Low-latency photonic packet switches with large number of ports

Speaker: Lin Wang

Research Advisor: Biswanath Mukherjee

D. Lucente, et al., “Low-latency photonic packet switches with large number of ports,” Networks and Optical Communications (NOC), 16th European Conference on. IEEE, pp. 5-7, 2011.

UCDAVIS

Challenge

- **Data center applications require sub-microsecond latency over interconnect link.**
- **Traffic load keeps increase rapidly.**
- **Power consumption grows with large-scaling data center.**
- **Optical communications is introduced as an alternative solution.**

Challenge

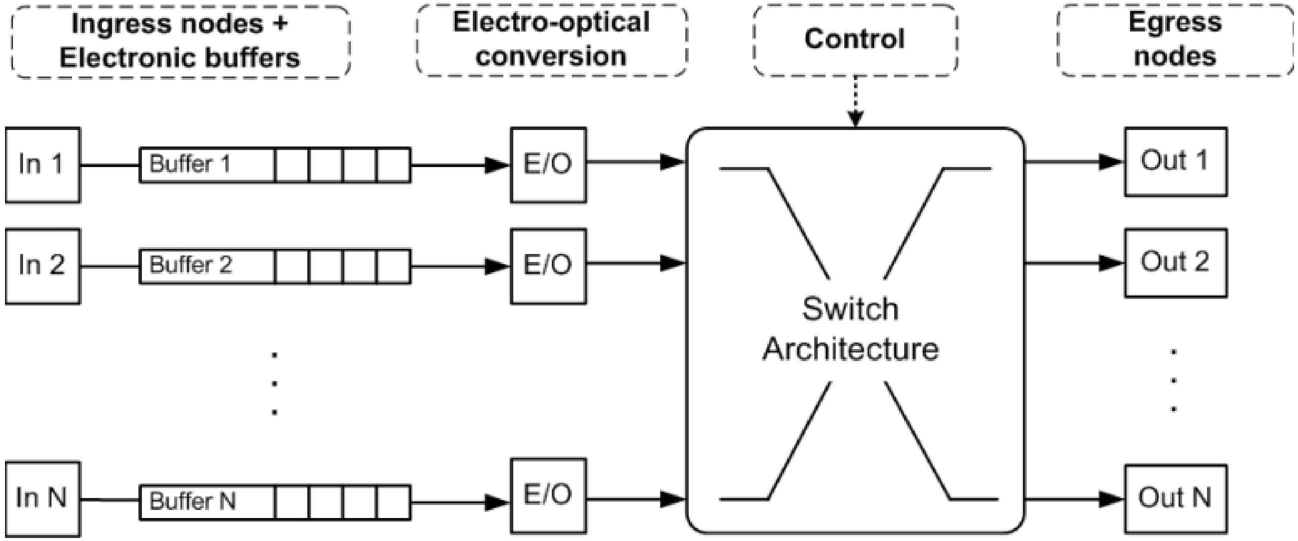
- **At present, many research projects focus on scaling optical packet switches to a large number of ports in order to meet the interconnectivity requirements.**
- **They overlooked the impact of the switch architecture on the performance of the node control.**
- **This paper gives an example of architecture that is controllable while scaling to over thousand of ports and supporting extremely low end-to-end latency**

Idea

- **They use highly distributed control and an modular structure, that can be reconfigured on a timescale in the order of few nanoseconds, regardless the number of input/output ports.**

- **The choice of the switch architecture has strong impact on the control complexity and on performance. Thus, they compare the results of different choices.**

System Description



- The ingress nodes contain queues of packets in (electronics) buffers.
- This system is time-slotted and operates in discrete time.
- At every time slot the switch controller handles all the requests for routing simultaneously.

Beneš architecture as the switch architecture.

Idea

- Allows any input to be connected to any output.

