Toward a Reliable Data Transport Architecture for Optical Burst-Switched Networks

Dr. Vinod Vokkarane
Assistant Professor, Computer and Information Science
Co-Director, Advanced Computer Networks Lab
University of Massachusetts Dartmouth, USA

April 28, 2006
Invited Main Speaker
3rd Workshop on Optimization of Optical Networks (OON) 2006
Montreal, QC
Presentation Outline

• Introduction to Optical Transport Paradigms

• Optical Burst Switching

• Reliable Data Transport in OBS
  - Loss Minimization Mechanisms
  - Loss Recovery Mechanisms

• Conclusion and Future Work
Applications Demands

Applications
- Voice Over IP
- Streaming Video
- Grid Computing
- Storage Area Networks
- Multimedia
- Data

Service Requirements
- High Bandwidth
- Dynamic Provisioning
- Reliability
- Low Latency

Optical Transport Paradigms
- Optical Circuit Switching
- Optical Packet Switching
- Optical Burst Switching

vvokkarane@umassd.edu
Optical Circuit Switching

• For each request,
  - Set-up a static circuit (lightpath)
  - Transfer data
  - Release connection

• Pros:
  - Suitable for smooth, longer-term, high-bandwidth applications

• Cons:
  - Long circuit set-up latency
  - Inefficient for short-term bursty applications
Optical Circuit Switching (cont.)

- Circuit switched networks optimized for **Voice**
- **Data**: Accounts for majority of total traffic
- Data tends to be **bursty**
- Static bandwidth allocation is **not** efficient
Optical Packet Switching

• A photonic packet contains header and payload

• Header is processed all-optically at each node

• Pros:
  - Statistical multiplexing of data
  - Suitable for bursty traffic

• Cons:
  - Very fast switching speeds (nanoseconds)
  - Synchronization
Optical Burst Switching

• Multiple IP packets assembled into a burst
• An out-of-band control header transmitted ahead of each data burst

• Pros:
  - Statistical multiplexing of data
  - Suitable for bursty traffic
  - Low data-transfer latency
  - Electronic control plane (practically feasible)
  - Optical data plane (high-speed)
# Motivation for OBS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical Circuit Switching</td>
<td>Low</td>
<td>High</td>
<td>Slow</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Optical Packet Switching</td>
<td>High</td>
<td>Low</td>
<td>Fast</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Optical Burst Switching</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

OBS combines the best of the two while avoiding their shortcomings
Presentation Outline

• Introduction to Optical Transport Paradigms

• Optical Burst Switching

• Reliable Data Transport in OBS
  - Loss Minimization Mechanisms
  - Loss Recovery Mechanisms

• Conclusion and Future Work
OBS Network Architecture

- Core
  - Signaling
  - Scheduling
  - Contention Resolution

- Edge
  - Burst Assembly
  - Routing

Input Traffic

Ingress Node

WDM Link

Core Node

Offset Time

Burst

Header

Output Traffic

Egress Node

Edge Node

Core Node

Input Traffic

Egress Node

Computer and Information Science Department
OBS Node Architecture

Control Wavelengths

Data Wavelengths

Scheduler

O/E/O

Switch

Control Packets

Control Packet Processing

Offset Time

Data Bursts

Adopted from Qiao
Burst Assembly

- Aggregate multiple (IP) packets going to the same destination into a single burst

- **Assembly Mechanisms**: Timer-based and Threshold-based

  - **Timer-based assembly**:
    - After a fixed timer interval, all the packets in the queue are framed into a single burst

  - **Threshold-based assembly**:
    - After a fixed length threshold is reached, all the packets in the queue are framed into a single burst.
Burst Assembly

Aggregate multiple (IP) packets going to the same destination into a single burst

A unique packet queue for every destination egress node

DST
0
1
N

IP Packet Queues

Time or length threshold is reached

A header is generated and sent out

Control Channel

Data Channel

Burst transmitted after offset time

Burst Assembly Node

Adopted from Qiao
Signaling Technique

- **One-way based (un-acknowledged) signaling**

- **Reservation Mechanism:** Based on the start of the reservation
  - **Immediate Reservation:** Immediately after the control heater
  - **Delayed Reservation:** At the start of the burst

- **Release Mechanism:** Based on the release of the reservation
  - **Implicit Release:** based on burst length information
  - **Explicit Release:** explicit release control packet used

**Tradeoff:** Efficiency vs. Simplicity
Just-Enough-Time (JET) Signaling

- Delayed Reservation and Implicit Release
- Header contains burst length, offset time, source, destination
- Offset time necessary for processing of header at intermediate nodes without buffering the data burst

- Just-In-Time (JIT): Immediate Reservation and Explicit Release
**Data Loss in OBS: Burst Contentions**

- **Contention** occurs when more than one burst attempts to go out of the same output port (or wavelength) at the same time.

