Exploiting Temporal Domain for a Transparent Fine-Grained Optical Network

Zhizhen Zhong
Tsinghua University & UC Davis
zhongzz14@mails.tsinghua.edu.cn, zzzzhong@ucdavis.edu
02 Feb. 2018
Networks Lab Group Meeting
Outline

- The need for a transparent fine-grained optical network
- Past solutions on transparent fine-grained optical switching
- Our plan: Optical Time Slice Switching (OTSS)
- OTSS use cases
- Conclusion and discussions
Outline

- The need for a transparent fine-grained optical network
- Past solutions transparent fine-grained optical switching
- Our plan: Optical Time Slice Switching (OTSS)
- OTSS use cases
- Conclusion and discussions
Transparent optical network: latency, energy and security

- **Electronic queuing and processing** delay increase **end-to-end latency**
- Introducing **all-optical switching** technology may reduce latency

---

![Diagram of Layer 3 Routing](image1)

- **Layer 3 Routing**
  - High energy-consumption
  - High latencies & latency jitters
  - High flexibility

![Diagram of Optical Bypass](image2)

- **Optical Bypass**
  - Low energy-consumption
  - Low latencies & latency jitters
  - Granularity?
Advantages on energy efficiency of optical switching

- **2010:**
  - 235 billion kWh

- **2020:**
  - 1 trillion kWh = 10 Three Gorges Power Stations

Electronic switching: 13.5W/10Gbps port
Optical switching: 0.24W/10Gbps port

DC power consumption

- Lighting: 5%
- Power supply: 10%
- IT: 45%
- Air conditioning: 40%
- Network devices: 23%
- Servers: 40%
- Storage: 37%

Total Power Consumption

---

Needs for a transparent fine-grained optical network

- Optical networks: enormous transmission bandwidth.
  - High-order modulation (PAM4): increase per-channel capacity.
  - Space-division-multiplexing: increase spatial channels.
- Mismatch between application demands and optical channel capacity.
  - Traffic grooming is the first proposal.
- Drawback of grooming: energy, latency, security, etc.
Rethinking Optical Networks

- Resource granularity, flexibility and usability.
- Match between request and resource. Avoid waste.
- Flexi-grid technologies: provides more channels with fine granularities.
Outline

- The need for a transparent fine-grained optical network
- **Past solutions transparent fine-grained optical switching**
- Our plan: Optical Time Slice Switching (OTSS)
- OTSS use cases
- Conclusion and discussions
Transparent connections vs. network scale

- Number of LPs **required vs. that can be offered.**
  - All-to-all communication.
  - Set up **Dedicated end-to-end lightpath** for each node pair.
- **Topologies:** 2-D/6-D torus, Butterfly, Fat tree
  Nodal degree: $d=4$ (for 6-D torus: $d=12$).

OPS & OBS

• Temporal-domain switching techniques, such as **OPS** and **OBS**, can provide much finer granularity than frequency-domain switching techniques.

• **OPS** requires optical buffer (not mature) to avoid collisions.

• For **OBS**, burst loss rate could be very high at a heavy load without buffer.
Outline

- The need for a transparent fine-grained optical network
- Past solutions transparent fine-grained optical switching
- **Our plan: Optical Time Slice Switching (OTSS)**
- OTSS use cases
- Conclusion and discussions
Build a transparent, bufferless, fine-grained, WDM-like network

- Why WDM can avoid collision?
  - Wavelength channels are separated by a global coordinate.
  - (frequency! All the same in different nodes)
- Time synchronization: a global coordinate in temporal domain
  - Definite time, all nodes are synchronized for a global coordinate.
- Temporally-statistical multiplexing for asynchronous transmission based on synchronized global time.
  - We call it: Optical Time Slice Switching (OTSS).
OTSS principle

- Designing a WDM-like TDM switching paradigm
  - WDM: all nodes have same frequency coordinate.
  - OTSS: all node should have same time coordinate!

