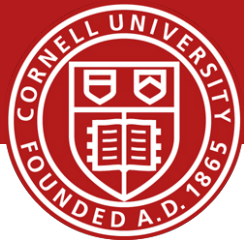


Alternative Switching Technologies: Optical Circuit Switches

Hakim Weatherspoon

Assistant Professor, Dept of Computer Science

CS 5413: High Performance Systems and Networking

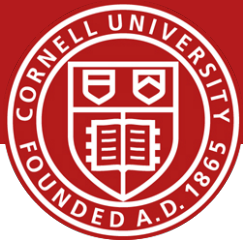
April 12, 2017

Agenda for semester

- Project
  - Continue to make progress.
  - **Intermediate project report 2 due TODAY, Wednesday, April 12th.**
  - **BOOM, next week, Wednesday, April 19**
  - **End of Semester presentations/demo, Wednesday, May 10**

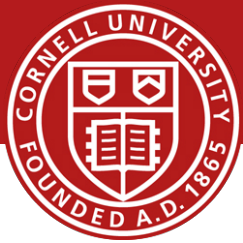
- Check website for updated schedule

Where are we in the semester?



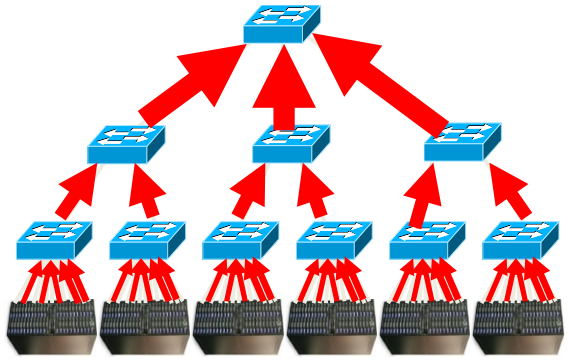
- Interested Topics:
  - SDN and programmable data planes
  - Disaggregated datacenters and rack-scale computers
  - Alternative switch technologies
  - Datacenter topologies
  - Datacenter transports
  - Advanced topics

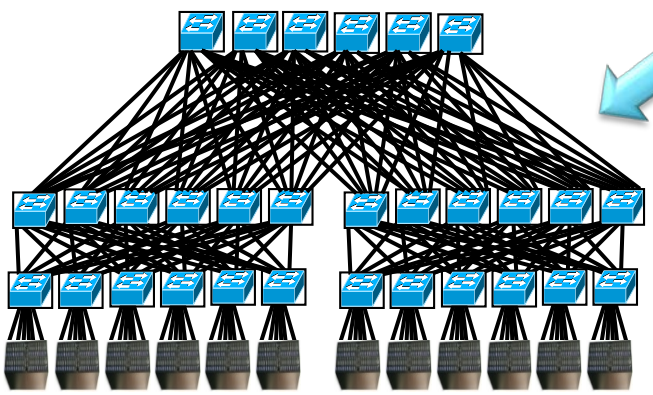
Goals for Today



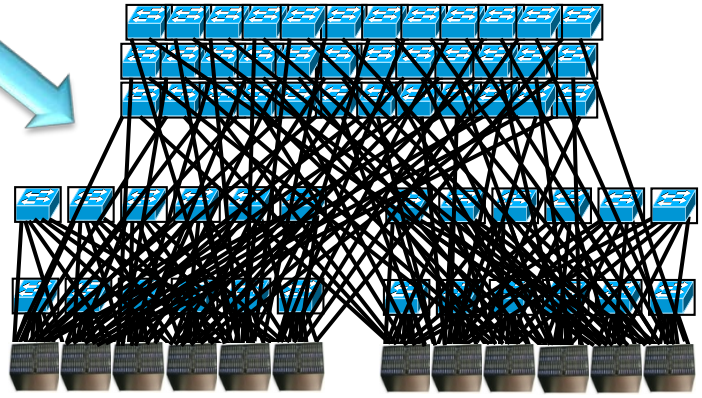
- c-through: part-time optics in datacenters
  - G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. S. Eugene Ng, M. Kozuch, M. Ryan. ACM SIGCOMM Computer Communication Review (CCR), Volume 40, Issue 4 (October 2010), pages 327-338.

Current solutions for increasing data center network bandwidth





FatTree

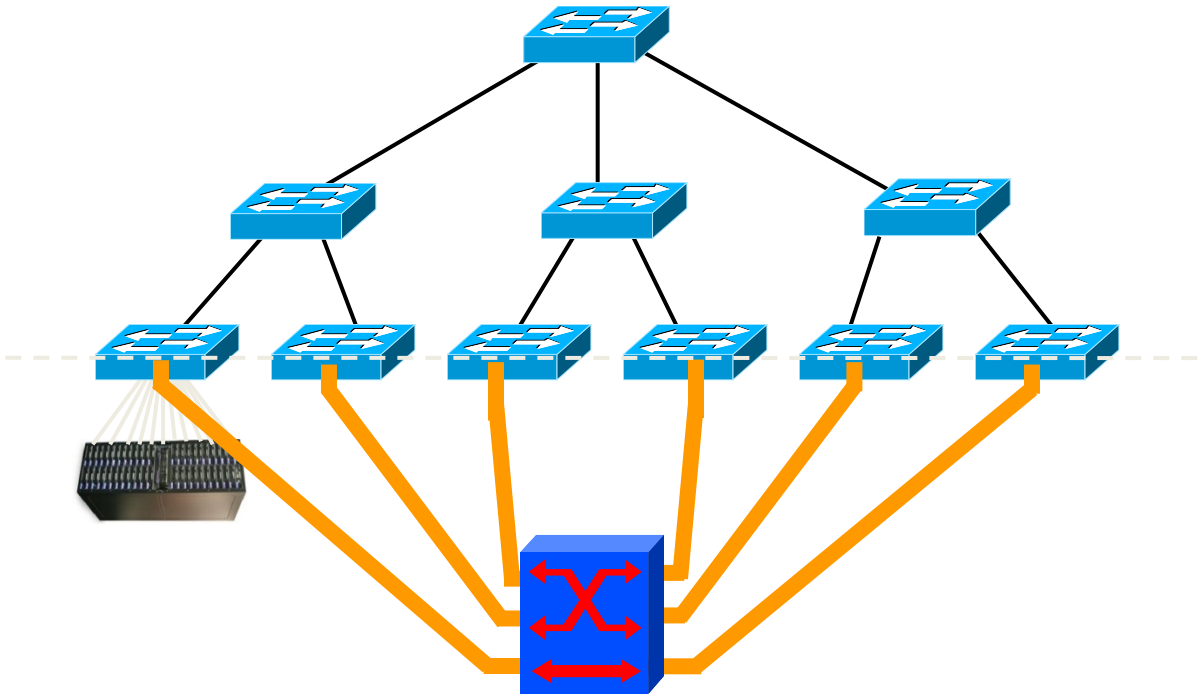


BCube

1. Hard to construct

2. Hard to expand

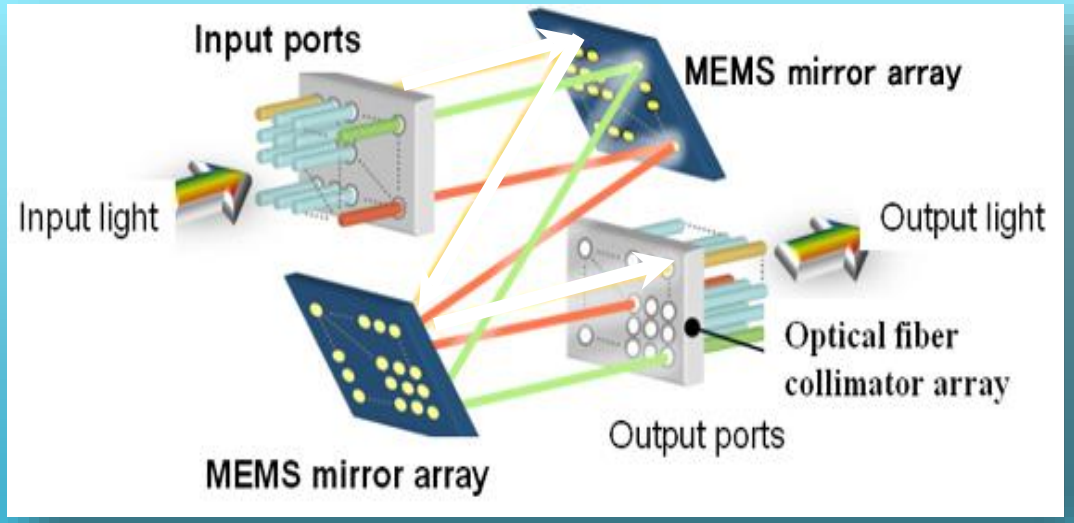
An alternative: hybrid packet/circuit switched data center network



