Delay-line storage in optical communications switching

David K. Hunter, University of Essex, UK
Ivan Andonovic, University of Strathclyde, UK
Mike C. Parker, Fujitsu Labs, UK
Stuart Walker, University of Essex, UK

A sampler of optical delay-line technologies

- Fibre or integrated optic waveguide delay-lines
  - 5 ns per metre = 50 bits per metre at 10 Gb/s
  - Fibre can get very bulky, and requires temperature stabilisation
- Ring resonator delays (filters) using conventional waveguides
  - Constructed from ring waveguide offset from main waveguide
- Photonic crystal delays (filters)
  - These are basically periodic dielectric structures
  - Promise ultra-compact photonic integrated circuits
    - Sharp waveguide bends possible
  - Loss presently a major problem
- Photonic crystal and ring resonator filters proposed for delays
  - Made from resonators in series
    - Resonators in parallel pose an addressing problem
  - We assume that up to 10 devices may be cascaded at present
    - Maximum of 1 bit per device implies up to 10 bits can be stored
- 3D stacked spiral waveguides in photonic crystals
  - Make use of tight waveguide bends
  - 1 mm² structure = 50 m waveguide = 2500 bits at 10 Gb/s
  - See Chutinan et al, University of Toronto
Basis for discussion

- What systems can be constructed, assuming the following devices will be viable in the near future:
  - 10-bit fixed delay?
    - 10 photonic crystal resonant filters in series
    - Due to present loss of say $\approx 20$ dB for 10 devices
  - 10-bit variable delay?
    - 10 tuned ring resonators in series
  - 2500-bit fixed delay?
    - 3D stacked spiral waveguide structure
- Further research and development is ongoing in all the above technologies
- The above figures only serve as examples
  - Longer delays will become feasible as each technology matures

---

Incoherent OCDMA encoder/decoders

- Fixed delays up to $O(100)$ bits useful for encoding and decoding chip sequences
- Exceeds our lower maximum of 10
  - Trade-off between pulse width, filter bandwidth and delay
  - Femtosecond pulses imply wide bandwidth
  - Better suited to waveguide delays
- Variable delay useful for placing a chip within a specific timeslot
- Chip sequences are normally sparse
  - Contain few chips (pulses) and many empty slots (no pulses)
  - 10-20 chips in over 100 slots
  - Code tuning speed requirement relatively slow (~msec)
Coherent OCDMA encoders/decoders

- Each chip is encoded with phase information
  - Chips uniformly spaced at output
- Fixed delays can be implemented as before but incremented serially
- Phase shift can be implemented by
  - Phase modulation
  - Variable delay e.g. ring resonator
- Improved correlation properties

Parallel structure TSI

- Idea of parallel structure optical TSI proposed in the 1980s
  - $n$ is number of timeslots per frame
  - For example, Thompson et al JLT Jan 1987
- Demultiplexer made from optical switches
  - Operate in fixed and repetitious fashion
  - Easier to implement control because of low signal bandwidth
  - Assume bit multiplexing – interchanging single bits
- After first (fixed) stage, pulses are aligned in time
- Electronically adjusted variable delays re-schedule pulses for output
  - Assume they can be reconfigured in nanoseconds
Dimensioning of TSI

- Fixed delays can be made using e.g. silica delay-lines
- Variable delays can be made using e.g. electro-optically controlled ring resonators
  - Each can delay by up to one bit period
  - Assume we can presently only cascade up to (say) 10 of these devices
- Thus $n - 1 \leq 10$ so $n \leq 11$
  - Implies a total of 110 such devices
- Suppose both stages are made from ring resonator devices
  - Only 6 timeslots per frame are allowed
    - $n - 1 + n - 1 = 2n - 2 \leq 10$
  - Total of 45 devices
- Suppose 2500-bit fixed delay without excessive loss or dispersion were possible
  - Larger and more elaborate systems could be constructed
  - Could have many bits per timeslot – advantageous for switching

Optical packet switching

- Each packet contains information in its header defining its destination
- Above is a very simple packet switch
  - 2 inputs and outputs
  - Buffer depth of 7 packets
  - Delay lines must buffer units of packets
    - Not units of single bits as before
  - Fixed delay-lines are sufficient
    - But we need some way of implementing them
- Suppose each packet is 424 bits long
  - As in an ATM cell
- Then $424 \times 7 = 2968$ bits
  - Much greater than our first limit of 10
  - Comparable to the second limit of 2500
  - However, this is a very simple example!
  - Careful networking and architectural design required
- This structure can also be used for CDMA code generation/detection
  - Chips switched at high speed instead of packets
It is generally agreed that trying to build a replacement for a present-day Internet router with entirely optical buffers is not reasonable.

However, there has been some new research at Stanford on dimensioning router buffers:

- Conventional wisdom states that the buffer size in bits on each router linecard should be $\text{RTT} \times \text{data rate}$.
  - Current router linecards can contain as much as 10 Gbits of buffer memory.
- Appenzeller et al. (SIGCOMM 2004) showed that this can be divided by $\sqrt{n}$.
  - $n$ is the number of TCP flows through the linecard.
- Enachescu et al. (ACM SIGCOMM CCR July 2005) showed that this can be further reduced to $O(\log W)$.
  - $W$ is the window size of each TCP flow.
  - Minor modifications to TCP are necessary.
- Assumes that all traffic is TCP.
  - Analysis extends to UDP very easily.