OTSS switch principle

OTSS principle: node switching

- The optical transmission channels are organized into repetitive OTSS frames.
- Each OTSS frame contains one or several variable-length time slice(s).
- When a time slice arrives, the switch controller sends control signals to the OTSS fabric at the precise time to direct the time slice to the expected output port.

OTSS data plane: Routing and Time slice allocation (RTA) problem

- **Time-slice contiguity constraint**: a request’s time slice should be contiguous on temporal domain.

- **Time-slice continuity constraint**: time slice should be continuous along different links.

- **Time slice shifting constraint**: signal propagation delay may result in time slice shifting in OTSS frame.

Flexi-grid wavelength switching and optical time slice switching
OTSS data plane: Routing and Time slice allocation (RTA) problem

Mathematical formulations and Algorithms design

Parameters:
- \( G(N, E) \): network topology in a unidirectional graph, where \( N \) and \( E \) denotes the set of nodes and MMF fiber links, respectively.
- \( R \): set of traffic requests.
- \( s_i, d_i, b_i \): source, destination, and bandwidth of traffic request \( r_i \in R \).
- \( X \): accumulated crosstalk threshold for direct-detection receivers.
- \( M \): set of supporting modes by a MMF.

Variables:
- \( \lambda_{ij}^{m,m} \): binary, which equals one if request \( r_i \) uses mode \( m \) and time slot \( t \) on fiber link \((i,j)\).
- \( \phi_{rmt}^{m,m} \): binary, which equals one if request \( r_i \) on mode \( m_1, m_2 \) have crosstalk on time slot \( t \) of fiber link \((i,j)\).
- \( \rho \): binary, which equals one if request \( r \) is accepted.

Objective: Maximize network throughput first, then minimize network resource usage.

Maximize: \( \eta_1 \sum_{i \in S} a_i - \eta_2 \sum_{r \in R} \sum_{m \in M} \sum_{i, j \in N} \lambda_{ij}^{m,m} \) (1)

