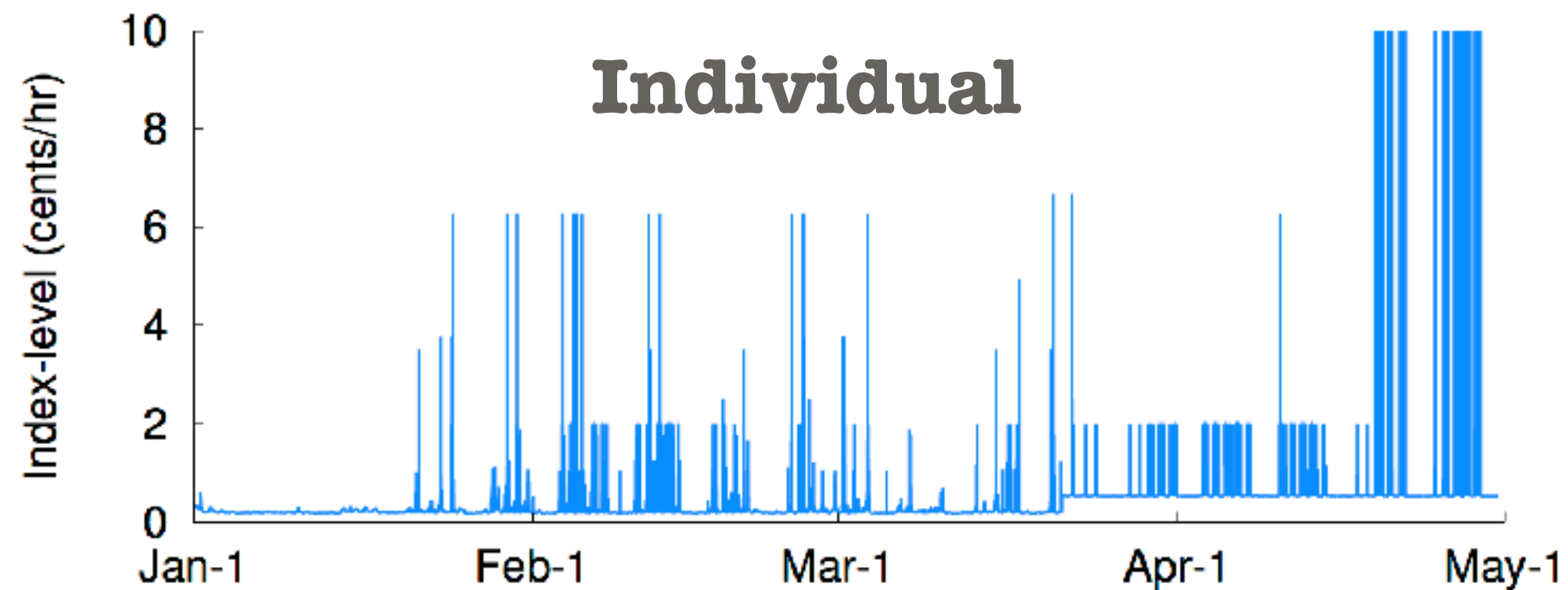
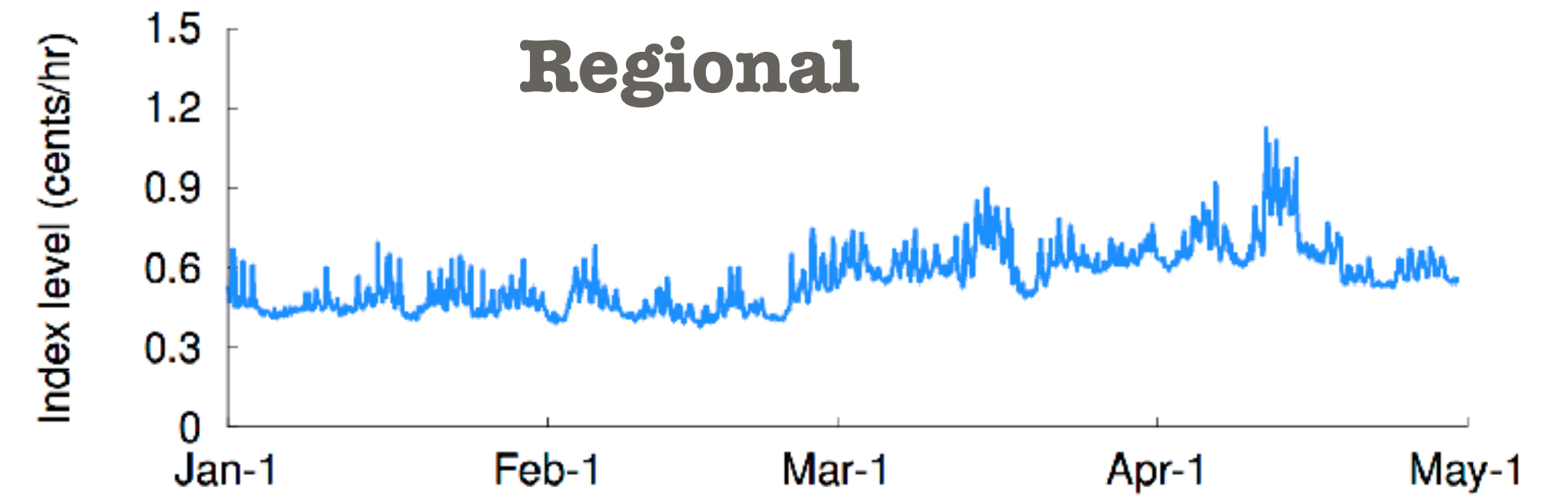