- Goal of this work:
  - Feasibility: software design that enables efficient use of optical circuits
  - Applicability: application performance over a hybrid network

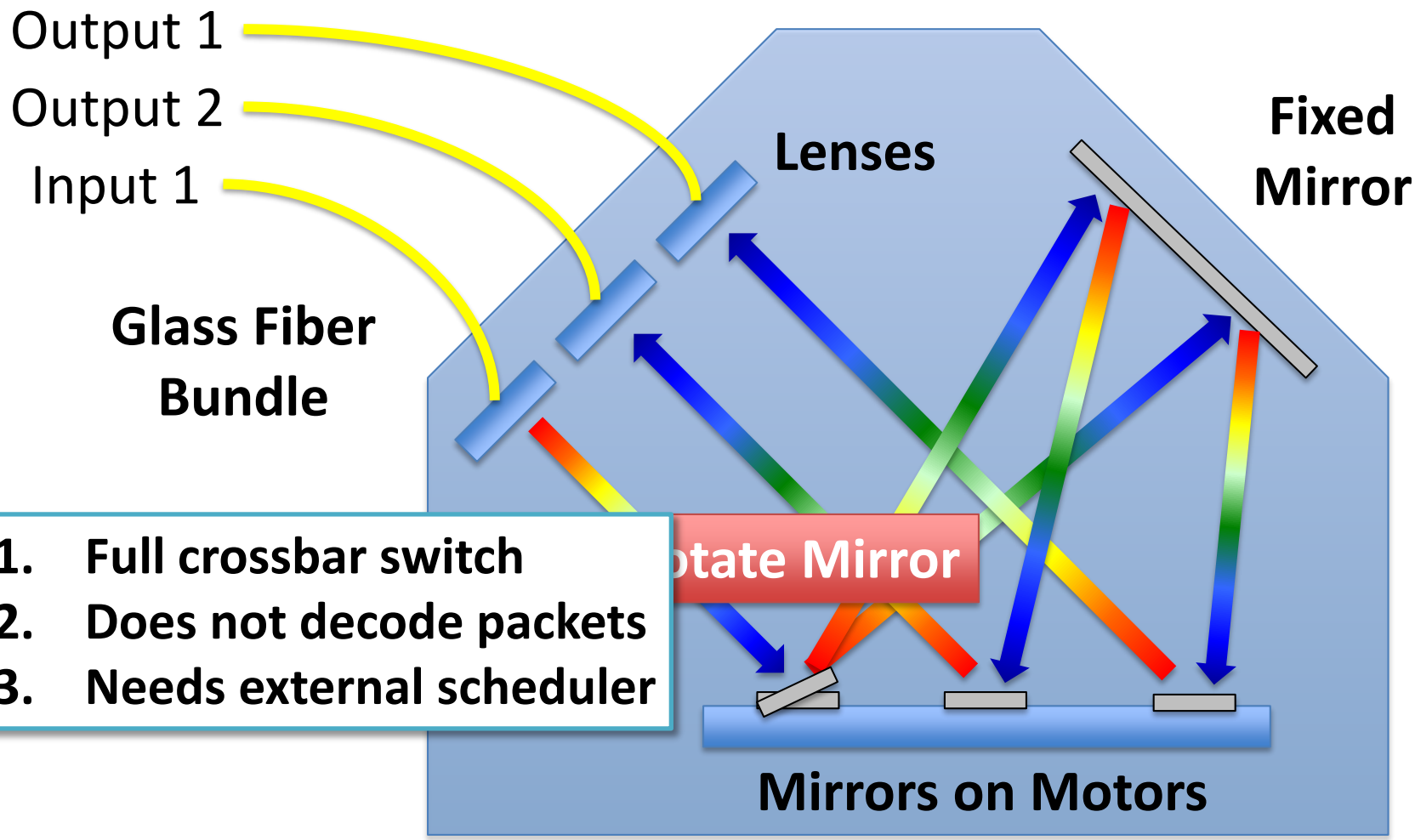
Optical circuit switching v.s. Electrical packet switching

	Electrical packet switching	Optical circuit switching
Switching technology	Store and forward	Circuit switching
Switching capacity	1 e	et, ect
Switching time	F	



