Efficient VM Placement with Multiple Deterministic and Stochastic Resources in Data Centers

Hao Jin[1], Deng Pan[2], Jing Xu[2] and Niki Pissinou[2]

[1]Department of Electrical and Computer Engineering, [2]School of Computing and Information Science

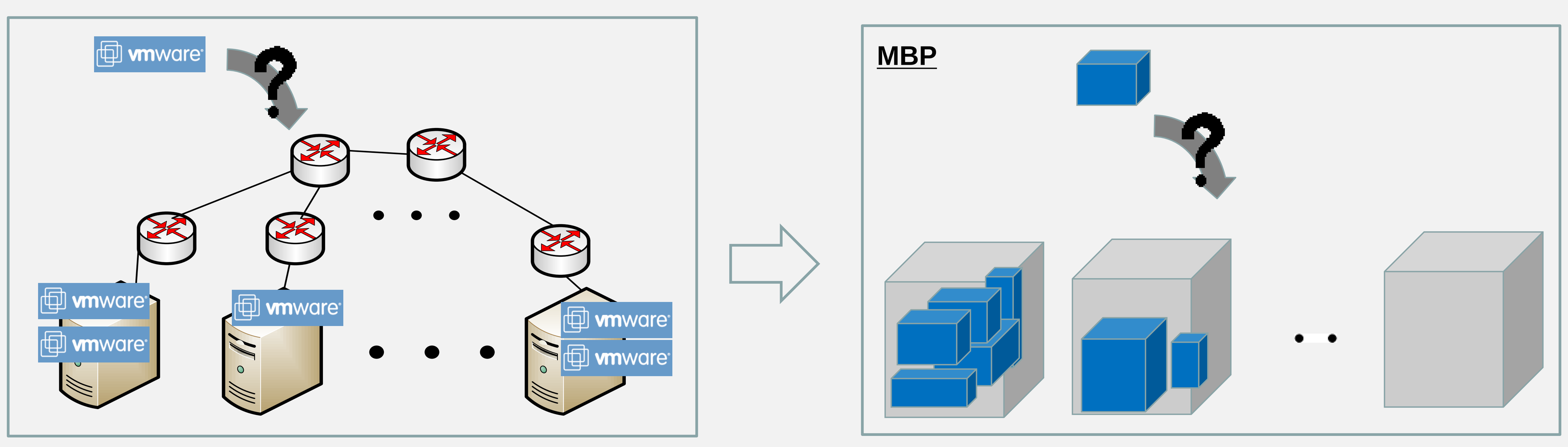
Florida International University

FIU FLORIDA INTERNATIONAL UNIVERSITY

National Science Foundation

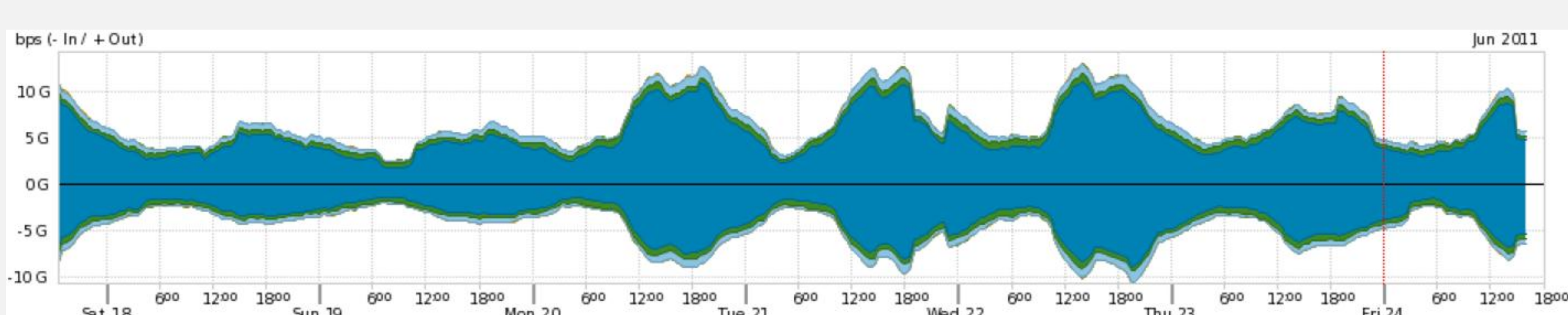
Introduction

- Virtual machine (VM) improves data center's efficiency and flexibility.
- VM placement needs to consider multiple server resource demands.
  - Including memory, CPU, power consumption, etc.
  - Can be modeled as Multidimensional Bin Packing (MBP) problem
  - Assume fixed-value resource demands – deterministic demands



