Dynamic Scheduling of TWIN: Time-domain Wavelength Interleaved Networks

Kevin Ross
Stanford University

Hot Interconnects Conference
August 21, 2003

Co-authors: N. Bambos (Stanford),
K. Kumaran, I.Saniee, I. Widjaja (Lucent)

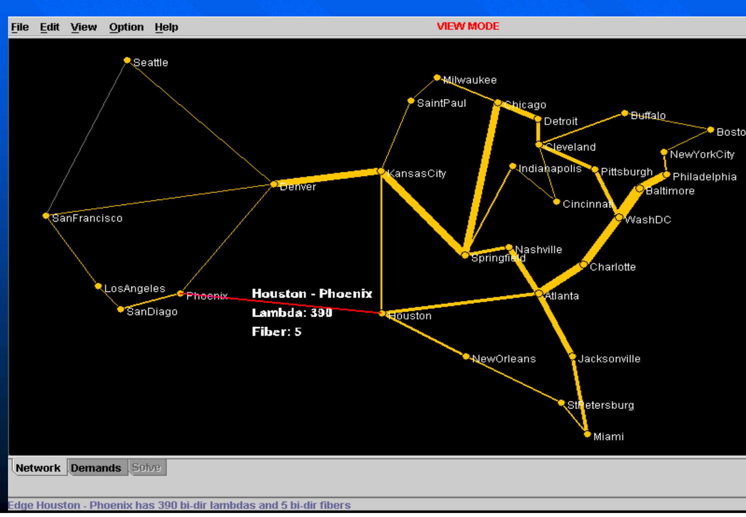
Talk Outline

- TWIN features
- Scheduling problem
- Batch schedules
- 100% throughput guarantee
- Simulation examples
- Conclusions and future research

Key Points

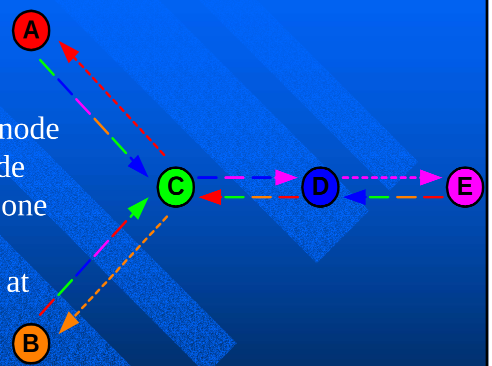
- Scheduling in TWIN is
  - Necessary
  - Interesting
- Adaptive batch schedules provide superior performance
- 100% throughput can be guaranteed in TWIN

Time-domain Wavelength Interleaved Networks (TWIN)



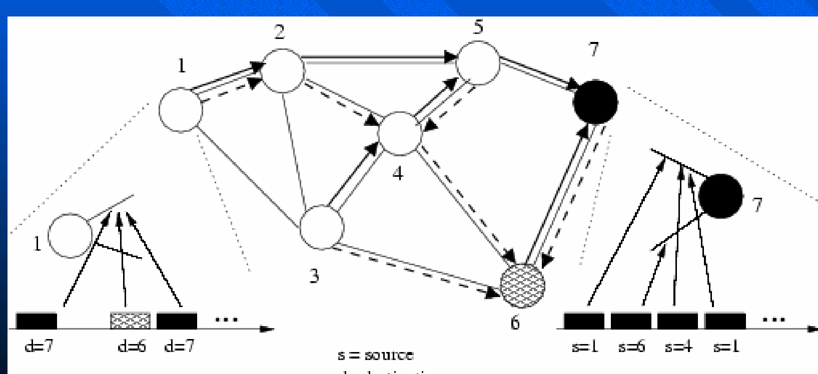
TWIN Features

1. End-to-end demands are only a fraction of wavelength capacity
2. Fixed paths between node pairs (path design)
3. Single fast tunable laser at each node
4. Single mode receiver at each node
   Each destination is assigned one (a set of) wavelengths
5. Each link can transmit one burst at each frequency per timeslot
6. Dynamic demand between node pairs
7. Global scheduler

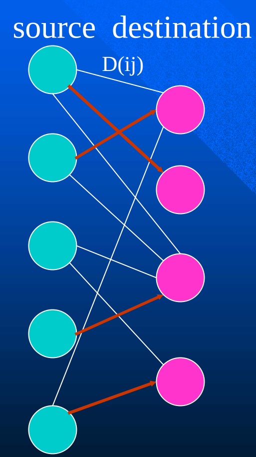


Scheduling Problem

- Source nodes send one burst per timeslot
- Destination nodes receive one burst per timeslot
- Internal nodes transmit one burst per frequency per timeslot

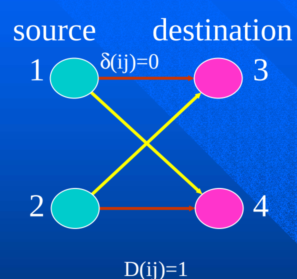


With Zero Propagation Delays, Problem Is Matching



- Each edge is assigned demand D(ij)
- Edges with common source or destination conflict
- A matching is a set of edges with no source or destination conflict