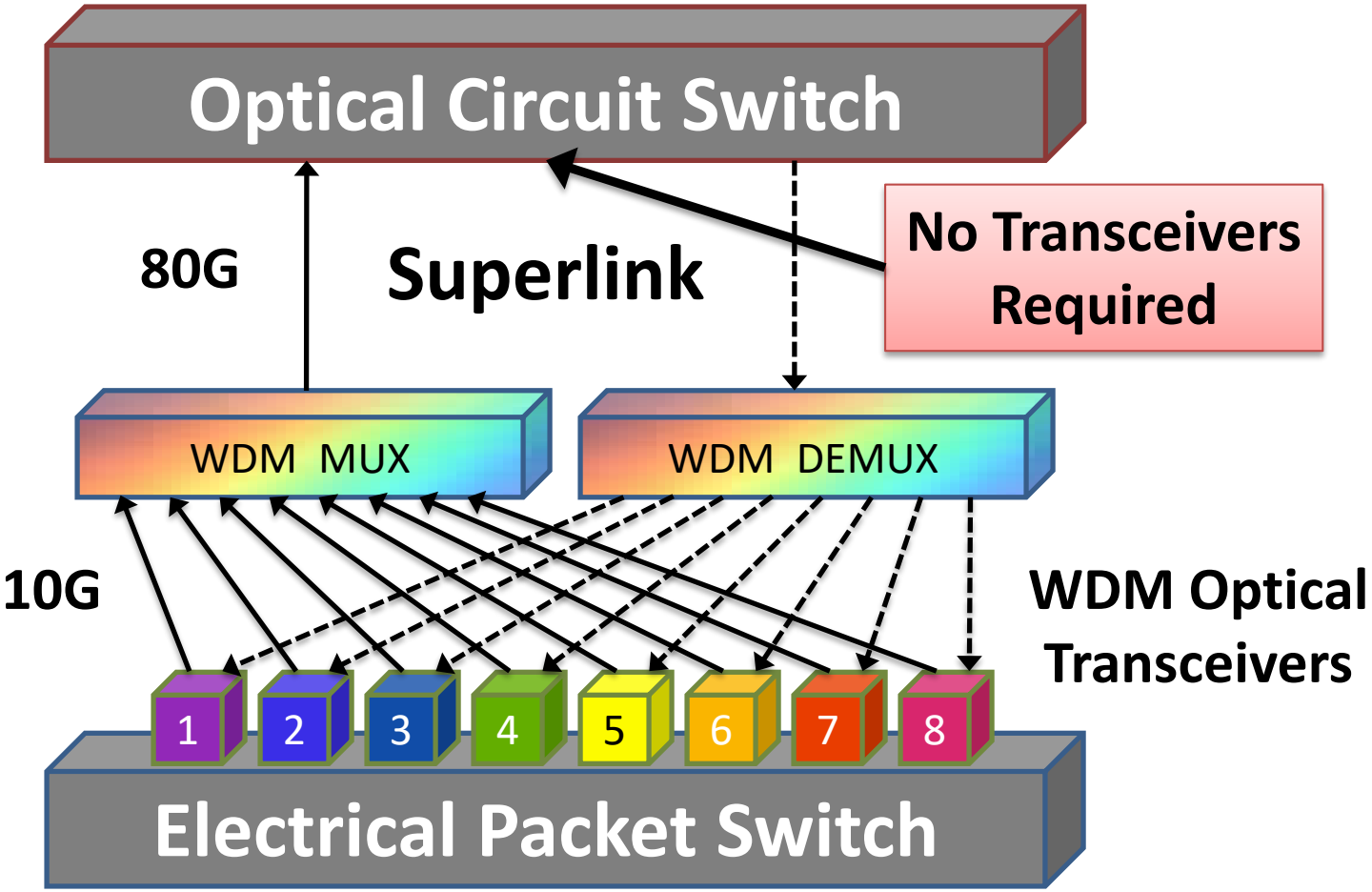
e.g. MEMS optical switch

Technology: Optical Circuit Switch

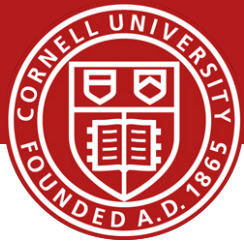


1. Full crossbar switch
2. Does not decode packets
3. Needs external scheduler

Wavelength Division Multiplexing



Optical circuit switching is promising despite slow switching time

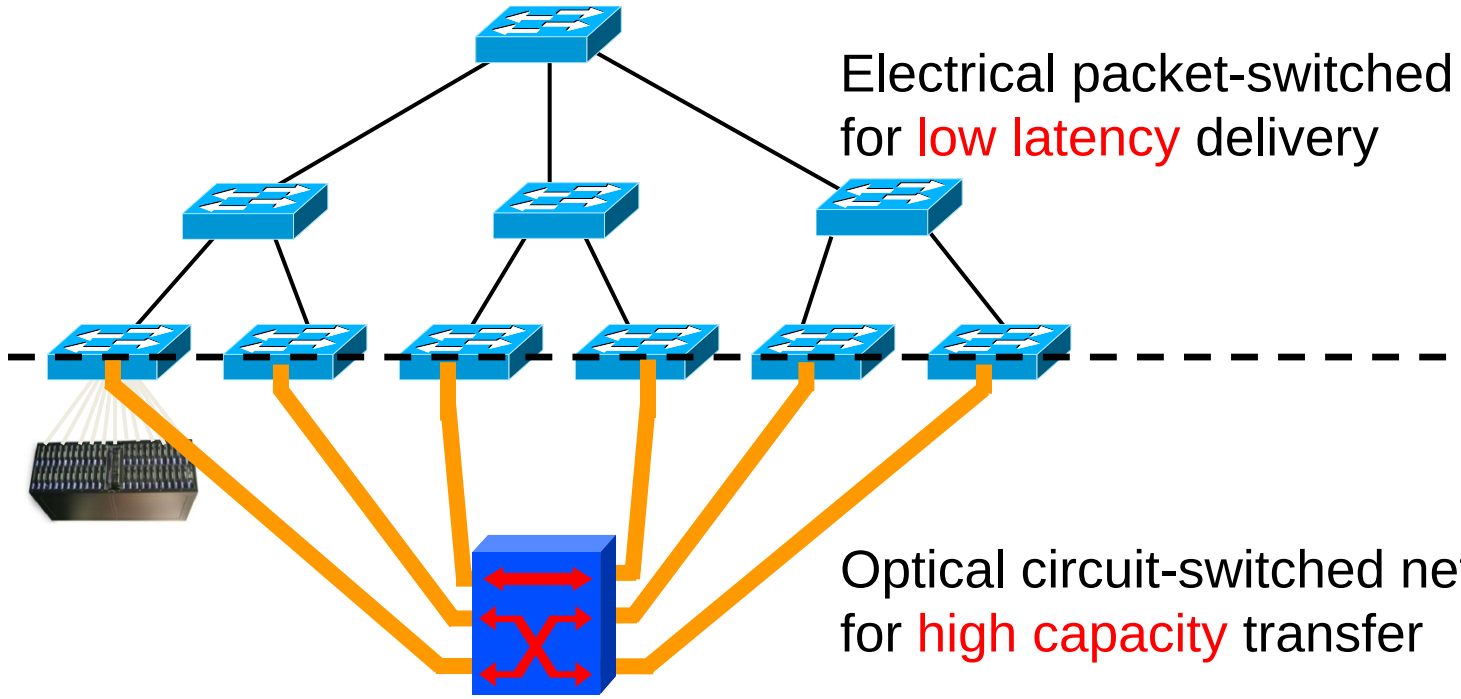


- [IMC09][HotNets09]: "*Only a few ToRs are hot and most their traffic goes to a few other ToRs. …*"

- [WREN09]: "*…we find that traffic at the five edge switches exhibit an ON/OFF pattern…* "

Full bisection bandwidth at packet granularity may not be necessary

Hybrid packet/circuit switched network architecture

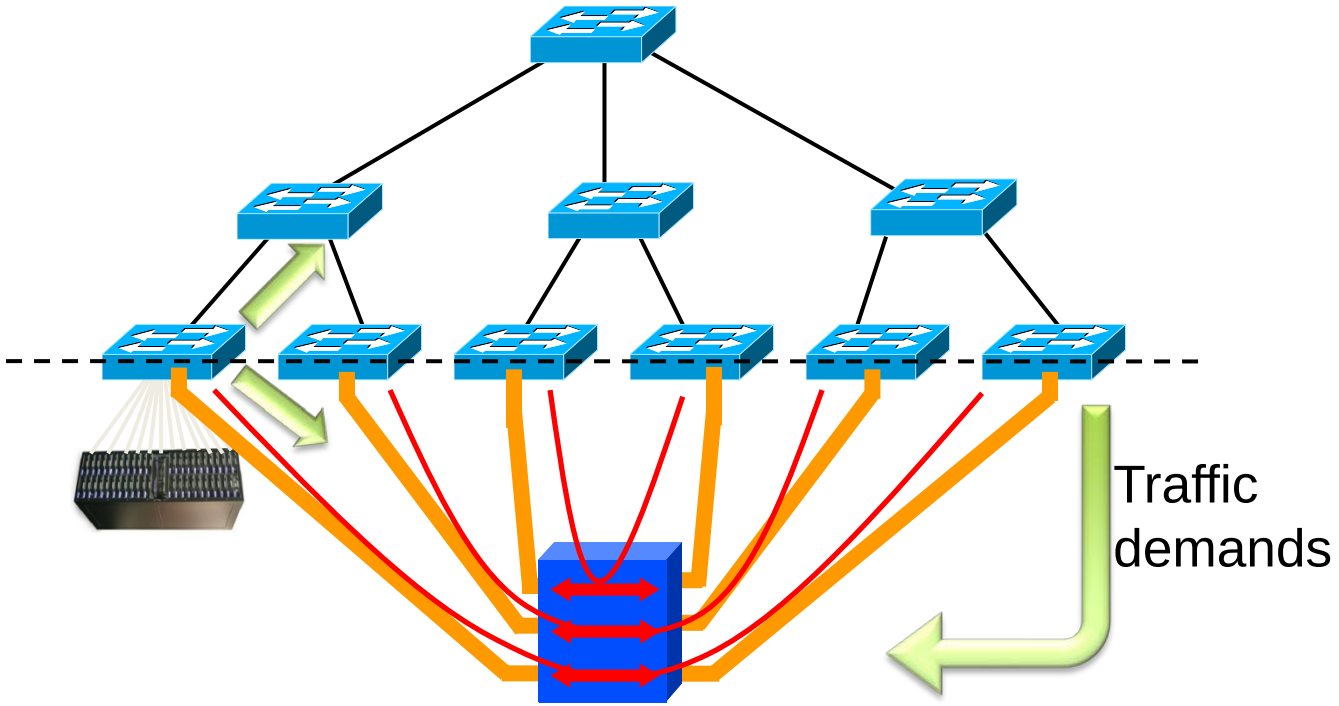


Electrical packet-switched network for low latency delivery

Optical circuit-switched network for high capacity transfer

- Optical paths are provisioned rack-to-rack
  - A simple and cost-effective choice
  - Aggregate traffic on per-rack basis to better utilize optical circuits

