清华大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

Performance 2020

# Heavy Traffic Analysis of Approximate Max-Weight Matching Algorithms for Input-Queued Switches

Yu Huang and Longbo Huang

IIIS, Tsinghua University

# Outline



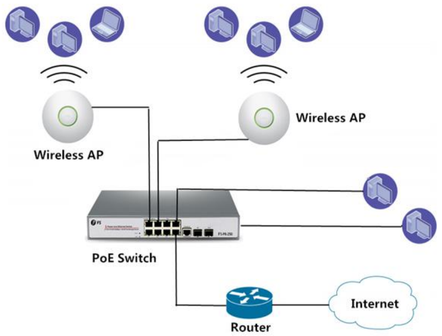
- Motivation
- System Model & Problem Settings
- Main Results
- Conclusion

# Motivation

- ❑ High-Speed Router
- ❑ Data Center Networks
- ❑ Wireless Networks

https://www.indiamart.com/infinityentp-hapur/network-switches.html

https://www.techiexpert.com/google-built-an-ai-to-help-keep-its-data-centers-cool/




https://community.fs.com/blog/power-over-ethernet-technology-poe-switch.html

**How to design a good policy ?**

# Motivation

清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

❖ **Shceduling Policy:** Max–Weight Matching (MWM)

- Throughput optimal
- Good delay performance
- Heavy traffic queue length optimal

❖ **Problem:**

- High complexity of computation: $O(n^3)$

**Consider a class of approximate MWM algorithms with lower complexity**

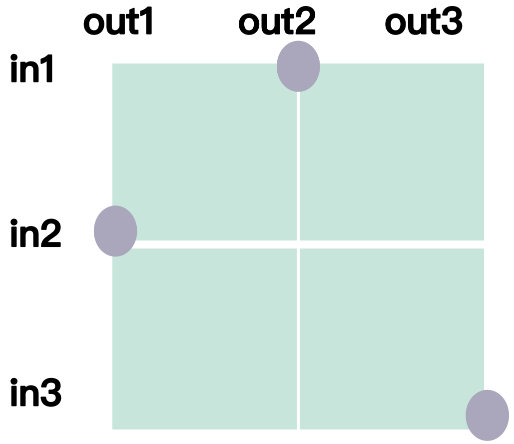
# System Model:



## ❖ n×n switch:

- Schedule process:

$$\mathcal{S} = \left\{ \mathbf{S} \in \{0,1\}^{n^2} : \sum_{i=1}^{n} S_{ij} \le 1, \sum_{j=1}^{n} S_{ij} \le 1, \forall i,j \in \{1,2,\dots,n\} \right\}$$




- Arrival process: bounded by $A_{max}$, I.I.D. Mean&Var: $\boldsymbol{\lambda}, \boldsymbol{\sigma}^2$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

## ❖ Heavy Traffic Setttings:

- Capacity Region

$$\mathcal{C} = \mathrm{Conv}(\mathcal{S})$$

$$= \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^{n^2} : \sum_{i=1}^{n} \lambda_{ij} \le 1, \sum_{j=1}^{n} \lambda_{ij} \le 1 \quad \forall i,j \in \{1,2,\dots,n\} \right\}$$

$$\boldsymbol{\nu} \in \partial\mathcal{C} \left[ \begin{array}{l} \sum_{j'=1}^{n} \nu_{ij'} = 1, \forall i \le n_1 \\ \sum_{i'=1}^{n} \nu_{i'j} = 1, \forall j \le n_2 \end{array} \right.$$

$$\epsilon \longrightarrow 0 \qquad \boldsymbol{\lambda}^{(\epsilon)} = \boldsymbol{\nu} - \epsilon\boldsymbol{\eta}$$

