On Arbitrating the Power-Performance Tradeoff in SaaS Clouds

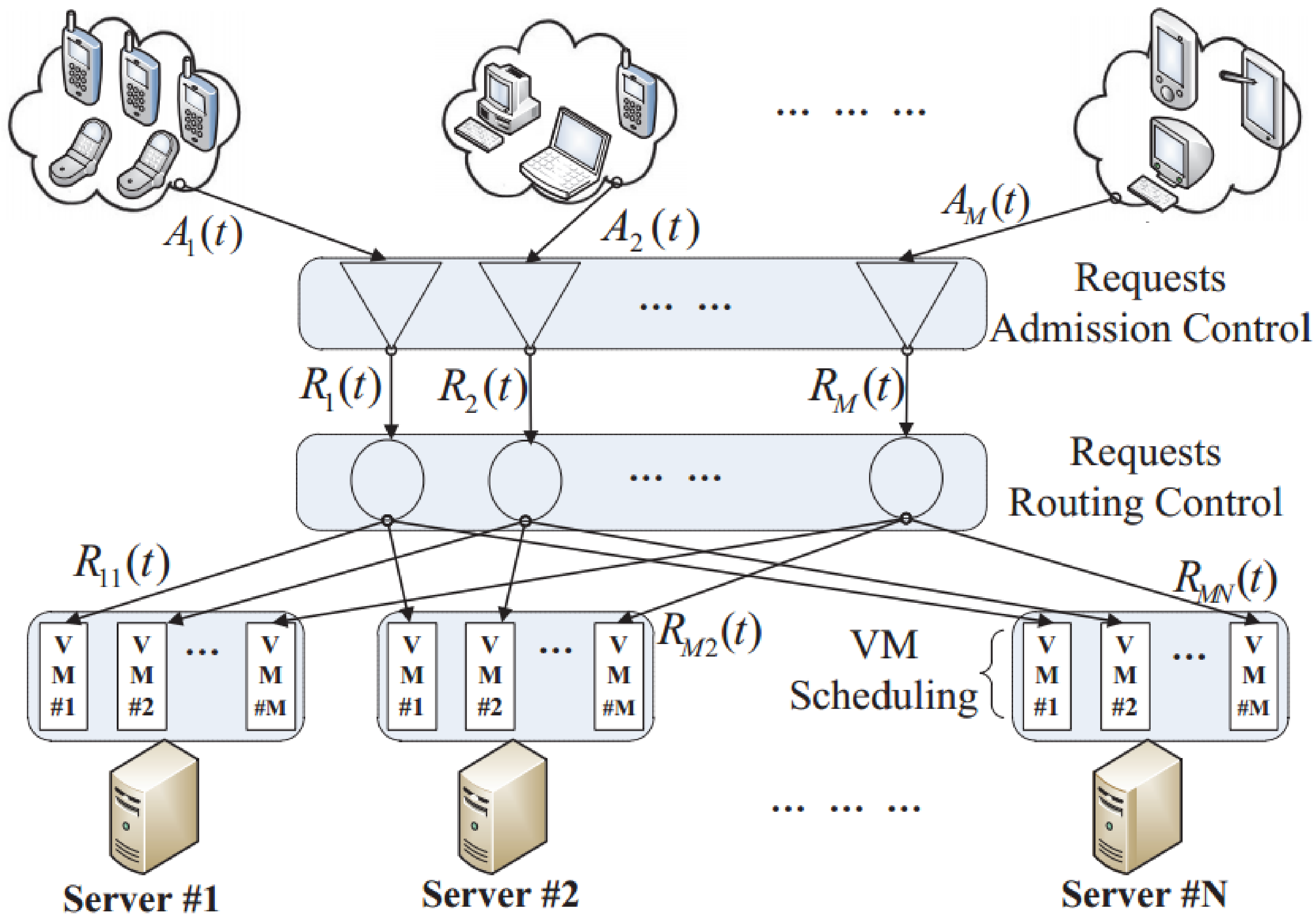
Presented by Xuanjia Qiu

Feb. 1, 2013

Why I pick this paper

- To appear in INFOCOM 2013, as a full paper
- Simple model and rich results
- Relevance to my current work

