TCP Performance in Hybrid Multigranular OBS Networks

Maurizio Casoni and Carla Raffaelli

(maurizio.casoni@unimore.it)

Department of Information Engineering
University of Modena and Reggio Emilia
Italy

Outline

- Introduction: Optical Burst Switching scenario
- Hybrid technology switch for supporting multi-granular connections
  - SOA
  - MEMS
  - Slow and fast paths
- Assembly algorithms
  - Mixed and per-flow
  - Time and volume based
- Investigated scenario and numerical results
  - TCP throughput and CWND behaviour
  - Ns2 simulation tool
- Conclusions

Optical Networks: Evolution

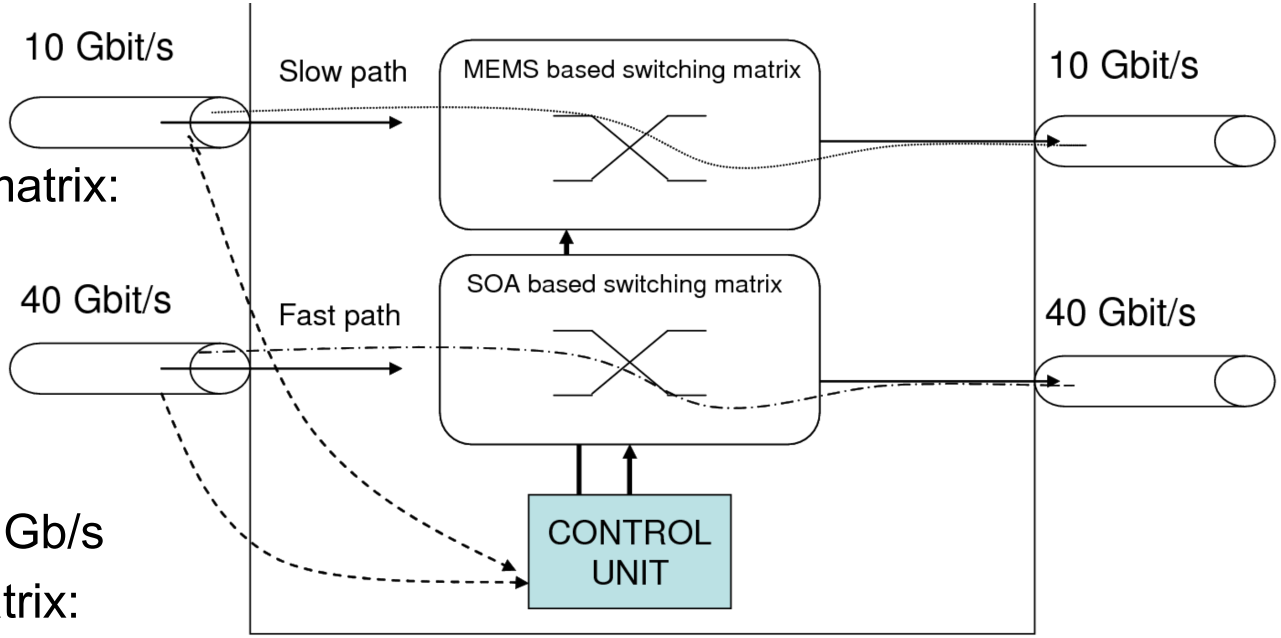
- DWDM technique
  - **Transmission rate in the range of Tbit/s**
- Architectural semplification
  - **From IP over ATM over SONET over WDM to IP over WDM**
- Need to exploit in an effective way the huge transmission bandwidth with IP traffic
  - **Wavelength Routing**
    - ✓ all-optical data network
    - ✓ Low flexibility for IP traffic
  - **Optical Packet Switching**
    - ✓ Ideal transfer mode for IP traffic
    - ✓ Severe technological constraints → not feasible in the short/middle term
      - Optical components immature
      - Optical buffers

Optical Burst Switching

Goal: better sinergy between the mature electronic technologies and the new optical tecnologies (mid-term solutions)