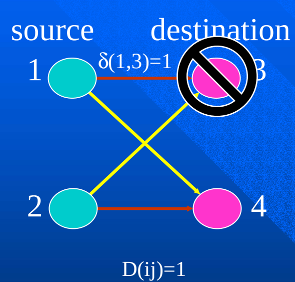
Scheduling Without Delay: Two Timeslots



No delay

With zero propagation delay, demand can be met in two timeslots

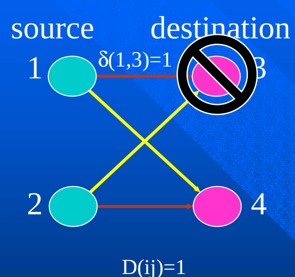
Scheduling With Delay: Two Timeslots



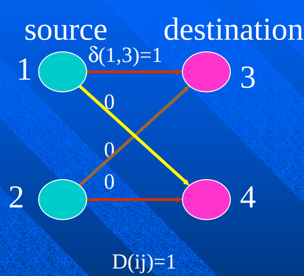
With delay

With nonzero propagation delay, demand cannot be met in two timeslots

Scheduling With Delay: Three Timeslots



No delay



With delay

With nonzero propagation delay, demand requires an extra timeslot

Scheduling With Propagation Delay

- No longer a matching on bipartite graph
- Schedule must take into account past and future to avoid conflicts at destination

Problem Notation

$A_{ij}(n) =$ Bursts arriving to source i destined for destination j in timeslot n

$D_{ij}(n) =$ Bursts waiting to be sent from source i destined for destination j in timeslot n

$S_{ij}(n) =$ Bursts sent from source i to destination j in timeslot n

$\delta_{ij} =$ delay sending source i to destination j

$\rho_{ij} =$ Long term arrival rate for source-destination pair (i,j)

$$D_{ij}(n+1) = D_{ij}(n) + A_{ij}(n+1) - S_{ij}(n+1)1_{\{D_{ij}(n)>0\}}$$

$$\rho_{ij} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} A_{ij}(k)}{n}$$

$$\sum_i \rho_{ij} \le 1\ \forall j, \sum_j \rho_{ij} \le 1\ \forall i$$

$$\rho_{ij}^{out} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} S_{ij}(k)1_{\{D_{ij}(n-1)>0\}}}{n}$$

Problem Formulation

Given $A_{ij}(n)$ and $D_{ij}(n-1)$, select $S_{ij}(n)$ to satisfy:

$$\sum_j S_{ij}(n) \le 1 \quad \forall i$$

$$\sum_i S_{ij}(n-\delta_{ij}) \le 1 \quad \forall j$$

$$\rho_{ij}^{out} = \rho_{ij}$$

•Arrival Traffic:

Only assumption: long term arrival rate

•Throughput:

arrival rate = departure rate

for any allowable traffic load

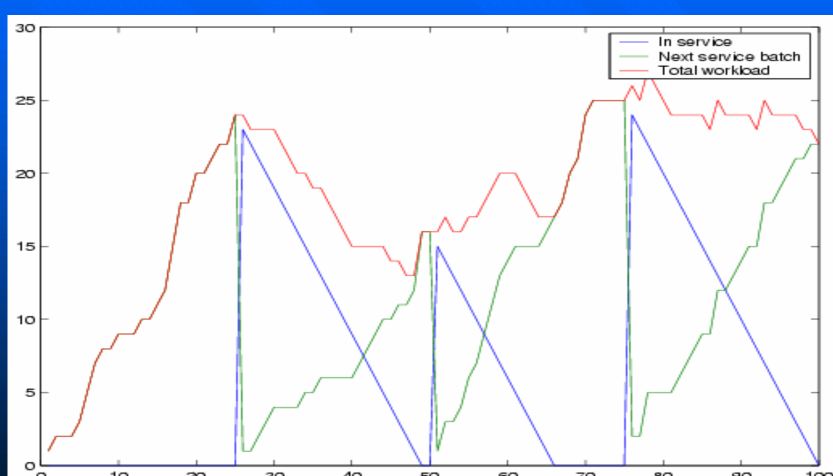
Scheduling Options

- Narrow-view scheduling
  - Consider current demand and conflicts only
  - Intuitive and simple
  - Unstable and unfair!
- Batch scheduling
  - Set schedule for batch of timeslots at discrete timeslots

