Fractional λ Switching: Node Design & Blocking Analysis

Viet-Thang Nguyen
Ph.D. Thesis Presentation
****
Advisor: Prof. Renato Lo Cigno
Co-advisor: Prof. Yoram Ofek

ICT School - University of Trento

March 01, 2007

Contents

UNIVERSITY OF TRENTO - Italy

Contents

UNIVERSITY OF TRENTO - Italy

Contents

Contents

UNIVERSITY OF TRENTO - Italy

Contents

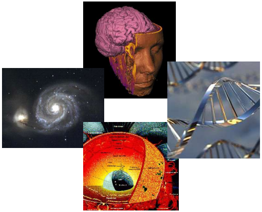
Contents

UNIVERSITY OF TRENTO - Italy

Problem 1: Streaming-oriented traffics threat the Internet

- Sources of streaming-oriented/real-time traffics: Joost, Inuk, YouTube, live streaming, VoD, IPtv, etc.
- App. 80% of the Internet traffic belongs to peer-2-peer applications.

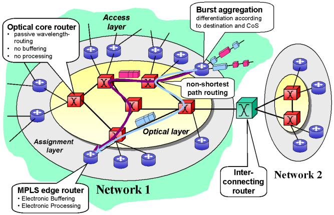
- If 35% link capacity is loaded by streaming traffic, link utilization starts degrade.
- Current router/switch arc. are not ready for the change: due to high OH and sophisticated QoS mechanisms.



FλS - A novel switching technology can solve the problem.

UNIVERSITY OF TRENTO - Italy

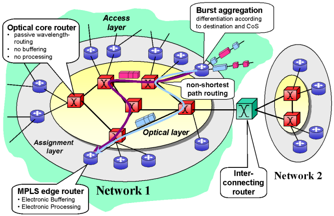
Problem 2: no sub-wavelength switching exists



Currently, it is the whole wavelength switching:

- A single WL capacity: 2.5-100 $Gbit/s$.
- Can accommodate a large num. of IP-traffic users.
- More bandwidth efficient to partition an optical channel into a num. of sub-channels.

Problem 2: no sub-wavelength switching exists



Currently, it is the whole wavelength switching:

- A single WL capacity: 2.5-100 $Gbit/s$.
- Can accommodate a large num. of IP-traffic users.
- More bandwidth efficient to partition an optical channel into a num. of sub-channels.

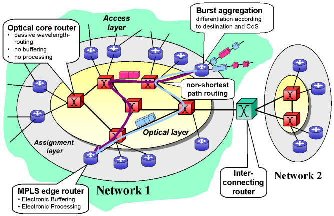
Some efforts to realize sub-WL switching:

- Optical burst switching (OBS): complex control plane, high collision of large pac. (bursts) → high blocking/low utilization.
- Optical packet switching (OPS): no Optical-RAM; all-optical processing not available.
- Some others: SONET/SDH, WDM+TDM, TSI-WDM: timing/synchornization issues; do not use pipeline forwarding; some are special cases of FλS.

FλS is a promising approach to realize sub-wavelength switching.

UNIVERSITY OF TRENTO - Italy

Problem 2: no sub-wavelength switching exists



Currently, it is the whole wavelength switching:

- A single WL capacity: 2.5-100 $Gbit/s$.
- Can accommodate a large num. of IP-traffic users.
- More bandwidth efficient to partition an optical channel into a num. of sub-channels.

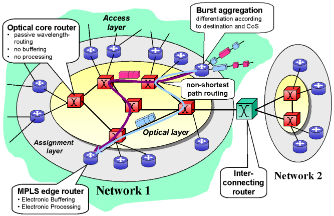
Some efforts to realize sub-WL switching:

- Optical burst switching (OBS): complex control plane, high collision of large pac. (bursts) → high blocking/low utilization.
- Optical packet switching (OPS): no Optical-RAM; all-optical processing not available.
- Some others: SONET/SDH, WDM+TDM, TSI-WDM: timing/synchornization issues; do not use pipeline forwarding; some are special cases of FλS.

FλS is a promising approach to realize sub-wavelength switching.