- **Unique to all-optical networks**
  - Traditional networks employ electronic buffering to resolve contentions.
  - Lack of optical buffers (cannot store light).

- **Drop Policy:**
  - One of the bursts will be dropped in its entirety.
  - Even though overlap between the bursts may be minimal.

---

**Diagram:**

- Original burst
- Contending burst
- Core Switch
- Drop Entire Burst

---

vvokkarane@umassd.edu

Computer and Information Science Department
TCP over OBS

- **Transmission Control Protocol (TCP)**
  - Majority of applications depend on TCP for reliable data transfer.
  - TCP assumes packet loss is always due to network congestion.
  - TCP congestion avoidance mechanisms reduce sending rate in the event of a packet loss.

- **OBS**
  - Random burst loss occurs even when network is NOT congested.

- **TCP over OBS**
  - **False Timeout** - time out when network is NOT congested.
  - TCP falsely reduces send rate - even when network is NOT congested.
  - Significantly degrades throughput of high-bandwidth apps.
Presentation Outline

• Introduction to Optical Transport Paradigms

• Optical Burst Switching

• Reliable Data Transport in OBS
  - Loss Minimization Mechanisms
  - Loss Recovery Mechanisms

• Conclusion and Future Work
Toward a Reliable OBS

* - Our proposed solution approaches marked in blue
Toward a Reliable OBS

* - Approaches discussed in this presentation in blue
Traditional Contention Resolution

- **Optical Buffering (FDLs)**
  - Achieved through Fiber Delay Lines
  - **Issues**: Limited buffer capacity and additional hardware cost

- **Wavelength Conversion**
  - Converting the wavelength of an incoming channel to another wavelength at the outgoing channel
  - **Issues**: Additional hardware cost

- **Deflection Routing**
  - Deflect contending bursts to alternate port
  - **Issues**: Higher delay and out-of-sequence delivery
Burst Segmentation

- When contention occurs, only overlapping segments are dropped
- Two Approaches: Head Dropping and Tail Dropping
- Details: Vokkarane and Jue [IEEE ICC 2002, New York]

Diagram:
- Core Switch
- Original burst
- Contending burst
- Dropped segments
- Head Dropping
- Tail Dropping
Evaluation Criteria

• Evaluation of proposed policies
  - Average end-to-end packet loss probability
  - Average end-to-end packet delay (hops)
  - TCP throughput

• Numerical Analysis
  - Analytical modeling – Markov models
  - Simulation results – Discrete-event simulations

vvokkarane@umassd.edu
Burst Segmentation: Analytical Loss Model

- Burst Arrivals: Poisson Process
- M/G/1/1 Queueing Model

Burst Length Distribution (After $k$ hops):

$$G_{l_{sd}}^{(k)}(t) = 1 - (1 - G_{l_{sd}}^{(k-1)}(t))(e^{-\lambda_{l_{sd}}^{(k)}})$$

$$= 1 - (1 - G_{l_{sd}}(t))e^{-\left(\sum_{i=1}^{k} \lambda_{l_{sd}}^{(i)}\right)t}.$$ 

End-to-End Packet Loss:

$$P_{loss} = \sum_{s} \sum_{d} \frac{\lambda_{sd}}{\lambda} P_{loss}^{sd}.$$ 

[Vokkarane: IEEE JSAC 2003, SPIE Optical Networks 2003]
Simulation Network

Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burst Arrivals</td>
<td>Poisson</td>
</tr>
<tr>
<td>Average Burst Length</td>
<td>100 $\mu$s (exponentially dist.)</td>
</tr>
<tr>
<td>Link Transmission Rate</td>
<td>10 Gb/s</td>
</tr>
<tr>
<td>Packet Length</td>
<td>1500 Bytes</td>
</tr>
<tr>
<td>Switching Time</td>
<td>10 $\mu$s</td>
</tr>
<tr>
<td>Optical Buffering</td>
<td>NO</td>
</tr>
<tr>
<td>Wavelength Conversion</td>
<td>NO</td>
</tr>
</tbody>
</table>

14-node NSFNET
Packet Loss Performance

![Graph showing Packet Loss Probability vs Load (in Erlang) for different methods: Drop, Segmentation, Deflection, and Deflect-First. The graph compares the performance of SDP, SDDP, DP, DDP, and DSDP.]
Average Number of Hops (~Delay)

Load (in Erlang) ----->

Average Number of Hops ------>

Deflect-First
Deflection
Segment-First
Segmentation
Drop

vvolkarane@umassd.edu
Toward a Reliable OBS
Reactive Loss Recovery: Burst Retransmission

