Data Center's Energy Savings for Data Transport via TCP on Hybrid Optoelectronic Switches

Artur Minakhmetov, Cédric Ware, and Luigi Iannone

LTCI, Télécom Paris, Institut Polytechnique de Paris

TuC3.3, IEEE IPC 2019 talk: October 1, 2019

Motivation for Optically-Switched Data Center Networks

- IT sector energy consumption **growing 9%/year**, currently 4% carbon emissions

Motivation for Optically-Switched Data Center Networks

- IT sector energy consumption **growing 9%/year**, currently 4% carbon emissions
  - Energy OpEx: about $1/3$ = data centers, of which **60% = switching & transport**
  - Currently: Electronic Packet Switching (EPS) over Optical Circuit Switching
  - Packets need Optical-Electrical-Optical conversion at every switch!
  - Packet Loss Ratio (PLR) $\simeq 0$

Motivation for Optically-Switched Data Center Networks

- IT sector energy consumption **growing 9%/year**, currently 4% carbon emissions
  - Energy OpEx: about $1/3$ = data centers, of which **60% = switching & transport**
  - Currently: Electronic Packet Switching (EPS) over Optical Circuit Switching
  - Packets need Optical-Electrical-Optical conversion at every switch!
  - Packet Loss Ratio (PLR) $\simeq 0$
- Optical or Hybrid Packet Switching (OPS/HOPS):
  - More efficient capacity use (packet mode)
  - Packets pass switches without OEO conversion
  - Reduces number of transceivers, can use burst mode $\Rightarrow$ less light emission (80% of transceiver power)

Motivation for Optically-Switched Data Center Networks

- IT sector energy consumption **growing 9%/year**, currently 4% carbon emissions
  - Energy OpEx: about $1/3$ = data centers, of which **60% = switching & transport**
  - Currently: Electronic Packet Switching (EPS) over Optical Circuit Switching
  - Packets need Optical-Electrical-Optical conversion at every switch!
  - Packet Loss Ratio (PLR) $\simeq 0$
- Optical or Hybrid Packet Switching (OPS/HOPS):
  - More efficient capacity use (packet mode)
  - Packets pass switches without OEO conversion
  - Reduces number of transceivers, can use burst mode $\Rightarrow$ less light emission (80% of transceiver power)
- OPS: High PLR due to contention and no practical optical buffers!

Motivation for Optically-Switched Data Center Networks

- IT sector energy consumption **growing 9%/year**, currently 4% carbon emissions
  - Energy OpEx: about 1/3 = data centers, of which **60% = switching & transport**
  - Currently: Electronic Packet Switching (EPS) over Optical Circuit Switching
  - Packets need Optical-Electrical-Optical conversion at every switch!
  - Packet Loss Ratio (PLR) $\simeq 0$
- Optical or Hybrid Packet Switching (OPS/HOPS):
  - More efficient capacity use (packet mode)
  - Packets pass switches without OEO conversion
  - Reduces number of transceivers, can use burst mode $\Rightarrow$ less light emission (80% of transceiver power)
- OPS: High PLR due to contention and no practical optical buffers!
- HOPS: use an electronic buffer to alleviate contention in OPS
  - OEO conversions only for buffering
  - Low PLR (though higher than EPS)

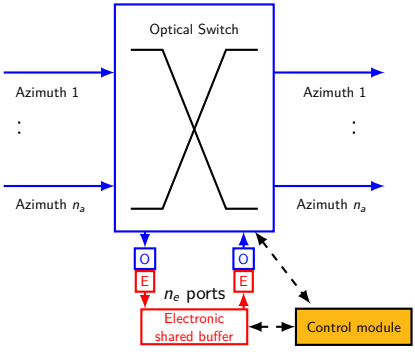
Motivation for Optically-Switched Data Center Networks