UNIVERSITY OF TRENTO - Italy

Problem 2: no sub-wavelength switching exists



Currently, it is the whole wavelength switching:

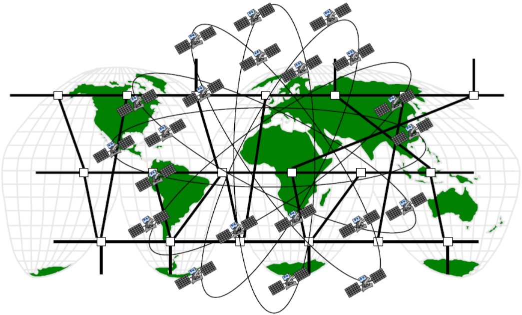
- A single WL capacity: 2.5-100 $Gbit/s$.
- Can accommodate a large num. of IP-traffic users.
- More bandwidth efficient to partition an optical channel into a num. of sub-channels.

Some efforts to realize sub-WL switching:

- Optical burst switching (OBS): complex control plane, high collision of large pac. (bursts) → high blocking/low utilization.
- Optical packet switching (OPS): no Optical-RAM; all-optical processing not available.
- Some others: SONET/SDH, WDM+TDM, TSI-WDM: timing/synchronization issues; do not use pipeline forwarding; some are special cases of FλS.

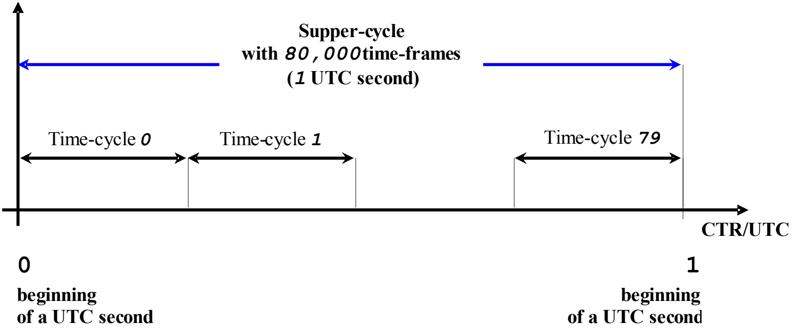
FλS is a promising approach to realize sub-wavelength switching.

Fractional Lambda Switching (FλS)



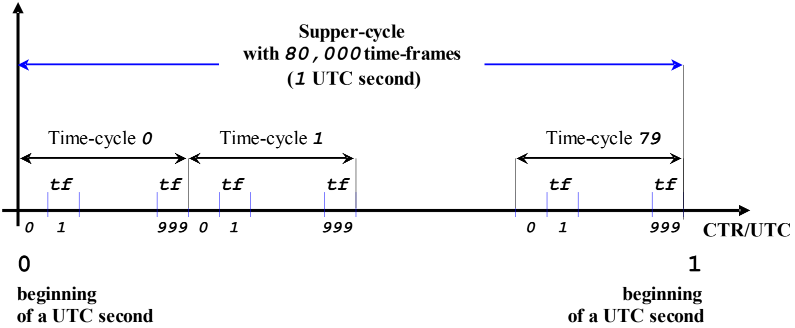
- Use a Common Time Reference (CTR) to guide traffics end-to-end;
- Enable the pipeline forwarding method;
- All network nodes are synchronized to Universal Time Coordinated (UTC).
- Header processing is eliminated;
- It is a jitter- loss- congestion-free networking paradigm.

FλS: Timing principle



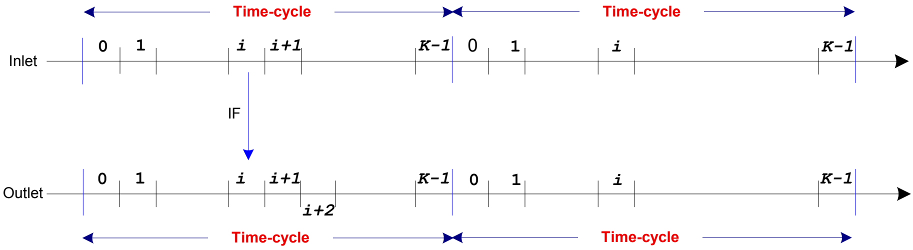
- A common-time-reference: UTC (from GPS, Galileo, Glonass);
- 1 UTC sec (1 super cycle) is split into time-cycles;

FλS: Timing principle



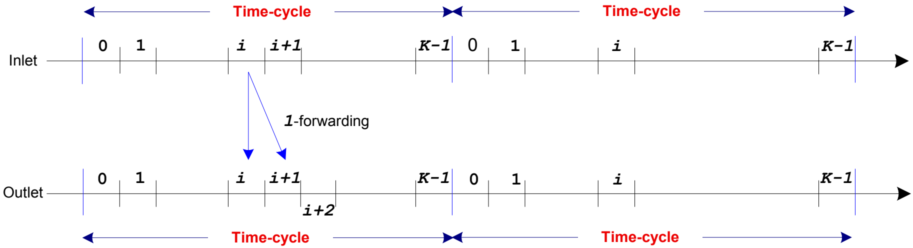
- 1 UTC sec (1 super cycle) is split into time-cycles;
- A time-cycle is split into multiple time-frames.