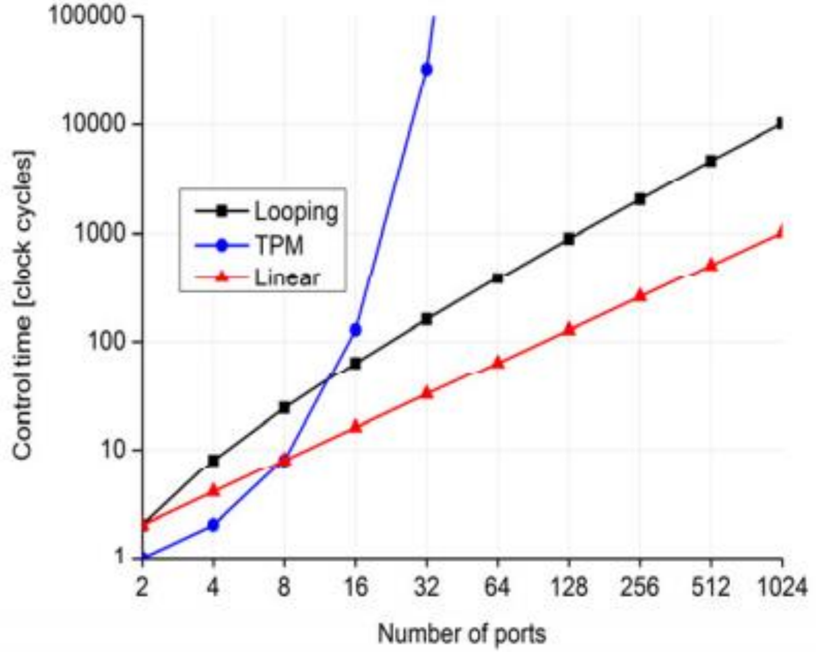
Problem:

- If the input state of the switch changes, the entire switch matrix needs to be reconfigured to establish a new connection map.

- Best known algorithm for the time required for reconfiguring a Beneš switch, scales as Nlog2N.

- Other slight improvements and multi-processor implementations have been considered, all implementations scale at least linearly with N.

Control time compare for Beneš architecture



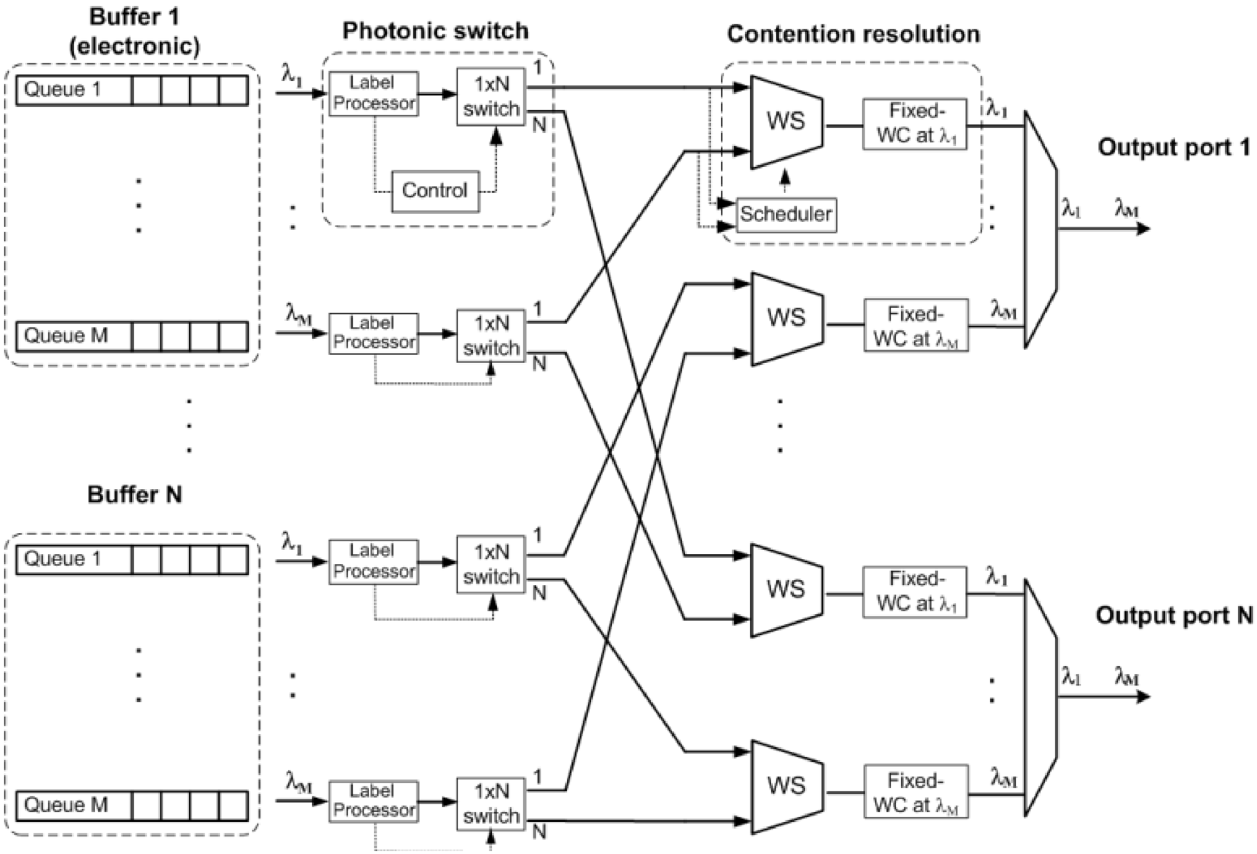
- Configuration time of a Beneš switch expressed in clockcycles according to looping algorithm, Trial Partition Machine [TPM] and an algorithm that scales linearly with the port-count.

