Optimal Scheduling for Wireless Assisted All-Optical Data Center Networks

Chang-Heng Wang, Prof. Tara Javidi
Department of Electrical and Computer Engineering, University of California, San Diego

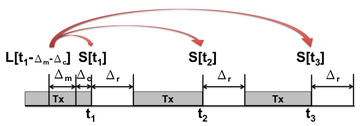
center for networked systems

Motivation

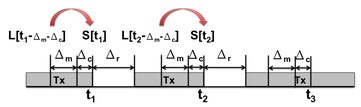
- Modern data centers usually consist of hundreds of thousands of servers and intensive data exchanges occur within data centre networks.
- Ever increasing data bandwidth requirement (40 Gbps, 100 Gbps, or beyond) and number of port counts become bottlenecks for traditional electronic data switch.
- **Optical switches** have the advantages in scalability and lower power consumption, and reduce the use of optical-electrical-optical (O/E/O) conversion needed in the conjunction of optical fibers and electronic switches. These advantages
- However, the down time of optical switches when performing schedule reconfiguration imposes a great challenge on optimal scheduling policy design.

Key Insights

1. Dynamic Scheduling:
   Dynamic scheduling utilizes the most recent queue information for each schedule selection.



(a) Quasi-Static Scheduling

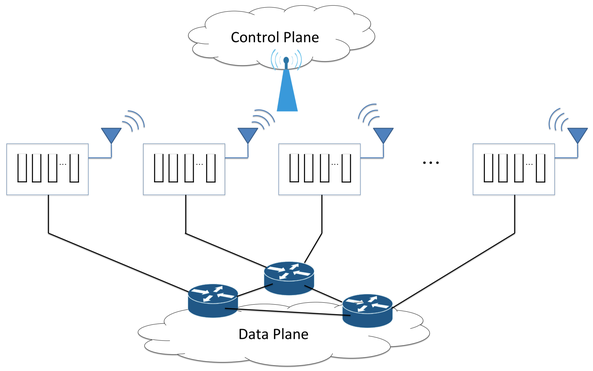


(b) Dynamic Scheduling

2. Queue Monitoring:
   Scheduling policies utilizing queue information could schedule network traffic without assumptions on traffic pattern.
   The freshness of the queue information then becomes an important factor for the performance.

System Overview

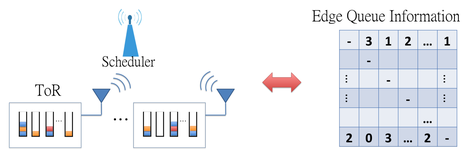
- Network architecture:



- System Parameters:
  - Monitoring Time Δ_m
  - Computation Time Δ_c
  - Reconfiguration Time Δ_r
- **Data plane**: Consists of optical switches and optical fiber interconnections, and is responsible for data exchange within the network.

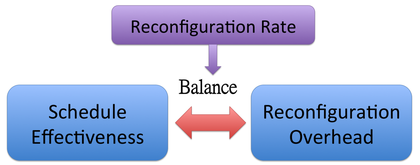
* Current feasible technology supports Δ_r as low as tens $\sim 20\ \mu s$.

- **Control plane**: Collects queue information from top of rack (ToR) switches through a wireless channel. The control plane computes schedules based on collected queue information and then disseminates the schedule to ToR switches and optical switches.



* For a network up to few hundreds of ToRs, 60 GHz technolgy supports $\Delta_m < 10\ ms$.

Scheduling Objectives



* For stability, duty cycle D and traffic load ρ need to satisfy

$$D = \frac{T - \Delta_r}{T} > \rho$$

Dynamic Scheduling Policies

- Based on queue information $\mathbf{L}[t]$, each schedule $\mathbf{S}$ has a weight $w[t] = \langle \mathbf{S}, \mathbf{L}[t] \rangle$ at time t
- Maximum Weight Matching (MWM) denotes the schedule $\mathbf{S}^*$ with the largest weight $w^*[t]$
- We propose two dynamic scheduling policies based on MWM:

Periodic Maximum Weight Matching

- Pick MWM schedule at schedule change
- Change Schedule every T microseconds

Adaptive Maximum Weight Matching

- Pick MWM schedule at schedule change
- Change Schedule when $w[t] < \gamma w^*[t]$

Simulation Setup

- Number of ToR switches: N = 32
- Link rate: B = 100 Gbps, packet size = 1.5Kb
- System parameters setup:
  - $\Delta_r = 20\ \mu s\ /\ 1 \sim 200\ \mu s$
  - $\Delta_m = 0\ /\ 0 \sim 20\ ms$
  - $\Delta_c \approx 0$

Simulation Results

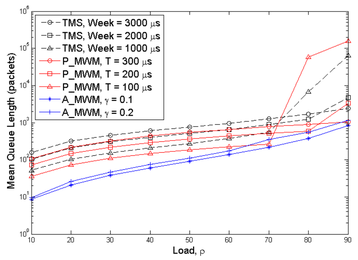


Fig.1: Performance under varying workload

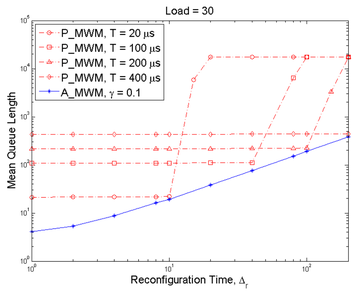


Fig.2: Performance v.s. reconfiguration time Δ_r

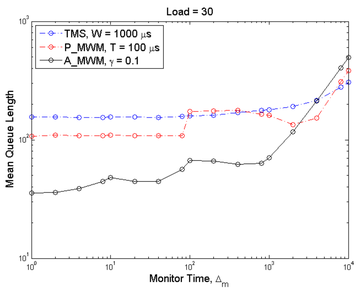


Fig.3: Performance v.s. monitoring time Δ_m

Conclusions

- Proposed network architecture separates the control plane and data plane, which enables improvement from technology advancement of both ends.
- Periodic MWM (P_MWM) performs well if the workload is known in advanced.
- Adaptive MWM (A_MWM) performs well even without prior knowledge of the load.
- Simulations showed that both reconfiguration time Δ_r and queue monitoring time Δ_m impact the performance.