- IT sector energy consumption **growing 9%/year**, currently 4% carbon emissions
  - Energy OpEx: about 1/3 = data centers, of which **60% = switching & transport**
  - Currently: Electronic Packet Switching (EPS) over Optical Circuit Switching
  - Packets need Optical-Electrical-Optical conversion at every switch!
  - Packet Loss Ratio (PLR) $\simeq 0$
- Optical or Hybrid Packet Switching (OPS/HOPS):
  - More efficient capacity use (packet mode)
  - Packets pass switches without OEO conversion
  - Reduces number of transceivers, can use burst mode $\Rightarrow$ less light emission (80% of transceiver power)
- OPS: High PLR due to contention and no practical optical buffers!
- HOPS: use an electronic buffer to alleviate contention in OPS
  - OEO conversions only for buffering
  - Low PLR (though higher than EPS) $\rightarrow$ **leverage TCP's congestion control**

Outline

- Motivation
- Hybrid Optical Packet Switching for Data Center Networks: Hybrid Switch + TCP CCA
  - Hybrid Switch – Device Level Solution
  - TCP CCAs – Network Level Solution for DCs
- Energy Savings for Data Transport in HOPS network
- Conclusions

- Hybrid switch = **all-optical** switch + shared **electronic** buffer
  - Switch has n_e Input/Output (I/O) ports to/from buffer
  - If packet is blocked: put into buffer
    - When output port becomes free, re-emit FIFO.



Hybrid Packet Switch Concept

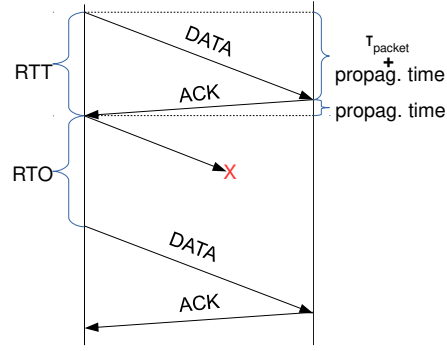
Fig. Source: W. Samoud, Performance Analysis of Hybrid Opto-Electronic Packet Switch, 2016

HOPS for DCN

TCP CCAs – DCs Network Level Solution

- TCP vs packet loss: acknowledge packets (ACK)
  - No ACK after Retransmission Time Out (RTO) → retransmit
  - Congestion Control Algorithms (CCAs) manage how many packets to send while waiting for ACK
  - TCP SACK: limits retransmissions

- TCP vs packet loss: acknowledge packets (ACK)
  - No ACK after Retransmission Time Out (RTO) → retransmit
  - Congestion Control Algorithms (CCAs) manage how many packets to send while waiting for ACK
  - TCP SACK: limits retransmissions
- TCP Stop-And-Wait (SAW)
  - For short ranges: one packet in flight
  - SAW: RTO ≈ round-trip time

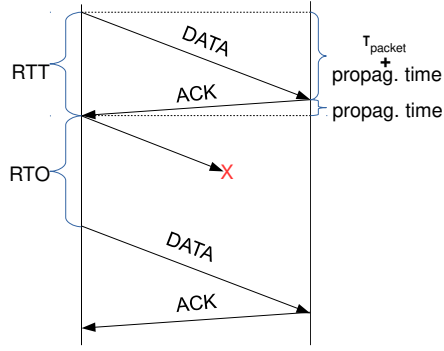


TCP SAW Working principle

HOPS for DCN

TCP CCAs – DCs Network Level Solution

- TCP vs packet loss: acknowledge packets (ACK)
  - No ACK after Retransmission Time Out (RTO) → retransmit
  - Congestion Control Algorithms (CCAs) manage how many packets to send while waiting for ACK
  - TCP SACK: limits retransmissions
- TCP Stop-And-Wait (SAW) → SAW-Longer (SAWL):
  - For short ranges: one packet in flight
  - SAW: RTO ≈ round-trip time
  - SAWL: add estimated buffer time to RTO



TCP SAW Working principle

A. Minakhmetov et al, Optical Networks Throughput Enhancement via TCP Stop-and-Wait on Hybrid Switches, OFC'18

Outline

- Motivation
- Hybrid Optical Packet Switching for Data Center Networks: Hybrid Switch + TCP CCA
  - Hybrid Switch – Device Level Solution
  - TCP CCAs – Network Level Solution for DCs
- Energy Savings for Data Transport in HOPS network
  - Metric for Optical-Electronic-Optical Conversions Reduction
  - Simulation Conditions
  - Network performance results
- Conclusions

