Dynamic Resource Allocation for Spot Markets in Clouds

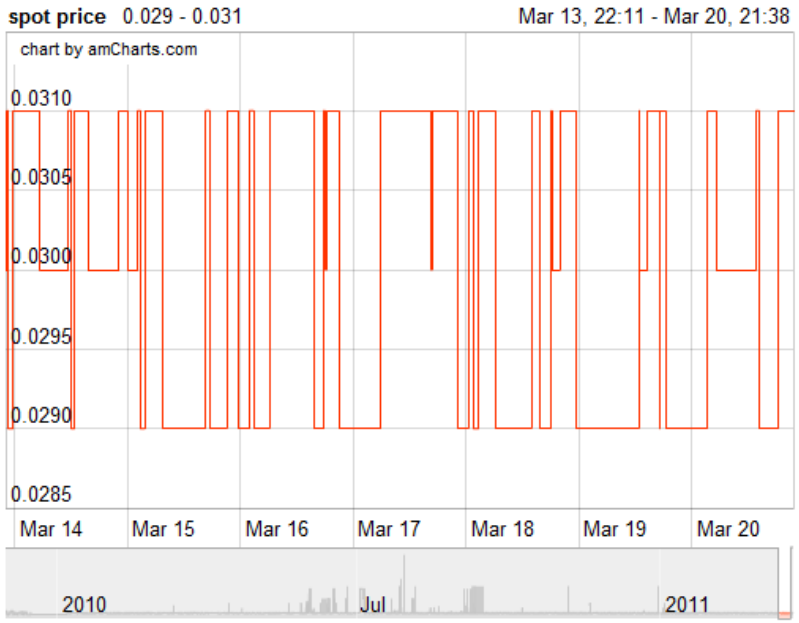
Qi Zhang, Eren Gurses, Jin Xiao, Raouf Boutaba

Introduction

- Cloud computing aims at providing resources to customers in an on-demand manner
  - A customer can purchase resources dynamically based on the current needs
- Typically, cloud providers employ usage-based pricing
  - A fixed unit price is specified for each type of VM offerings
- However, fixed pricing schemes lack incentives to encourage desirable customer behavior
  - Low demand results in poor resource utilization
  - high demand leads to revenue loss and customer dissatisfaction
- Market-based resource allocation is gaining popularity
  - Let the price fluctuates with supply and demand

Amazon EC2 Spot Instance Service

- Launched on Dec. 15, 2009
- Multiple VM types per availability zone
- Customers submit requests with bidding prices
- Spot price fluctuates with supply and demand
- Instances may be terminated with prior notice



Price of a single m1.small linux instance in US-East-1 between Mar. 14- Mar. 20, 2011

Motivation

- Multiple spot markets sharing the same resource pool
- As request arrival can be highly volatile, sometimes certain markets may be "hotter" than others
- A static allocation strategy can lead to situations where markets are over-supplied or under-supplied
  - Over-supplying a market causes poor resource utilization
  - Under-supplying a market leads to low income and customer dissatisfaction
- How to dynamically allocate resources to spot markets?

