# Priority-aware Scheduling for Packet Switched Optical Networks in Datacenter

**Speaker: Lin Wang**

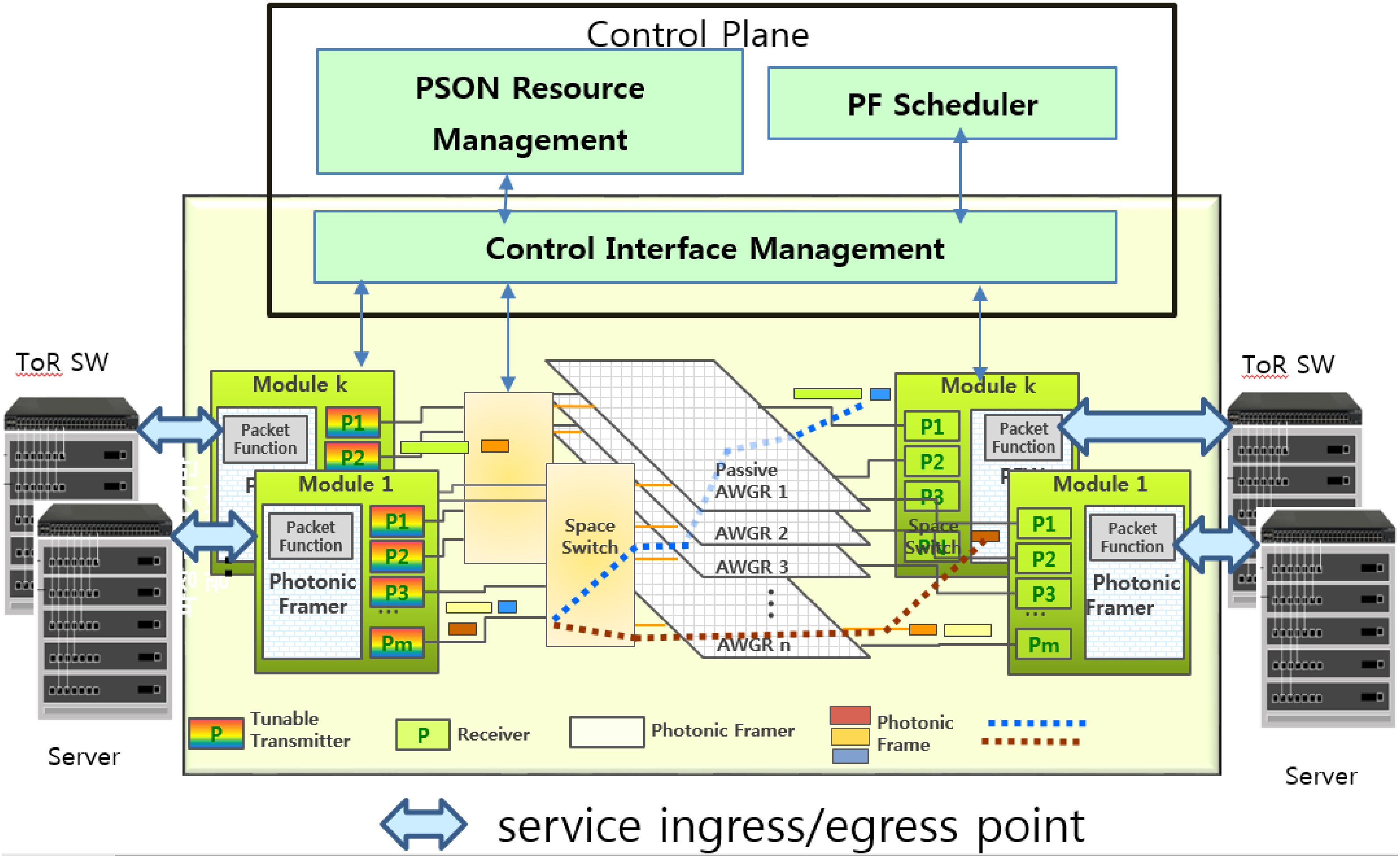
Research Advisor: Biswanath Mukherjee

UCDAVIS

## PSON architecture

- **Switch architectures and centralized controller**
- **Scheduling algorithm design**
- **Traffic generation**
- **Simulation set up**
- **Results evaluation**
- **Current work**