On Arbitrating the Power-Performance Tradeoff in SaaS Clouds



Assumption

- One data center
- N homogeneous servers
- Each server is virtualized to M VMs to serve M types of heterogeneous application
- Applications have diverse request arrival rates and workload (but identical for requests from an application)
- CPU-bounded application only
- Cost -> Power consumption -> CPU load

- Time slot t=0,1,2,3,…
- $A_i(t)$ requests arrive at the data center (*i* is the type of application), i.i.d. over time slots. $A_i(t) <= A_i^{max}$

Control Decisions

- Admission Control: $R_i(t)$, $0 <= R_i(t) <= A_i(t)$
- Routing Control: $R_{ij}(t)$
  - One request queue for each application(i.e., VM) in each server.
  - Constraint: $R_i(t) = \sum_{j=1}^{N} R_{ij}(t)$
- Scheduling of VMs (on or off)

$$a_{ij}(t) = \begin{cases} 1, & \text{if the } i\text{-th VM on server } j \text{ is running,} \\ 0, & \text{if the } i\text{-th VM on server } j \text{ is idle.} \end{cases}$$

$$Q_{ij}(t+1) = \max[Q_{ij}(t) - a_{ij}(t), 0] + d_i R_{ij}(t).$$

Modeling Cost

- Power Cost

$$P(s) = \alpha s^v + (1 - \alpha)$$

  – Normalize CPU load: $s_j(t) = \frac{\sum_{i=1}^{M} a_{ij}(t)}{M}$
  – Power Cost: $\text{Price} \cdot \text{PUE} \cdot \sum_{j=1}^{N} p_j$
    - (PUE: Power Usage Effectiveness)
- Revenue of throughput $g(r_i) = \log(1 + d_i r_i)$
- Aggregation:

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{A}} g(r_i) - \beta \sum_{j \in \mathcal{S}} p_j \\ \text{s.t.} \quad & 0 \le r_i \le \lambda_i, r_i \le N/d_i, \quad \forall i \in \mathcal{A}, \end{aligned}$$

Idle VMs cannot be shut down to further power saving

Lyapunov Optimization

- Original problem

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{A}} g(r_i) - \beta \sum_{j \in \mathcal{S}} p_j \\ \text{s.t.} \quad & 0 \le r_i \le \lambda_i, r_i \le N/d_i, \quad \forall i \in \mathcal{A}, \end{aligned}$$

- After transformation

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{A}} g(\gamma_i) - \beta \sum_{j \in \mathcal{S}} p_j && (6) \\ \text{s.t.} \quad & \gamma_i \le r_i, \quad \forall i \in \mathcal{A} && (7) \\ & 0 \le r_i \le \lambda_i, \quad \forall i \in \mathcal{A} && (8) \\ & r_i \le N/d_i, \quad \forall i \in \mathcal{A}. && (9) \end{aligned}$$

$$H_i(t+1) = \max[H_i(t) - R_i(t), 0] + \gamma_i(t),$$

$$0 \le \gamma_i(t) \le A_i^{\max}.$$

- Lyapunov function

$$L(\boldsymbol{\Theta}(t)) = \frac{1}{2}\Big[\sum_{i\in\mathcal{A}} d_i^2 H_i^2(t) + \sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{S}} Q_{ij}^2(t)\Big].$$

- Drift-minus-profit

$$\Delta(\boldsymbol{\Theta}(t)) - V\mathbb{E}\Big\{\sum_{i\in\mathcal{A}} g(\gamma_i(t)) - \beta\sum_{j\in\mathcal{S}} P_j(t)|\boldsymbol{\Theta}(t)\Big\}.$$