# Related Work:

清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

- **MWM:**
  [Tassiulas et al, 1992], [McKeown et al, 1999], [Georgiadis et al, 2006], [Basu et al, 2019] …
- **Low-Complexity Policy:**
  [Tassiulas, 1998], [Keslassy et al, 2001], [Shah et al, 2002], [Giaccone et al, 2003], [Lin et al, 2006], [Ross et al, 2007], [Gupta et al, 2007], [Lin et al, 2009] …
- **Heavy Traffic:**
  [Eryilmaz et al, 2012], [Maguluri et al, 2016], [Wang et al, 2017], [Maguluri et al, 2018], [Zhou et al, 2020] …

❖ **Remark:** Our work differs in

i. Extend the approximate MWM to an **expeted** sense
ii. Consider a **general** case: arbitrary number of ports are saturated
iii. Develop a novel communication efficient algorithm with good delay and throughput

# Main Results:

清華大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

- Expected 1-APRX
- Heavy Traffic Analysis
- Communication-Efficient Algorithm: MWM-AU

# Expected 1–APRX



**Weight of scehdule:** $$W_{\mathbf{S}}(t) \triangleq \langle \mathbf{Q}(t), \mathbf{S} \rangle = \sum_{i,j} Q_{ij}(t) S_{ij}$$

**MWM:** $$\mathbf{S}^*(t) \in \arg\max_{\mathbf{S} \in \mathcal{S}} \langle \mathbf{Q}(t), \mathbf{S} \rangle$$

**Expected 1–APRX:** $$\mathbb{E}\left\{ W_\pi(t) | \mathbf{Q}(t) \right\} \geq W^*(t) - f\left(W^*(t)\right)$$

**Remark:**

- Motivated by 1–APRX in [Shah et al, 2002]
- Containing a class of randomized policies e.g., TASS[Tassiulas, 1998], batch MWM [Ross et al, 2007]
- Expected 1–APRX achieves 100% throughput

# Heavy Traffic Results: SSC

清华大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

**Cone** $\mathcal{K}_{n_1 n_2} \triangleq \Big\{\mathbf{x} \in \mathbb{R}^{n^2} : \mathbf{x} = \sum_{i=1}^{n_1} w_i \mathbf{e}^{(i)} + \sum_{j=1}^{n_2} \widetilde{w}_j \widetilde{\mathbf{e}}^{(j)}$

$w_i \in \mathbb{R}^+ \text{ for } 1 \leq i \leq n_1, \widetilde{w}_j \in \mathbb{R}^+ \text{ for } 1 \leq j \leq n_2\Big\}$

**Theorem 1** For any fixed $\beta > 0$, and $0 < \epsilon \leq \nu'_{\min}/4(1+2\beta)\|\boldsymbol{\eta}\|$ each system with the steady state queue lengths vector satisfies:

$$\mathbb{E}\left[\left\|\overline{\mathbf{Q}}^{(\epsilon)}_{\perp\mathcal{K}}\right\| - \beta\left\|\overline{\mathbf{Q}}^{(\epsilon)}_{\|\mathcal{K}}\right\|\right] \leq M_\beta$$

Prior: e.g. [Maguluri et al, 2018]

$$\mathbb{E}\left[\|\mathbf{Q}_{\perp\mathcal{K}_{n1n2}}\|^r\right] \leq M_r$$

Our case:

$$\mathbb{E}\left[\left\|\overline{\mathbf{Q}}^{(\epsilon)}_{\perp\mathcal{K}}\right\|\right] / \mathbb{E}\left[\left\|\overline{\mathbf{Q}}^{(\epsilon)}\right\|\right] < \beta$$

# Main Idea of Proof: Drift Method

清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

## Drift function:

$$W(\mathbf{Q}) \triangleq \max\{\|\mathbf{Q}_{\perp\mathcal{K}}\| - \beta \left\|\mathbf{Q}_{\|\mathcal{K}}\right\|, 0\}$$

## Remark:

- Inspired by [Wang et al, 2017]
- Drift function used for MWM, e.g., $W(\mathbf{Q}) \triangleq \|\mathbf{Q}_{\perp\mathcal{K}}\|$ [Maguluri et al, 2018] cannot work for expected 1–APRX

