Dynamic Resource Allocation and Power Management in Virtualized Data Centers

IFIP'2010

Rahul Urgaonkar@usc
{Ulas C. Kozat, Ken Igarashi}@docomolabs
Michael J. Neely@usc

Outline

- Model
- Algorithm
- Analysis
- Evaluation
- Conclusion

Outline

- Model
- Algorithm
- Analysis
- Evaluation
- Conclusion

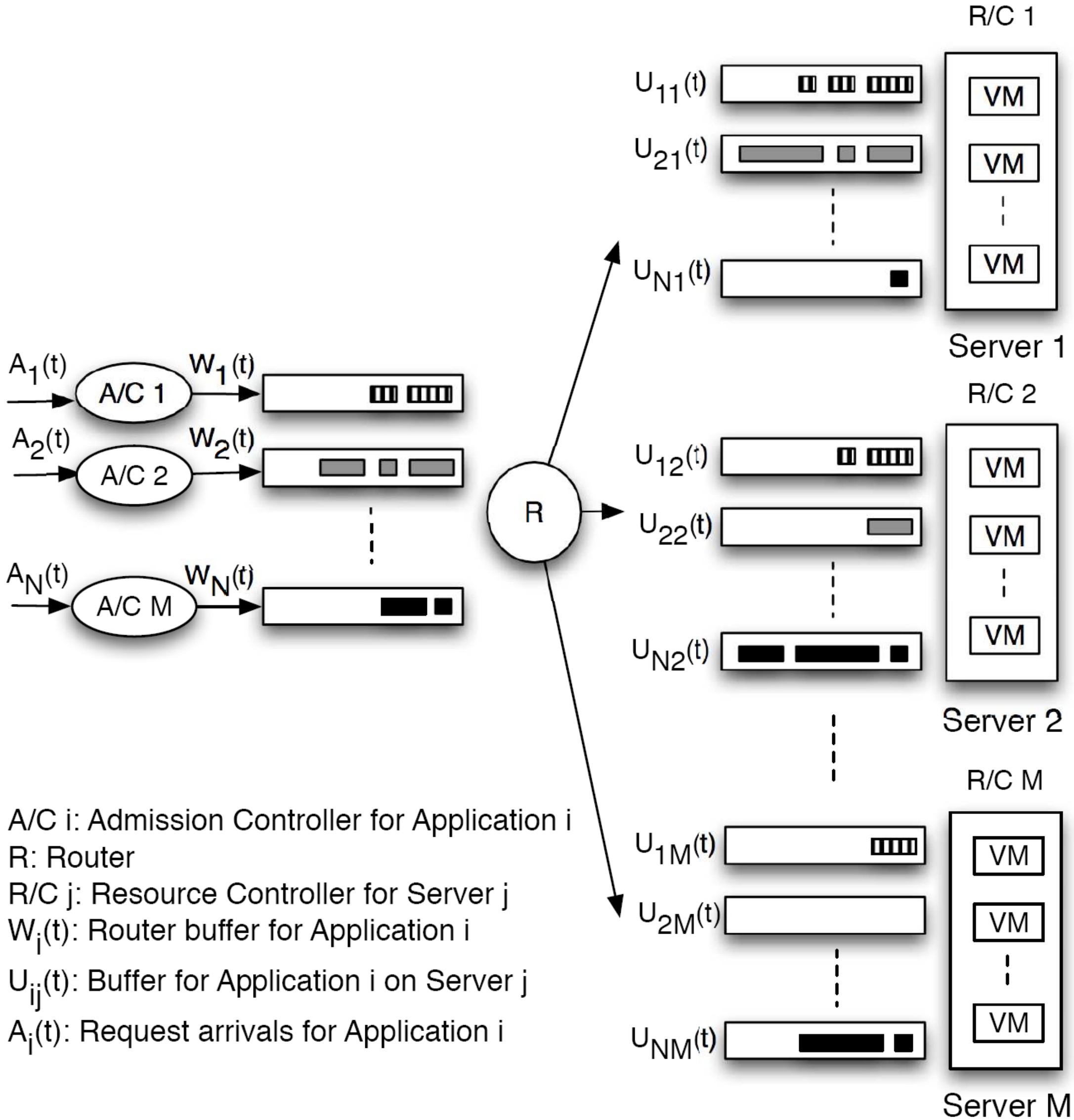