# PSON Architecture

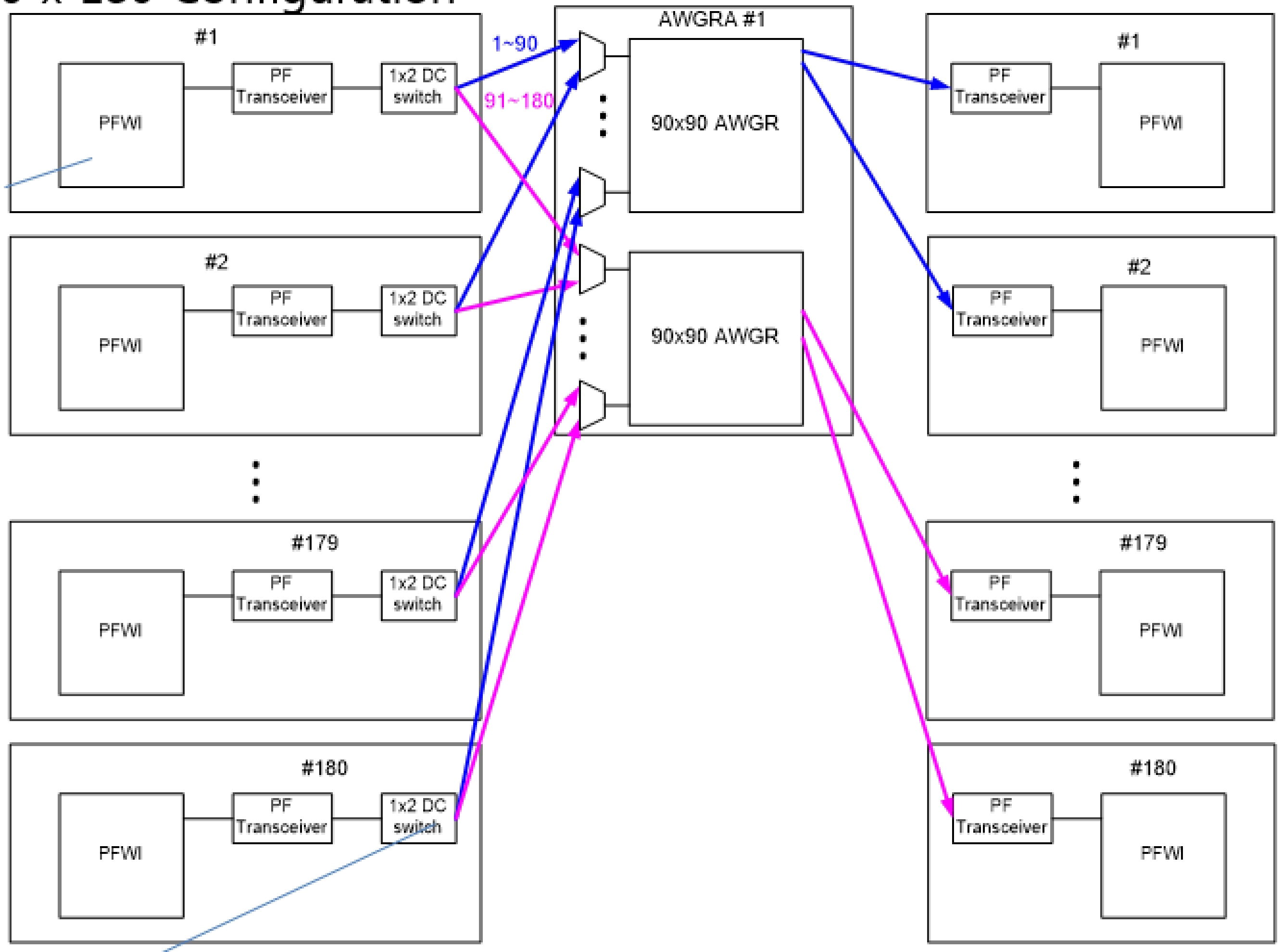



A control plane manages tunable transmitters, photonic framers and space switches for data plane with optical switch fabrics (AWGR)

# PSON data plane (with optical switch fabric)

- 180 x 180 Configuration

Photonic Framer




Space switch: Optical path switch to switch optical signal between module and AWGR at sub micro-second speed.