# Heavy Traffic Result: Upper Bound

清華大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

**Subspace**

$$\mathcal{S}_{n_1 n_2} \triangleq \left\{ \mathbf{x} \in \mathbb{R}^{n^2} : \mathbf{x} = \sum_{i=1}^{n_1} w_i \mathbf{e}^{(i)} + \sum_{j=1}^{n_2} \widetilde{w}_j \widetilde{\mathbf{e}}^{(j)} \text{ where} \right.$$

$$\left. w_i \in \mathbb{R} \text{ for } 1 \le i \le n_1, \widetilde{w}_j \in \mathbb{R} \text{ for } 1 \le j \le n_2 \right\}$$

**Theorem 2: For any fixed weight vector $\boldsymbol{\alpha} \in \mathbb{R}^{n^2}$, the steady state queue lengths vector satisfies:**

$$\epsilon \left( \mathbb{E}[\langle \overline{\mathbf{Q}}^{(\epsilon)}, \boldsymbol{\alpha} \rangle] - \underline{(\|\boldsymbol{\alpha}\| + 2n^2 \min\{n_1 + n_2, n\})\mathbb{E}\left[ \left\| \overline{\mathbf{Q}}^{(\epsilon)}_{\perp S} \right\| \right]} \right) \le \underline{\frac{1}{2} \left\langle (\boldsymbol{\sigma}^{(\epsilon)})^2, \boldsymbol{\zeta} \right\rangle} + B(\epsilon)$$

**Remark:**

- Collapse to $\mathcal{K}_{n_1 n_2}$ $\longrightarrow$ Collapse to $\mathcal{S}_{n_1 n_2}$
- Upper bound for weighted queue length $\epsilon \mathbb{E}[\langle \overline{\mathbf{Q}}, \boldsymbol{\alpha} \rangle]$ is close to $\frac{1}{2} \langle \boldsymbol{\sigma}^2, \boldsymbol{\zeta} \rangle$ in the heavy–traffic limit.

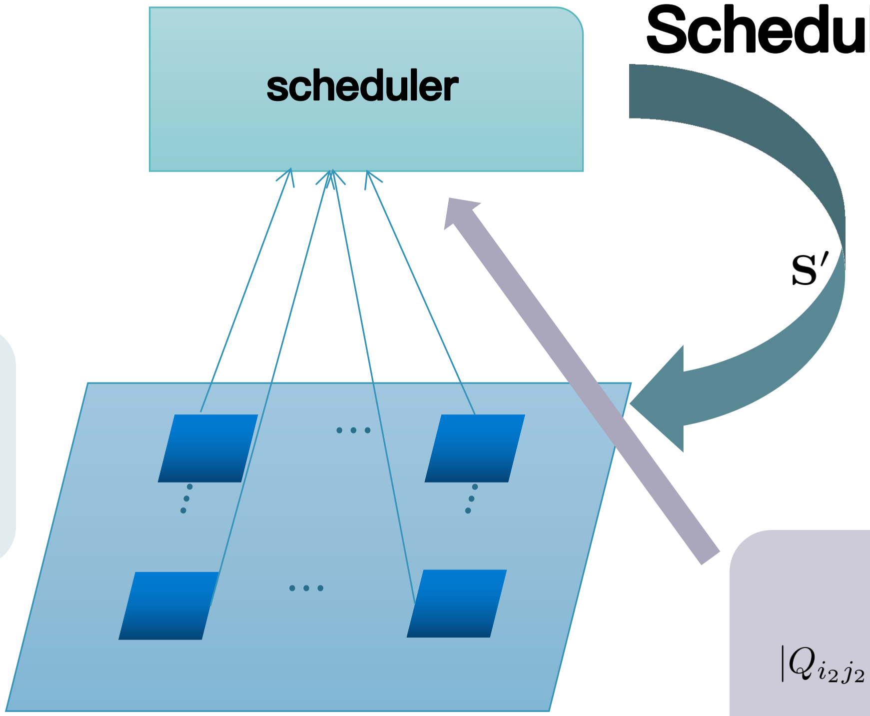
# Communication-Efficient Algorithm: MWM-AU

