Motivation

- Commercial workloads favor SMPs with snooping
- Directory-based protocols don’t dominate
- However, SMPs have problems
  - Networks that perform ordered broadcasts are expensive and unable to take advantage of the increasing level of integration
  - Synchronous broadcast is a poor match for emerging interconnection options
- Key observation: order only needs to be observed by transaction processing, not network delivery.
- Implementing snooping on switched network.

Timestamp Snooping Networks

- Everything happens in logical time. Only relative order matters.
- Ordering time (OT): The logical time when this transaction should be processed.
- Guarantee time (GT): The local current time at a node in the logical time space. Doesn't have to be the same at all nodes.
  - At GT, no transaction whose OT ≤ GT will arrive later

Transaction Processing

- Two virtual networks: broadcast address network and data network
- Address network operations
  - Assign OT: Source stamps transaction OT = Source GT + estimated time before being processed at destination
  - Broadcast: Network broadcasts the transaction
  - Compute GTs: Switches advance their GTs
  - Destination Operation: Destinations Process the transactions in OT order – “Bus” arbitration done distributively by destination nodes
A Special Case

- Each node has access to a global clock
- GT = clock time
- OT = Conservative estimate of when the transaction can be serviced
  - Travel time + queuing time at destination
- All processors will process the same transaction at the same time
- Same as synchronous broadcast on a bus!

One Network Implementation

- Maintain OTs and GTs implicitly
- Source assigns slack – time to wait at the destination before it is processed
- Token: clock ticks in logical time
- Switches exchange tokens and update slack
- Initially each input port of a switch has 1 or more tokens

Token Passing Example

Snooping Protocol

- Support any subset of MOESI states
  - Hard to implement owned and shared signals
  - Eliminate them and add states to memory
- Optimizations enabled:
  - Peeking at buffered transactions and prefetching
  - Processing other processors’ early transactions to blocks currently in stable states S, I, or not present.
Evaluation

- Target System
  - 16-node SPARC
  - Solaris 7
  - Simulated by extended Simics (with memory hierarchy)
  - Commercial workload

- MSI Protocols
  - TS-Snoop
  - DirClassic
  - DirOpt

- Network
  - four indirect radix-4 butterflies
  - direct 4x4 2D torus

Performance – Runtime

Normalized Runtime with Butterfly (left) and Torus (right)

Performance – Link Traffic

Normalized Link Traffic with Butterfly (left) and Torus (right)

Summary and Discussion

- SMPs with snooping protocols suitable for commercial workload but bus a bottleneck
- Implementing MOESI snooping protocol on switched network
  - Timestamp each transaction
  - Allow out of order delivery and arbitrary broadcast
  - Destinations enforce total order by processing transactions in timestamp order
  - Trade bandwidth for lower latency
- But why compare against directory protocols?