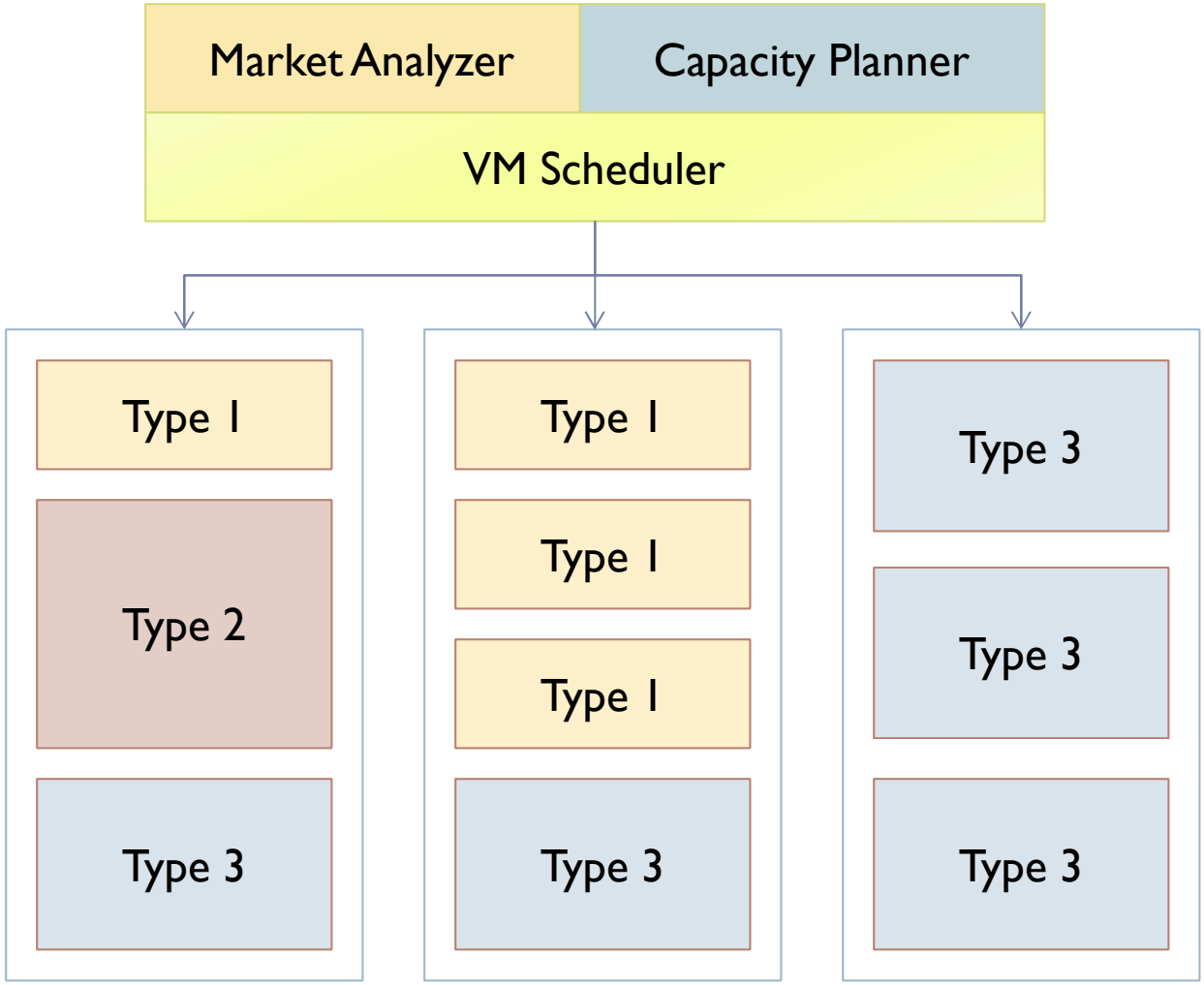
Contribution

- We propose a framework that dynamically adjust supply of spot markets to maximize total revenue

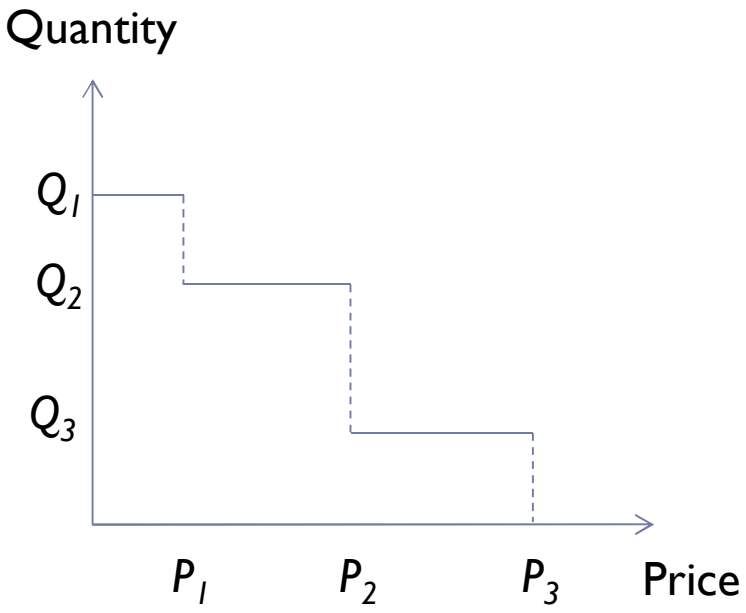
- Challenges
  - Need to predict future demand for every spot market
  - Need to determine the allocation strategy that optimizes revenue

- Our solution
  - Predicting future demand using an autoregressive (AR) model
  - Compute expected spot price and allocation for each market to maximize total revenue
  - Schedule VMs according to expected price