# Scheduling Algorithm for PSON

- **Iterative Round Robin algorithm.**

**Step 1: *Request.***

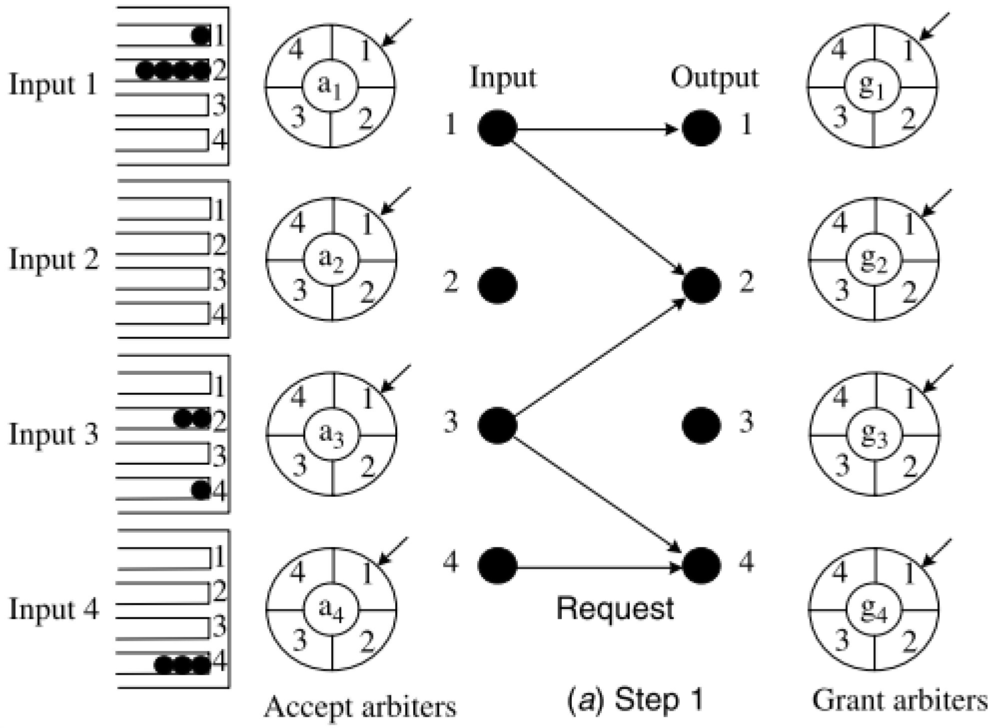
**Step 2: *Grant.***

**Step 3: *Accept.***

# Scheduling Algorithm for PSON

- **Step 1. Request**

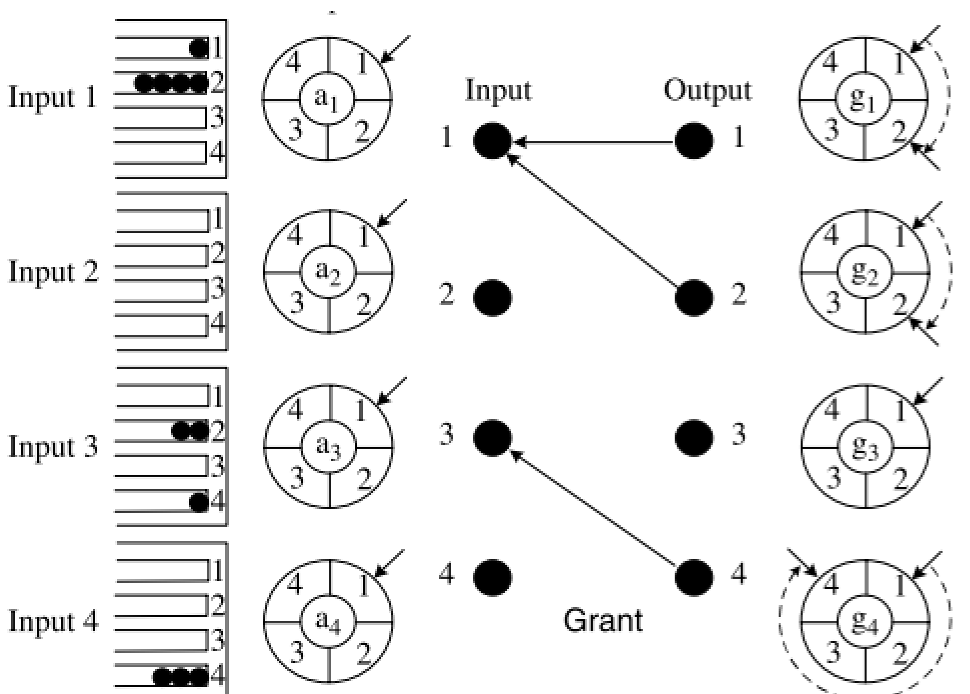
Each unmatched input sends a request to every output for which it has a queued cell.