- **Objective**
  - To recover from burst loss when network is not congested

- **Basic idea**
  - Retransmit lost bursts at source nodes
  - Stop retransmission when $T_r > \text{Delay Constraint} (\delta)$

[Zhang and Vokkarane: IEEE GLOBECOM 2005]
Analysis for Burst Retransmission

Objective
- Analyze the average burst loss probability in the network

Basic idea
- No retransmission for bursts blocked by retransmission buffers
  
  \[ P = P_b P_c + (1 - P_b) (P_c)^{R+1} \]

Diagram:
- Edge Node
- Retransmission Buffer
- OBS Network
- Burst Contention Probability \( P_c \)
- Buffer Blocking Probability \( P_b \)
- No Retransmission
- Retransmission \( (1 - P_b) \)
- Burst Loss \( (P_c)^R \)
- Burst Loss Probability \( P \)
Simulation Assumptions

- No. of wavelengths on each link is 4
- Transmission rate on a wavelength is 10 Gb/s
- Burst arrival is Poisson
- Traffic are uniformly distributed
- Average burst length are 100 µs
Burst Loss Probability
Analysis and Simulation Results for Burst Loss Probability

![Graph showing burst loss probability vs load](image)
Average Burst Delay

![Graph showing Average Burst Delay vs Load with different scenarios]

- W/O Ret
- Ret $\delta = 2T_p$ and $P_b = 0.1$
- Ret $\delta = 2T_p$ and $P_b = 0.001$
- Ret $\delta = 4T_p$ and $P_b = 0.1$
- Ret $\delta = 4T_p$ and $P_b = 0.001$
Buffer Capacity at Edge Nodes

![Graph showing buffer capacity at edge nodes with different load levels and parameters.](image)
TCP Throughput

![Graph showing TCP Throughput with different load and total TCP throughput values.](Image)
Performance of TCP Versions

Throughput

Num. of Timeouts

Total TCP Throughput (Mbps)

Total Number of TO Events

Load

Load
Burst Retransmission

• **Pros**
  - Reduce burst loss probability
  - Correctly indicate network congestion (min FTOs)
  - Significantly improve TCP throughput

• **Cons**
  - Additional electronic buffers at edge nodes
  - Longer delay for retransmitted bursts
Proactive Loss Recovery: Forward Redundancy (FR)

- Some or all the original packets of a burst are copied and sent in the forward direction from source to destination
- Receiver can recover from selective packet loss in the forward direction

FR Policies
- Serial or Parallel FR
- Partial (< 100%) or Complete (>= 100%) FR
- Same or Disjoint path - Protection
Serial Forward Redundancy (SFR)

We have evaluated SFR [IEEE GridNets 2005, IEEE WOCN 2006]
Simulation Assumptions

- No. of wavelengths on each link: 8
- Transmission rate: 10 Gbps per wavelength
- Burst arrivals: Poisson
- Packet size: 1250 byte [10 Kb]
- Fixed burst length: 100 packets [1Mb]
- Traffic: uniformly distributed
- Switching time = 10 μs
Packet Loss Performance

- **Drop Entire Burst**
- **Segmentation**

Log Scale

[Graph showing Packet Loss Probability vs. Load (in Erlang)]

- Triangle: Segmentation
- Circle: Drop
Packet Loss Probability

![Graph showing Packet Loss Probability vs Load](image-url)
End-to-End Packet Delay

- Baseline
- Segmentation
- 10% SFR
- 20% SFR
- 50% SFR
- 100% SFR
- RET Pb = 0.1 and δ = 2Tp
- RET Pb = 0.001 and δ = 2Tp

Average Packet Delay (ms) vs Load

vvolkarane@umassd.edu

Computer and Information Science Department
Analytical Loss Model Results
Packet Loss Probability

- 100% SFR
- 200% SFR
- 500% SFR

Load

Packet Loss Probability
TCP Throughput

![Graph showing TCP Throughput]
Presentation Outline

• Introduction to Optical Transport Paradigms

• Optical Burst Switching

• Reliable Data Transport in OBS
  - Loss Minimization Mechanisms
  - Loss Recovery Mechanisms

• Conclusion and Future Work
Conclusion

• OBS Network
  - Promising optical core data-transport paradigm
  - Suited for delay-sensitive applications

• Loss Minimization and Loss Recovery Mechanisms
  - Evaluated several new mechanisms
  - Proposed mechanisms significantly improves the reliability of data transfer over OBS networks

• Future Work
  - Develop dynamic mechanisms
  - Impact on newer high-speed TCP versions
Thank You

http://www.cis.umassd.edu/~vvokkarane/