- **Switching granularity between WR and OPS**
  - *Burst concept:* aggregation of IP packets with common features (e.g. destination and QoS), considered as the basic optical unit
- **Time and <u>space separation</u> of data and control (header) fields**
  - Control packet employs dedicated channel and precedes the relative data burst
    - All-optical network, buffer-less and data trasparent
    - Hybrid opto-electronic network for control signals (*out-of-band signaling*)
  - Simplification of the electronic processing of the control packets at intermediate nodes
  - Reduction of the opto-electronic functionalities required to router

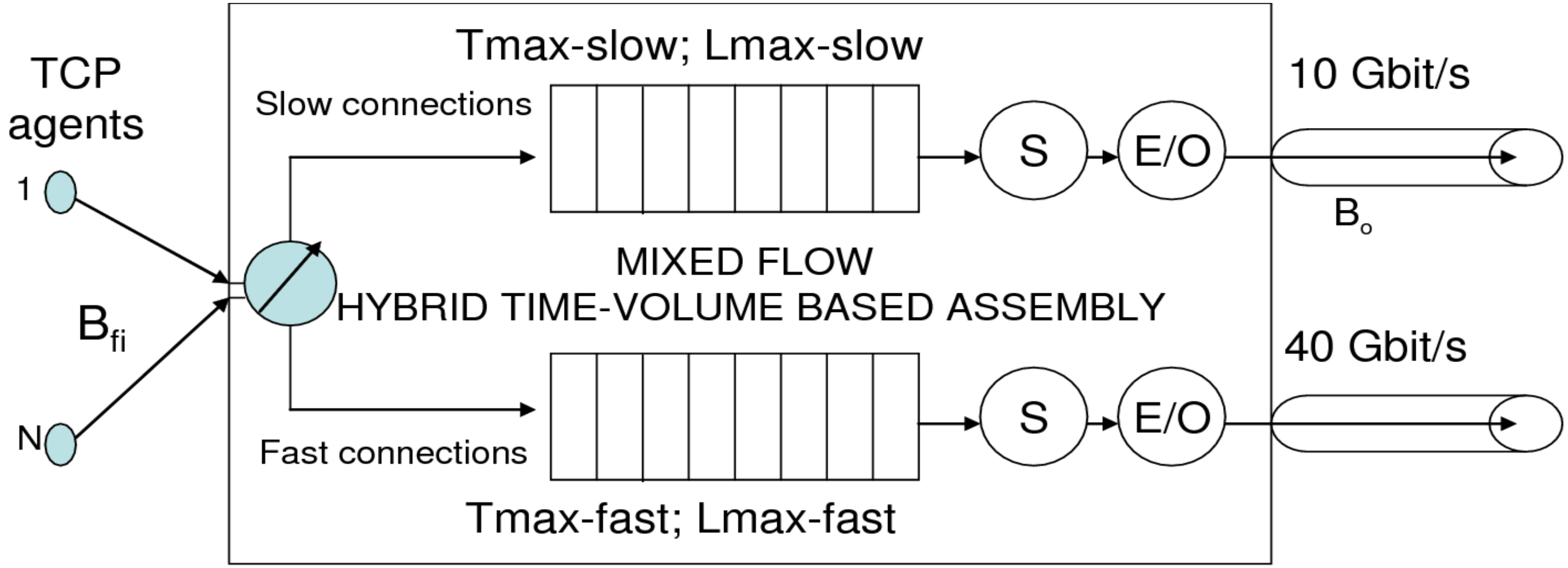
OBS CORE NODE



- **MEMS** based switching matrix:
  - reliable and flexible
  - high integration
  - relatively low cost
  - setup times: some ms
  - connected to I/O at 10 Gb/s
- **SOA** based switching matrix:
  - high performance
  - high costs
  - setup times: ns-some μs
  - connected to I/O at 40 Gb/s

- **Fast connection** (path):
  sequence of bursts switched in core nodes
  with fast optical technologies (short setup times)
- **Slow connection** (path):
  sequence of bursts switched in core nodes
  with slower optical technologies (long setup times)

OBS EDGE NODE



1. Input interface cards
   - IP datagram classification (fast or slow) and forwarding the assembly queue
2. Burst assembly unit
   - Mixed flow time-volume based
3. Output interface
   - Scheduling and E/O conversion