FλS: immediate forwarding



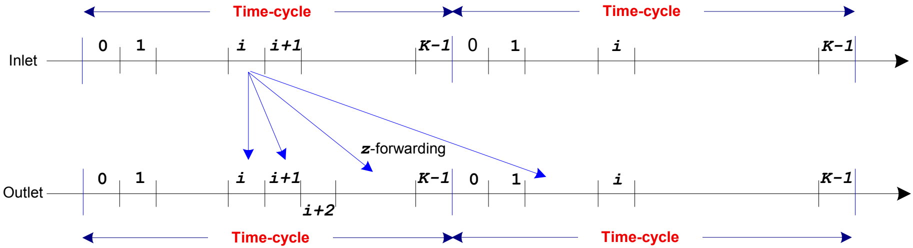
- IF scheme ⇔ zero scheduling delay
- No buffer is required.

FλS: 1-forwarding



- NIF schemes: scheduling delay is nonzero;
- 1-forwarding: 1 TF buffering is required.

FλS: *z*-forwarding

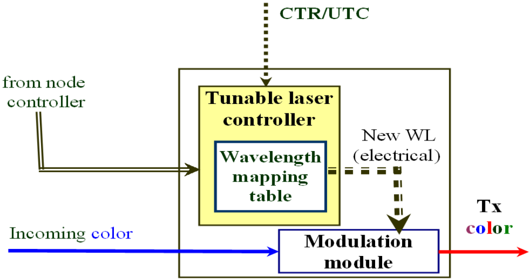


- Maximum z TFs delay for scheduling;
- The content of TF can be delayed for k_z TFs prior to being forwarded, $0 \leq k_z \leq z$.

Contents

Time-driven tunable Laser for wavelength swapping

- Advanced devices allow fast wavelength swapping;
- Important property: nonstop bit-stream.



- Various mechanisms to tune the transmitted WL.
- Very difficult to obtain stable operation, wide range tuning: no drift, thermal stabilization, etc.
- Slow tuning products are used in SONET/SDH/DWDM systems;

Node design criterions

Designs are compared:

- Low hardware complexity;
- High *scheduling feasibility*;

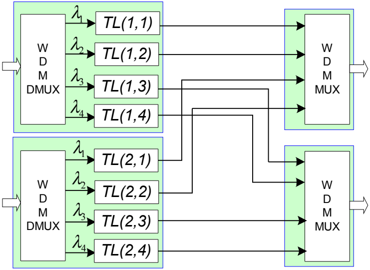
Scheduling feasibility:

- a func. of z, K, C, and N for a given h;
- measures the max. num. of *distinct schedules* that are available using time and wavelength swapping.
- relates to blocking performance.

UNIVERSITY OF TRENTO - Italy

Fixed-connection based design: FC-FλS

- No switching fabric;
- A set of fixed connections from in-ports to out-ports;
- Least cost but poor performances;



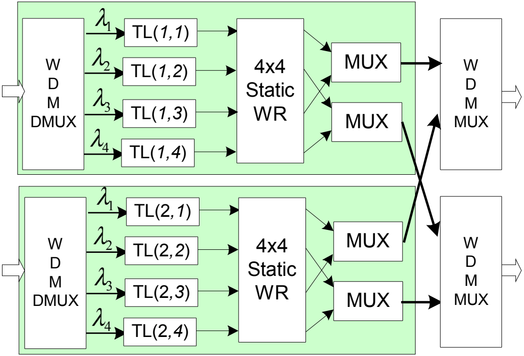
Pros:

- No switching fabric ⇒ low HW cost;
- Low operational overhead (only TLs are active devices).

and cons:

- Rigid routing;
- Low scheduling flexibility ⇒ high blocking probability ⇒ low throughput.

Wavelength-router based design: WR-FλS



- Not a new design;
- Diff. in-ports must use dif. sets of channels to reach the same out-port;
- Sets are fixed and depend on permutation patterns.