# Scheduling Algorithm for PSON

**Step 2: *Grant.***

If an output receives multiple requests, it chooses the one that appears next in a fixed RR schedule starting from the highest priority element. The grant pointer gi is incremented (module N) to one location beyond the granted input if and only if the grant is accepted in step 3 of the first iteration.

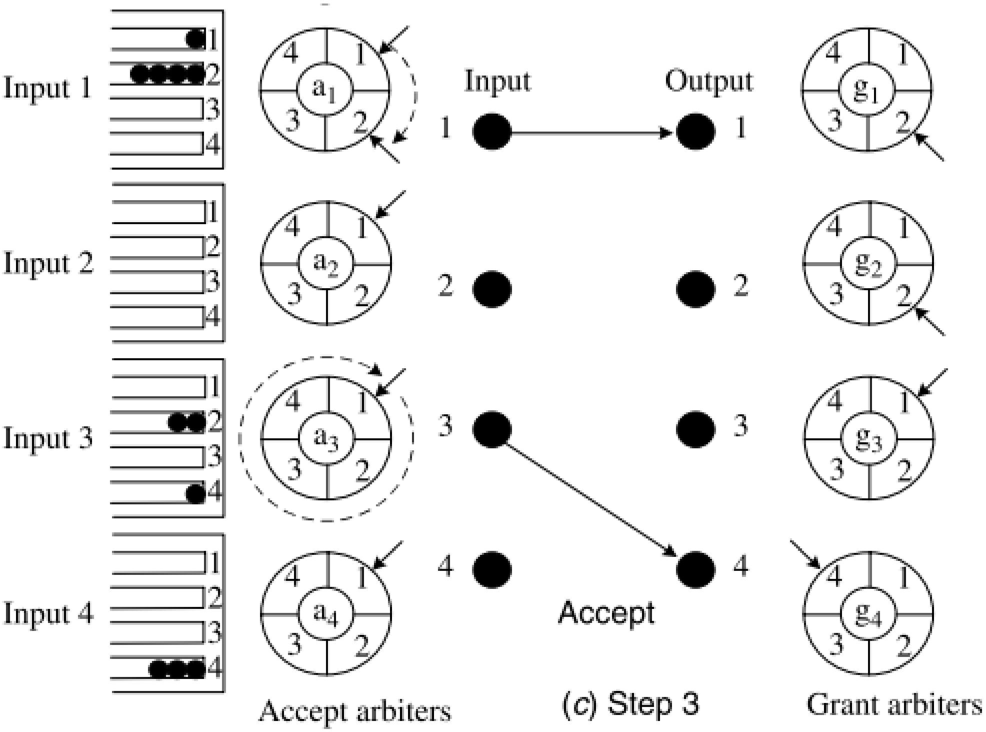



(*b*) Step 2

# Scheduling Algorithm for PSON

**Step 3: *Accept.***

If an input receives multiple grants, it accepts the one that appears next
in a fixed, round-robin schedule starting from the highest priority element. The pointer aj is incremented (modulo N) to one location beyond the accepted output. The accept pointers ai are only updated in the first iteration