Tmax-slow > Tmax-fast

Lmax-slow > Lmax-fast

PERFORMANCE EVALUATION

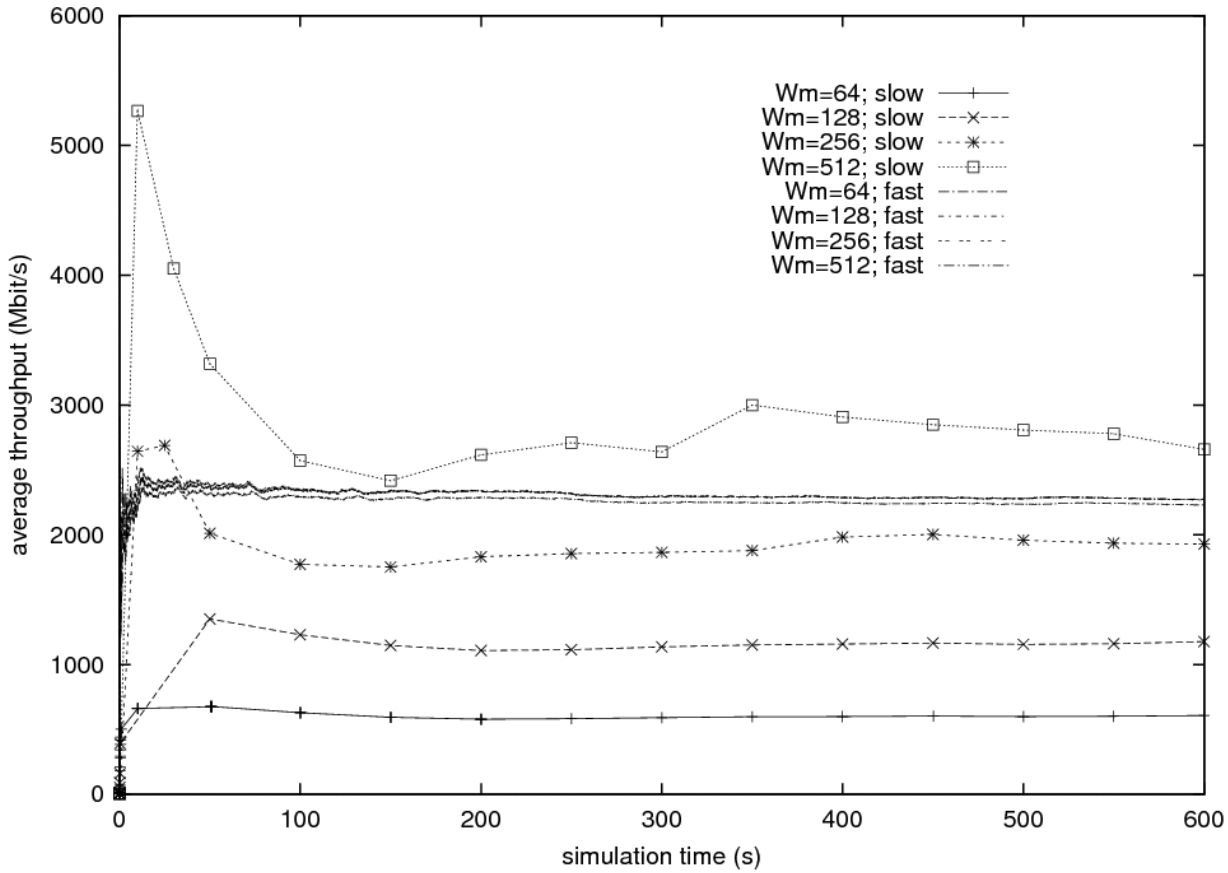
- **TCP SACK**
  - SACK option: the receiver informs the sender about the successfully received segments
  - sender retransmits lost segments only
- **Throughput**
  - Measure of the variability of the bandwidth usage over a given time-scale
- **Average throughput**
  - Amount of successfully transmitted bytes over a given time interval
- **Aggregated average throughput**
  - Average throughput over all active TCP flows, provided with a 100 Mb/s access bandwidth each

AGGREGATE AVERAGE THROUGHPUT

- MSS = 512 bytes
- Burst loss = $P_l = 10^{-3}$

TCP over slow paths

- N = 80
- B_o = 10 Gbit/s
- T_{max} = 10 ms
- L_{max} = 10 MB
- Best for W_m=512