- Minimizing upper bound

$$\begin{aligned}
\Delta(\boldsymbol{\Theta}(t)) - V\mathbb{E}\Big\{\sum_{i\in\mathcal{A}} g(\gamma_i(t)) - \beta\sum_{j\in\mathcal{S}} P_j(t)|\boldsymbol{\Theta}(t)\Big\} \le B_1 \\
-\sum_{i\in\mathcal{A}} \mathbb{E}\{V g(\gamma_i(t)) - d_i^2 H_i(t)\gamma_i(t)|\boldsymbol{\Theta}(t)\} \quad (15) \\
-\sum_{i\in\mathcal{A}} \mathbb{E}\Big\{d_i^2 H_i(t) R_i(t) - \sum_{j\in\mathcal{S}} d_i R_{ij}(t) Q_{ij}(t)|\boldsymbol{\Theta}(t)\Big\} \quad (16) \\
-\sum_{j\in\mathcal{S}} \mathbb{E}\Big\{\sum_{i\in\mathcal{A}} Q_{ij}(t) a_{ij}(t) - V\beta P_j(t)|\boldsymbol{\Theta}(t)\Big\}. \quad (17)
\end{aligned}$$

Auxiliary Variable Selection

Request Admission Control and Routing

VM Scheduling

Auxiliary Variable Selection

$$
\begin{aligned}
\max_{\gamma_i(t)} \quad & V\log(1+d_i\gamma_i(t)) - d_i^2 H_i(t)\gamma_i(t) \\
\text{s.t.} \quad & 0 \le \gamma_i(t) \le A_i^{\max}, \forall i \in \mathcal{A}.
\end{aligned}
$$

- Differentiating the objective function,

$$
\gamma_i(t) = \begin{cases} 0, & H_i(t) > \dfrac{V}{d_i} \\ \dfrac{V}{d_i^2 H_i(t)} - \dfrac{1}{d_i}, & \dfrac{V}{d_i^2 A_i^{\max} + d_i} \le H_i(t) \le \dfrac{V}{d_i} \\ A_i^{\max}, & H_i(t) < \dfrac{V}{d_i^2 A_i^{\max} + d_i} \end{cases} \quad (19)
$$

Request Admission Control and Routing

$$\begin{aligned} \max_{R_i(t), R_{ij}(t)} \quad & d_i^2 H_i(t) R_i(t) - d_i \sum_{j \in \mathcal{S}} R_{ij}(t) Q_{ij}(t) \quad (20) \\ \text{s.t.} \quad & 0 \le R_i(t) \le A_i(t), \forall i \in \mathcal{A}, \\ & R_i(t) = \sum_{j \in \mathcal{S}} R_{ij}(t). \end{aligned}$$

- Making Routing first, by assuming $R_i(t)$ is known

$$\begin{aligned} \min_{R_{ij}(t)} \quad & d_i \textstyle\sum_{j \in \mathcal{S}} R_{ij}(t) Q_{ij}(t) \qquad (21) \\ \text{s.t.} \quad & \textstyle\sum_{j \in \mathcal{S}} R_{ij}(t) = R_i(t), \forall i \in \mathcal{A}. \end{aligned}$$

$$\min_{R_{ij}(t)} \quad d_i \sum_{j \in \mathcal{S}} R_{ij}(t) Q_{ij}(t) \tag{21}$$
$$\text{s.t.} \quad \sum_{j \in \mathcal{S}} R_{ij}(t) = R_i(t), \forall i \in \mathcal{A}.$$

- Solution
  - Dispatch as many admitted requests as possible to the VM with shortest backlogged queue:

$$R_{ij}(t) = \begin{cases} R_i(t), & j = j_i^*, \\ 0, & \text{else}, \end{cases}$$

- Complexity reduced by the solution by "Power of the two choices", but compromising the rigor of optimality analysis