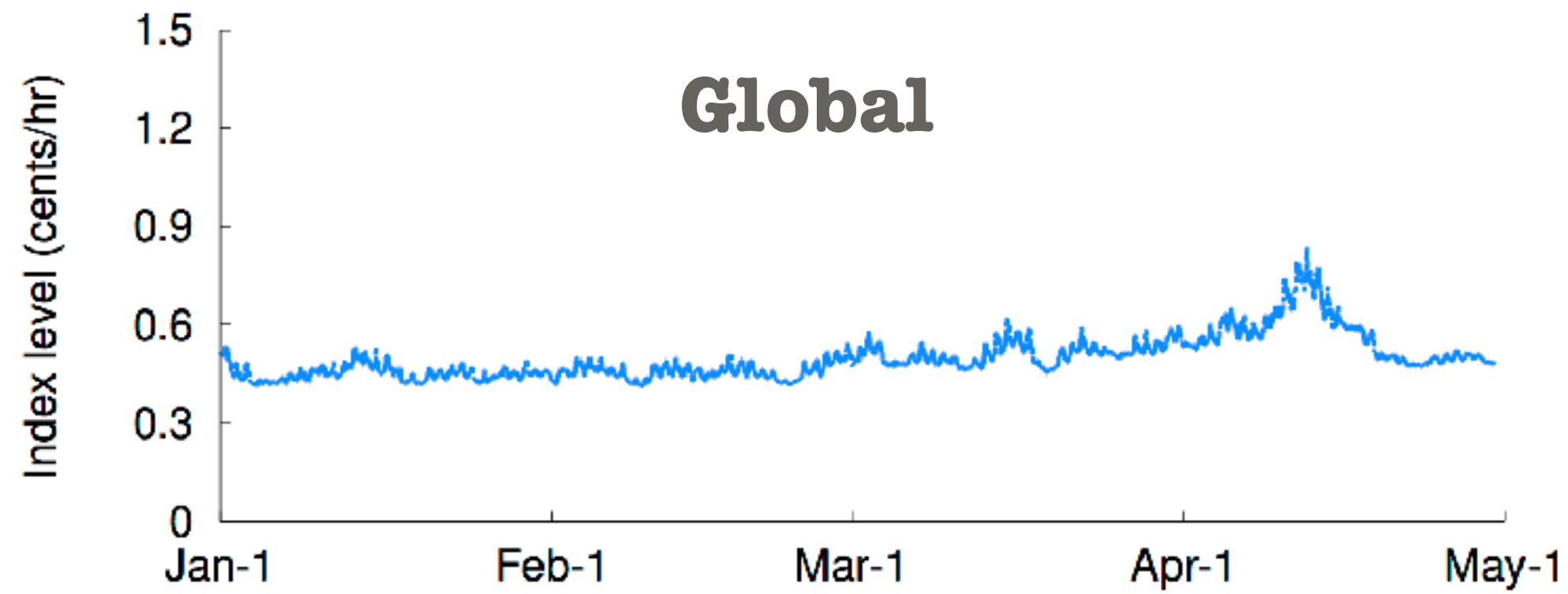