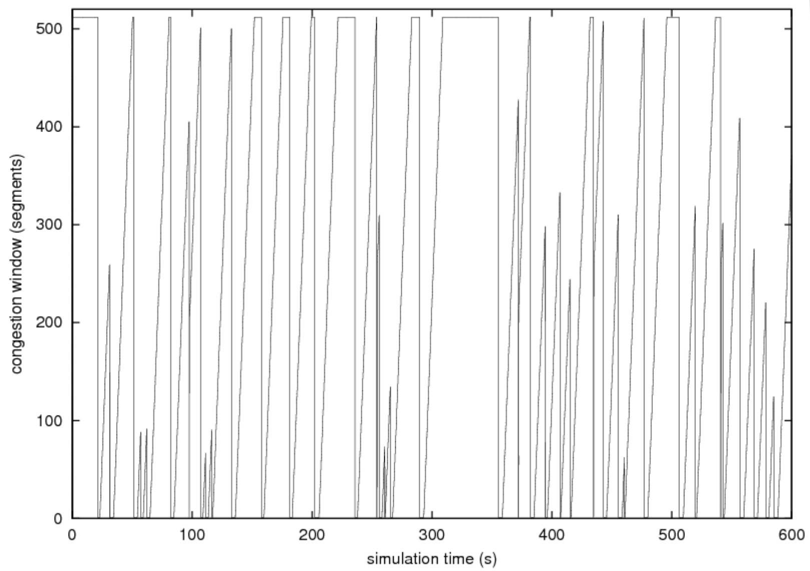


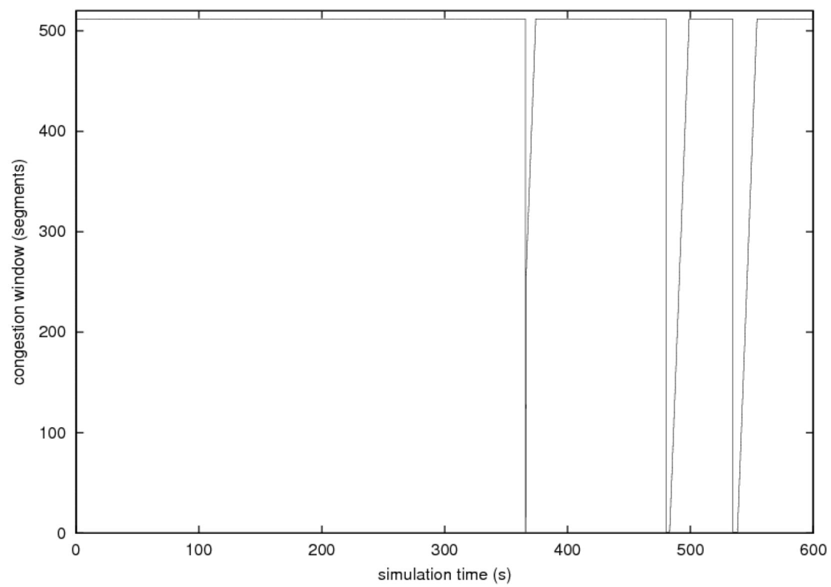
Ideal throughput

MSS x Wm X N / RTT = 5.5 Gbit/s

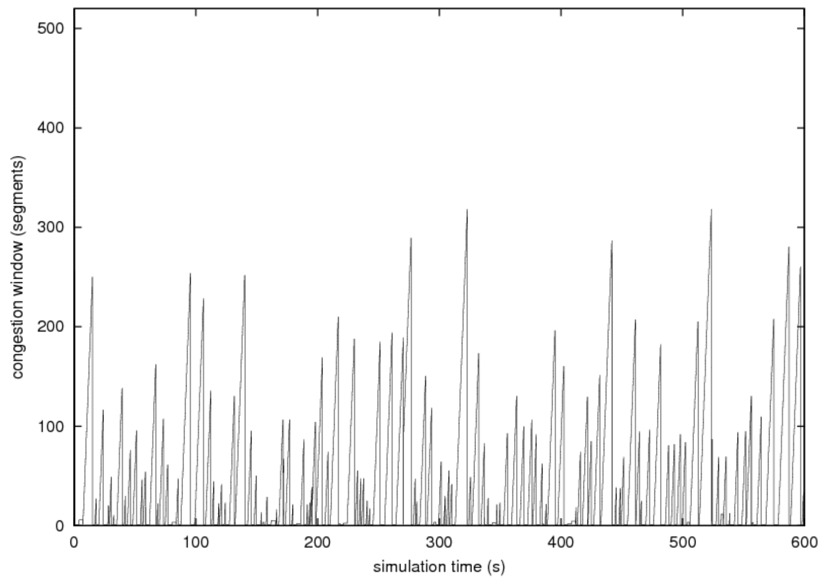
Different behaviour slow vs. fast

CWND FOR SLOW CONNECTIONS



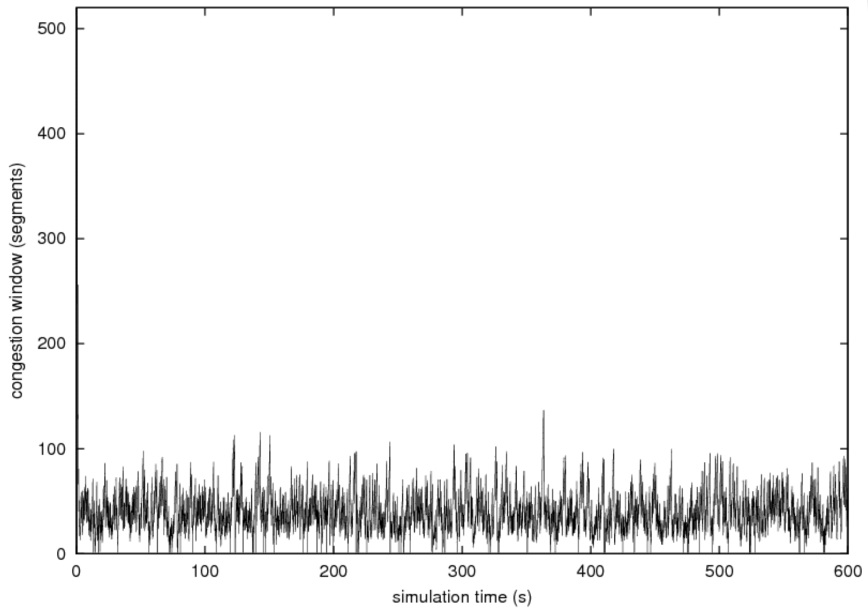