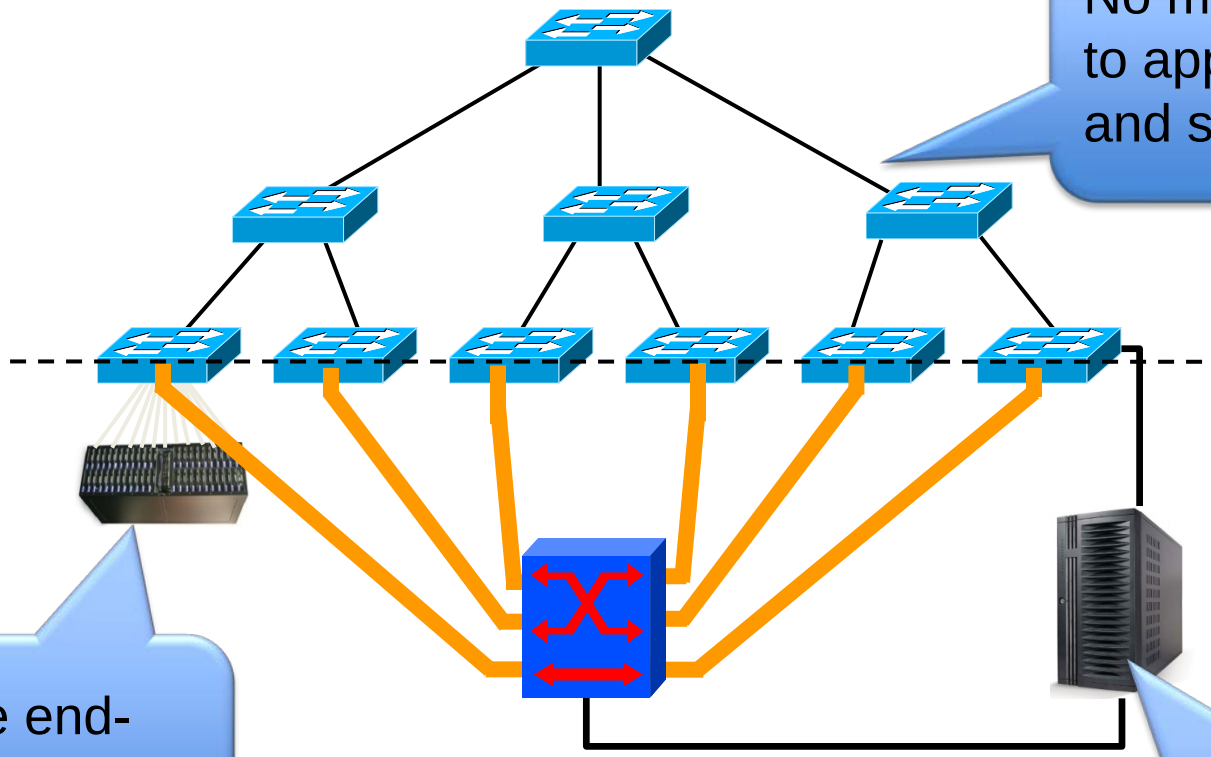
Design requirements



- Control plane:
  - Traffic demand estimation
  - Optical circuit configuration

- Data plane:
  - Dynamic traffic de-multiplexing
  - Optimizing circuit utilization (optional)

c-Through (a specific design)