Stochastic Demand Challenge

- Some resource demands are bursty and time varying.
  - Such as network bandwidth demand
- Model such demand as stochastic process – stochastic demand
  - Service level agreement (SLA)
- Challenge: No fixed value can be used in placement calculation



Goal

- Effective VM placement to minimize the number of servers
  - Consider multiple resource demands
  - Consider both deterministic and stochastic resource demands
    - Need to satisfy SLA for stochastic demands

Problem Formulation

- Multidimensional Stochastic VM Placement (MSVP) problem
- Model demands by using $N(\mu(v_i), \sigma^2(v_i))$ normal distribution
- Calculate equivalent demand based on SLA to quantify the stochastic demands:

$$\text{Total Equivalent Demand} = \sum_{i \in U} \mu(v_i) + \Phi^{-1}(\alpha)\sqrt{\sum_{i \in U} \sigma^2(v_i)}$$

- $\Phi^{-1}(\alpha)$ is the quantile of standard normal distribution $N(1,0)$ when SLA is equal to α.
- Calculate server resource utilization:

Deterministic:

$$\forall p \in P, R_p(s) = \frac{\sum_{i \in U} D_p(v_i)}{C_p(s)}$$

Stochastic:

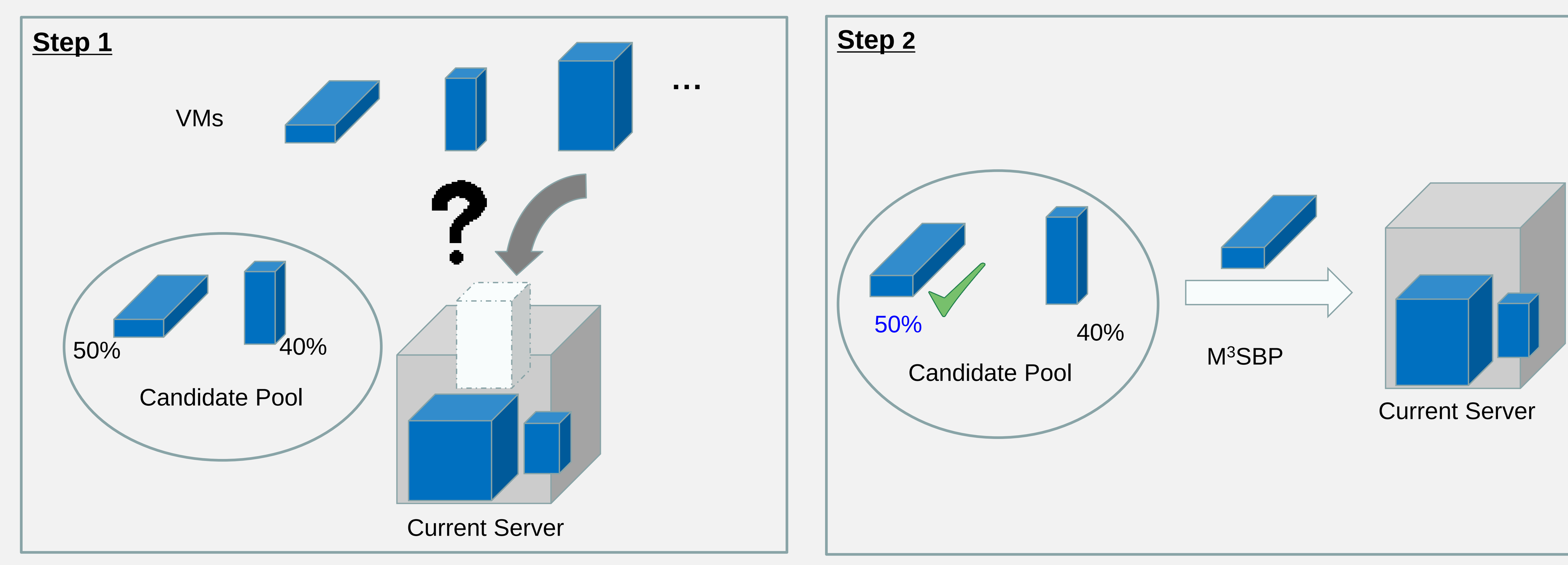
$$\forall q \in Q, R_q(s) = \frac{\sum_{i \in U} \mu_q(v_i) + \Phi^{-1}(\alpha)\sqrt{\sum_{i \in U} \sigma_q{}^2(v_i)}}{C_q(s)}$$