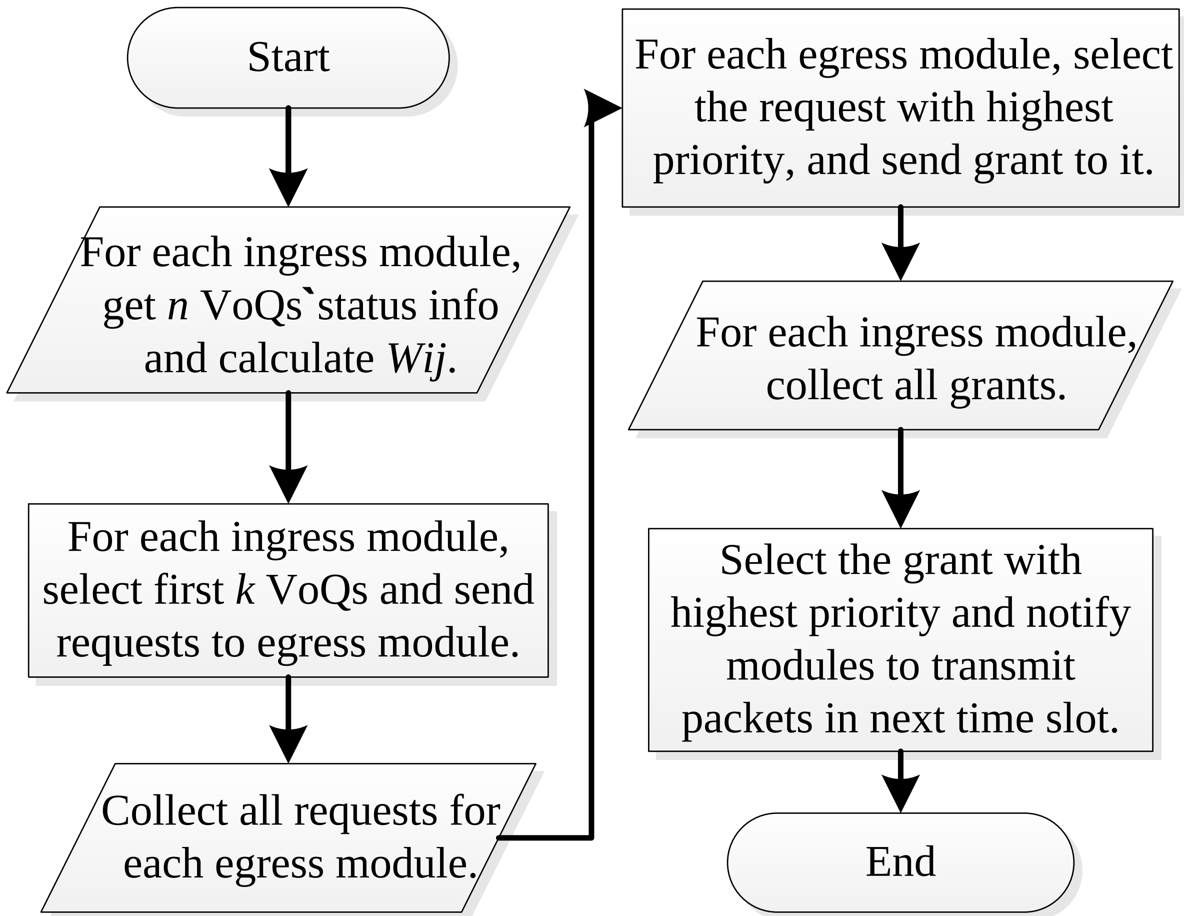
# Scheduling Algorithm for PSON

- **Priority-aware scheduling algorithm.**
- Modify Iterative Round Robin Scheduling
- Each ingress module maintains status information and gets priority values for *n* VoQs based on their status information. The priority value is calculated based on a combination of four strategies: longest queue fist (LQF), largest number of packets first (LNPF), oldest packet first (OPF), and less switching first (LSF) using the following weighted function:

$$W_{ij} = l_{ij} * w_l + p_{ij} * w_p + d_{ij} * w_d + s_{ij} * w_s$$

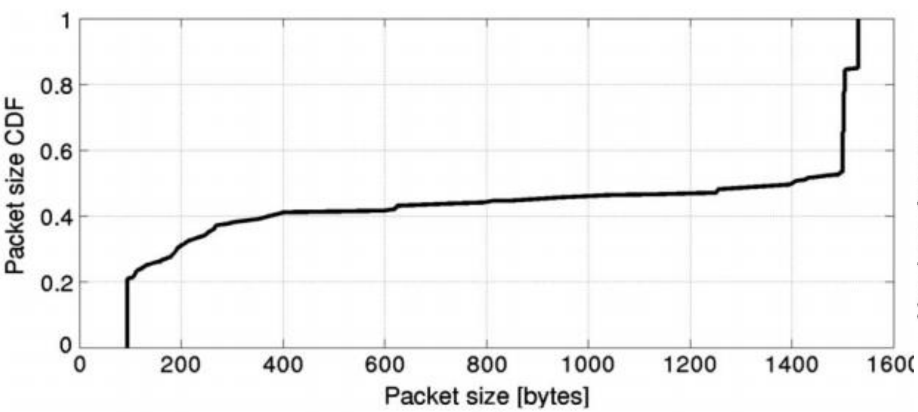
- We do not send all VoQ request in each module but choose first k VoQs with highest priority.

