ATOMIC-2

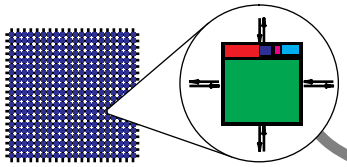
Production Use of a Gigabit LAN

Joe Touch, Hong Xu, Ted Faber, Annette DeSchon, Avneesh Sachdev
USC/ISI HPCC Division
http://www.isi.edu/atomic2

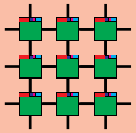
This work is supported by the Advanced Research Projects Agency through Ft. Huachuca contract #DABT63-93-C-0062 entitled "Netstation Architecture and Advanced Atomic Network". The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Department of the Army, the Advanced Research Projects Agency, or the U.S. Government. We receive additional support from Calren's ARC Consortia and GTE's SCAN Project.

ATOMIC History

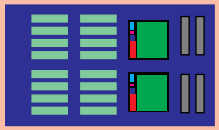
A Mesh Supercomputer becomes an Gigabit LAN

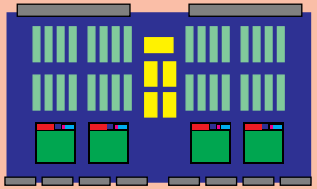


Cal Tech's Mosaic Supercomputer



3x3 Switch



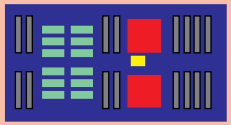


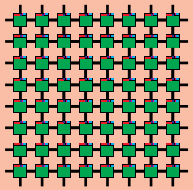
Sun S-bus Intf. *Sun VME Intf.*

Prototype ATOMIC LAN

Memoryless Mosaic processor
Dual-processor host interfaces
Lab-sized proof-of-concept

1992





Cable Driver *8x8 Switch*

ATOMIC LAN

"Address Consultant" Source-routing
Windowed differential cable drivers
LAN-sized proof-of-concept

1993

Netstation

ATOMIC LAN as a host backplane

ATOMIC-2

Production Use of the ATOMIC LAN

ATOMIC Components

Myricom-built hardware

Sun SPARC SBus Host Interface

8-port switch 10” x 10” x 1”

- *Uses 36-conductor twisted-pair cable (0.5-1.0” dia.)*
- *Byte-wide transmission (80 Mbps x 8 = 640 Mbps)*

Outline

Protocols

File Server

Gateway

Security and Authentication

ISI has the largest production stand-alone ATOMIC LAN

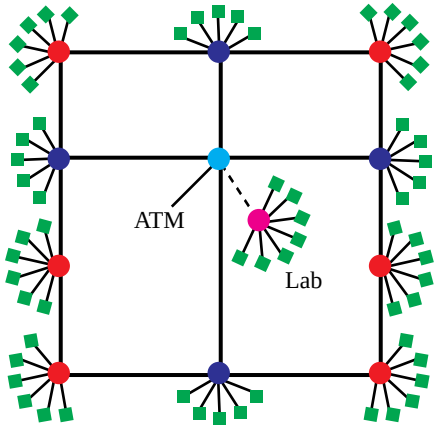
ISI's ATOMIC LAN

Total of 65 hosts

- *All HPCC Div. Sun 10s & 20s (entire floor)*

Two level topology

- *Allows intermediate switch failure*
- *Relies on dynamic source route updates*



LAN Topology - 64 hosts

Installation progress:

- *Approx. 20 workstations in daily use*
- *Preparing to deploy ATM/ATOMIC gateway*
- *Investigating Fast Ethernet/ATOMIC gateway*

Managing ATOMIC

New Hardware

- *What metrics to watch?*
- *What timescale is important?*
- *What are the acceptable values of those metrics?*

Tools

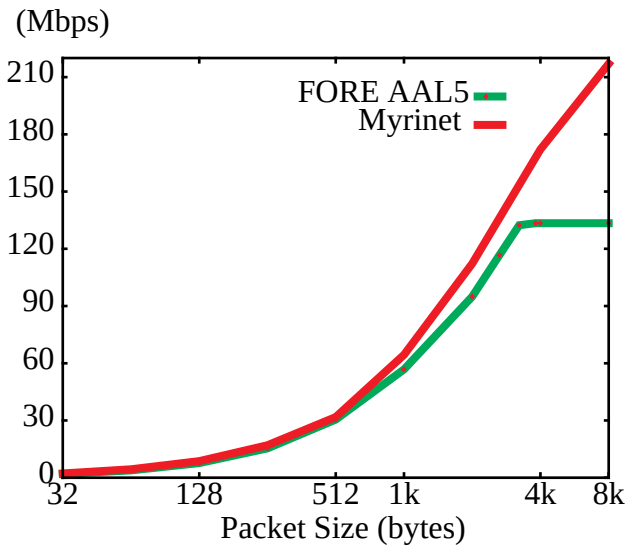
- *Short term interactive network monitor*
- *Long term network monitor and summary tool*

Bandwidth to the User

Protocols

Bandwidth Measurements

- *Hardware:*
  - SPARC 20/71
  - Fore SBA-200
  - Myricom LANai 2.3
- *Native transport*
  - 210 Mbps over ATOMIC
  - 136 Mbps over ATM
- *Kernel protocols (UDP)*
  - 150 Mbps over ATOMIC
  - 133 Mbps over ATM