Compute-time is 50% more expensive
in Brazil than Canada



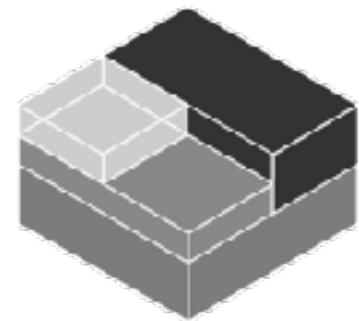
Worldwide spot market is remarkably stable
with ~80% discount from on-demand avg

Indices at Different Market Granularities

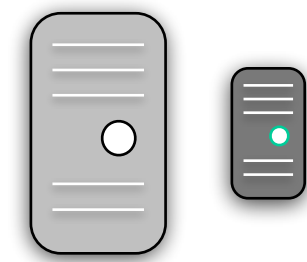


Price prediction is easier and more accurate at higher **market-level** than **individual** server level

Flexible Applications



Containerized



Scalable



No geographical
constraints

can benefit from Server Trading

- Spot Markets prices are *dynamic*
- Many *price inversions* exist
- Provider always “*buys*” back servers

... but Trading incurs Transaction Cost

Memory state and disk migration, Unused server time, Fault-tolerance overhead

Choosing the Best Server

$$\text{Sharpe ratio} = \frac{E [R_i - R_{\text{free}}]}{\sigma_i}$$