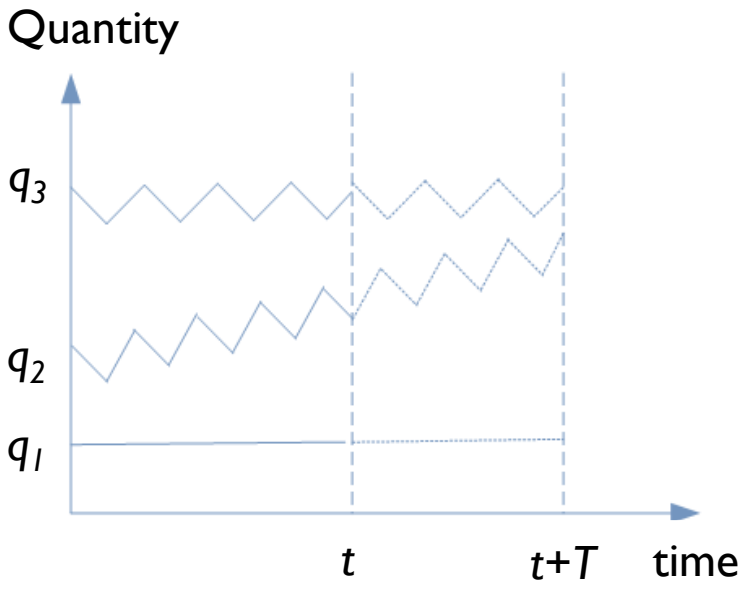
System Architecture



Demand Prediction



Demand curve



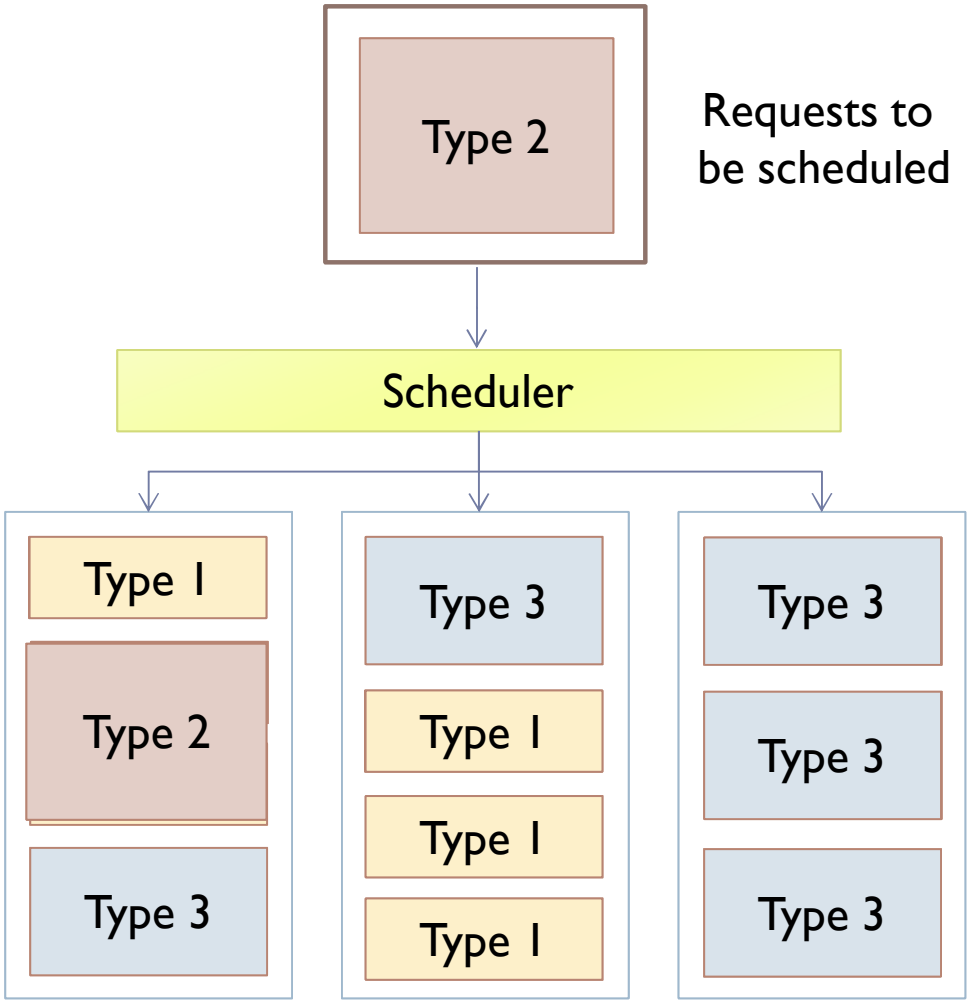
Predicting future demand curve

- Demand curve can be modeled as a non-increasing, piecewise-linear function
- Predicting future demand curve using autoregressive (AR) functions