Constraints:
1. \( \sum_{j \in J} \sum_{m \in M} \lambda_{ij}^{m,m} - \sum_{j \in J \setminus \{i\}} \sum_{m \in M} \lambda_{ij}^{m,m} = \rho \cdot b_i = s_i, \forall r_i \in R \) \( \rho \cdot b_i = d_i, \forall r_i \in R \) \( 0_i = s_i, d_i \) (2)
2. \( \sum_{j \in J} \lambda_{ij}^{m,m} = \sum_{j \in J \setminus \{i\}} \lambda_{ij}^{m,m}, \forall r_i \in R, t_i \in T \) (3)
3. \( \sum_{j \in J} \lambda_{ij}^{m,m} + \lambda_{ij}^{m,m} - \sum_{j \in J \setminus \{i\}} \lambda_{ij}^{m,m} = \sum_{j \in J \setminus \{i\}} \lambda_{ij}^{m,m}, \forall r_i \in R, t_i \in T, z \in N \setminus \{s_j, d_j\} \) (4)
4. \( \sum_{j \in J} \sum_{m \in M} \sum_{i \in I} \lambda_{ij}^{m,m} = \sum_{j \in J \setminus \{i\}} \sum_{m \in M} \lambda_{ij}^{m,m}, \forall r_i \in R, t_i \in T \) (5)
5. \( \sum_{j \in J} \sum_{m \in M} \sum_{i \in I} \lambda_{ij}^{m,m} = \sum_{j \in J \setminus \{i\}} \sum_{m \in M} \lambda_{ij}^{m,m}, \forall r_i \in R, t_i \in T, z \in N \setminus \{s_j, d_j\} \) (6)
6. \( \sum_{r \in R} \lambda_{ij}^{m,m} \leq 1, \forall m \leq M, t_i \in T, (i,j) \in E \) (7)
7. \( \sum_{i \in I} \sum_{j \in J} \sum_{m \in M} \lambda_{ij}^{m,m} - \sum_{i \in I} \sum_{j \in J \setminus \{i\}} \sum_{m \in M} \lambda_{ij}^{m,m} \leq 2, \forall r \in R, m \in M, (i,j) \in E \) (8)
8. \( \sum_{i \in I} \sum_{j \in J} \sum_{m \in M} \lambda_{ij}^{m,m} - \sum_{i \in I} \sum_{j \in J \setminus \{i\}} \sum_{m \in M} \lambda_{ij}^{m,m} \leq 1, \forall r \in R, (i,j) \in E \) (9)
9. \( \sum_{i \in I} \sum_{j \in J} \sum_{m \in M} \lambda_{ij}^{m,m} - \lambda_{ij}^{m,m} - \sum_{i \in I} \sum_{j \in J \setminus \{i\}} \sum_{m \in M} \lambda_{ij}^{m,m} \leq 1, \forall r \in R, (i,j) \in E \) (10)
10. \( \sum_{i \in I} \sum_{j \in J} \sum_{m \in M} \sum_{l \in L} \lambda_{ij}^{m,m} \cdot d(i,j), Y(m_1, m_2) \leq X, \forall r \) (11)
11. \( \sum_{i \in I} \sum_{j \in J} \sum_{m \in M} \sum_{l \in L} \lambda_{ij}^{m,m} \cdot d(i,j), Y(m_1, m_2) \leq X, \forall r \) (12)
12. \( \beta_{rmt}^{m,m} \leq \beta_{rmt}^{m,m}, \forall 1 \leq t \leq T, \forall r, m_1, m_2, \forall j \in J \setminus \{i\} \) (13)
13. \( \beta_{rmt}^{m,m} \leq \beta_{rmt}^{m,m}, \forall 1 \leq t \leq T, \forall r, m_1, m_2, \forall j \in J \setminus \{i\} \) (14)
14. \( \beta_{rmt}^{m,m} \leq \beta_{rmt}^{m,m}, \forall 1 \leq t \leq T, \forall r, m_1, m_2, \forall j \in J \setminus \{i\} \) (15)
OTSS control plane: unified control and signaling

- A software-defined unified control architecture with time synchronization.
- RSVP signaling for distributing different temporally switching command to different nodes.

Enabling technology: time synchronization

- **Time synchronization: mature**
  - precision should be finer than the smallest time slice in OTSS.
  - GPS receiver: 50ns.
  - IEEE 1588 or 1588v2: 100ns

- **Fast optical switch: mature**
  - PLZT high speed switch: 10ns switching speed.
  - MO high speed switch: 100us switching speed.
OTSS experimental validation

Experimental demonstration

Time slice switching (separation and aggregation)  Transmission and control signal delay

Outline

- The need for a transparent fine-grained optical network
- Past solutions transparent fine-grained optical switching
- Our plan: Optical Time Slice Switching (OTSS)
- OTSS use cases
- Conclusion and discussions
Overview of switching paradigms

- **Optical packet/burst switching**: delay-sensitive flows with relatively small sizes (query, coordination and control state messages).

- **Optical circuit/\(\lambda\) switching**: delay-insensitive bandwidth-hungry data transfer (file backup and virtual machine migration).

- **Optical time slice switching**: combine the advantages of capacity and energy consumption of optical fiber/\(\lambda\) switching and the flexibility of electrical packet switching.