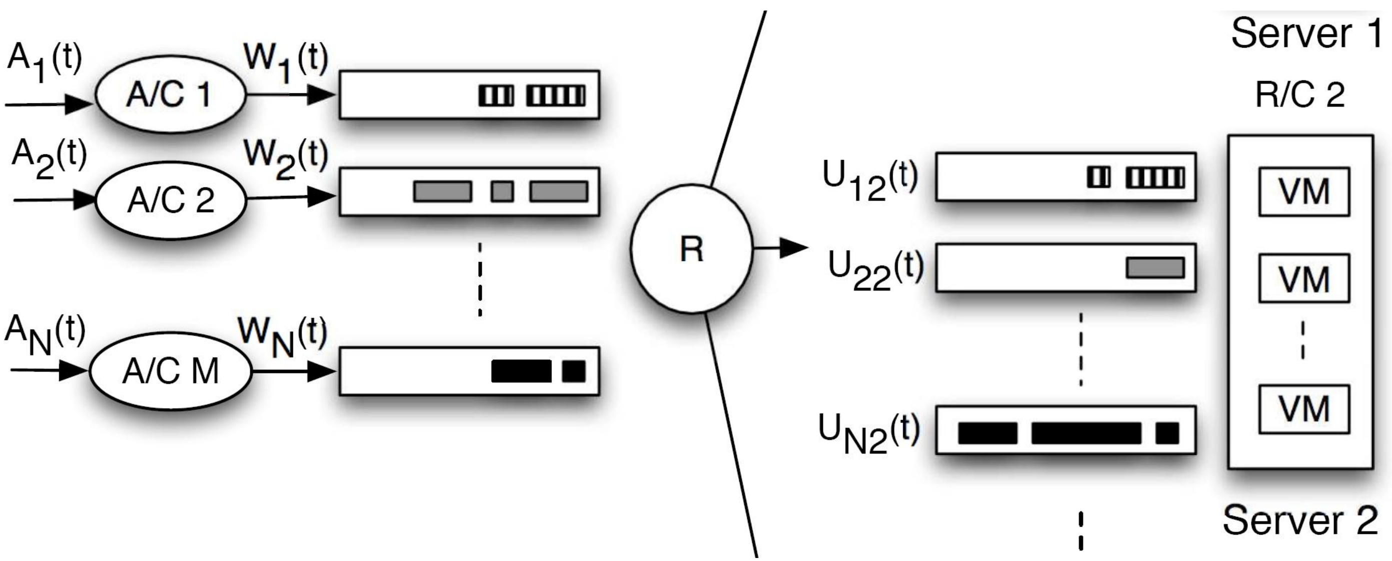
Fig. 1. Illustration of the Virtualized Data Center Architecture.

Queuing Dynamics

- $R_i(t)$: the number of requests out of $A_i(t)$ admitted into the Router
- $W_i(t)$: the backlog of router buffer
- $R_{ij}(t)$: the number of requests for application i routed to server j