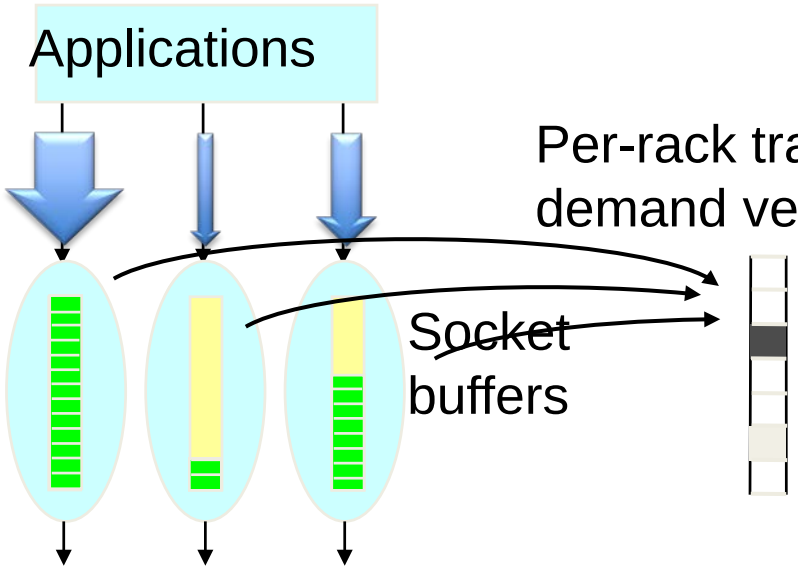


No modification to applications and switches

Leverage end-hosts for traffic management

Centralized control for circuit configuration

c-Through - traffic demand estimation and traffic batching

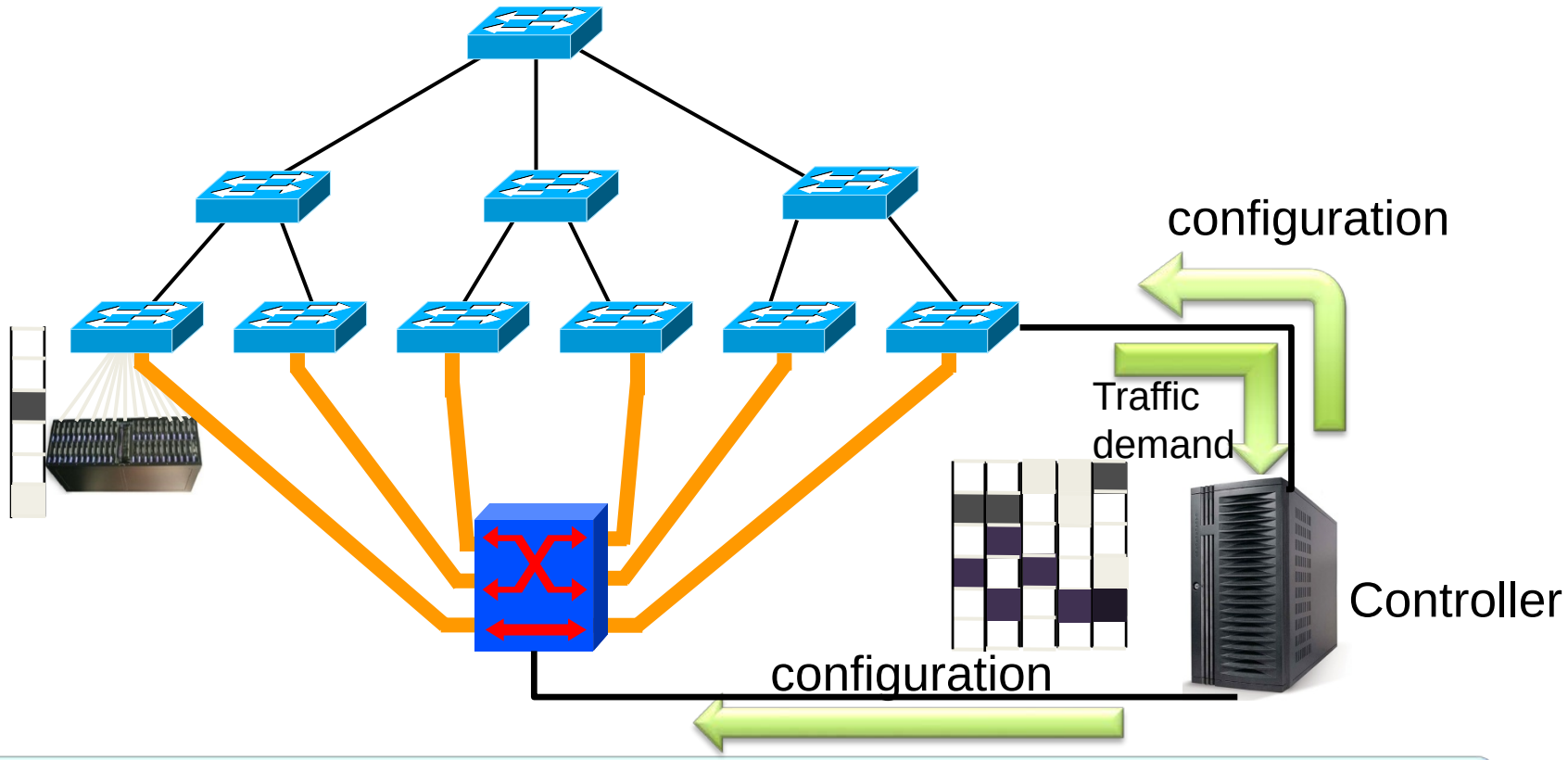


1. Transparent to applications.

2. Packets are buffered per-flow to avoid HOL blocking.

- Accomplish two requirements:
  - Traffic demand estimation
  - Pre-batch data to improve optical circuit utilization

c-Through - optical circuit configuration

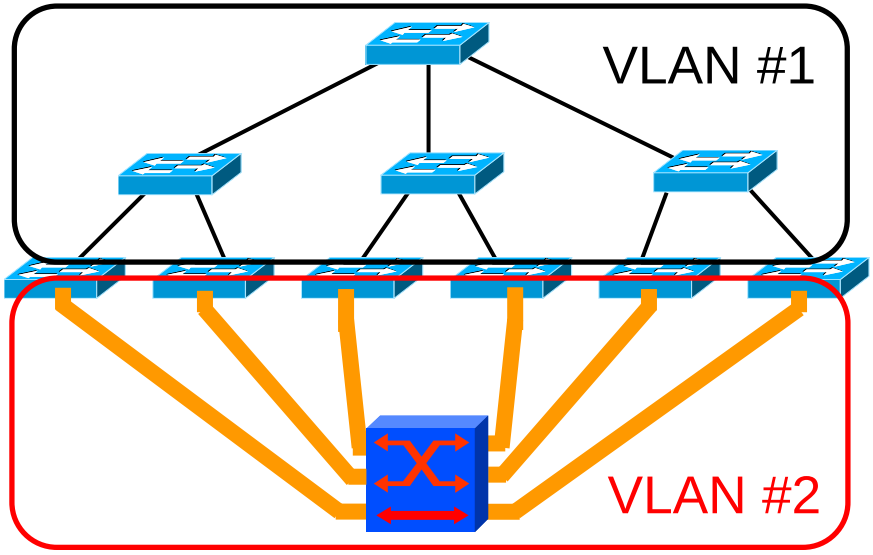


Use Edmonds' algorithm to compute optimal configuration

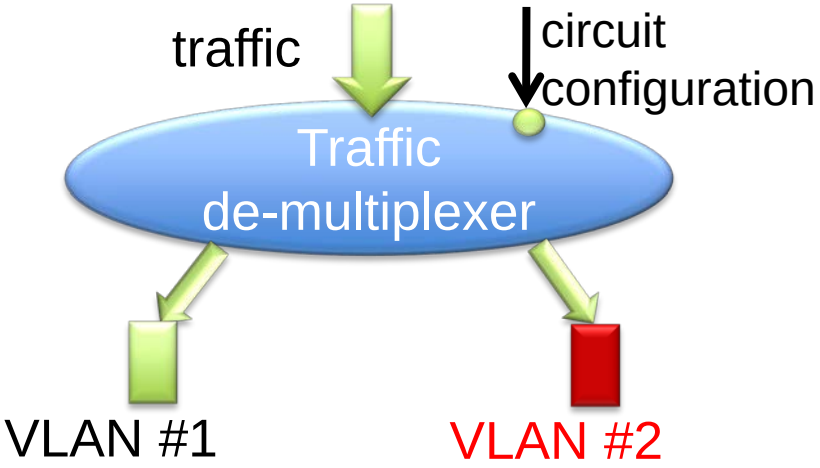
Many ways to reduce the control traffic overhead

c-Through - traffic de-multiplexing

- VLAN-based network isolation:
  - No need to modify switches
  - Avoid the instability caused by circuit reconfiguration