---

*Buffer can improve OBS performance*
1. Fine-grained Communications for Smart Grid

- Small bandwidth
- Stringent latency

<table>
<thead>
<tr>
<th>Application</th>
<th>Paradigm</th>
<th>Bandwidth</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teleprotection</td>
<td>P2P</td>
<td>~ 500 Kb/s</td>
<td>8-10 ms</td>
</tr>
<tr>
<td>Load Shedding for Underfrequency</td>
<td>P2P &amp; HS</td>
<td>~ 500 Kb/s</td>
<td>10 ms</td>
</tr>
<tr>
<td>SCADA</td>
<td>P2P &amp; HS</td>
<td>~ 800 Kb/s</td>
<td>100-200 ms</td>
</tr>
<tr>
<td>Smart Metering</td>
<td>HS</td>
<td>~ 500 Kb/s</td>
<td>250-1000 ms</td>
</tr>
<tr>
<td>File Transfer</td>
<td>Random</td>
<td>200-1000 Mb/s</td>
<td>≥ 1000 ms</td>
</tr>
</tbody>
</table>

1. Fine-grained Communications for Smart Grid

- We study the 10 ms latency bound for security-related smart-grid traffic.
- Those requests larger than 10 ms, can not be served.
- Compare OTSS and flexi-grid, using the same amount of spectrum.

2. Transparent Fine-grained MDM Network

- OTSS + Mode Division Multiplexing
- OTSS provide fine-grained transparent channels breaking mode-coupling wavelength channels constraint.
- Can be adopted in datacenters.

Zhizhen Zhong, Nan Hua, Yufang Yu, Zhongying Wu, Juhao Li, Haozhe Yan, Shangyuan Li, Ruijie Luo, Jialong Li, Yanhe Li, and Xiaoping Zheng, “Throughput Scaling for MMF-Enabled Optical Datacenter Networks Based on Time-Slicing-Based Crosstalk Mitigation,” to be presented, OFC, Mar. 2018.
Background

Datacenter is the basic infrastructure for future information-based society.

1. More datacenters being built
2. Datacenter itself evolves to be larger

Cloud Computing
5G mobile networks
Video streaming
Interconnected car
Internet of things
Motivation

Roadmap to Ubiquitous Datacenters: Single-Mode or Multi-Mode for intra-DC interconnection networks?

Intra-DC optical networks:
- Short reach (100m~1km fiber length)
- Large connectivity (millions of network nodes)
- Large channel capacity (10Gb/s, 100Gb/s, 400Gb/s)

SMF and MMF are both mature technologies. Cost matters!

For future widely-located large-scale cloud/fog datacenters, MMF is a better choice.

<table>
<thead>
<tr>
<th>Cisco 10G SFP+ Transceiver*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Mode SFP+</td>
<td>$7</td>
</tr>
<tr>
<td>Multi-Mode SFP+</td>
<td>$6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cisco 40G SFP+ Transceiver*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Mode QSFP+</td>
<td>$340</td>
</tr>
<tr>
<td>Multi-Mode QSFP+</td>
<td>$55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cisco 100G SFP+ Transceiver*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Mode QSFP28</td>
<td>$2800</td>
</tr>
<tr>
<td>Multi-Mode QSFP28</td>
<td>$400</td>
</tr>
</tbody>
</table>

*Price from Fiberstore: [www.fs.com](http://www.fs.com)
Motivation

MMF suffers severe modal crosstalk in direct detection transmission systems.

• Modal crosstalk accumulate along propagation path, and can induce OSNR degradation at the receiver end [4,5].
• Such crosstalk must be prevented, as it cannot be fully undone by electrical signal processing after direction detection[6].

Is there a way to apply MMF into DC?

Time-Slicing-Based MDM

- **Basic idea**: stagger utilized modes in temporal domain via synchronized time slices.
- **Design Principle**: avoiding using modes with high crosstalk in the same time slice, while changing utilized modes in different time slices by switching at selected switching points, assisted by precise synchronized time.
Theoretical Analysis

Modal crosstalk, caused by random mode coupling, depending on fiber fabrication, imperfections, bending or twisting [7].

Accumulated crosstalk [8,9]:

\[ XT = \tanh h z \]

where \( h \) is the mode coupling parameter (m\(^{-1}\)) and \( z \) is the fiber length (m).


