Switch Architectures For Small-buffered Optical Packet Switched Networks

Onur Alparslan, Shin'ichi Arakawa, Masayuki Murata

Osaka University

4 July 2007

Outline

- Buffering Problem in OPS networks
- Objective
- Proposed Solutions
- Simulation Results
- Conclusions

Buffering Problem in OPS Networks

- **According to a rule-of-thumb, an output link of a router needs a buffer sized at *B = RTT x BW***
  - Huge buffer size requirement due to ultra-high speed of optical networks
- **Electronic RAM is not a feasible solution**
  - O/E/O conversion is hard at ultra-high speed of optical networks
- **FDLs (fiber-delay-lines) have limitations**
  - FDLs can provide small amount of buffering with fixed delays
  - Bulky devices
- **Optical RAM**
  - Still under research
  - Do not expect it to have large capacity soon

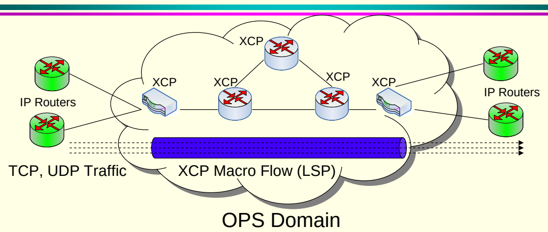
Objective

- **Designing an all-optical OPS network architecture that can achieve high utilization and low packet drop rate by using small FDL buffers**
- **Showing the buffer requirements**

Proposed Solutions 1/2

- **Preventing wavelength over-utilization**
  - Apply XCP-based congestion control
    - XCP is a new congestion control algorithm specifically designed for high-bandwidth and large-delay networks.
  - Carefully select XCP parameters
  - Control maximum wavelength utilization ratio by XCP
- **OPS Architecture**
  - Time-slotted WDM OPS network
  - Variable length IP packets enter OPS domain without aggregation

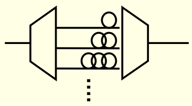
Proposed Solutions 2/2



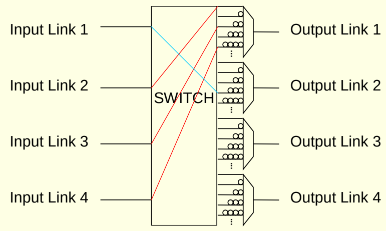
- **Burstiness**
  - Establish macro flows between edge nodes
  - Assign incoming TCP, UDP traffic to macro flows (similar to XCP-CSFQ, TeXCP)
  - Apply leaky bucket pacing to macro flows according to XCP flow rate at edge node
  - Possible to use LSPs for controlling macro flows if GMPLS is available

Architecture

- **FDL Architecture**
  - Single stage FDL set with B delay lines
  - FDL length distribution increases linearly (x,2x,3x,4x….) where x is FDL granularity
    - » FDL granularity of 3 means a FDL set of (3 slots, 6 slots, 9 slots, 12 slots….)
  - Voids occur between packets inside FDLs when x>1

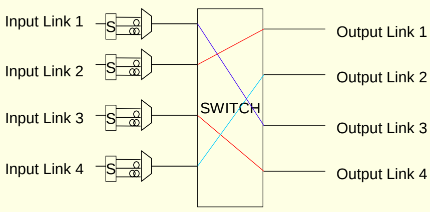


Output Buffering



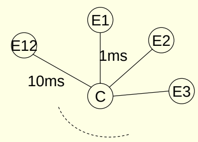
- Switch size is large because all delay lines are connected to a single switch
- Fabrication of a single big switch may be costly

Input Buffering with FDLs



- Smaller Switches

Star Topology Simulations

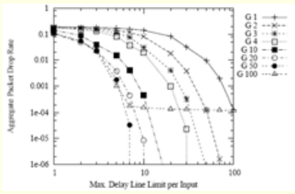


- 12 links star topology
- Wavelength speed is 1Gbit/s
- Slot size is 52Bytes
- Each edge node sends traffic to all other edge nodes
- 30% XCP target utilization
- Realistic packet size distribution