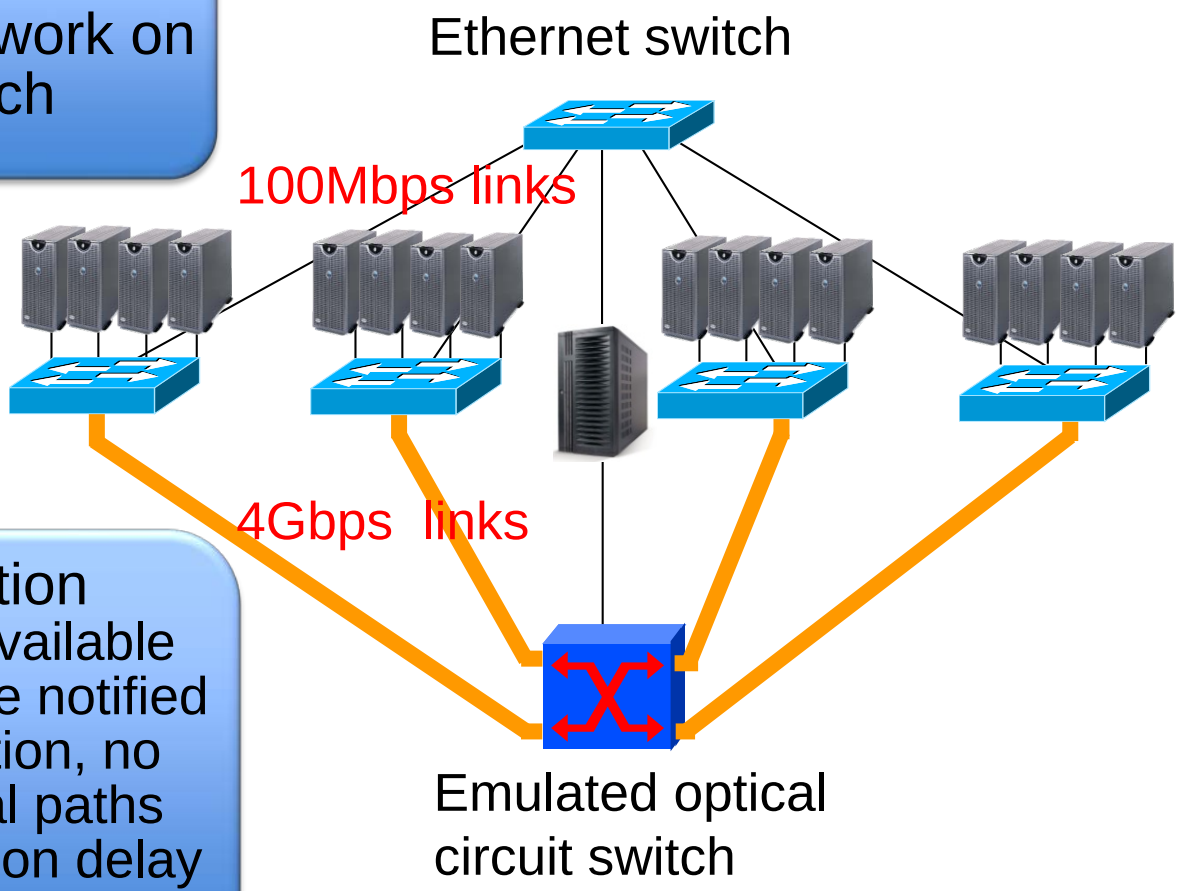


- Traffic control on hosts:
  - Controller informs hosts about the circuit configuration
  - End-hosts tag packets accordingly



Testbed setup

- 16 servers with 1Gbps NICs
- Emulate a hybrid network on 48-port Ethernet switch

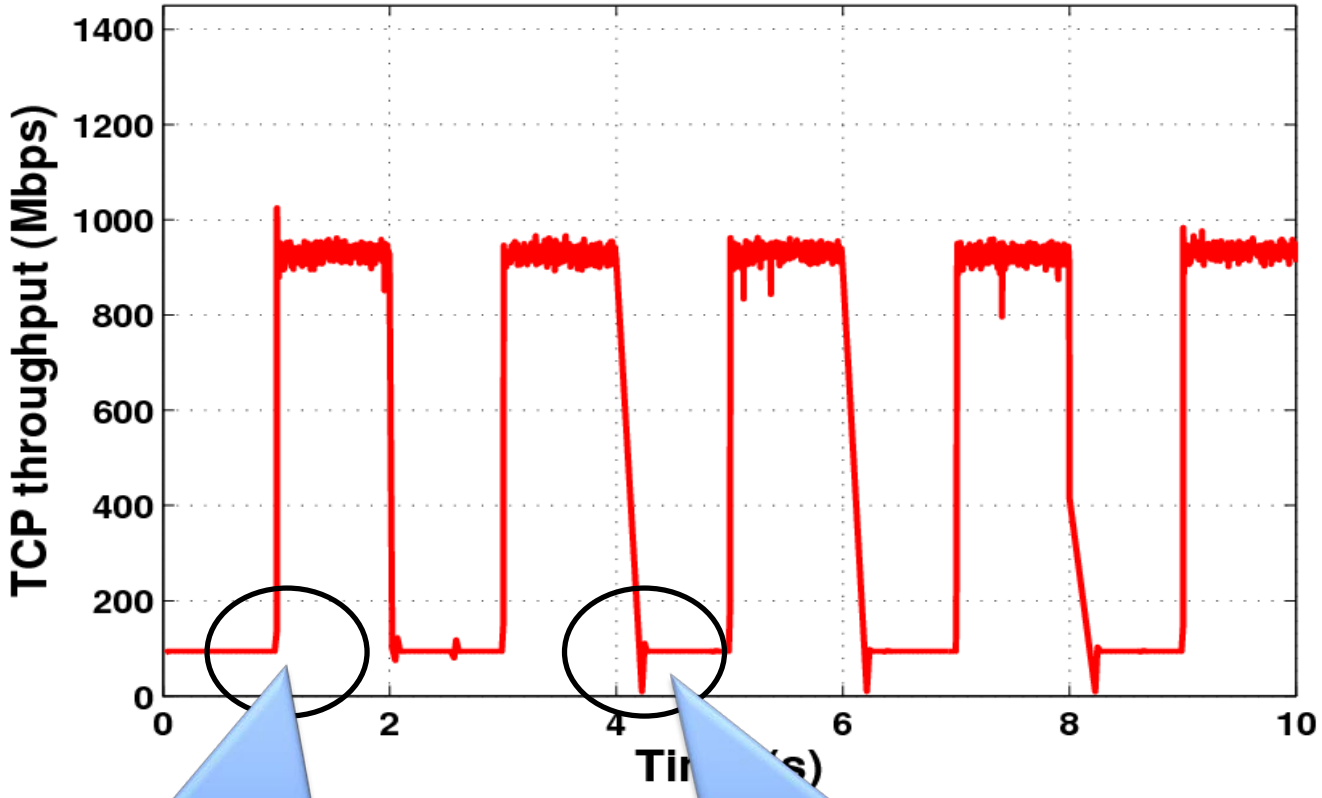


- Optical circuit emulation
  - Optical paths are available only when hosts are notified
  - During reconfiguration, no host can use optical paths
  - 10 ms reconfiguration delay

Evaluation

- Basic system performance:
  - Can TCP exploit dynamic bandwidth quickly?
  - Does traffic control on servers bring significant overhead?
  - Does buffering unfairly increase delay of small flows?
- Application performance:
  - Bulk transfer (VM migration)?
  - Loosely synchronized all-to-all communication (MapReduce)?
  - Tightly synchronized all-to-all communication (MPI-FFT) ?

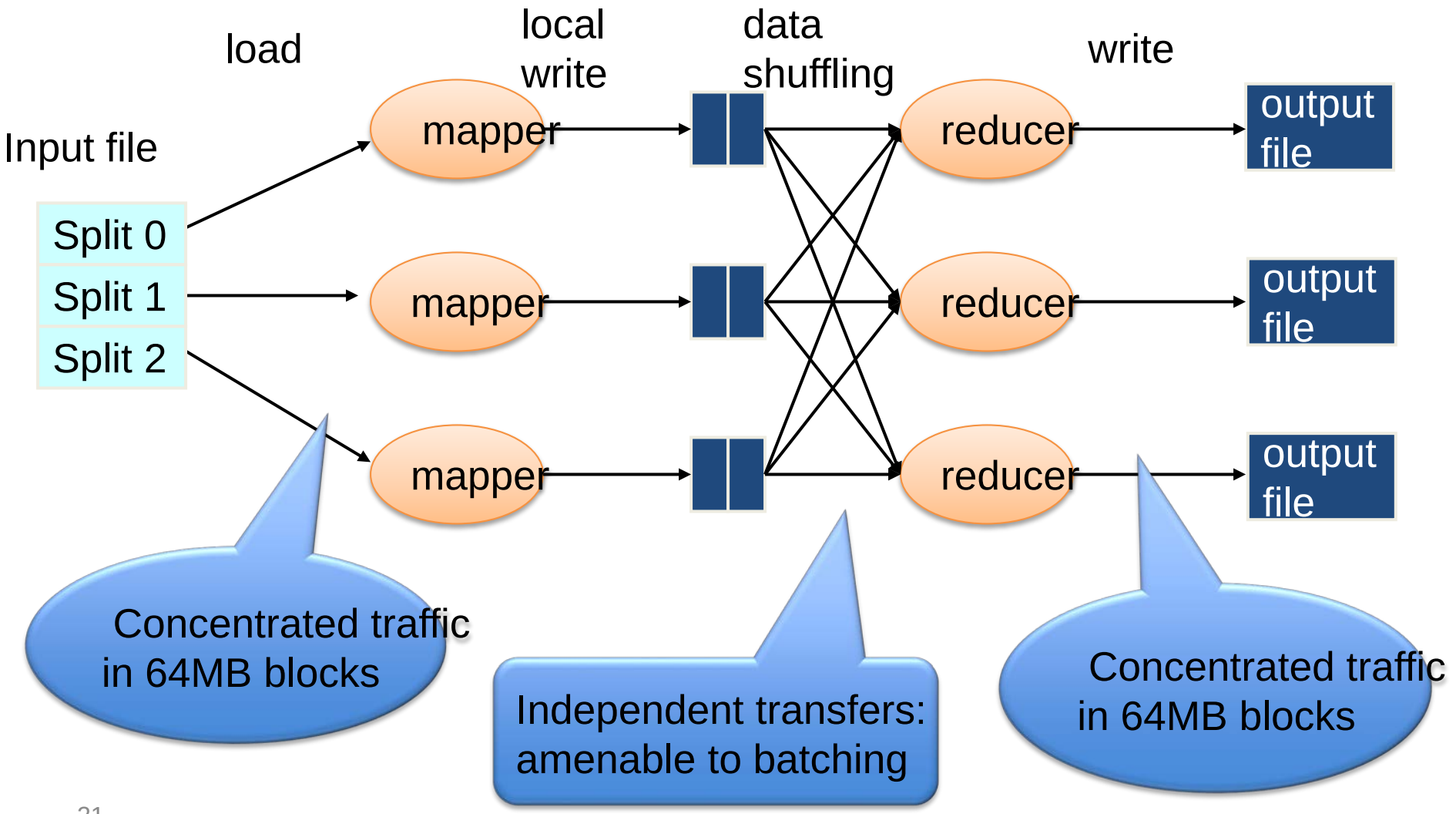
TCP can exploit dynamic bandwidth quickly