$$0 \leq R_i(t) \leq A_i(t)$$

$$W_i(t+1) = W_i(t) - \sum_j R_{ij}(t) + R_i(t)$$

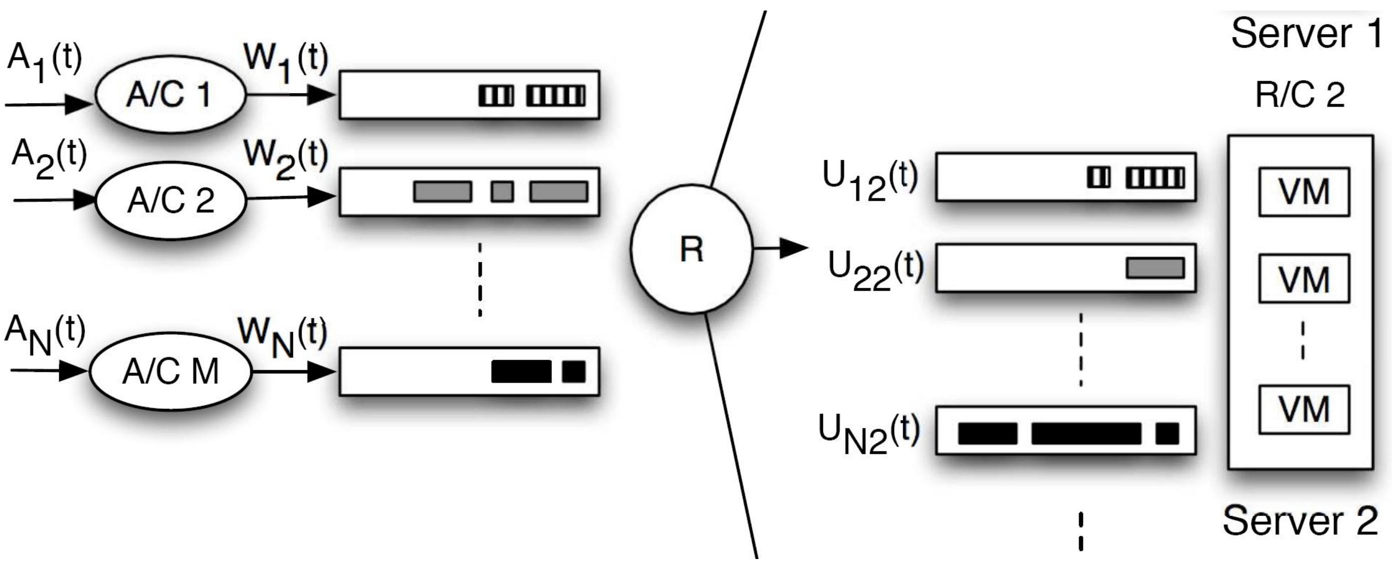


Queuing Dynamics

- $U_{ij}(t)$: the backlog of server j for application i
- $\mu_{ij}(I_j(t))$: the service rate provided to appilcation I on server j taking control action $I_j(t)$

- $I_j(t)$: the particular control decision at server j

$$U_{ij}(t+1) = \max[U_{ij}(t) - \mu_{ij}(I_j(t)), 0] + R_{ij}(t)$$



Control Objective

$$r_i^\eta = \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{R_i^\eta(\tau)\}$$

$$e_j^\eta \triangleq \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\left\{P_j^\eta(\tau)\right\}$$

Maximize: $\sum_{i\in\mathcal{A}} \alpha_i r_i^\eta - \beta \sum_{j\in\mathcal{S}} e_j^\eta$

Subject to: $0 \le r_i^\eta \le \lambda_i \ \forall\, i \in \mathcal{A}$

$I_j^\eta(t) \in \mathcal{I}_j \ \forall\, j \in \mathcal{S}, \ \forall t$

$\boldsymbol{r} \in \Lambda$

Power-frequency relationship is approximated by a quadratic model: $P(f) = P_{min} + c(f - f_{min})^2$

- A: the set of applications
- S: the set of servers
- η: the decision policy
- r_i^η: average expected rate of admitted requests for app I under policy η
- e_j^η: average expected power consumption of server j under policy η
- α_i, β: non-negative weights

Outline

- Model
- Algorithm
- Analysis
- Evaluation
- Conclusion

Data Center Control Algorithm (DCA)

- Admission Control
  - Which requests will be served?
- Routing
  - Which server will a request be served at?
- Resource Allocation
  - CPU policy (frequency, voltage)?
  - Which servers will be shut down?

Admission Control

- V is a control parameter that is input to the algorithm

$$\text{Maximize:} \qquad R_i(t)[V\alpha_i - W_i(t)]$$

$$\text{Subject to:} \qquad 0 \leq R_i(t) \leq A_i(t)$$

Routing

- *Join the Shortest Queue*
  - Choose the server j’ which has the smallest queue backlog $U_{ij'}(t)$ If $W_i(t) > U_{ij'}(t), R_{ij'}(t) = W_i(t)$, else $R_{ij}(t) = 0$

Resource Allocation 1

- CPU policy (t != n T)

$$\text{Maximize: } \sum_i U_{ij}(t)\mathbb{E}\left\{\mu_{ij}(I_j(t))\right\} - V\beta P_j(t)$$

$$\text{Subject to: } I_j(t) \in \mathcal{I}_j, P_j(t) \geq P_{min}$$

Resource Allocation 2

- Server ON/OFF ($t = nT$)

$$\mathcal{S}^*(t) = \underset{\mathcal{S}(t)\in\mathcal{O}}{\operatorname{argmax}} \Big[\sum_{ij} U_{ij}(t)\mathbb{E}\left\{\mu_{ij}(I_j(t))\right\} - V\beta \sum_j P_j(t) + \sum_{ij} R_{ij}(t)(W_i(t) - U_{ij}(t)) \Big]$$

$$\text{subject to: } j \in \mathcal{S}(t), I_j(t) \in \mathcal{I}_j, P_j(t) \geq P_{min}$$