Solution

- Present a non-blocking optical packet switch architecture that allows highly distributed control.
- This approach allows controlling each input channel independently of the others and leads to evident advantages in terms of control complexity and thus lower end-to-end latency.

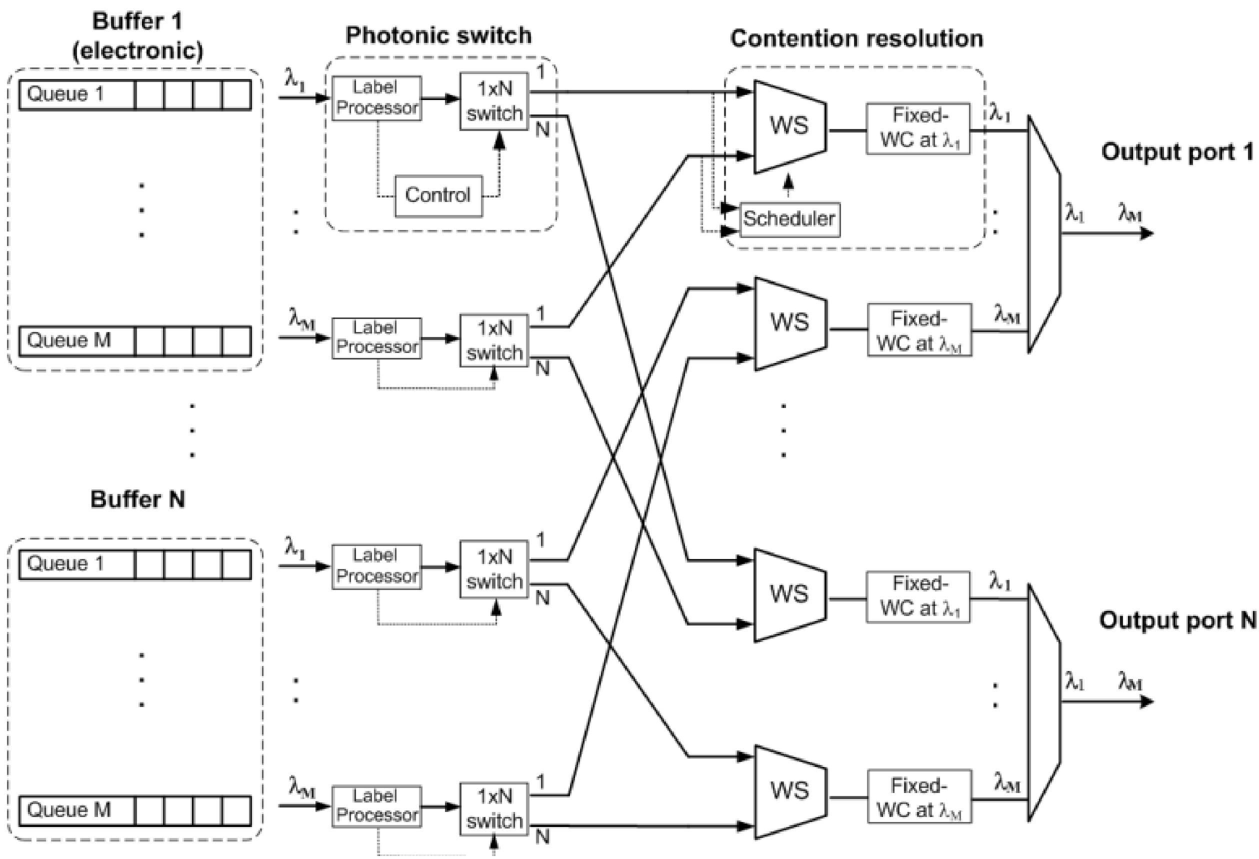
Solution

- Contention Resolution Block (CRB) at each of its outputs.
- Each of the NM WDM wavelength channels that input the switch, are fed into a 1xN optical switch.