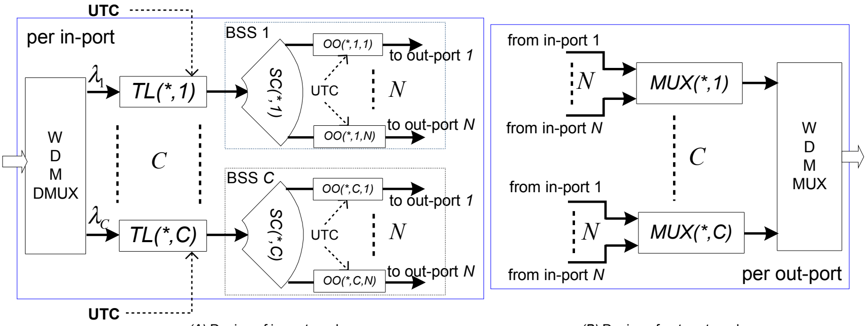
Pros:

- Low operational overhead (only TLs are active devices).
- Fabric: contention-free (nature of static WRs);

and cons:

- Low scheduling flexibility.
- Scheduling flexibility depends on the connection ratio $r = C/N$.

Broadcast-then-select design: BS-FλS

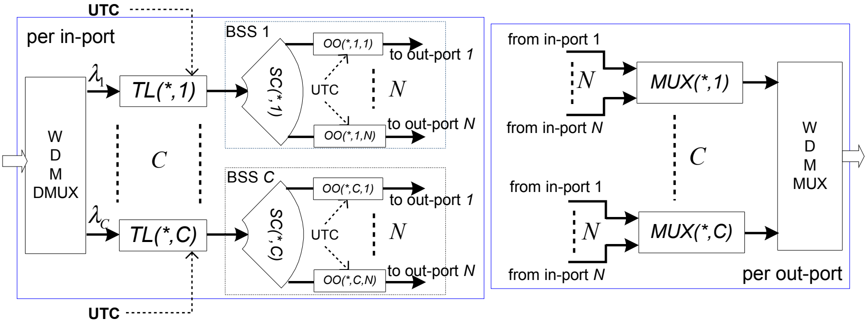


(A) Design of in-port card (B) Design of out-port card

- 1 tunable laser + 1 BSS block per input channel;
- A BSS comprises of one star-coupler and N ON/OFF switching elements;
- *Important*: If $N = C$, the HW complexity (i.e. num. of switching elements) is Clos equivalency:

$$CN^2 = N'\sqrt{N'}$$

BS-FλS: strictly non-space-blocking design



(A) Design of in-port card

(B) Design of out-port card

Pros:

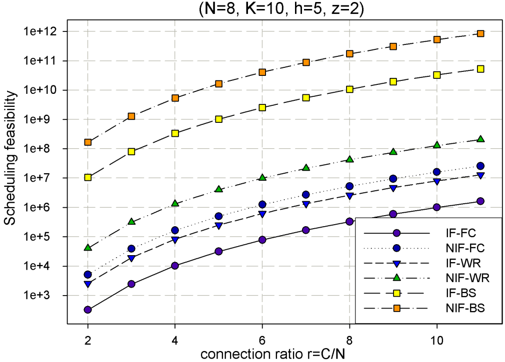
- Strictly non-space-blocking;
- Full routing adaptivity;
- Complexity level equals to Clos network.

and cons:

- High overhead: tunable lasers and ON/OFF elements are controlled.

Brief comparisons

	Hardware				Scheduling Feasibility		Routing Adapt.
	N_{TL}	N_{WR}	N_{SC}	N_{OO}	IF scheme	NIF scheme	
FC	NC	- -	- -	- -	$K\left(\frac{C}{N}\right)^h$	$K\left(\frac{C}{N}\right)^h(z+1)^{h-1}$	None
WR	NC	N	- -	- -	$K\left(\frac{C}{N}\right)^h N$	$K\left(\frac{C}{N}\right)^h(z+1)^{h-1}N$	Partial
BS	NC	- -	NC	N^2C	$K\left(\frac{C}{N}\right)^h N^h$	$K\left(\frac{C}{N}\right)^h(z+1)^{h-1}N^h$	Full

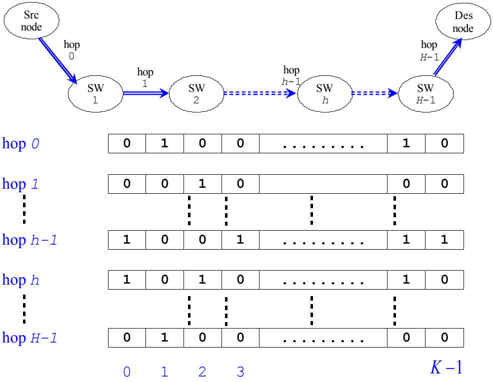


- *Scheduling feasibility* - func. of z, K, C and N for a given h, measures the max. num. of *distinct schedules* that are available using time and wavelength swapping.
- *Routing adapt.* indicates the freedom of changing the WL routing.