- DCA: unfinished requests are dropped when the server turns inactive
- DCA-M: unfinished requests are rerouted to other active servers

Outline

- Problem and Model
- Algorithms
- Analysis
- Evaluation
- Conclusion

Performance Analysis

1) The worst case queue backlog for each application Router buffer $W_i(t)$ is upper bounded by a finite constant W_i^{max} for all t:

$$W_i(t) \leq W_i^{max} \triangleq V\alpha_i + A_i^{max} \tag{14}$$

Similarly, the worst case queue backlog for application i on any server j is upper bounded by $2W_i^{max}$ for all i, t:

$$U_{ij}(t) \leq 2W_i^{max} = 2(V\alpha_i + A_i^{max}) \tag{15}$$

Performance Analysis

2) The time average utility achieved by the DCA algorithm is within BT/V of the optimal value:

$$\liminf_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \Big[\sum_{i\in\mathcal{A}} \alpha_i \mathbb{E}\left\{R_i(\tau)\right\} - \beta \sum_{j\in\mathcal{S}} \mathbb{E}\left\{P_j(\tau)\right\} \Big] \geq \sum_{i\in\mathcal{A}} \alpha_i r_i^* - \beta \sum_{j\in\mathcal{S}} e_j^* - \frac{BT}{V} \tag{16}$$

Lyapunov optimization framework

- Constructing an appropriate Lyapunov function of the queue backlogs
- Defining the Lyapunov drift
- Minimizing the drift over all control policies

- *If the drift is bounded, the performance is bounded*

Proof: Lyapunov Drift

- Lyapunov function:

$$L(\boldsymbol{Q}(t)) \triangleq \frac{1}{2}\Big[\sum_{i\in\mathcal{A},j\in\mathcal{S}} U_{ij}^2(t) + \sum_{i\in\mathcal{A}} W_i^2(t)\Big]$$

- Lyapunov drift:

$$\Delta(\boldsymbol{Q}(t)) \triangleq \mathbb{E}\left\{L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t)) | \boldsymbol{Q}(t)\right\}$$

$$B \triangleq \frac{\sum_i (A_i^{max})^2 + NM\mu_{max}^2}{2}$$

$$\Delta(t) - V\mathbb{E}\left\{\sum_i \alpha_i R_i(t) - \beta \sum_j P_j(t)|\boldsymbol{Q}(t)\right\} \leq B$$

$$- \sum_{ij} U_{ij}(t)\mathbb{E}\left\{\mu_{ij}(I_j(t))|\boldsymbol{Q}(t)\right\} + V\beta \sum_j \mathbb{E}\left\{P_j(t)|\boldsymbol{Q}(t)\right\}$$

$$- \sum_{ij} \mathbb{E}\left\{R_{ij}(t)(W_i(t) - U_{ij}(t))|\boldsymbol{Q}(t)\right\}$$

$$- \sum_i \mathbb{E}\left\{R_i(t)(V\alpha_i - W_i(t)|\boldsymbol{Q}(t)\right\} \quad (19)$$

