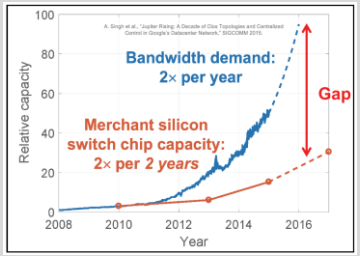
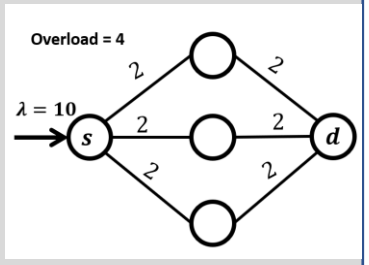
Queueing Delay Minimization in Overloaded Networks via Rate Control

Motivation

Network Overload: demand > capacity

- Occurs more frequently in datacenter due to increasing demand-capacity gap [1]
- Non-economic to provision capacity for bursty traffic (e.g. 8x than usual [2])



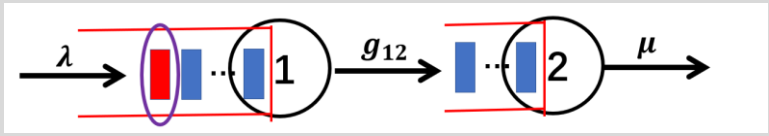


Q: Optimal rate control to minimize queueing delay under overload?

Contributions

1) Prove routing policies to minimize average and max delay simultaneously in single-hop networks.
2) Generalize the delay-optimal policies to multi-stage networks, e.g., Clos; Fat-tree.
3) Show 10% ↓ in $\bar{D}_{avg}$, 50% ↓ in $\bar{D}_{max}$ on Clos structure with different fan-in fan-out structures, compared to max-rate serving.

Delay Model



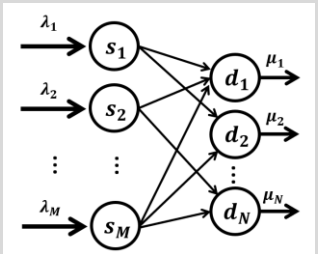
Delay of **red** packet arrived to node **1** at t

$$D_1(t) = \frac{q_1(t)}{g_{12}} + \frac{q_2\left(t + \frac{q_1(t)}{g_{12}}\right)}{\mu}$$

Main Results

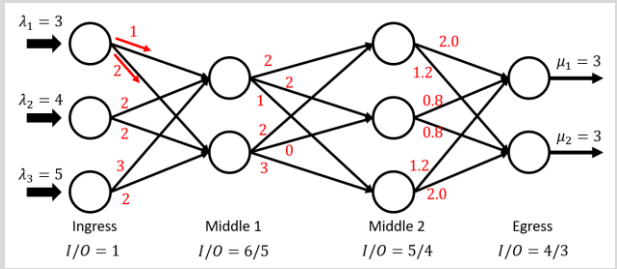
Proportional Policy Design To minimize $\bar{D}_{avg}$ & $\bar{D}_{max}$

Single-hop



$$\begin{cases} \frac{\sum_{k=1}^{N} g_{ik}(t)}{\sum_{k=1}^{N} g_{jk}(t)} = \frac{\lambda_i}{\lambda_j}, \forall i \neq j \\ \frac{\sum_{k=1}^{M} g_{ki}(t)}{\sum_{k=1}^{M} g_{kj}(t)} = \frac{\mu_i}{\mu_j}, \forall i \neq j \\ \sum_{i=1}^{M} g_{ij}(t) \geq \mu_j, \forall j = 1, \ldots, N \end{cases}$$

Clos: (keep same I/O ratio of nodes in same layer)



Extension: Fat-tree; Queue-based policy

Delay Metrics

$\bar{D}_{avg}$ & $\bar{D}_{max}$

$\bar{D}_i$: Mean delay of packets sent to s_i in [0, T]

$$\bar{D}_i = \frac{1}{T}\int_0^T D_i(t)\,dt$$

Average delay

$$\bar{D}_{avg} = \sum_{i=1}^{N} \frac{\lambda_i}{\sum_{k=1}^{N} \lambda_k} \bar{D}_i$$

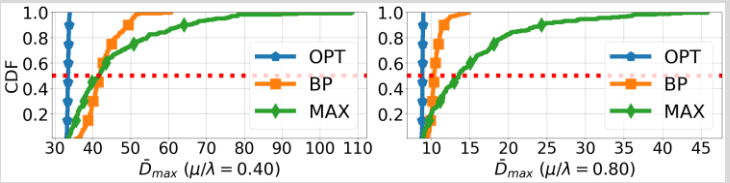
Max ingress delay

$$\bar{D}_{max} = \max_{i=1,2,\ldots,N} \bar{D}_i$$

Simulation

Our delay-optimal policy **minimizes** & **well balances** delay

Exp: 16 x 12 x 8 x 6



Clos Topology	Policy	$\bar{D}_{avg}$, Gap Mean	$\bar{D}_{avg}$, Gap Max	$\bar{D}_{max}$, Gap Mean	$\bar{D}_{max}$, Gap Max	$\bar{D}_{max}/\bar{D}_{avg}$ Mean	$\bar{D}_{max}/\bar{D}_{avg}$ Max
15x12x9x12x15	OPT	1.12	1.16	1.12	1.16	1.00	1.01
	BP	1.34	1.93	1.37	2.11	1.02	1.15
	MAX	1.49	2.28	1.52	2.34	1.02	1.07
9x12x15x12x9	OPT	1.12	1.16	1.12	1.16	1.00	1.00
	BP	1.53	2.70	1.56	2.71	1.02	1.09
	MAX	1.45	2.74	1.47	2.76	1.01	1.07
12x12x12x12x12	OPT	1.12	1.16	1.12	1.16	1.00	1.00
	BP	1.41	2.49	1.44	2.72	1.02	1.09
	MAX	1.51	2.64	1.54	2.70	1.02	1.07

[1] Singh, Arjun, et al. "Jupiter rising: A decade of clos topologies and centralized control in google's datacenter network." *ACM SIGCOMM computer communication review* 45.4 (2015): 183-197.
[2] Zhang, Yiwen, et al. "Aequitas: admission control for performance-critical RPCs in datacenters." *Proceedings of the ACM SIGCOMM 2022 Conference*. 2022.

X. Wu, D. Wu and E. Modiano, "Queueing Delay Minimization in Overloaded Networks via Rate Control," in Allerton Conference 2022. (complete version to be submitted to IEEE Transactions on Networking)