Contents

UNIVERSITY OF TRENTO - Italy

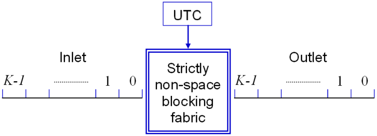
Time-blocking analysis: the problem



- H hops: indexed from 0 to $H-1$;
- Hop load: a available TFs (sym. '1', b busy TFs (sym. '0'), $a+b=K$;
- Distribution of TFs is uniform, independent between hops;
- Nodes (switches) are strictly non-space-blocking (e.g., the BS-FλS design.)

Blocking is defined as the occurrence in which transmission resources are available (i.e., some available TFs at all hops), but there is no schedule.

A single switch blocking analysis



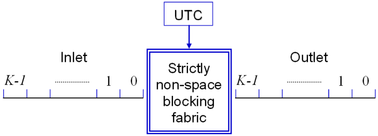
k	0	1	2	3	4	5	6	7	8	9	10	11
inlet	0	1	0	0	1	1	0	0	0	1	0	0
outlet	1	0	0	0	0	0	0	1	0	0	1	0

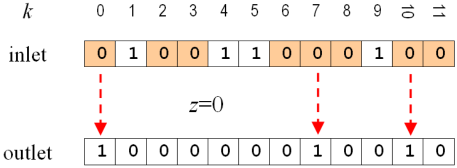
$z=0$

Zero scheduling delay:

$$p_{blk} = \binom{b}{a} \Big/ \binom{K}{b}$$

A single switch blocking analysis

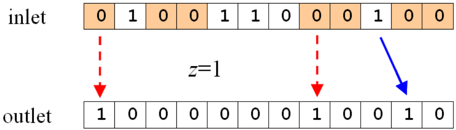




Zero scheduling delay:

$$p_{blk} = \binom{b}{a} \Big/ \binom{K}{b}$$

Nonzero scheduling delay: a complex counting problem.



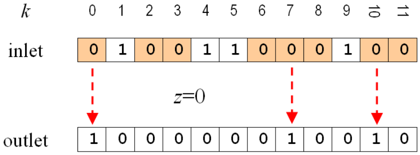
A single switch analysis: the approach

Place all avai. TFs of the outlet into <u>blocked positions</u> generated by the inlet.

Num. of *blocked* positions generated by the inlet depends on:

- z-forwarding scheme.
- distribution of *available* TFs at the inlet.

$z = 0$



UNIVERSITY OF TRENTO - Italy

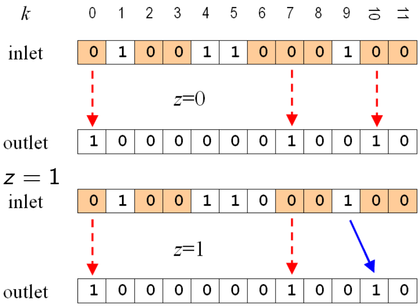
A single switch analysis: the approach

Place all avai. TFs of the outlet into blocked positions generated by the inlet.

Num. of *blocked* positions generated by the inlet depends on:

- z-forwarding scheme.
- distribution of *available* TFs at the inlet.

$z = 0$



To derive p_{blk}, we find num. of comb. generating exactly:

- a *blocked* positions.
- $a+1$ *blocked* positions.
- ...
- K *blocked* positions.

Arrangements beans in a circle (not exactly the necklace problem in combinatorics) + under various counting constraints.

UNIVERSITY OF TRENTO - Italy

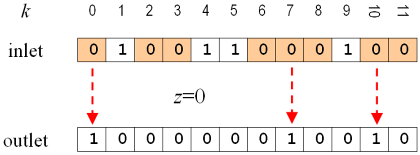
A single switch analysis: the approach

Place all avai. TFs of the outlet into <u>blocked positions</u> generated by the inlet.

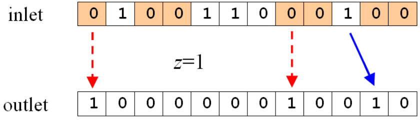
Num. of *blocked* positions generated by the inlet depends on:

- z-forwarding scheme.
- distribution of *available* TFs at the inlet.