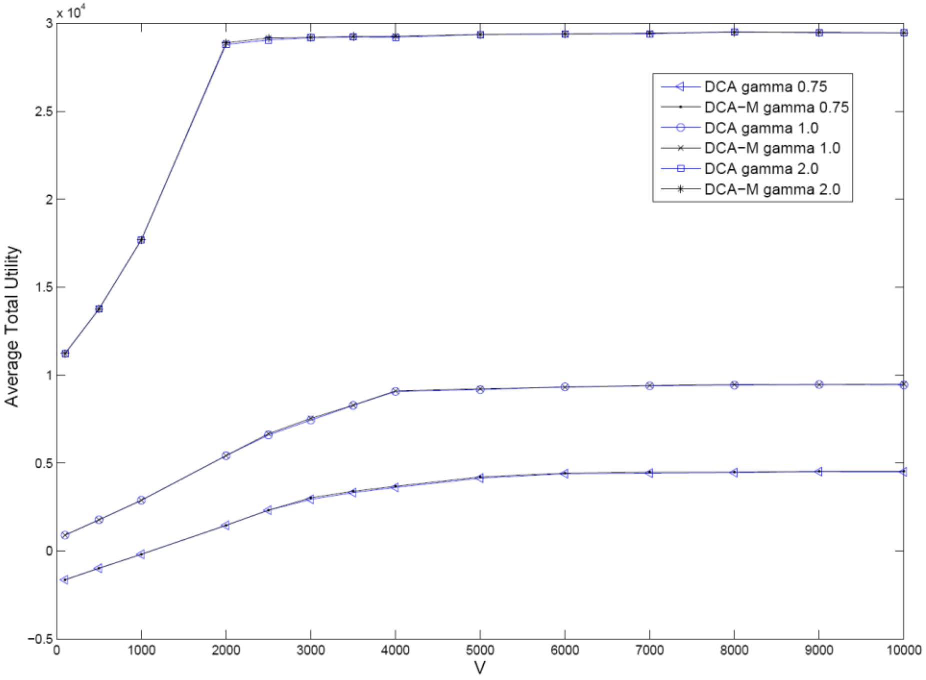
Outline

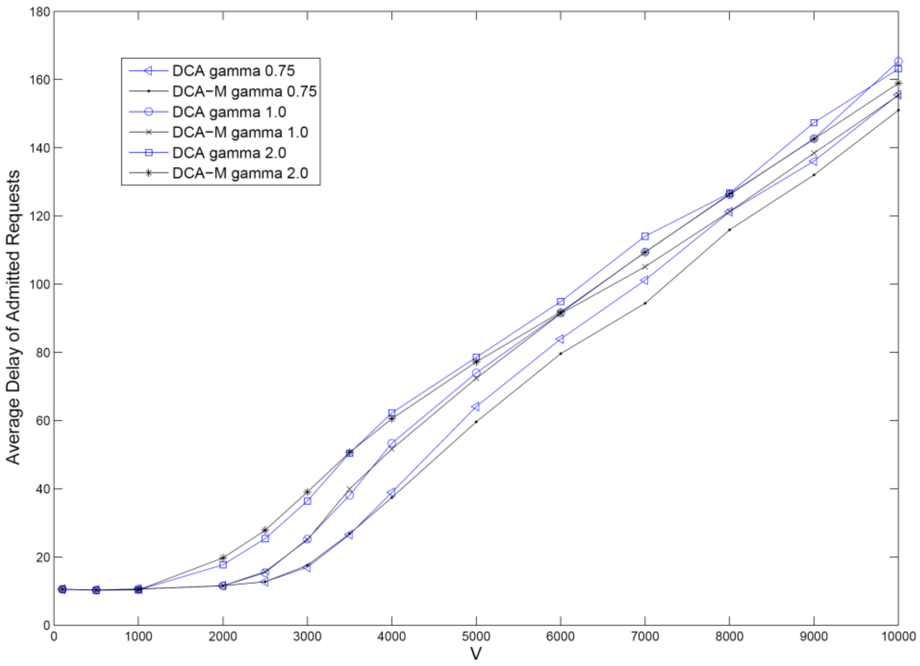
- Problem and Model
- Algorithm
- Analysis
- Evaluation
- Conclusion

Experiments

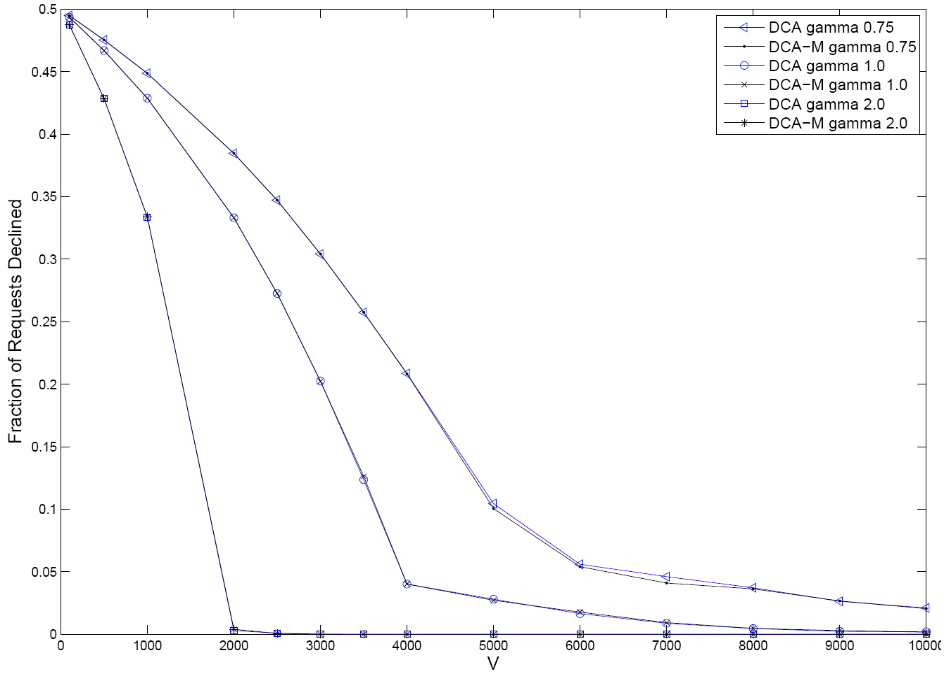
- 100 servers and 10 applications
- Frame size T = 1000 slots
- Application requests: uniformly random distributed in $[0, 2\lambda_i]$
- $\alpha_i = \alpha, \gamma = \frac{\alpha}{\beta}$