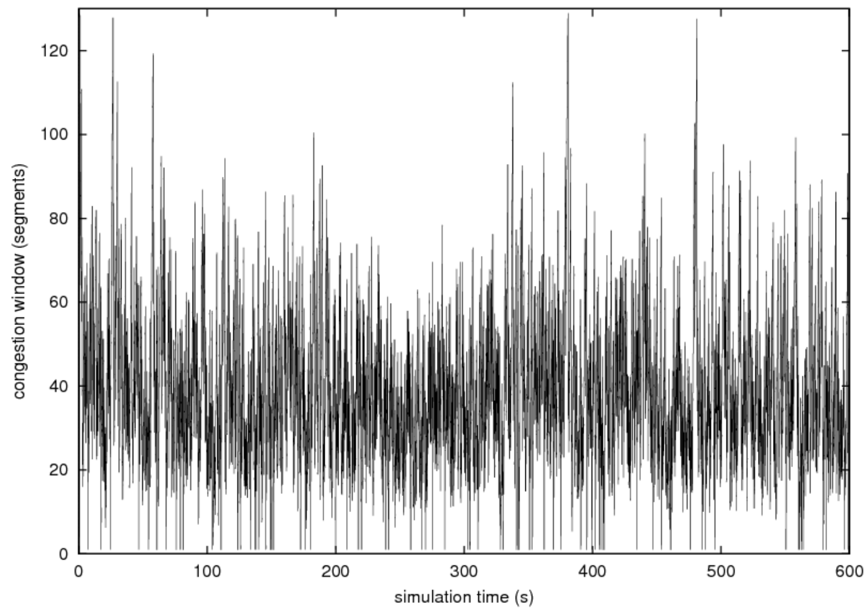
- Burst loss = 10^{-3}, 10^{-4} ,10^{-2}
- CWND = 512



- Losses detected by RTO expiration
- Av. No. of segments per burst = 12360
- i.e. each burst loss makes a TCP flow loose 154 segments (N=80)