- These 1xN switches require a local control, and they thus operate independently from each other.



Solution (Continue)

- The CRB is implemented by using Wavelength Selectors (WSs) and Fixed Wavelength Converters (FWCs).
- The function of the CRB is to solving contentions between packets destined to the same output port by employing wavelength conversion.

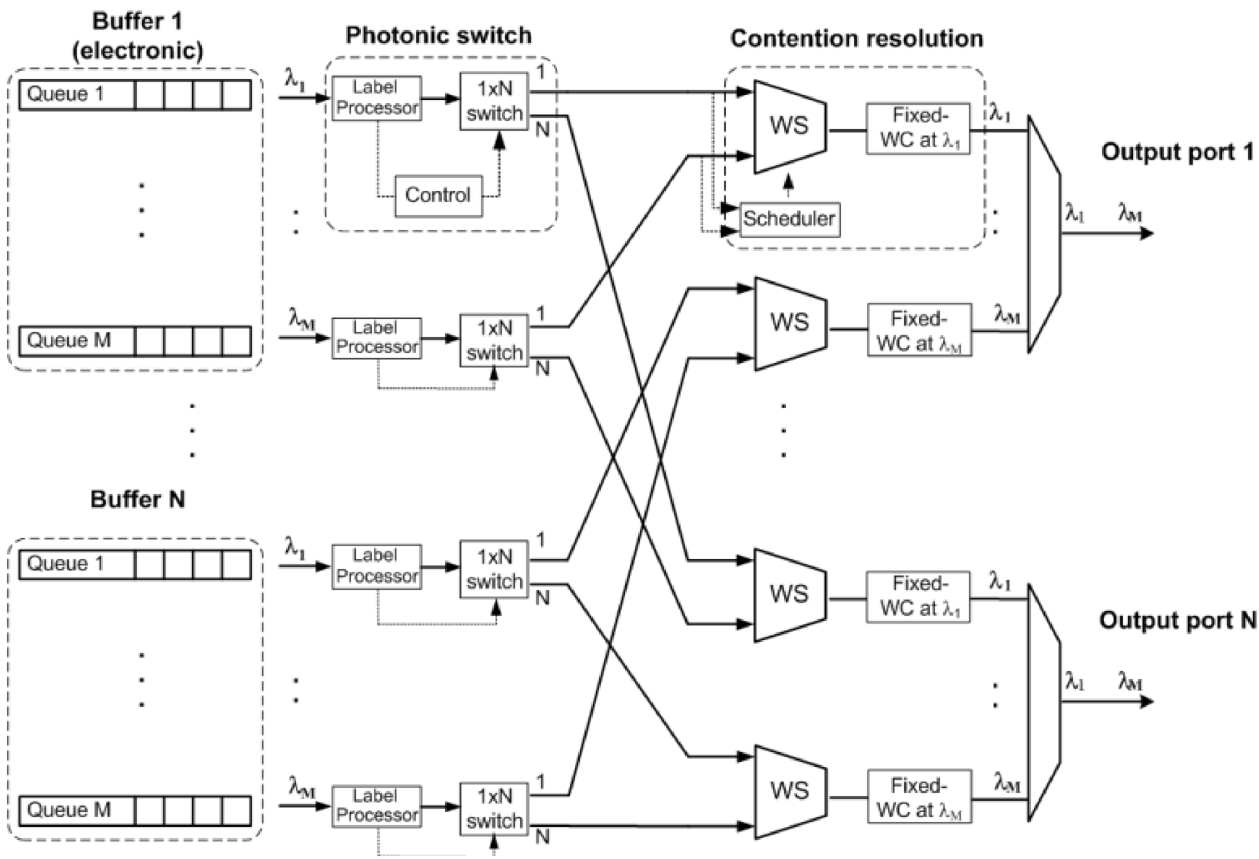


UC DAVIS
UNIVERSITY OF CALIFORNIA

Solution (Continue)