$z = 0$



$z = 1$



To derive p_{blk}, we find num. of comb. generating exactly:

- a *blocked* positions.
- $a + 1$ *blocked* positions.
- ...
- K *blocked* positions.

Arrangements beans in a circle (not exactly the necklace problem in combinatorics) + under various counting constraints.

UNIVERSITY OF TRENTO - Italy

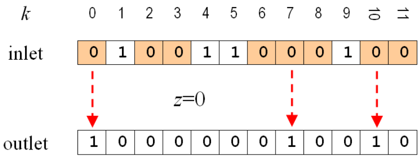
A single switch analysis: the approach

Place all avai. TFs of the outlet into blocked positions generated by the inlet.

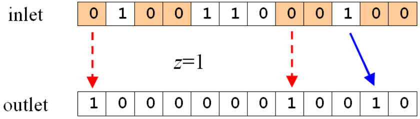
Num. of *blocked* positions generated by the inlet depends on:

- z-forwarding scheme.
- distribution of *available* TFs at the inlet.

$z = 0$



$z = 1$



To derive p_{blk}, we find num. of comb. generating exactly:

- a *blocked* positions.
- $a + 1$ *blocked* positions.
- ...
- K *blocked* positions.

Arrangements beans in a circle (not exactly the necklace problem in combinatorics) + under various counting constraints.

UNIVERSITY OF TRENTO - Italy

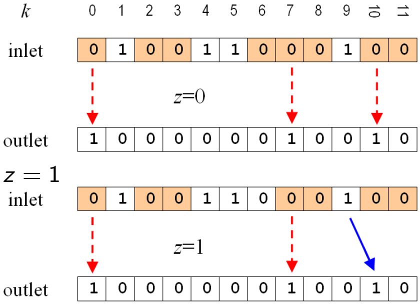
A single switch analysis: the approach

Place all avai. TFs of the outlet into <u>blocked positions</u> generated by the inlet.

Num. of *blocked* positions generated by the inlet depends on:

- z-forwarding scheme.
- distribution of *available* TFs at the inlet.

$z = 0$

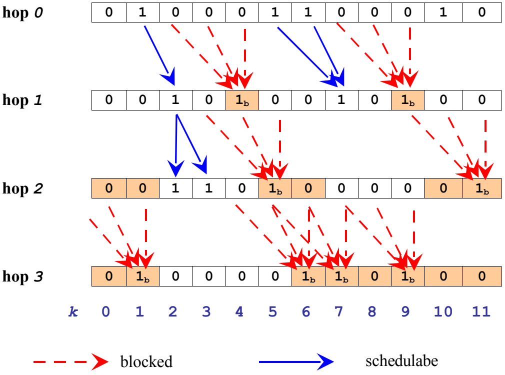


To derive p_{blk}, we find num. of comb. generating exactly:

- a *blocked* positions.
- $a+1$ *blocked* positions.
- ...
- K *blocked* positions.

Arrangements beans in a circle (not exactly the necklace problem in combinatorics) + under various counting constraints.

Multi-hop blocking: an example



An example with $K = 12$, $b = 8$, $a = 4$ and $z = 2$.

- Total number of possible combinations is huge: $C_T = \binom{K}{b}^H$.
- How many combinations, C_{blk}, make blocking? (i.e., no *schedule* can be found)?
- Blocked TFs ('1_b'): those are available but blocked.

Exact solution: feasible for small parameters

The exact solution requires:

Step 1 Knowledge of all possible combination patterns;

Step 2 the hop-based transition probability matrix.

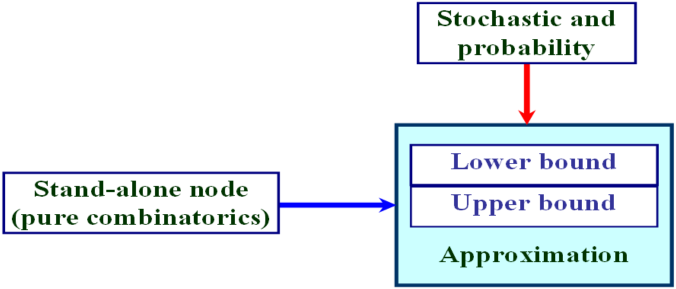
Number of combinations must be examined in **Step 1** is extremely huge.

For example, at the point of 50% load of a time-cycle $K = 128$:

- requires a knowledge of $1,741,360$ patterns
- Transition probabilities between all pair of patterns: a matrix of $1,741,360^2$ cells.