清华大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

$$\mathbf{Q}' = \left[Q'_{ij}\right]$$

**Scheduling:**

$$\mathbf{S}'(t) = \arg\max_{\mathbf{S}\in\mathcal{S}} \sum_{i,j} Q'_{ij}(t) S_{ij}$$




✖ **Update** $Q_{i_1 j_1}$

$$|Q_{i_1 j_1}(t) - Q'_{i_1 j_1}(t-1)| \leq g\left(Q_{i_1 j_1}(t)\right)$$

g sub-linear, non-decreasing

concave function

**Update** $Q_{i_2 j_2}$

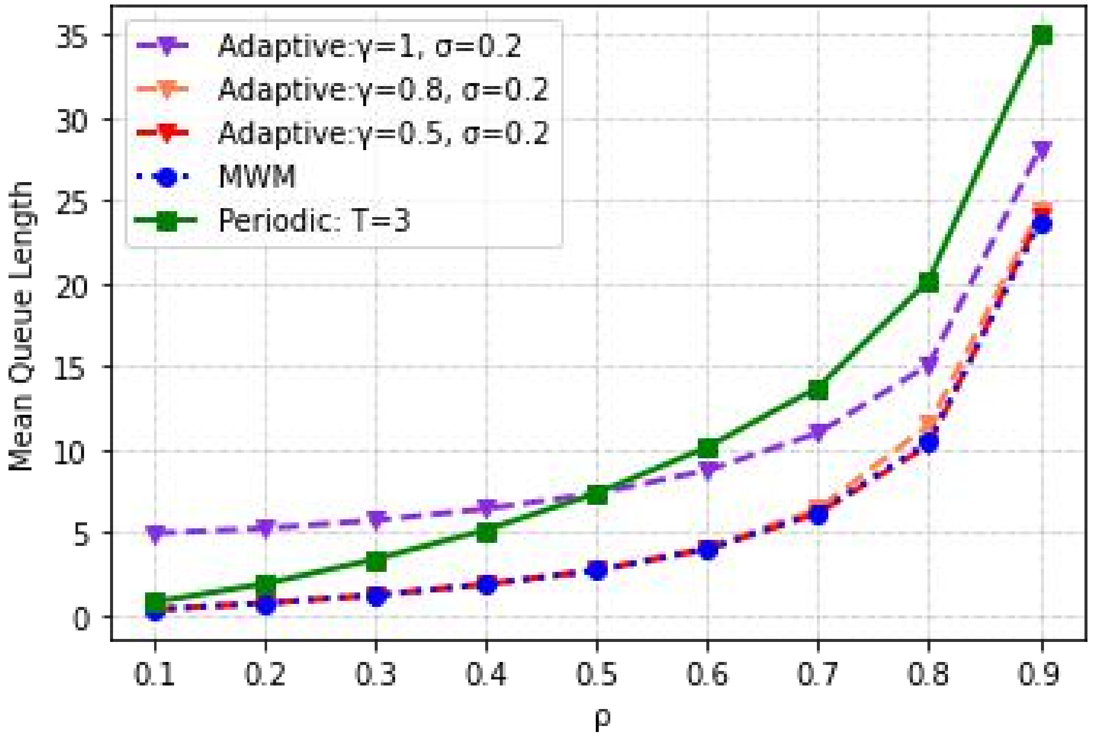
$$|Q_{i_2 j_2}(t) - Q'_{i_2 j_2}(t-1)| > g\left(Q_{i_2 j_2}(t)\right)$$