Process:

- At every time step there is a fixed probability of a packet to arrive at one of the 1xN switches.
- The switch control contains a label processor to forward the packet.
- If more than one packet is sent to the same output port, the scheduler drives the WS to select one of these packets.
- The link is equipped with flow-control, thus the unselected packets have to be retransmitted.

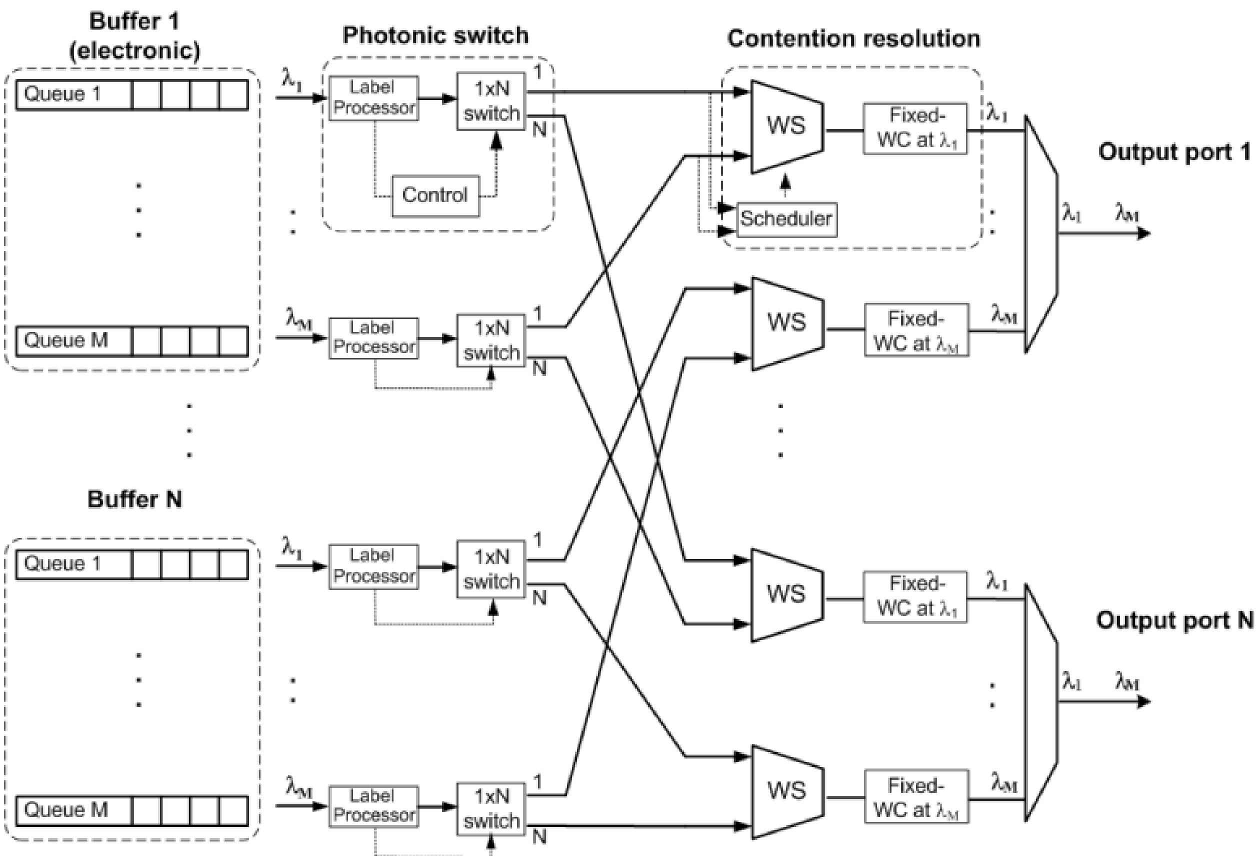


UCDAVIS UNIVERSITY OF CALIFORNIA

Solution (Continue)