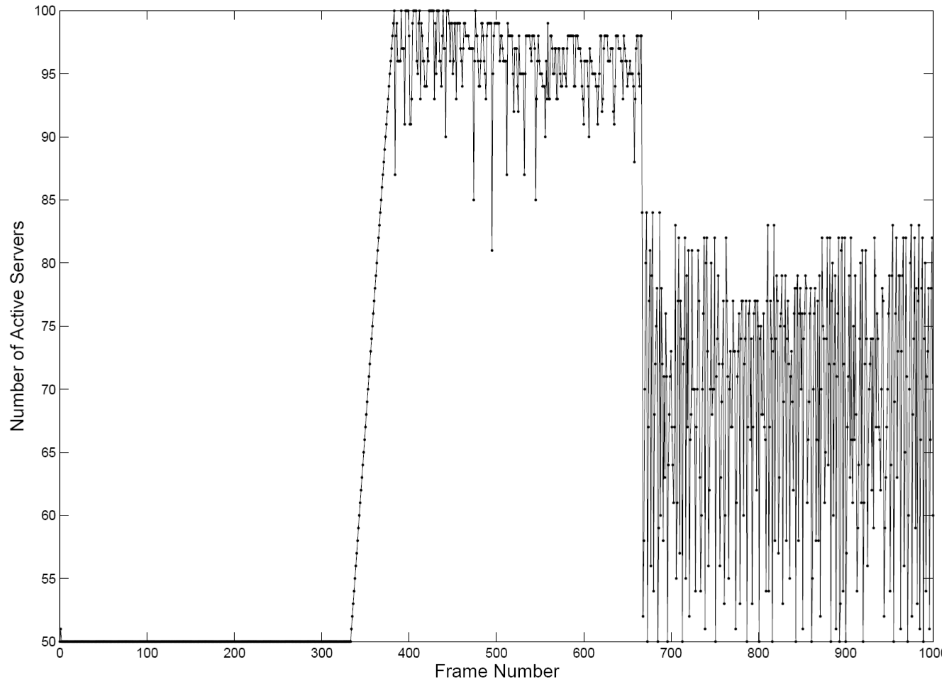
Evaluation 1





Evaluation 2





Conclusion

- Design and algorithm should follow the analysis framework
- Compare the different features of the algorithms in evaluation

Theorem 5.4. (Lyapunov Optimization) If there are positive constants V, ϵ, B such that for all timeslots t and all unfinished work matrices $\boldsymbol{U}(t)$, the Lyapunov drift satisfies:

$$\Delta(\boldsymbol{U}(t)) - V\mathbb{E}\{g(\boldsymbol{R}(t))\,|\boldsymbol{U}(t)\} \leq B - \epsilon\sum_{i=1}^{N} U_i(t) - Vg^*, \tag{5.19}$$

then time average utility and congestion satisfies:

$$\limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{i=1}^{N}\mathbb{E}\{U_i(\tau)\} \leq \frac{B + V(\overline{g} - g^*)}{\epsilon}, \tag{5.20}$$

$$\liminf_{t\to\infty} g(\overline{\boldsymbol{r}}(t)) \geq g^* - \frac{B}{V}, \qquad \overline{\boldsymbol{r}}(t) \triangleq \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\{\boldsymbol{R}(t)\}.$$

where $\overline{\boldsymbol{r}}(t)$ is defined in (5.18), and $\overline{g}$ is defined:

$$\overline{g} \triangleq \limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\{g(\boldsymbol{R}(\tau))\}.$$