Energy Savings for Data Transport

Metric for Electronic-Optical Conversions Reduction

- Burst Transceiver can spend $>$**80 %** of power on Tx $\rightarrow$ EO conversions predominant

Energy Savings for Data Transport

Metric for Electronic-Optical Conversions Reduction

- Burst Transceiver can spend $>$**80 %** of power on Tx $\rightarrow$ EO conversions predominant
- Metric to measure EO conversions:

$$\text{Bit transport energy factor} = \frac{Data_{pckt}[B] \times EO_{data} + Ack_{pckt}[B] \times EO_{ack}}{Payload[B]}$$

  - Defined as how many bits should be physically emitted to ensure delivery of one bit

Energy Savings for Data Transport

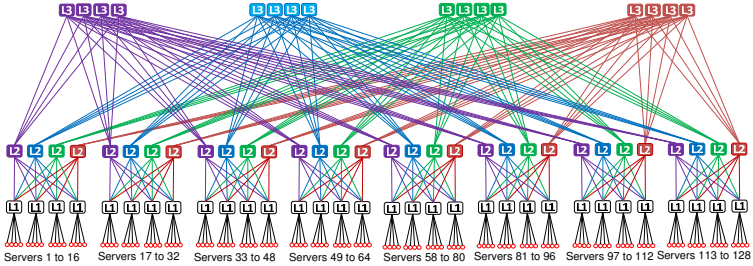
Metric for Electronic-Optical Conversions Reduction

- Burst Transceiver can spend $>$**80 %** of power on Tx $\rightarrow$ EO conversions predominant
- Metric to measure EO conversions:

$$\text{Bit transport energy factor} = \frac{Data_{pckt}[B] \times EO_{data} + Ack_{pckt}[B] \times EO_{ack}}{Payload[B]}$$

  - Defined as how many bits should be physically emitted to ensure delivery of one bit
  - Takes into account RTO re-transmissions induced by TCP CCA
  - Takes into account EO conversions induced by buffer of a Hybrid Switch
  - Estimates energy consumption by multiplying with [J/b] of a transmitters used
- "Transmission energy cost" measures BTEF under varying network load

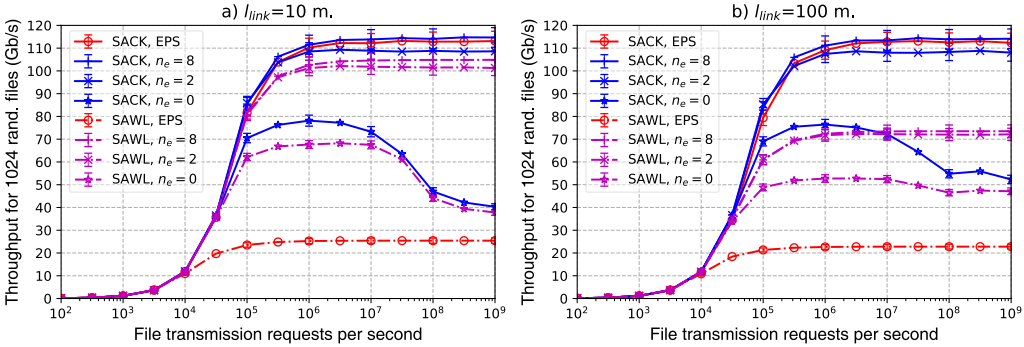
Data Center Simulation Conditions



8-ary fat-tree DC network (related to **Facebook DC** network)