- MSVP formulation:

$$\begin{aligned} & \text{minimize } |S| \\ \text{s.t.} \quad & \forall p \in P, \forall s \in S, R_p(s) \le 1 \\ & \forall q \in Q, \forall s \in S, R_q(s) \le 1 \end{aligned}$$

Algorithm

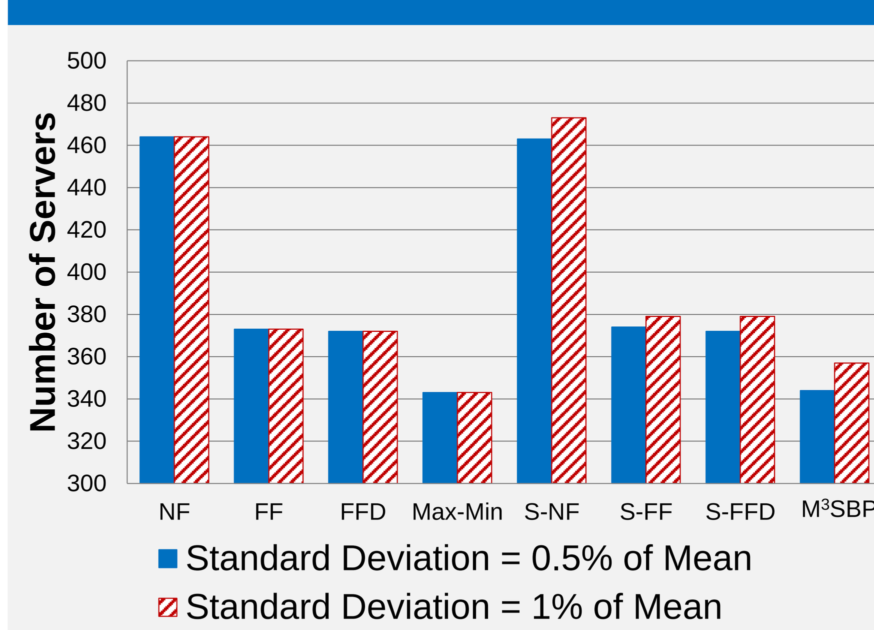
- Max-Min Multidimensional Stochastic Bin Packing (M^3SBP)
  - Step 1:
    - Find candidate VMs that can be potentially put in current server
    - Employ Total Equivalent Demand for stochastic resources
    - Record the minimum resource utilization
    - If no VM is found, open a new server.
  - Step 2:
    - Compare the minimum utilizations recorded in step 1
    - Choose the VM that maximizes the minimum utilization