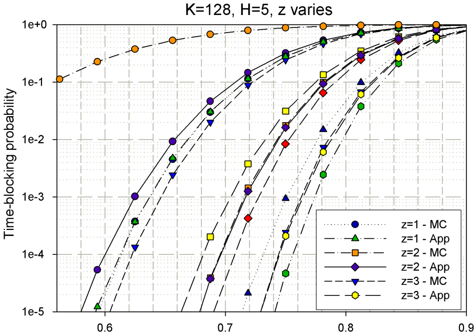
UNIVERSITY OF TRENTO - Italy

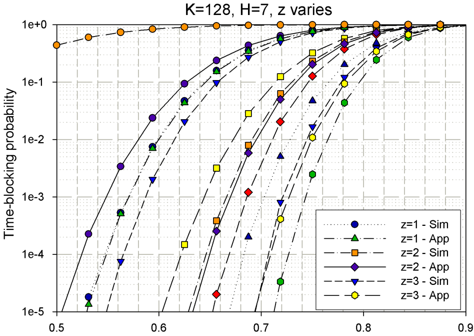
Upper and lower bounds



- Stand-alone node analysis helps obtain fundamental results that are later used to analyze multi-hop cases.
- Use stochastic and probability to obtain bounds.

Sample results





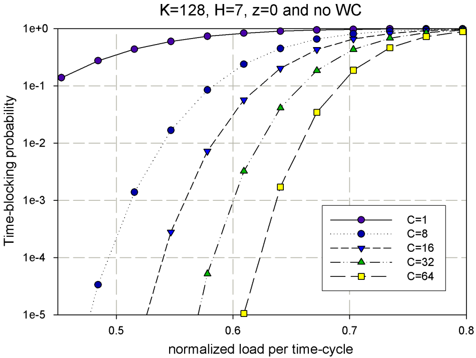
- Very close approximation:

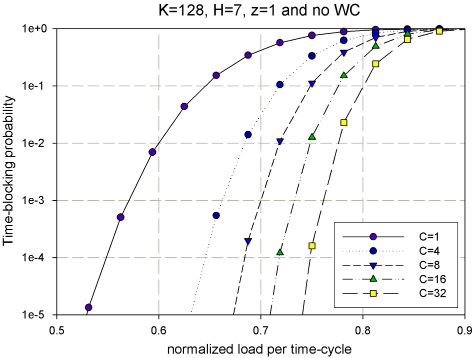
$$p_{blk} = \sqrt{p_{up} p_{lo}}$$

- As z is increases, errors slightly increase.
- Great gains of blocking probability: $z = 1$ vs. $z = 0$.
- Errors of bounds increase as a longer route is considered, but the approx. is still accurate.

UNIVERSITY OF TRENTO - Italy

Multi-channel: zero vs. nonzero scheduling delay



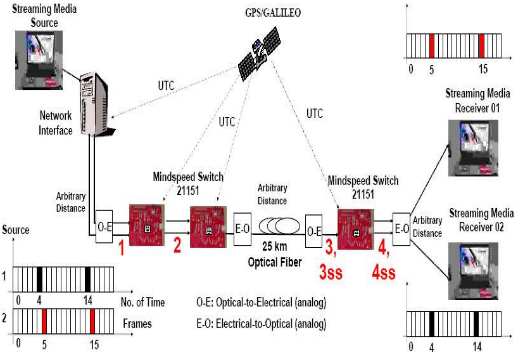


- Multi-channel + nonzero scheduling delay = great improvement of blocking performances.
- If there is WL conversion, performances must be (much) better.

UNIVERSITY OF TRENTO - Italy

Contents

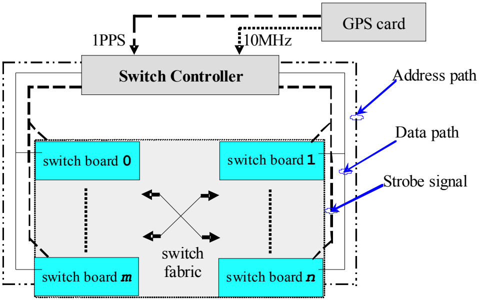
Prototype overview



- Electronics switching: very cost effective compared to the all-optical approach.
- Main components of the prototype:
  1. TDP router: the interface between conventional IP networks and TDS networks;
  2. FPGA-based controller: the "brain" of the switch;
  3. Streaming media and connections.