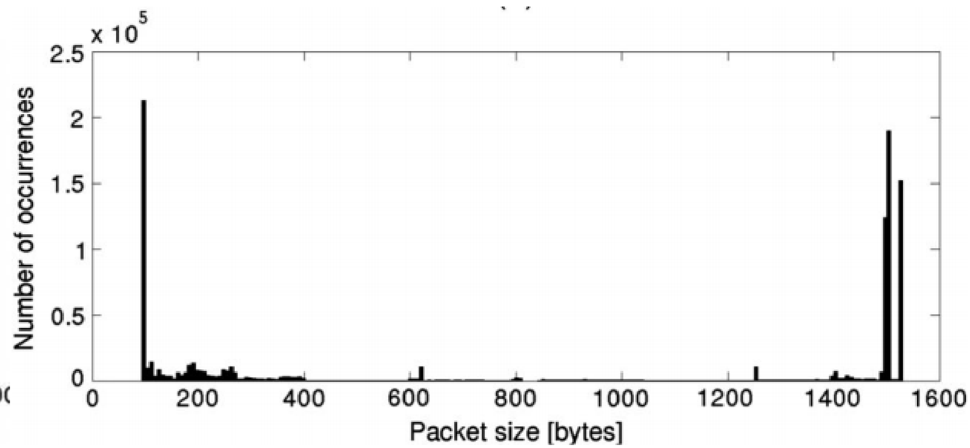
# Scheduling Algorithm for PSON



# Traffic generation

- Each of module receives the input traffic generated by 80 simulated servers.
- The amount of traffic load is normalized and can be scaled from 0 to 1.
- Packet length in real scenarios is mostly found to be a bimodal distribution around 40 bytes and 1500 bytes . [1]-[3]

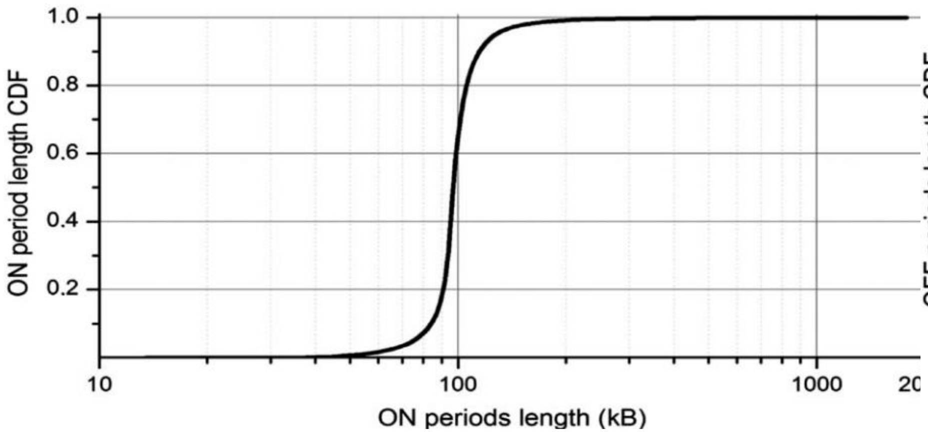


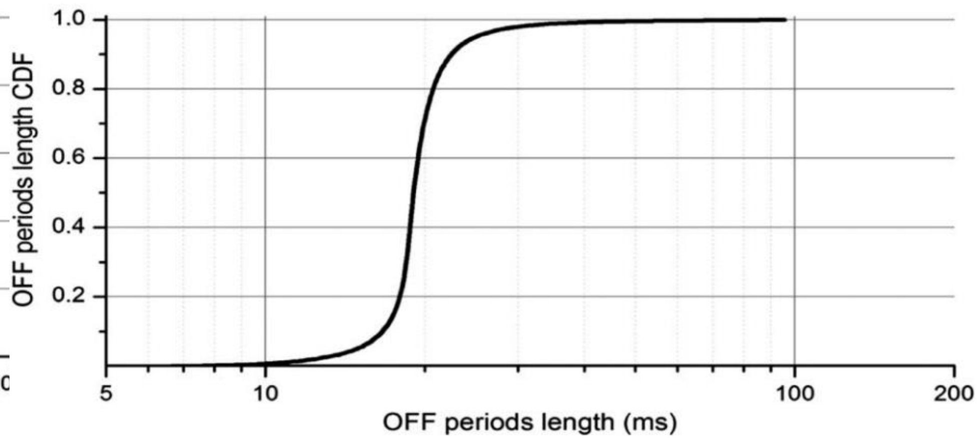




[1] T. Benson, A. Anand, A. Akella, and M. Zhang, “Understanding data center traffic characteristics,” Comput. Commun. Rev., vol. 40, no. 1, pp. 92–99, 2010.
[2] T. Benson, A. Akella, and D. A. Maltz, “Network traffic characteristics of data centers in the wild,” in Proc. Internet Measurement Conf. (IMC), Melbourne, Australia, Nov. 2010, pp. 267–280.
[3] S. Kandula, S. Sengupta, A. Greenberg, A. Patel, and R. Chaiken, “The nature of datacenter traffic: measurements & analysis,” in Proc. of the 9th ACM SIGCOMM Internet Measurement Conf. (IMC’09), 2009, pp. 202–208.