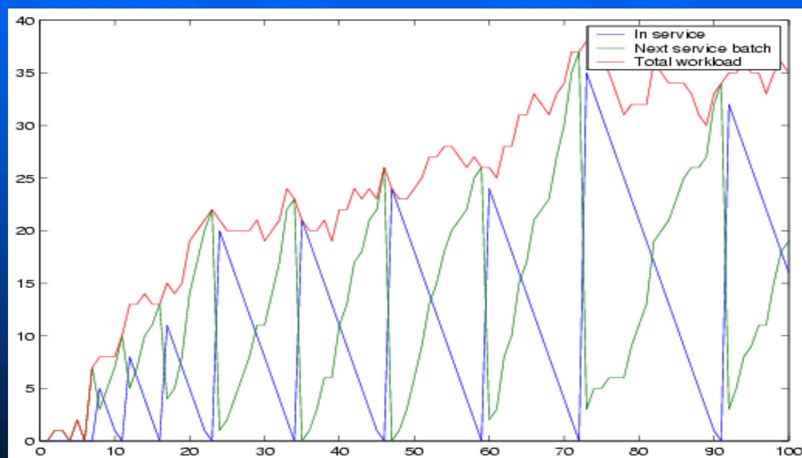
Batch Schedules

- At beginning of new batch T_N schedule waiting bursts to be served in that batch
- During batch service, arriving bursts are buffered for following batch
- Batches can be:
  - Fixed length
  - Adaptive length (triggered by demand)

Fixed Batch Lengths



Adaptive Batch Lenths



Analytical Results

- Fixed batch lengths:
  - Given a fixed batch size, one can construct an allowable arrival trace for which the batch schedule is unstable.
- Adaptive batch lengths:
  - Given some conditions on the scheduling within a batch, adaptive batch schedules can guarantee 100% throughput.

Proof Outline

$$\rho = \rho^{out}$$

$$\Leftarrow \lim_{n\to\infty} \frac{\sum_{k=1}^{n} A(k)}{n} = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} S(k)}{n}$$

$$\Leftarrow \lim_{n\to\infty} \frac{D(n)}{n} = 0$$

$$\Leftarrow \lim_{N\to\infty} \frac{T_N - T_{N-1}}{T_N} = 0$$

$$\Leftarrow \lim_{N\to\infty} \frac{B(D(T_N))}{T_N - T_{N-1}} \le 1$$

$\rho =$ long term arrival rate matrix
$\rho^{out} =$ long term departure rate matrix
$A =$ arrival matrix
$S =$ service matrix
$D =$ demand matrix
$T_N =$ start of Nth batch
$B(D) =$ batch time to serve demand D

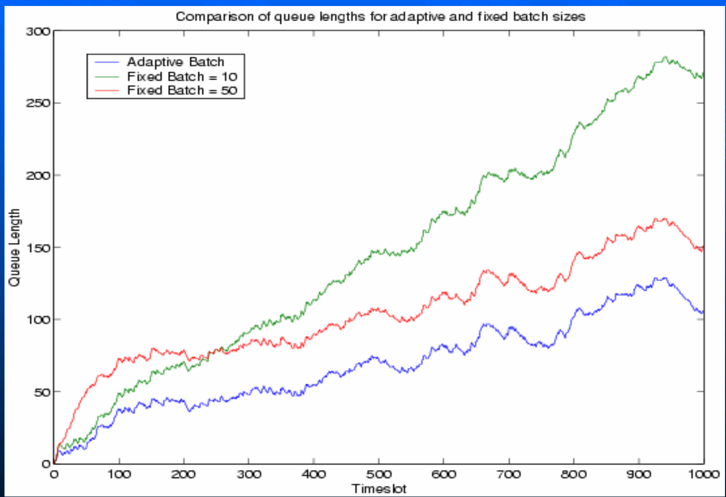
Simple Batch Schedule

- At $T_{N,}$ the start of each batch:
  - Use minimal schedule assuming zero propagation delay
  - Wait sufficient time between service to avoid arrival conflicts
- Effect of delay on schedule goes to zero as demand increases

Simulation Example

- Extremely simple network
  - Two source nodes
  - One destination node
- Compare fixed batch (long and short length) with adaptive batch policies

Simulation Example
Single Trace



Conclusions

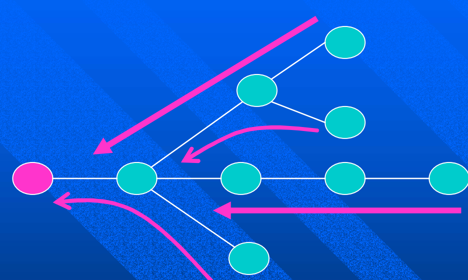
- Propagation delays lead to a new scheduling problem
- Batch scheduling is a good approach in TWIN
- 100% throughput can be guaranteed using adaptive batch lengths but not using fixed batch lengths
- Adaptive batches also lead to better quality of service in simulated results

Future Work

- Network extensions
  - Tunable receivers, capacity variation
- Quantifying QoS
  - Buffer size, delay
- Extending the class of throughput-maximizing algorithms
- Practical implementation issues
  - Centralized or decentralized control
  - Distributed versions of algorithm

Questions…

Receiver Tree



Bursts which don’t conflict at arrival
don’t conflict on internal links

Tree Path Assumption Satisfied by

- Any spanning tree
- Set of shortest path trees
- Ring