Zhizhen Zhong, Nan Hua, Yufang Yu, Zhongying Wu, Juhao Li, Haozhe Yan, Shangyuan Li, Ruijie Luo, Jialong Li, Yanhe Li, and Xiaoping Zheng, “Throughput Scaling for MMF-Enabled Optical Datacenter Networks Based on Time-Slicing-Based Crosstalk Mitigation,” to be presented, OFC, Mar. 2018.
Optimization Setup

- IBM CPLEX solver
- Accumulated XT threshold: -13dB
- Fiber length: 100m
- Traffic: uniformly generated between edge switch pairs, 1Gb/s
- Modal channel: 10Gb/s
- OTSS frame: 20ms
- Min time slice: 5ms
- Only on 1550nm wavelength

Tab. 1, crosstalk (dB/100m) of selected modes.

<table>
<thead>
<tr>
<th>XT</th>
<th>(m_1)</th>
<th>(m_2)</th>
<th>(m_3)</th>
<th>(m_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m_1)</td>
<td>-</td>
<td>-26.0</td>
<td>-21.2</td>
<td>-43.0</td>
</tr>
<tr>
<td>(m_2)</td>
<td>-17.7</td>
<td>-</td>
<td>-15.8</td>
<td>-19.7</td>
</tr>
<tr>
<td>(m_3)</td>
<td>-19.5</td>
<td>-14.3</td>
<td>-</td>
<td>-15.6</td>
</tr>
<tr>
<td>(m_4)</td>
<td>-21.5</td>
<td>-16.7</td>
<td>-17.5</td>
<td>-</td>
</tr>
</tbody>
</table>
Optimization Results

- Time-Slicing MDM can achieve at least 60% higher throughput than conventional MDM scheme.
- This throughput increase comes from flexible allocation of time slices.
- This increase will be more significant when adopting more wavelength and modes.

Zhizhen Zhong, Nan Hua, Yufang Yu, Zhongying Wu, Juhao Li, Haozhe Yan, Shangyuan Li, Ruijie Luo, Jialong Li, Yanhe Li, and Xiaoping Zheng, “Throughput Scaling for MMF-Enabled Optical Datacenter Networks Based on Time-Slicing-Based Crosstalk Mitigation,” to be presented, OFC, Mar. 2018.
Experimental Demonstration

- 4-mode MDM transmission systems of OM3 MMF with core/cladding diameter of 50/125μm.
- Xilinx VC709 FPGA generate PRBS codes at 10Gb/s at 1550nm wavelength.
- OTSS frame length 20ms, min time slice 500μs, guard interval 50μs.
- Magneto-electronic (MO) 2 × 2 optical switch.

Zhizhen Zhong, Nan Hua, Yufang Yu, Zhongying Wu, Juhao Li, Haozhe Yan, Shangyuan Li, Ruijie Luo, Jialong Li, Yanhe Li, and Xiaoping Zheng, “Throughput Scaling for MMF-Enabled Optical Datacenter Networks Based on Time-Slicing-Based Crosstalk Mitigation,” to be presented, OFC, Mar. 2018.
Overview of my recent publications on OTSS


Outline

- The need for a transparent fine-grained optical network
- Past solutions transparent fine-grained optical switching
- Our plan: Optical Time Slice Switching (OTSS)
- OTSS use cases
- Conclusion and discussions
Conclusion

- Time-division-multiplexing (TDM) is the main way to a transparent fine-grained optical network.
- Time synchronization over network is the key to bufferless switching and networking.
- OTSS is proposed as transparent fine-grained bufferless networking paradigm for next-generation optical networks.
- New problems arises as how to allocate resource for OTSS networks, and novel solutions have been proposed.
- Several use cases where OTSS show significant advantages over conventional networks.
Thank you for attention!

Zhizhen Zhong
Tsinghua University & UC Davis
zhongzz14@mails.tsinghua.edu.cn, zzzhong@ucdavis.edu
02 Feb. 2018
Networks Lab Group Meeting
Mice flows are numerous: 99% of flows are smaller than 100MB. However, more than 90% of bytes are in flows between 100MB and 1GB.