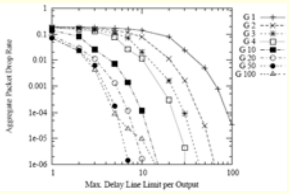
Simulation Results

- **Aggregate packet drop rate in the core**

Input Buffering



Output Buffering

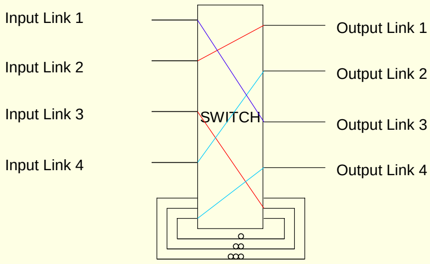


- **Similar fiber delay line requirements**
- **Input buffering uses smaller switches**

Recent Research and Preliminary Results

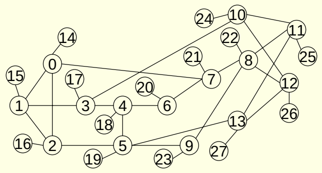
- **Preliminary results show that**
  - Shared buffering with XCP control further decreases the FDL requirements
  - FDL requirement decreases as slot size increases
    - » However using bigger slot size decreases utilization efficiency due to padding for small packets
    - » Trade-off
- **Evaluating throughput of TCP with the proposed OPS architecture on mesh topology**

Shared Buffering



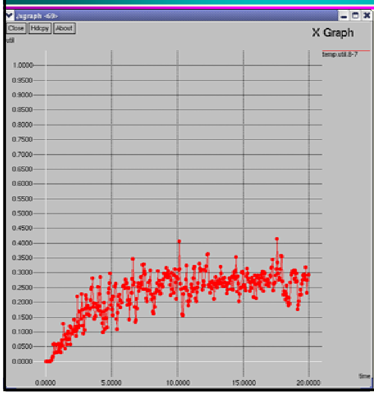
- Only single circulation is allowed in the shared buffer

TCP Traffic Simulation (Preliminary Results)



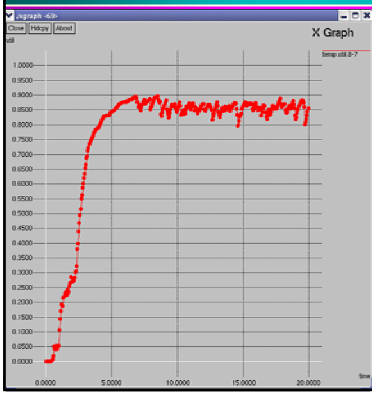
- NSFNET topology with 28 nodes (14 edge + 14 core) and 35 links
- Wavelength speed 1Gbit/s
- Slot size is 1500Bytes
- Apply TCP traffic among all edge node pairs
- TCP data packets are 1500Bytes, TCP ACK packets are 40Bytes
- XCP target utilization 90%
- Core nodes have Shared FDL buffering with only 7 delay lines per switch with granularity of 1 slot
  - Max. buffer capacity per link is 7 packets
  - Rule-of-thumb requires 20,834 packet buffer size per link

Link Utilization (Without XCP control)



- **Utilization of link from node 8 to node 7**
- **Around 25% utilization**
- **High packet drop rate**

Link Utilization (Our Architecture)



- **Apply XCP controlled pacing to incoming TCP packets at edge nodes of OPS domain**
- **Around 85% utilization**
- **Lower packet drop rate**
- **Nodal degree is 5 and switch has only 7 delay lines shared buffer, so delay line per link is only 1.4**

Conclusions

- **Possible to decrease buffer requirements of OPS core routers by**
  - applying XCP pacing at edge nodes
  - XCP-based utilization control at core nodes in the OPS network
- **Input and output buffering have similar fiber delay line requirements when utilization is low**
- **Shared buffering with proposed XCP based pacing may further decrease the buffer requirements**
  - Possible to get more than 3 times higher TCP throughput with very small buffers

Thank you