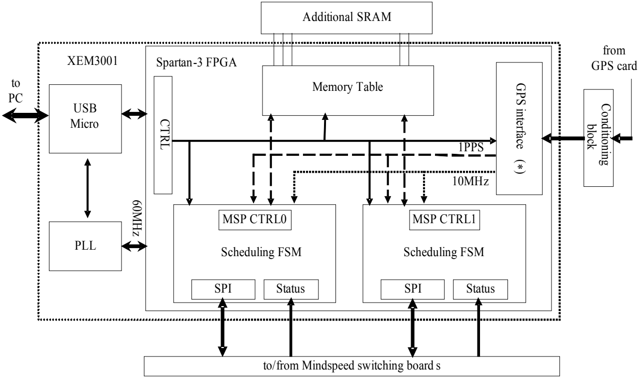
UNIVERSITY OF TRENTO - Italy

TDS node prototype: block diagram



- GPS receiver provides UTC time-of-day;
- An FPGA-based controller;
- Switching fabric: connected Mindspeed chips.

FPGA-based controller: functional diagram



(∗) Can be configured for simulating GPS clocks (1PPS and 10MHz) derived from 60MHz

- Communication with PC via USB interface;
- Communication with Mindspeed switching board via SPIs.
- Can be extended using additional SRAM.

UNIVERSITY OF TRENTO - Italy

Experimental results

- IP-FLOW project demonstrations in CER'06, Infocom'06, IST'07.
- Extra experience: emulation of MAN network (looping video streams through 75 km of fibers connecting 5 switches.)
- Online demo (upon request): `http://dit.unitn.it/ip-flow/`

Contents

UNIVERSITY OF TRENTO - Italy

Conclusions

The research concentrates on two theoretical aspects of FλS:

1. Node designs based on the use of tunable lasers:
   - FC-FλS: low cost design, the fabricless;
   - BS-FλS: strictly non-space blocking design;
   - With a specific condition, BS-FλS has a nice property: the complexity is equivalent to Clos interconnection network.
2. Time-blocking analysis:
   - Exact solution is possible for small parameters.
   - Complete bounds are derived and proved (hence a very closed estimation).
   - Very good blocking performances if there is scheduling delay and/or multi-channel per hop.

The prototype proves the practical aspect of TDS:

- Low cost from off-the-self components.
- Scalability to multi $Tbit/s$ switching.
- Show the advance of electronics switching, compared to the optical one.

... and future works

Time-blocking performance has been intensively studied, however it has been pointed out in the thesis that:

- Closer bounds are possible with more complex analysis.
- For multi-channel cases: wavelength conversion is not considered, and there is room for the extension of the analysis.
- The complete blocking analysis includes also space-blocking switches (i.g., the Banyan fabric): the twist of space-blocking and time-blocking, very challenging issue.

UNIVERSITY OF TRENTO - Italy

Thank you!

Working papers:

1. On editing: "Multi-hop Blocking Analysis for Time Driven Switching," with M. Telek, R. Lo Cigno, and Y. Ofek., plan to submit to ACM/IEEE Trans. on Net.
2. V.T. Nguyen, R. Lo Cigno, and Y. Ofek, "Blocking Analysis of Time-Driven Switching," submitted to the IEEE HPSR 2007.

Refereed articles:

1. V.T. Nguyen, R. Lo Cigno, and Y. Ofek, "Design and Analysis of Tunable Laser based Fractional Lambda Switches," submitted to IEEE Trans. of Comm. (revised).
2. IP-FLOW project paper, "Scalable Switching Testbed not Stopping the Serial Bit Stream", accepted to appear in Proc. of IEEE ICC, 2007.
3. D. Agrawal, M. Corra, V.T Nguyen and Y. Ofek, " UTC based Controller for Scalable Time Driven Switching," in Proc. of IEEE GLOBECOM, 2006.
4. V.T. Nguyen, R. Lo Cigno and Y. Ofek, "Design and Analysis of Tunable Laser based Fractional Lambda Switching (FLS)," in Proc. of IEEE INFOCOM, 2006.
5. IP-FLOW project paper, "Ultra Scalable UTC-based Pipeline Forwarding Switch for Streaming IP Traffic," in Proc. of IEEE INFOCOM (Poster/Demo), 2006.
6. V.T. Nguyen, R. Lo Cigno, M. Baldi, and Y. Ofek, " Wavelength Swapping using Tunable Lasers for Fractional Lambda Switching," Proc. of IEEE LANMAN, 2005.

UNIVERSITY OF TRENTO - Italy