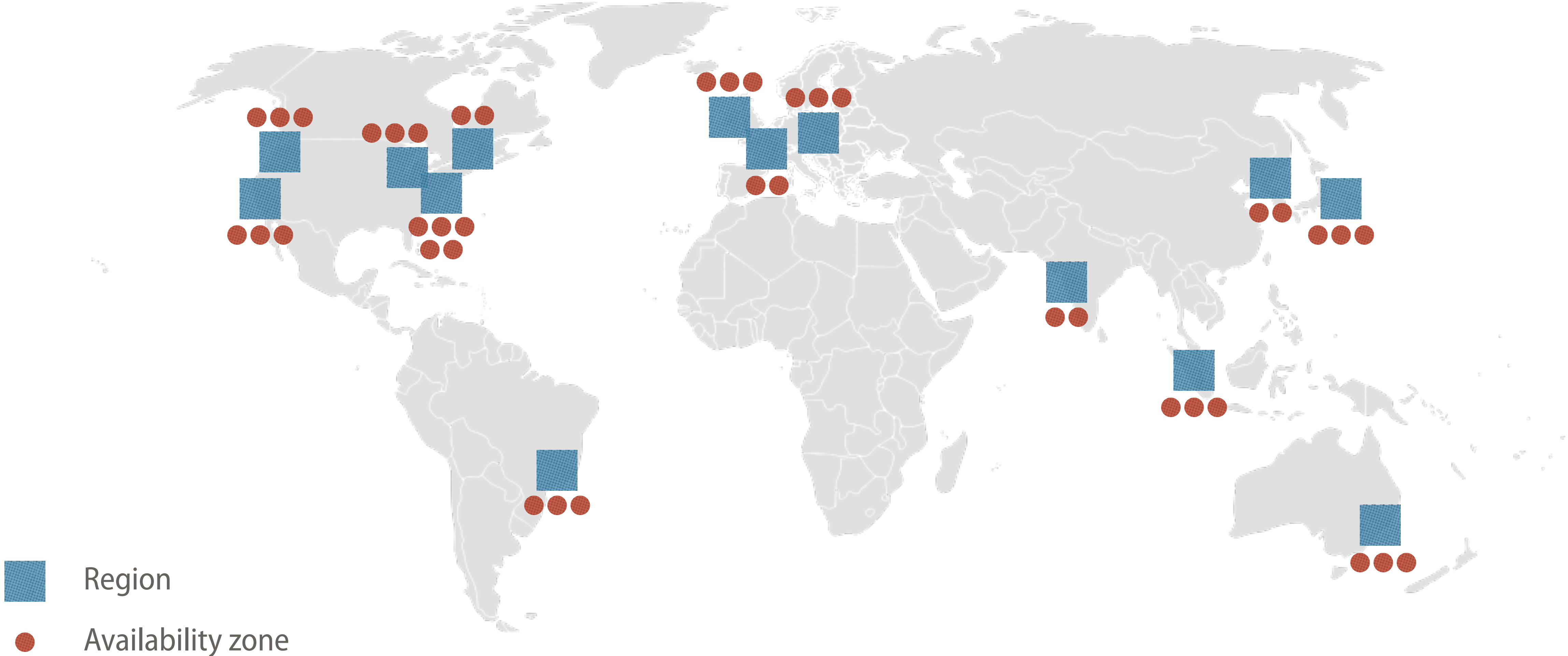
$R_i = \text{Asset's return}$

$R_{\text{free}} = \text{Risk free return}$

$\sigma_i = \text{Std. deviation of returns}$

Chooses the server that has not only low price but also low volatility

Scope of Server Trading in EC2

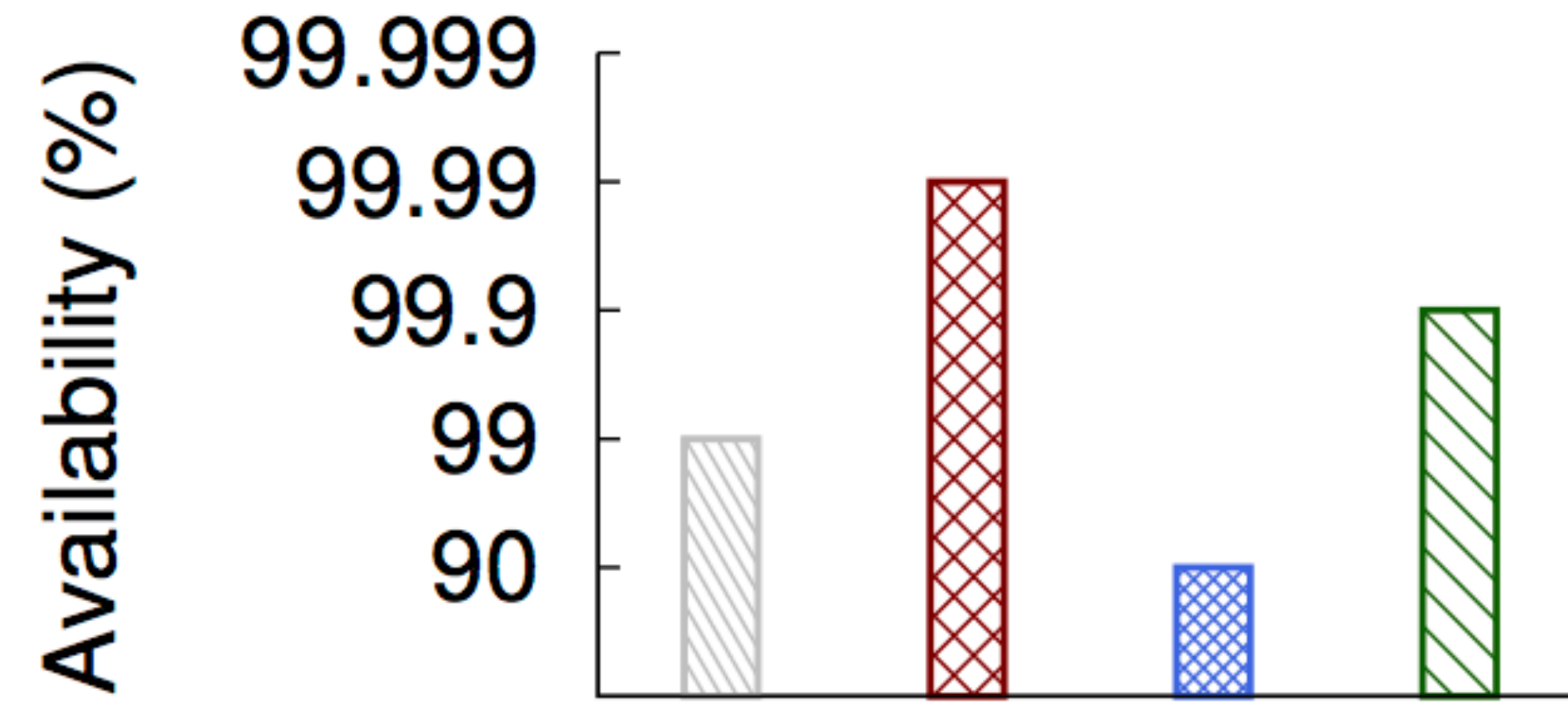
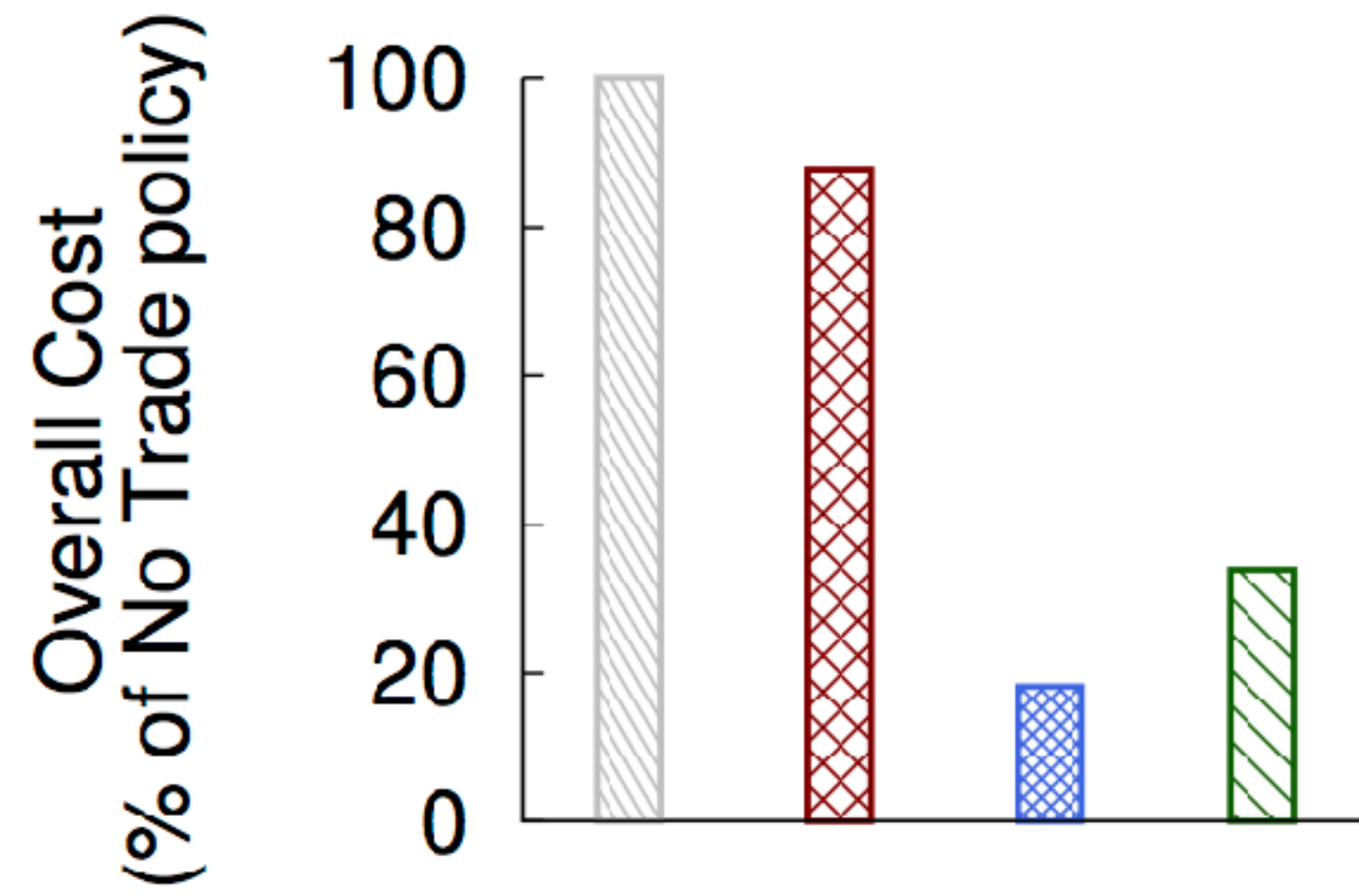






World map: https://commons.wikimedia.org/wiki/File:BlankMap-World_gray.svg

Server Trading Policies

| Policy | Server Choice | Trading | Trading Cost |
|--------------------------------|---|-------------------|------------------------|
| Market-based No Trading | Globally best server | No | 0 |
| Market-based Local Trading | Globally best server | Within the zone | Fixed (120s) |
| Market-based Global Trading | Globally best server | Anywhere globally | Proportional (1-4m/GB) |
| Index-based Global Trading | Globally best zone, then locally best server | Within the zone | Fixed (120s) |

Evaluation



- Market-based No Trading 
- Market-based Local Trading 
- Market-based Global Trading 
- Index-based Global Trading 

To Conclude...

Spot price prediction is an active research topic

Prior works have focused on individual servers, we introduce **market-based indices**

Flexible applications can trade servers

We demonstrate trading based on market-based achieve **best cost-performance tradeoff**

Future work

Defining **application-specific indices**

Using indices for **benchmarking** spot-based systems

Thank you!

Supreeth Shastri

shastri@umass.edu

UMassAmherst