CWND FOR FAST CONNECTIONS

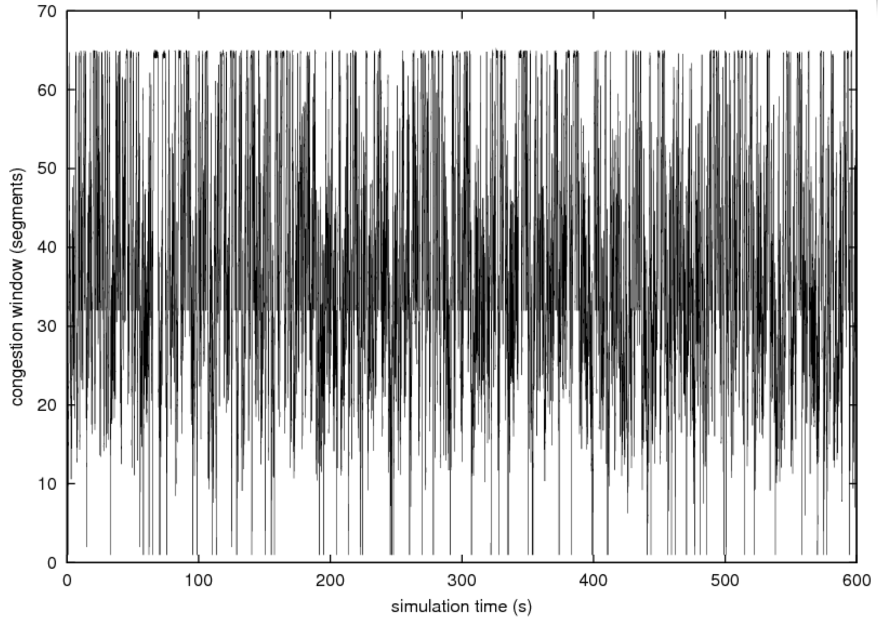


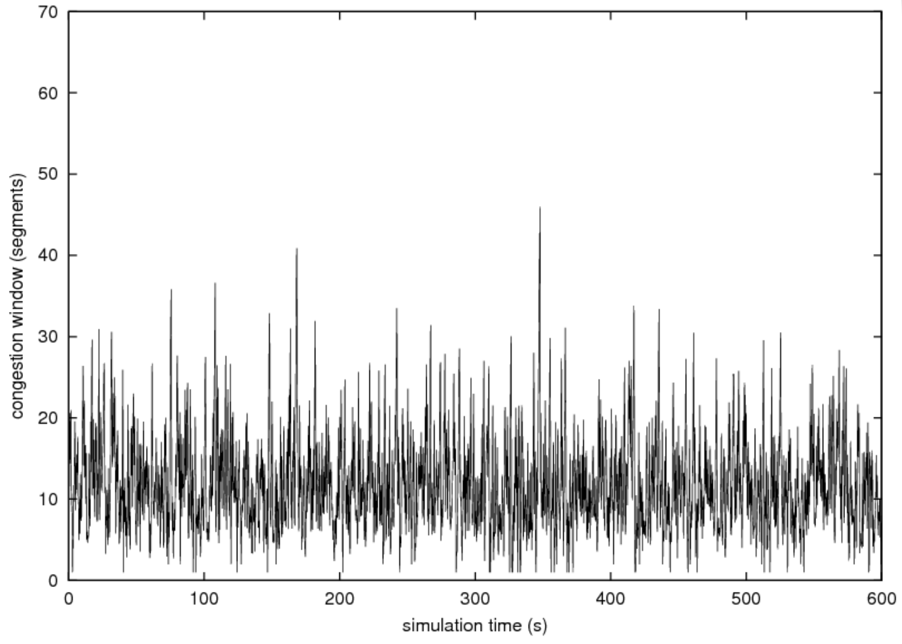


- **TCP over fast paths**
- N = 320
- Bo = 40 Gbit/s
- Tmax = 0.5 ms
- Lmax = 25 KB

- Throughput independent of W_m
- No gain by increasing W_m
- Burst loss = 10^{-3}, W_m = 512, 128
- Losses detected by TD
- Av. No. of segments per burst = 44
- Each source has at most 1 segment per loss
- But, losses are very frequent so CWND can not fully open