- Combine the results:

$$\begin{aligned}\max_{R_i(t),R_{ij}(t)} \quad & d_i^2 H_i(t)R_i(t) - d_i \sum_{j\in\mathcal{S}} R_{ij}(t)Q_{ij}(t) \quad (20)\\ \text{s.t.} \quad & 0 \le R_i(t) \le A_i(t), \forall i \in \mathcal{A},\\ & R_i(t) = \sum_{j\in\mathcal{S}} R_{ij}(t).\end{aligned}$$

$$R_{ij}(t) = \begin{cases} R_i(t), & j = j_i^*,\\ 0, & \text{else},\end{cases}$$

- Get

$$\begin{aligned}\max_{R_i(t)} \quad & d_i^2 H_i(t)R_i(t) - d_i R_i(t)Q_{ij_i^*}(t) \qquad (23)\\ \text{s.t.} \quad & 0 \le R_i(t) \le A_i(t), \forall i \in \mathcal{A}.\end{aligned}$$

VM Scheduling

$$
\begin{aligned}
\max_{a_{ij}(t)} \quad & \sum_{i \in \mathcal{A}} Q_{ij}(t) a_{ij}(t) - V \beta P_j(t) \\
\text{s.t.} \quad & a_{ij}(t) \in \{0, 1\}, \forall i \in \mathcal{A}, \forall j \in \mathcal{S},
\end{aligned}
$$

$$P_j(t) = \alpha\left(\frac{\sum_{i=1}^{M} a_{ij}(t)}{M}\right)^v + (1 - \alpha)$$

- Greedy algorithm: search from VM with most backlogged queue to the least, until the growth in the sum of backlogs falls below the growth of the power consumption for a certain VM

Analytical Results

Theorem 1: For arbitrary arrival rates of application requests $(\lambda_1(t), \lambda_2(t), ..., \lambda_M(t))$ (possibly exceeding the processing capacity of a datacenter), a datacenter using the **OCA** algorithm with any $V \geq 0$ (the stability-profit tradeoff parameter defined in Sec. III-A1) can guarantee that all the actual and virtual queues are strongly stable over time slots:

$$H_i(t) \leq \frac{V}{d_i} + A_i^{\max}, \forall i \in \mathcal{A}, \tag{26}$$

$$Q_{ij}(t) \leq V + 2d_i A_i^{\max}, \forall i \in \mathcal{A}, \forall j \in \mathcal{S}. \tag{27}$$

Meanwhile, the gap between its achieved time averaged profit and the optimal profit ξ^* is within B_1/V:

$$\liminf_{t \to \infty} \Big\{ \sum_{i \in \mathcal{A}} g(r_i) - \beta \sum_{j \in \mathcal{S}} p_j \Big\} \geq \xi^* - \frac{B_1}{V}, \tag{28}$$

where $\xi^* = \sum_{i=1}^{M} g(r_i^*) - \beta \sum_{j=1}^{N} p_j^*$, r_i^* and p_j^* are the optimal solution to **Problem** (5), and B_1 is a finite constant parameter defined in Lemma 1.