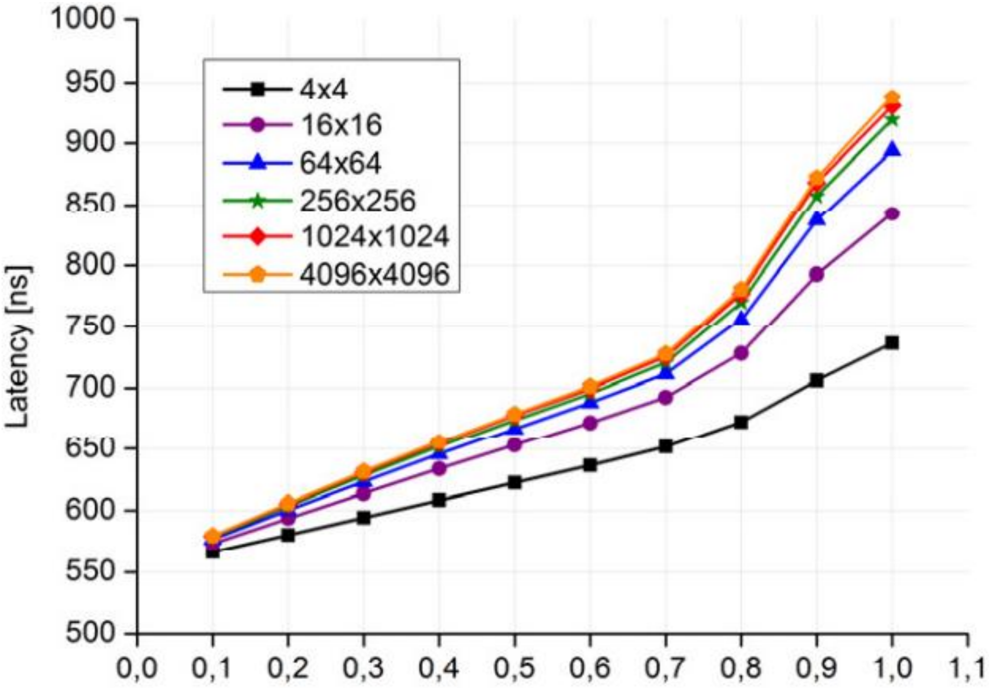
Why less configuration time?

- The modular structure of this architecture and the decoupled implementation of the CRBs allow a highly distributed control.
- This implies that the reconfiguration time of the entire NMxNM switching matrix is limited by the configuration time of a single 1xN switch.