Computing Expected Allocation

- Goal: determine the expected price and allocation of resources to spot markets to maximize total revenue
- Simple case: Prices are fixed
  - This problem is a variant of the NP-hard multiple knapsack problem (MKP)
- Real case: Prices are not fixed
  - Much harder than MKP, as objective function is non-linear
  - By approximating the revenue using a concave function, the problem can be reduced to a MKP

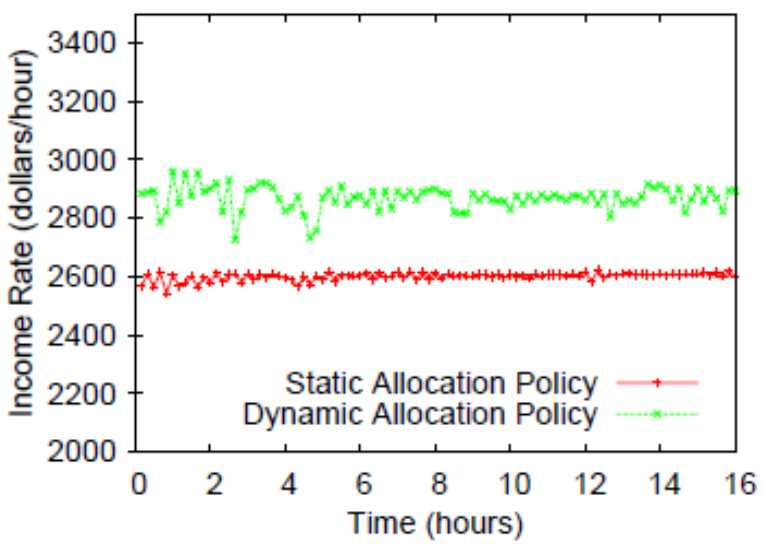
Scheduling Algorithm

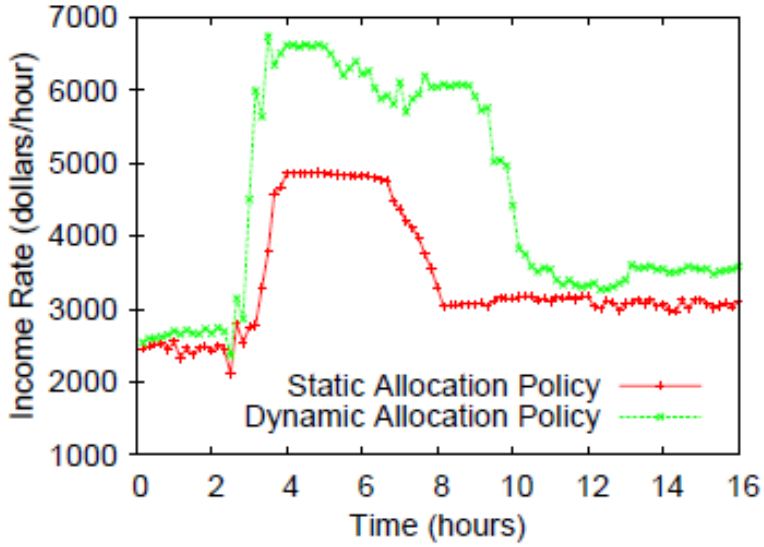


Experiment Setup

- Implemented the scheduler using CloudSim
  - Modified default resource allocation policies
- Workload
  - Non-homogenous poisson process with artificial high and low arrival periods
  - Bidding price and running time are generated from normal distributions
- Scheduling policies
  - Static allocation for each individual market
  - Our dynamic allocation scheme

Experiments





Policy	Metric	Income	Revenue Loss	Net Income
Static	Mean	67030.44	399.01	66631.43
	Std.	13573.32	172.45	13400.87
Dynamic	Mean	78026.33	3398.36	74627.97
	Std.	15173.28	1083.63	14089.65

- Dynamic Resource Allocation for Spot Markets in Clouds

Conclusion

- Market-based resource allocation mechanisms provide economic incentives to encourage desirable customer behavior

- We have presented a framework that dynamically adjust supply for different spot markets, with the goal of maximizing total revenue
  - Practical and applicable for any market-based cloud environment that uses uniform price scheme

Thanks!

Questions?