## Traffic generation (Cont.)

- Packet arrival times are modeled matching ON/OFF periods.
- ON/OFF periods follows Pareto distribution.
- ON periods follow the same length distribution regardless of load.
- OFF periods is proportional to the chosen simulation load value.

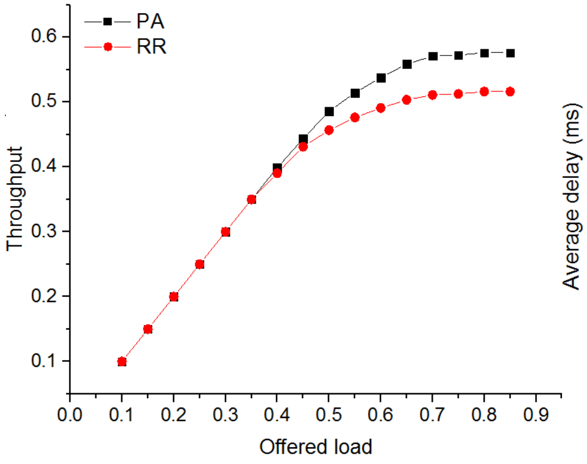


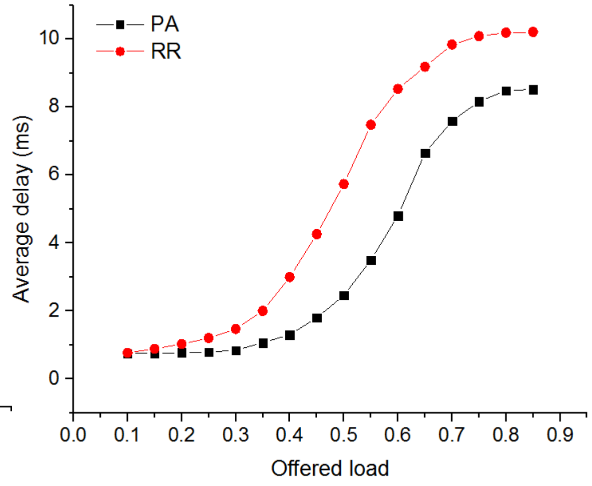



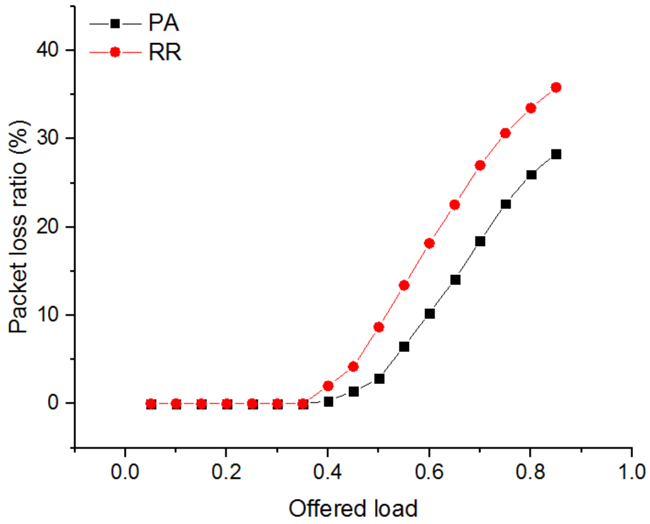

[1] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," Comput. Commun. Rev., vol. 40, no. 1, pp. 92–99, 2010.
[2] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," in Proc. Internet Measurement Conf. (IMC), Melbourne, Australia, Nov. 2010, pp. 267–280.
[3] S. Kandula, S. Sengupta, A. Greenberg, A. Patel, and R. Chaiken, "The nature of datacenter traffic: measurements & analysis," in Proc. of the 9th ACM SIGCOMM Internet Measurement Conf. (IMC'09), 2009, pp. 202–208.