Evaluation

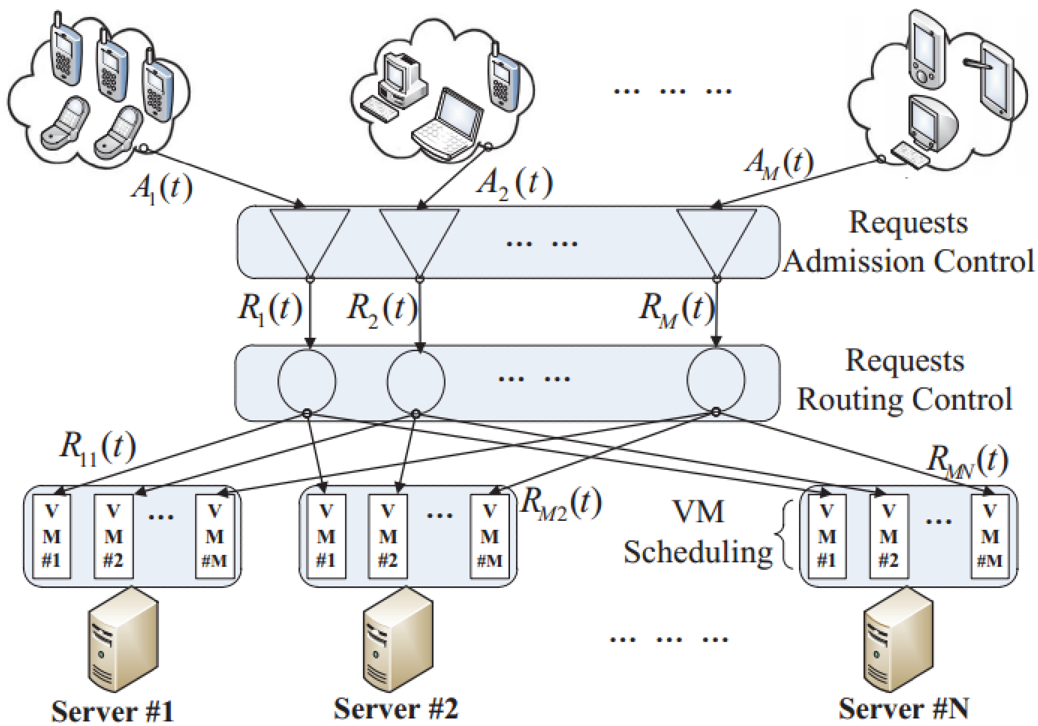
- Assumption and input
  - 100 servers, 10 VMs on each server (corresponding 10 heterogeneous applications)
  - Number of newly arrived requests in each time slot is assumed to be uniformly and randomly distributed within $[0, A_i^{max}]$

TABLE II: Request Arrival Rates and Sizes of Different Applications.

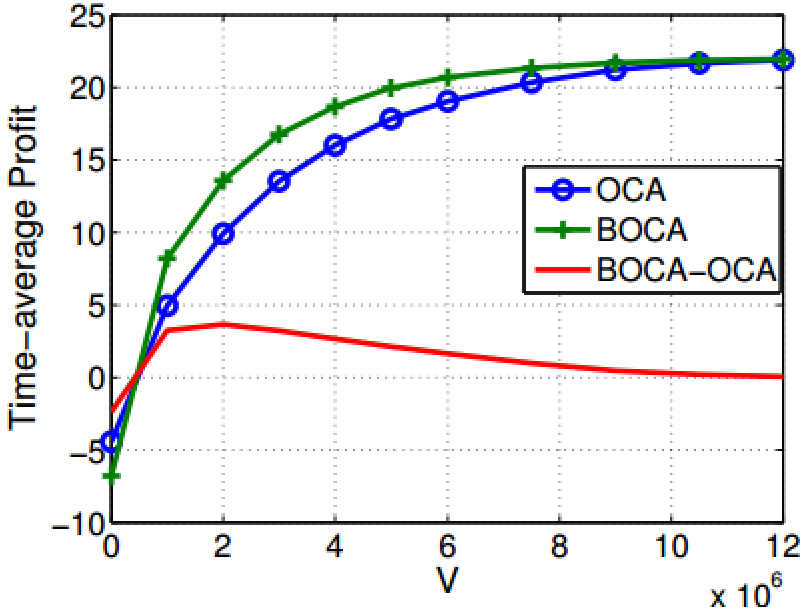
App i	1	2	3	4	5	6	7	8	9	10
$\lambda_i(\times 10^3)$	2.5	2	3.5	2	3	2	2.75	2.4	2.6	2.8
$d_i(\times 10^{-2})$	2	3	2	4	3	5	4	5	5	5
$d_i\lambda_i(\times 10)$	5	6	7	8	9	10	11	12	13	14

Buffer or not buffer?

- Store the newly admitted requests in a buffer before they are routed to VMs



Buffer or not buffer?