**Proposition 1:** $W_a(t) \geq W^*(t) - 2ng(W^*(t)/n)$ i.e., **MWM-AU** belongs to **expected 1-APRX**

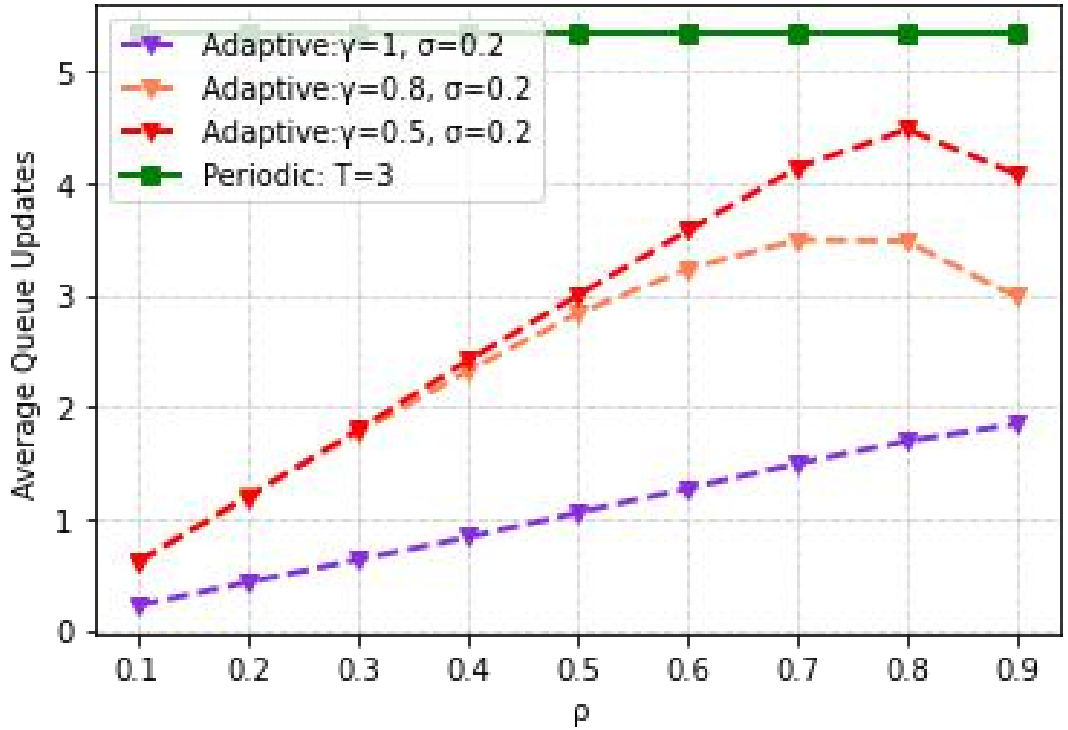
- **Throughput optimal**
- **Upper bound:** $\epsilon\mathbb{E}[\langle \overline{\mathbf{Q}}^{(\epsilon)}, \boldsymbol{\alpha}\rangle] \leq \frac{1}{2}\left\langle (\boldsymbol{\sigma}^{(\epsilon)})^2, \boldsymbol{\zeta}\right\rangle$