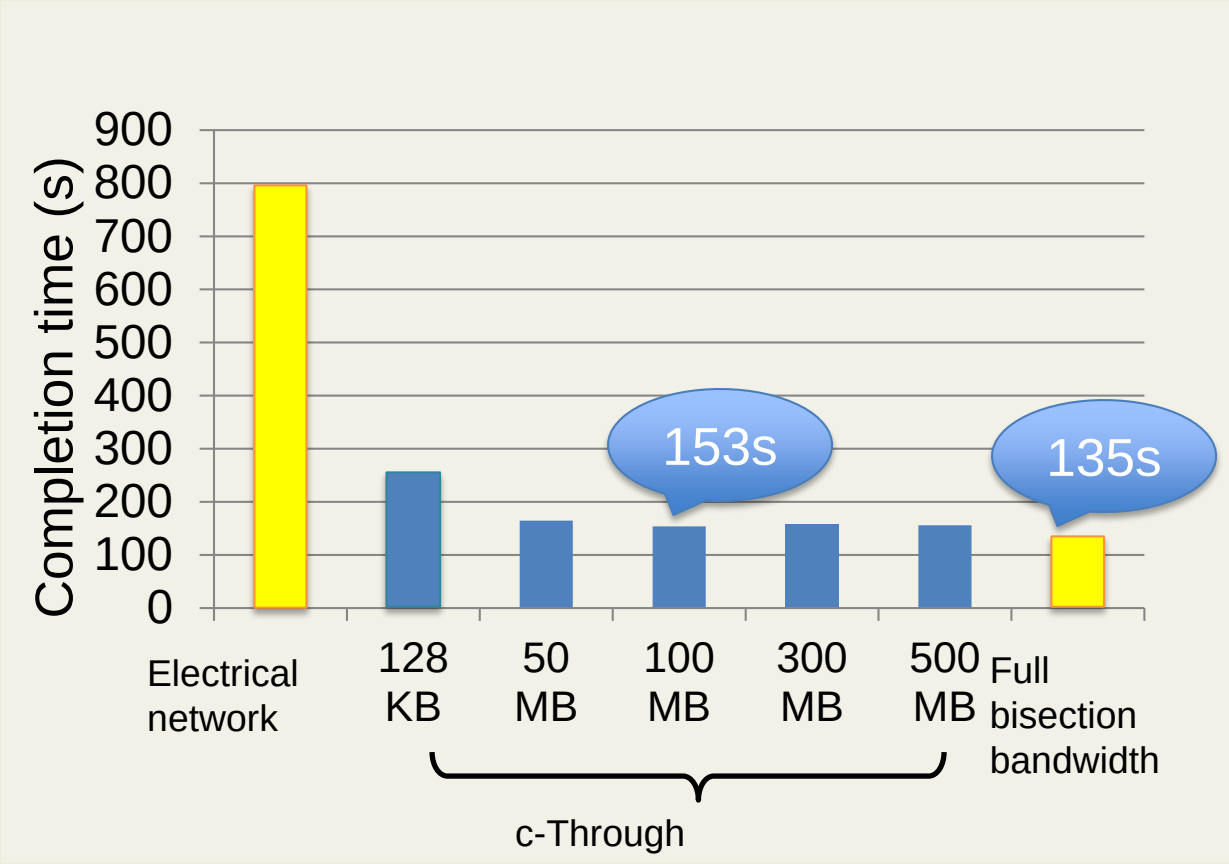
Throughput ramps up within 10 ms

Throughput stabilizes within 100ms

MapReduce Overview

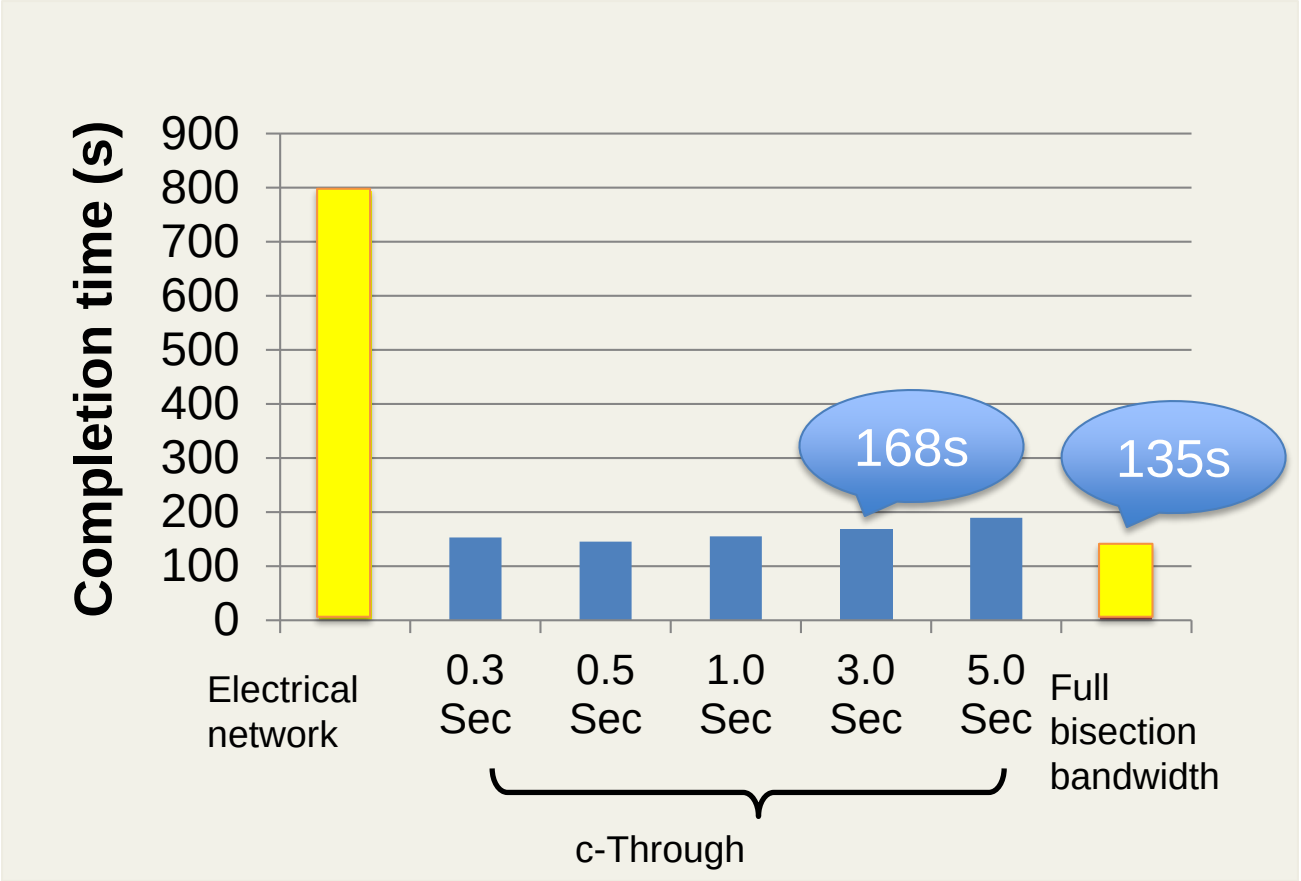


MapReduce sort 10GB random data



c-Through varying socket buffer size limit (reconfiguration interval: 1 sec)

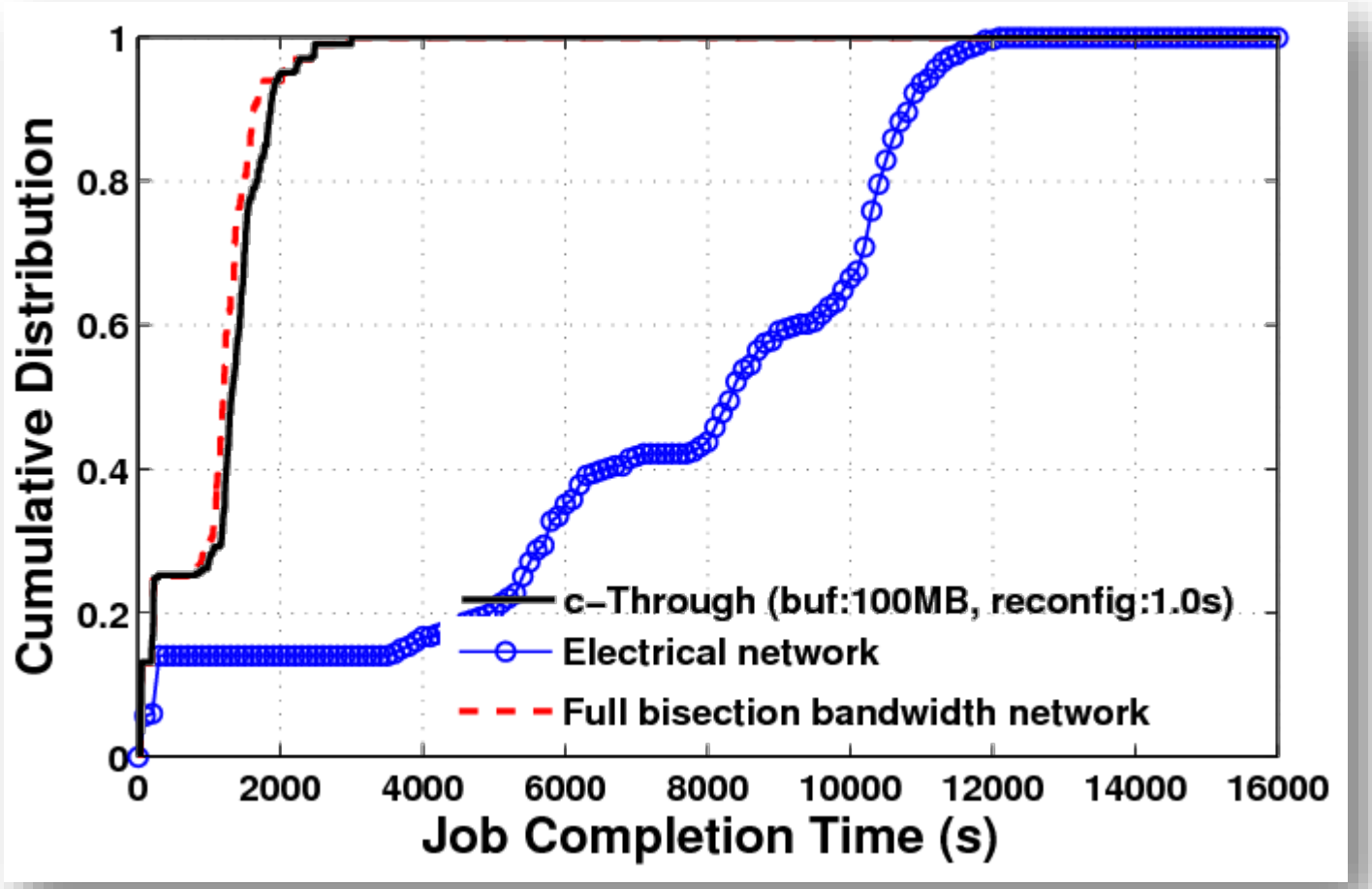
MapReduce sort 10GB random data



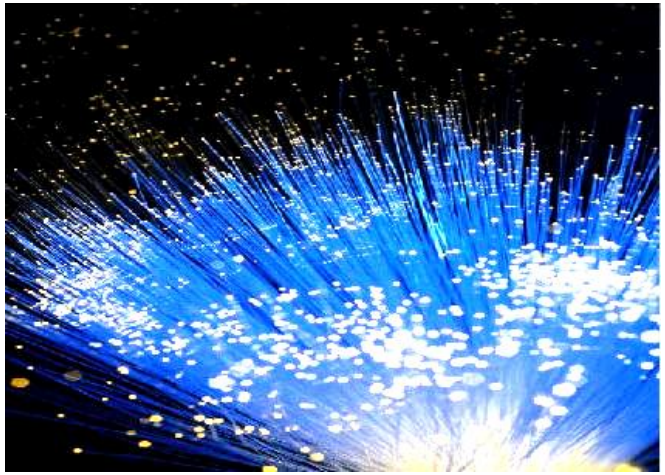
c-Through varying reconfiguration interval
(socket buffer size limit: 100MB)

Yahoo Gridmix benchmark

- 3 runs of 100 mixed jobs such as web query, web scan and sorting
- 200GB of uncompressed data, 50 GB of compressed data



Summary



- Hybrid packet/circuit switched data center network
  - c-Through demonstrates its feasibility
  - Good performance even for applications with all to all traffic

- Future directions to explore:
  - The scaling property of hybrid data center networks
  - Making applications circuit aware
  - Power efficient data centers with optical circuits

Picture from Internet websites.

Related Work

	Link Technology	Modifications Required	Working Prototype
Helios (SIGCOMM '10)	Optics w/ WDM 10G-180G (CWDM) 10G-400G (DWDM)	Switch Software	Glimmerglass, Fulcrum
c-Through (SIGCOMM '10)	Optics (10G)	Host OS	Emulation
Flyways (SIGCOMM '11, HotNets '09)	Wireless (1G, 10m)	Unspecified	
IBM System-S (GLOBECOM '09)	Optics (10G)	Host Application; Specific to Stream Processing	Calient, Nortel
HPC (SC '05)	Optics (10G)	Host NIC Hardware	

Agenda for semester

- Project
  - Continue to make progress.
  - **Intermediate project report 2 due TODAY, Wednesday, April 12th.**
  - **BOOM, next week, Wednesday, April 19**
  - **End of Semester presentations/demo, Wednesday, May 10**

- Check website for updated schedule