UCDAVIS
UNIVERSITY OF CALIFORNIA

# Results and analysis

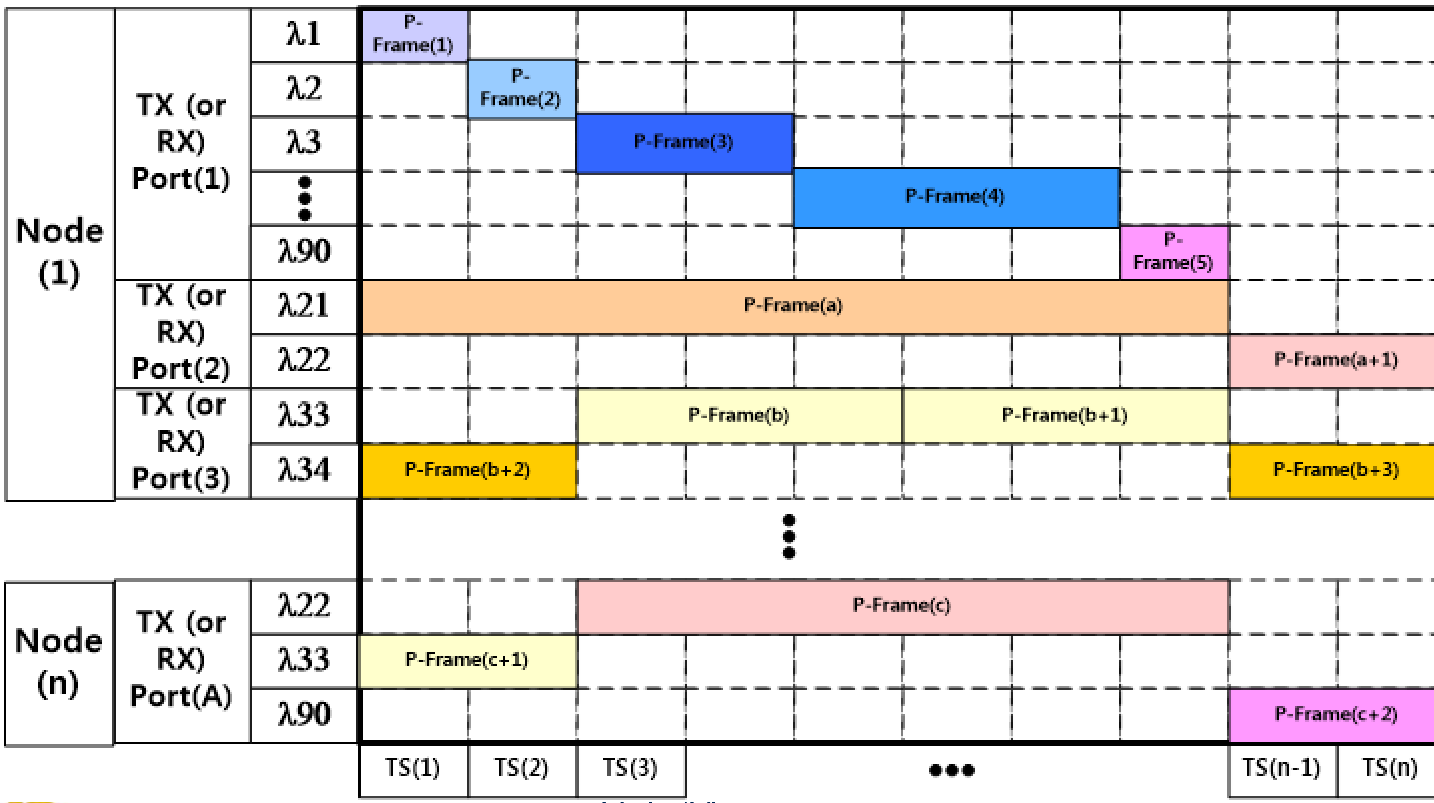


# Current work

- **Scheduling algorithm design**
  - ❖ Photonic Frames can have variable lengths as shown below.
  - ❖ Each module(Node) can have multiple Tx (or Rx).
  - ❖ Consider other scheduling methods besides round robin.




- **Performance Evaluation**
  - ❖ Effect of frame size
  - ❖ Effect of buffer size
  - ❖ Effect of offset time

UC DAVIS
UNIVERSITY OF CALIFORNIA


UC DAVIS
UNIVERSITY OF CALIFORNIA

amlwang@ucdavis.edu