# Simulations

清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University



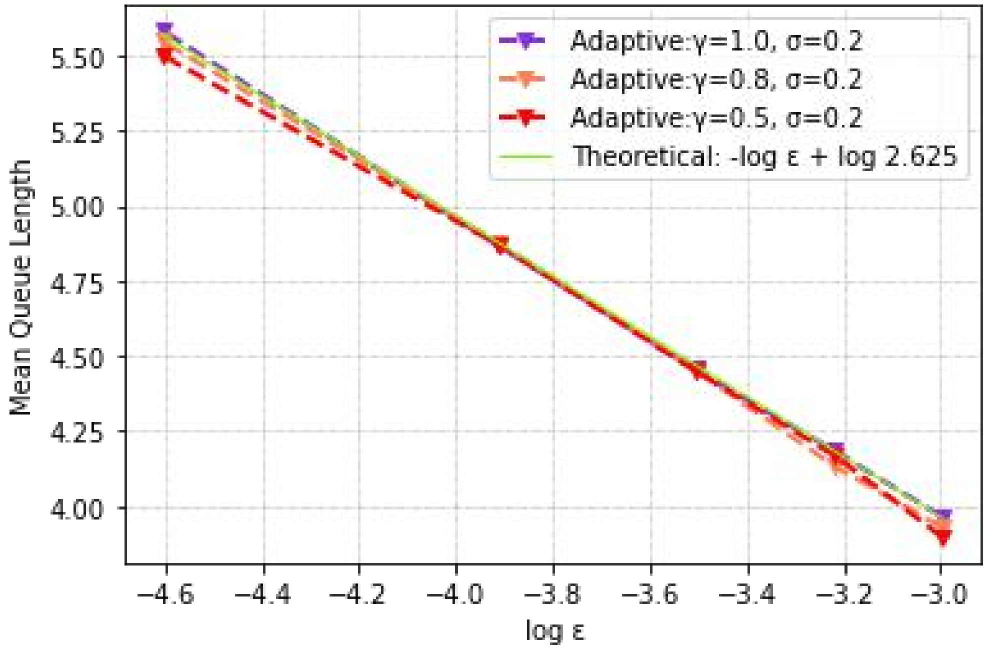

delay performance



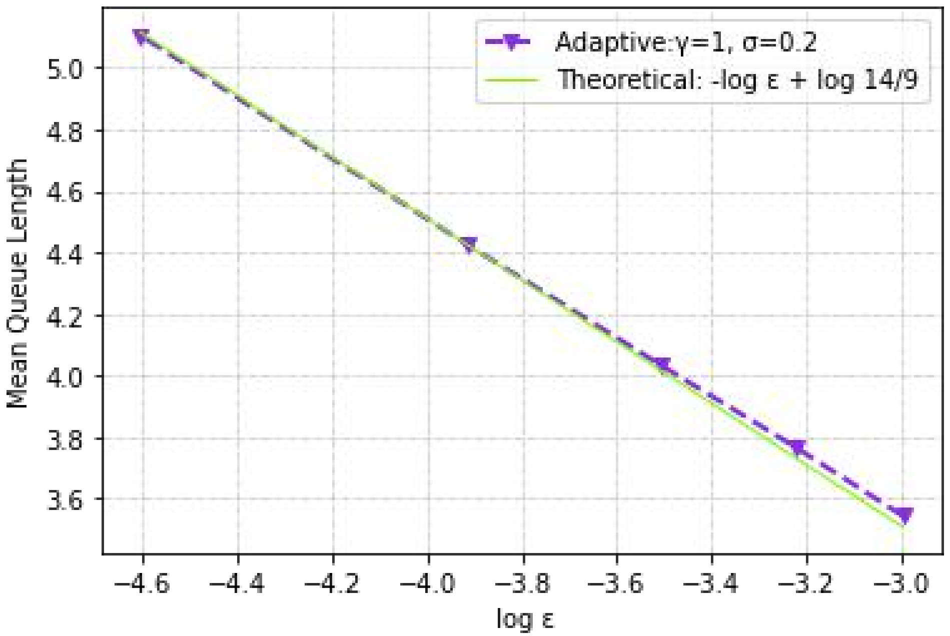

communication frequency

# Simulations: Heavy Traffic

清华大学 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University




uniform traffic




non-uniform traffic

$$\epsilon\mathbb{E}[\langle \overline{\mathbf{Q}}^{(\epsilon)}, \boldsymbol{\alpha}\rangle] \leq \frac{1}{2}\left\langle (\boldsymbol{\sigma}^{(\epsilon)})^2, \boldsymbol{\zeta} \right\rangle$$

Theoretical: $-\log\epsilon + \log(\frac{1}{2}\left\langle (\boldsymbol{\sigma}^{(\epsilon)})^2, \boldsymbol{\zeta} \right\rangle)$

# Conclusions:

清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

- **Expected 1–APRX**
  I. Extend 1–APRX to an expected sense
  II. Contains a large class of low–complexity policies

- **Heavy Traffic Analysis**
  I. Establish a state–space collapse result
  II. Obtain an upper bound for the weighted queue length

- **Communication–Efficient Algorithm : MWM–AU**
  I. Significantly reduce communication frequency
  II. Achieve the same delay performance as MWM

# THANK YOU!

Yu Huang
IIIS, Tsinghua University
E-mail: y-huang20@mails.tsinghua.edu.cn