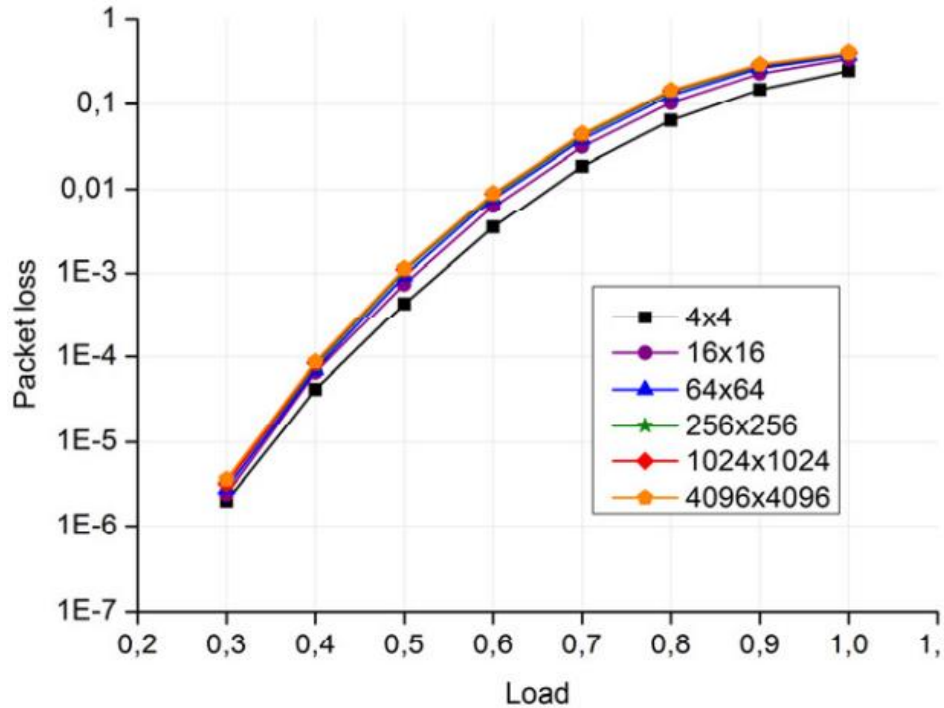
UCDAVIS UNIVERSITY OF CALIFORNIA

Simulation Result



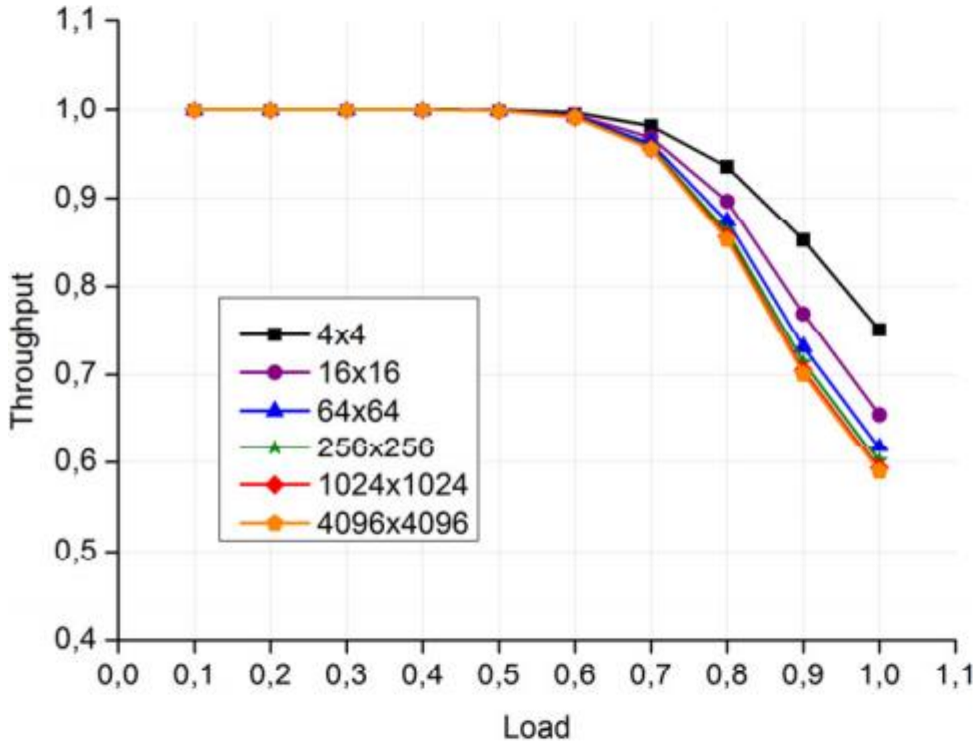
- A length link of 50 meters and packet duration of 40 nanoseconds.
- At every time slot there is a fixed probability that a packet is sent to the 1xN switch.
- The input buffer capacity is set to 14 packets, in agreement with the rule-of-thumb B = RTT x Packet-rate (where B represents the amount of buffering, expressed in number of packets, and RTT stands for Round Trip Time).

Simulation Result



- A length link of 50 meters and packet duration of 40 nanoseconds.
- At every time slot there is a fixed probability that a packet is sent to the 1xN switch.
- The input buffer capacity is set to 14 packets, in agreement with the rule-of-thumb B = RTT x Packet-rate (where B represents the amount of buffering, expressed in number of packets, and RTT stands for Round Trip Time).

Simulation Result



- A length link of 50 meters and packet duration of 40 nanoseconds.
- At every time slot there is a fixed probability that a packet is sent to the 1xN switch.
- The input buffer capacity is set to 14 packets, in agreement with the rule-of-thumb B = RTT x Packet-rate (where B represents the amount of buffering, expressed in number of packets, and RTT stands for Round Trip Time).

UC DAVIS UNIVERSITY OF CALIFORNIA

amlwang@ucdavis.edu