Optical router re-visited:

- Enachescu’s results suggest that an IP router linecard could be built with a buffer depth of only 20 datagrams.
- Assume datagram length of 1500 bytes.
  - Then maximum delay is $20 \times 1500 \times 8 = 240,000$ bits.
  - If byte-parallel transfer is used, this reduces to 30,000 bits.
  - Still well in excess of both our maxima!
    - 10 bits and 2500 bits.
- Optical fibre stores 50 bits per metre at 10 Gb/s.
  - 240,000 bits = 4.8 km.
  - Temperature stabilisation required.
Key features of SOAPS

- Reduced core optical buffering, due to edge processing
  - Smoothing and “slot shifting” at the network edge
  - Each individual stream consists of regularly spaced slots
    - Reduces core contention

- Fixed-length “slot” structure
  - Encoded using Coarse WDM (CWDM)

- Electronics for:
  - Edge switching and buffering
  - Control, packet scheduling and header processing throughout the network

- Optics for core switching and minimal buffering
  - Simulations have demonstrated successful operation with 10 slots’ delay per core switch port
  - Slots could be as large as 2500/10 = 250 bits with 3D structure
Adapted re-negotiated service

- Re-negotiated service shapes traffic flowing into the optical network into a relatively regular pattern
  - Regularly spaced slots, like ATD
  - After each “interval”, user re-negotiates bandwidth with network

- Problems for conventional RCBR:
  - If the stream requires new bandwidth, which is not available, re-negotiation fails
  - Once the network is saturated, the original user-driven scheme will make the network even more congested
    - Because it tries to satisfy source bandwidth requests triggered by TCP

- Our adapted re-negotiated service relieves these problems
  - Considers both user requirements and network conditions
  - The network conditions considered are much more detailed than TCP
    - Obtained via flooding of core node status information
    - TCP considers only packet loss as a sign of congestion

Re-negotiation failure rate

- Norm – re-negotiated service, only considering user requirements
- Nkt-level – re-negotiated service, considering both user requirements and network conditions
When there is no contention, the shaped CBR streams passing through each core node do not coincide.

When contention occurs (see above), it is necessary to shift the timing of one or more CBR sequences.

Two places where shifting can take place:
- At edge switch – electronic buffer is available
- At core switch – no electronic buffer is available

Our slot shifting scheme predicts possible contended slots and shifts them at edge switch:
- During current interval $T$, calculate possible contention during next interval $T$
- Prediction is not perfect so some core contention resolution still necessary

Performance of CASS:
- Non-sft – scheme without contention-aware slot shifting implemented
- Sft-aggr – scheme implementing contention-aware slot shifting with edge connection aggregation
- Sft-non-aggr – scheme implementing contention-aware slot shifting without edge connection aggregation
- Buffer storage for 10 slots provided for each core node port
### Congestion control

- Congestion control in core must react quickly
  - Large amount of data might be lost before TCP can react, due to high optical speeds
- Congestion control should be TCP aware
- Must be based on optical-level time slot
  - Because this is used for transport in the core
  - Must take design of core network into account
- Control function is implemented at edge
  - Simplifies implementation of core switches

### Congestion control algorithm

1. Loss reported to edge
   - Reduce rate, depending upon number of losses
   - new rate = max(negotiated rate – loss rate, reserved rate)
   - Continues until interval with no loss; go to step 2
2. Zero loss reported for first time during last interval
   - Want to test to see if any further network capacity is available
   - new rate = previous sending rate + half of all previous accumulated losses from step 1
   - If the next slot also has zero loss, go to step 3
3. Positive network situation
   - An aggressive policy is pursued
   - new rate = min(previous slot sending rate + exponential increase component, maximum possible rate)
   - Algorithm only adjusts transmission rate
   - Re-transmissions are performed by TCP itself
Congestion control performance

- Loss plot and throughput plot demonstrate the effectiveness of this technique

Conclusions

- Optical fibre is a simple but problematic technology for optical delays
  - Becomes bulky for long lengths
  - Requires temperature stabilisation

- Optical resonant filter technologies presently have loss problems
  - Presently cannot cascade them beyond about 10 devices
  - Bandwidth x delay < 1 for each element
  - To avoid addressing problems, must be organised into delay-lines

- 3D stacked spiral waveguide structures
  - Feasible systems may be possible, with careful networking and architectural design
  - Thousands of bits’ delay (or more) may be possible

- Planar (2D) waveguide technologies, especially silica-on-silicon
  - Typically up to 10 bits delay at 10 Gb/s

- Techniques to reduce length of delay-lines
  - Spectrally efficient coding (DPSK)
    - 2 or more bits/s per Hz
    - Does not alter main conclusions
  - Parallel delays
    - Require more optical switches