CWND FOR FAST CONNECTIONS





- Burst loss = 10^{-3} and 10^{-2}, W_m = 64
- W_m = 64 seems enough for optimize performance
- Losses detected by TD

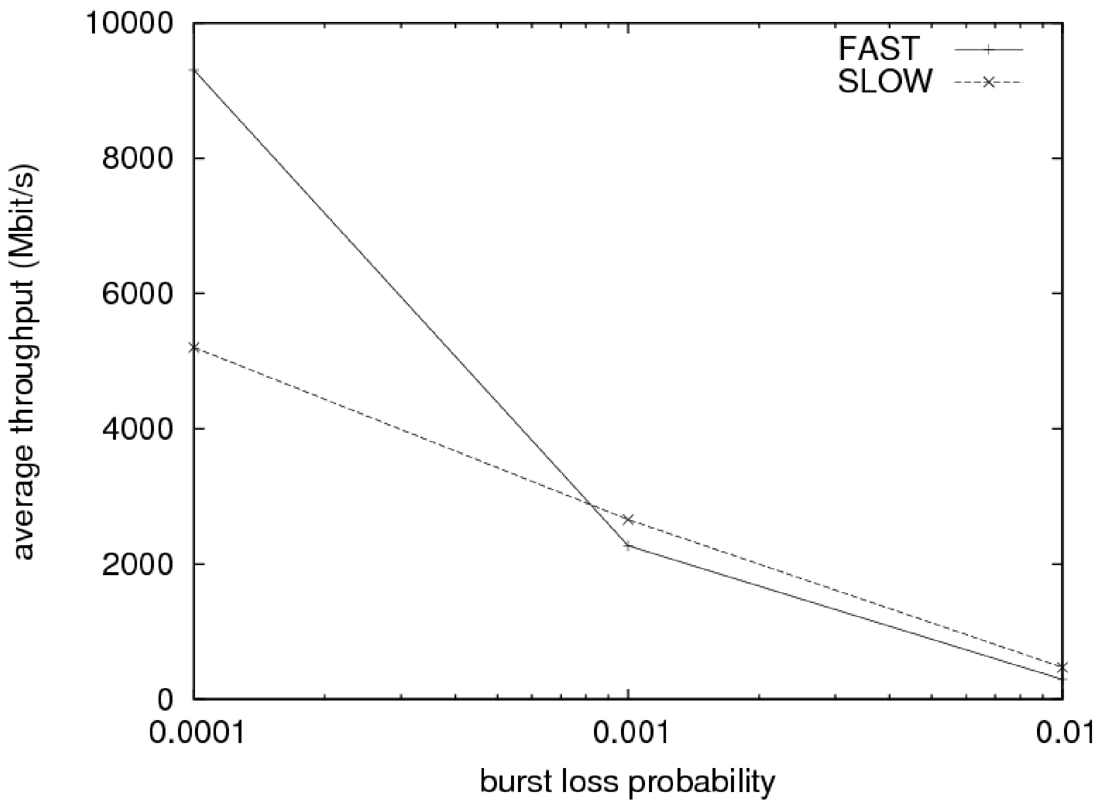
AGGREGATE AVERAGE THROUGHPUT

Fast paths:

- $W_m = 64$

Slow paths:

- $W_m = 512$



When burst loss is very low fast connections provide better throughput but they are very sensitive to losses so that already for burst loss = 10^{-3} performance remarkably degrade

Conclusions

- Performance evaluation of TCP with hybrid technologies
- Availability of fast and slow paths for supporting multi-granular connections
- Different switching set up times imply different offsets times
- Fast and slow paths get best throuhgput for different values of W_m
  - For fast paths there is not gain at increasing W_m above 64 segments
  - For slow paths the best W_m is 512
  - Different loss detection mechanisms: RTO vs. TD
- For low burst losses (e.g. lower than 10-3) fast paths give remarkably better throughput than slow paths, otherwise the thorouhgput is almost the same, despite of different transmission rates and technologies employed

THANK YOU FOR YOUR ATTENTION

maurizio.casoni@unimore.it
casoni@ieee.org

http://www.dii.unimore.it/~mcasoni

… suggestions are very very welcome