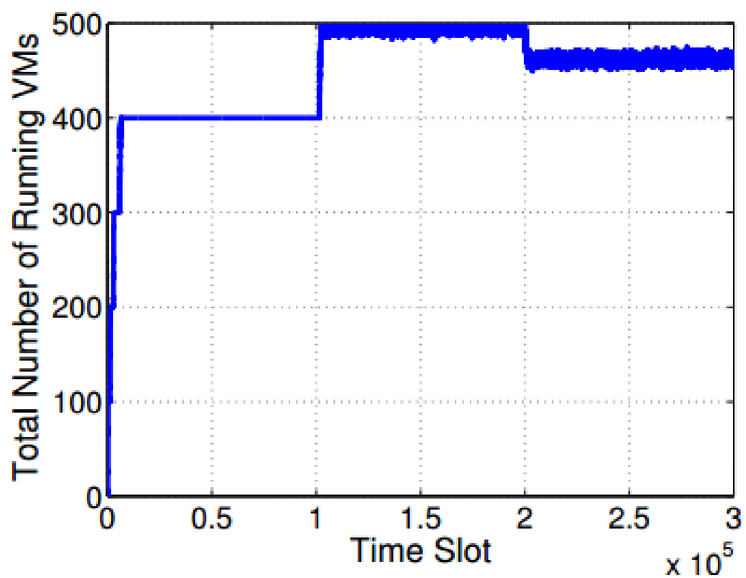
$$L(\Theta(t)) = \frac{1}{2}\Big[\sum_{i=1}^{M} d_i^2 H_i^2(t) + \sum_{i=1}^{M} d_i^2 L_i^2(t) + \sum_{i=1}^{M}\sum_{j=1}^{N} Q_{ij}^2(t)\Big]$$

$$\liminf_{t\to\infty} \left\{ \sum_{i=1}^{M} g(r_i) - \beta \sum_{j=1}^{N} p_j \right\} \geq \xi^* - \frac{B_2}{V}$$

- $B_2 > B_1$
- Buffering is no better than no-buffering.
- Reasoning: too many queues, reducing the weight on performance

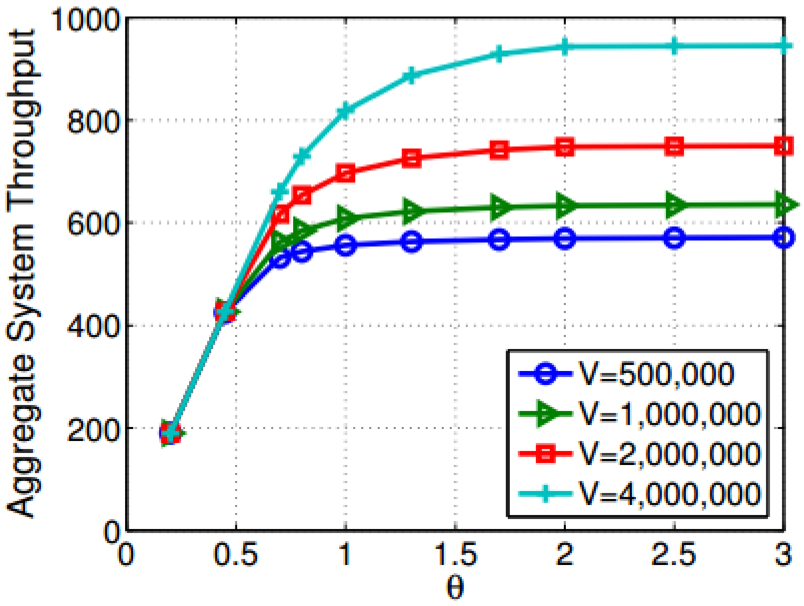
Selected Evaluation Results

- In response to bursty request arrivals
  - Vary the mean request arrival rates to be 0.5, 1, and 1.5 times of the original



Selected Evaluation Results

- Effectiveness of admission control



Selected Points that we can learn from

- Model
  - Workload of requests are normalized according to the VM processing capacity in one time slot, and <1
- Solution
  - Dig the optimality structure of sub-problem first
  - Use greedy algorithm to solve non-linear, non-differentiatable optimization problem
- Evaluation
  - Compare different queueing models (buffer against no-buffer)

Q&A