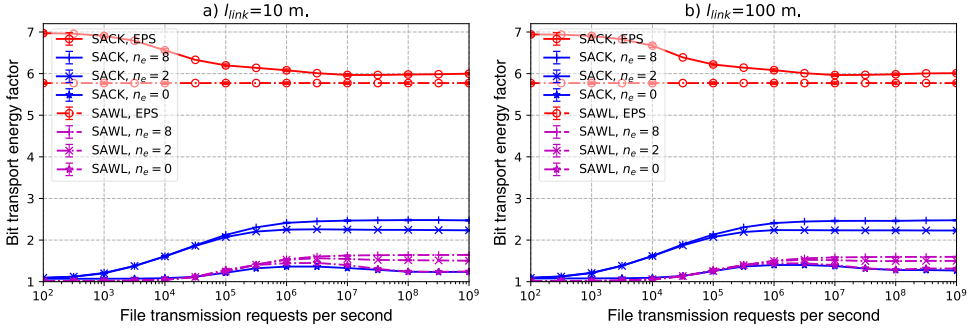
- Two cases of $l_{link} = \{10, 100\} m$.
- 4 switch types: $n_e = 0$ (OPS), $n_e = 2$ and $n_e = 8$ (HOPS), all-electronic switch (EPS).
- Two TCP CCAs: TCP SAWL and TCP SACK.
- File transmission through TCP connection, with packet size = 9 kB on 10 Gbit/s bit-rate.
- Load – mean number of file transmission requests/s (req/s) in Poissonian process.
- Throughput (Gbit/s) and Transmission energy cost studied as function of load.

Network performance: Throughput



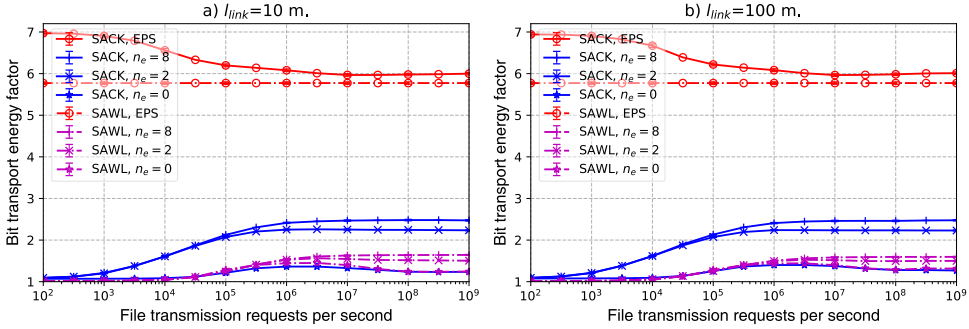
- Performance on high load of OPS drops, HOPS holds with **few n_e** and **outperform** OPS.
- EPS+SAWL performs poorly due to high latency, invoked by store-and-forward mode.
- SACK outperforms SAWL **by only 10%** for $l_{link} = 10$ m and 50% for 100 m on HOPS.
- SACK on HOPS, $n_e = 2$ is very close to EPS, and with $n_e = 8$ **outperforms** EPS.

Network performance: Transmission energy cost (1/2)



- EPS performs **the worst** in terms of energy consumption (all packets OEO).
- Worst case of HOPS outperforms best case of EPS by **factor more than 2**.
- OPS performs **the best** energy-wise (but not throughput-wise).
- No change for different l_{link}.

Network performance: Transmission energy cost (2/2)



- SACK + HOPS consumes $\approx \times 1.5$ more than SAWL + HOPS.
- SAWL + OPS consumes least energy, but as well has lowest throughput.
- SAWL + HOPS, $n_e = 2$ is a trade-off solution for $l_{link} = 10m$ DCN:
  - Throughput: SAWL + HOPS, $n_e = 2$ outperformed **by only 10%** by SACK+EPS.
  - Energy: SAWL + HOPS, $n_e = 2$ saves **4 times** than SACK+EPS.

Outline

- Motivation
- Hybrid Optical Packet Switching for Data Center Networks: Hybrid Switch + TCP CCA
  - Hybrid Switch – Device Level Solution
  - TCP CCAs – Network Level Solution for DCs
- Energy Savings for Data Transport in HOPS network
  - Metric for Optical-Electronic-Optical Conversions Reduction
  - Simulation Conditions
  - Network performance results
- Conclusions

Conclusions

- **HOPS** = **robust solution** in OPS data center network with **few n_e**.
- HOPS + SACK delivers **best throughput**, better than EPS +SACK, and energy consumption **reduced by factor of 2** at least.
- HOPS + SAWL delivers **only 10% lower** throughput than best, but help reduce energy consumption energy consumption **by factor of 4**.
- TCP CCA + hybrid switches = solution for making OPS in data center a reality.