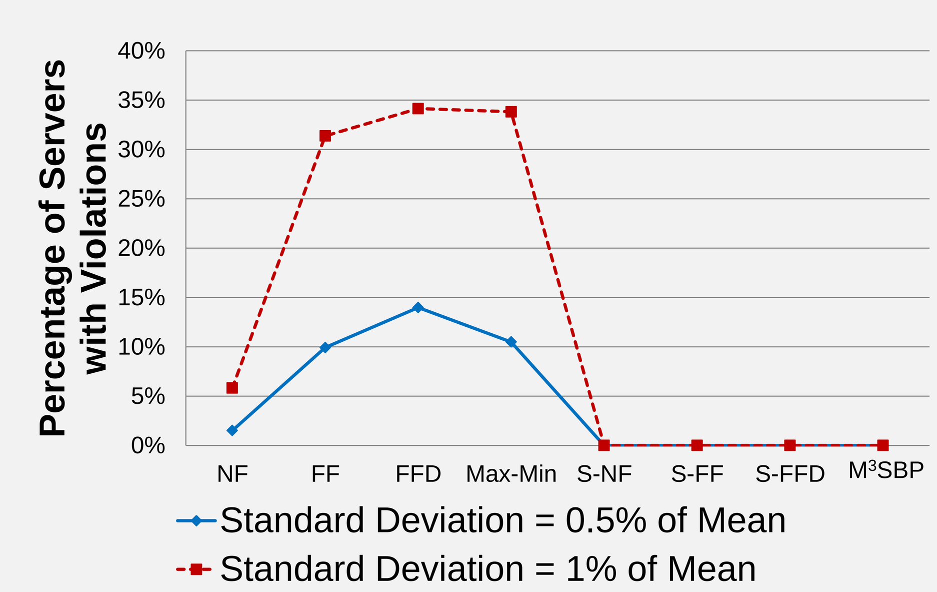


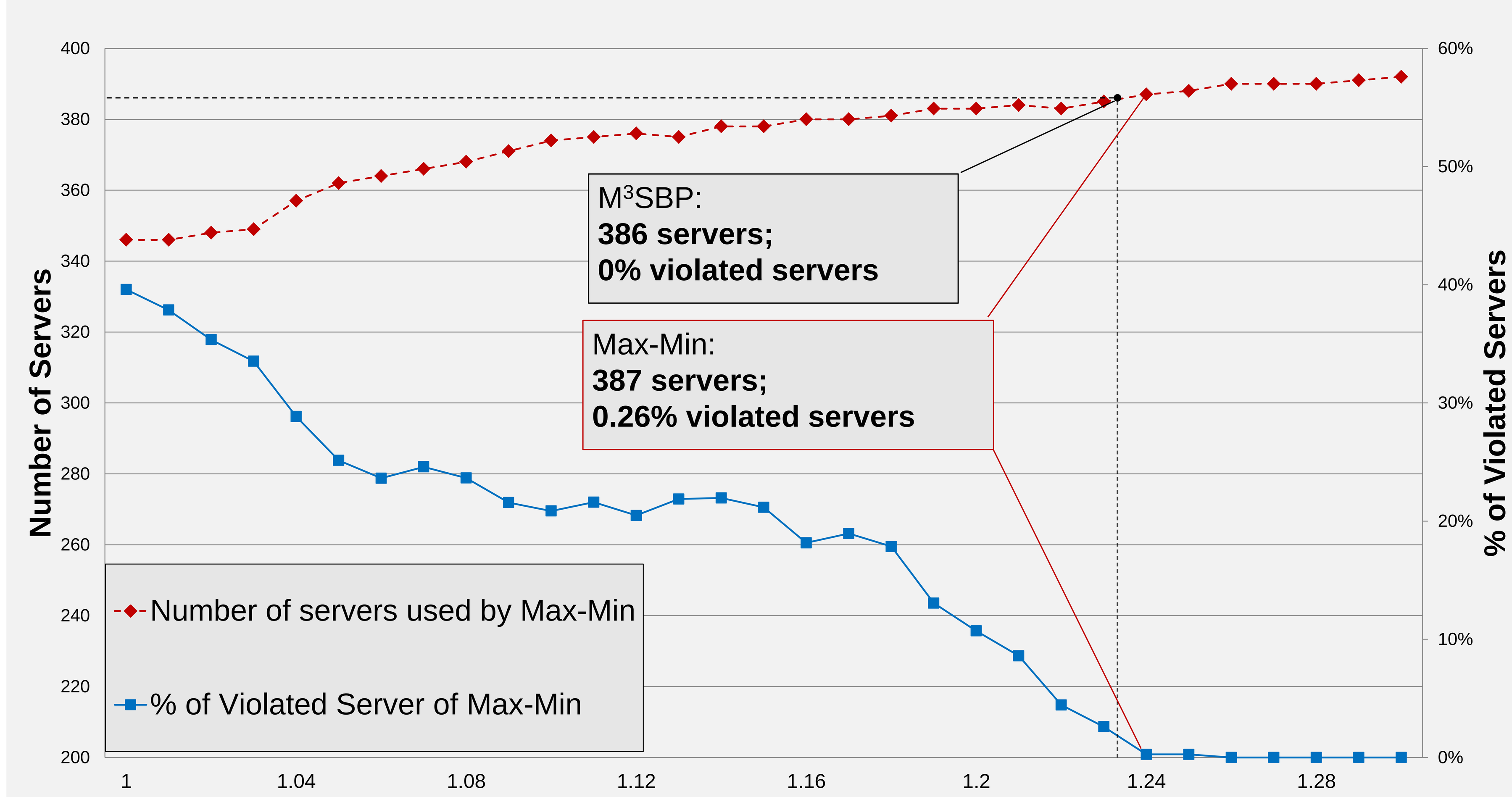
Performance Evaluation

- Simulation configurations
  - VMs with 4 different resource demands
    - Deterministic: memory, CPU and power consumption
    - Stochastic: bandwidth
      - Standard deviation to represent burst level
  - 2000 VMs and 99.99% SLA
- Compare M^3SBP with benchmark algorithms

Results







Conclusion

- M^3SBP – An effective VM placement algorithm
  - Calculate total equivalent demands for stochastic resources
  - Maximize the minimum utilization among all server resources
  - Satisfy SLA requirement with fewest servers