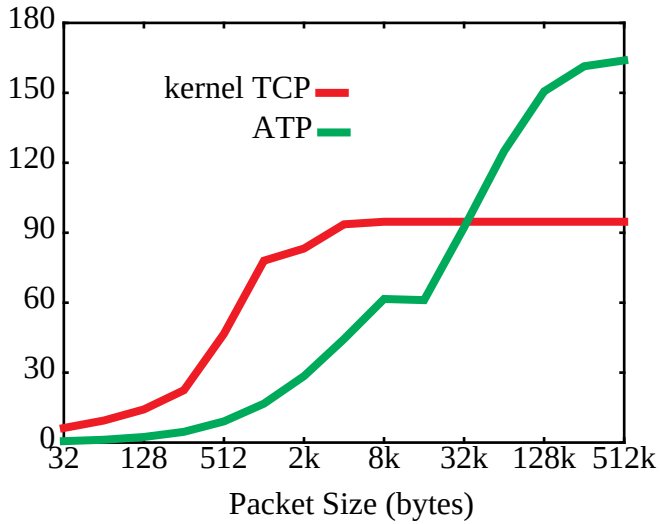


ATOMIC Protocols

Protocols

Dual-stack Protocols

- *Atomic Transport Protocol (ATP)*
  - RPC send/receive style reliable protocol
  - Designed to facilitate PVM over ATOMIC
  - Tuned down to device DMA
  - Implemented in user space
- *PVM decomposition*
  - ATP protocol for data
  - Kernel TCP/IP for control
  - Dual-stack driver
- *PVM speedup*
  - 65 Mbps using system protocols
  - 164 Mbps using ATP



ATPng

Protocols

A LAN transport protocol

- *Full duplex, reliable transport within ATOMIC*
- *Leverage DMA techniques from ATP*

Compatibility Library for TCP applications

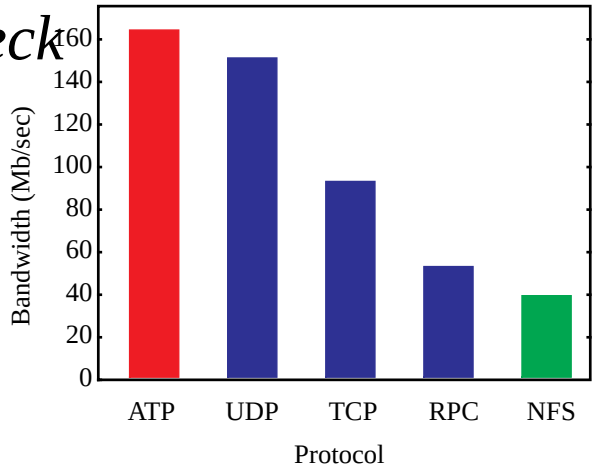
- *Relinking most applications will provide TCP compatibility*
- *Compatibility at the socket level*
- *Library will use new transport or fall back to TCP*

Network File Systems

File Server

Goals

- *Remove network/protocol bottleneck*
- *Aggregate user requests fairly*



Studies

- *AFS and NFS studied*
- *Protocol limits throughput*

Hardware

- *Texas Memory Systems SAM-200 RAM disk*
  - 140 Mbps, 4KB blocks, SPARC 20/71
- *Classic disk/network bottlenecks are removed*

File System Protocol Issues

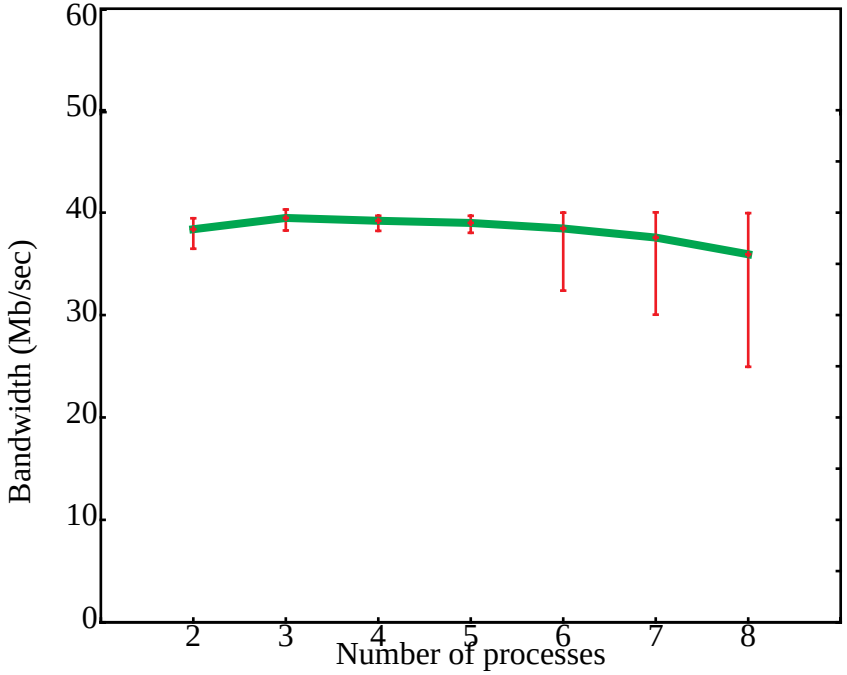
File Server

Protocol/System bottlenecks

- *Serial RPC*
- *Parallelism from multiple processes*
- *xdr adds a copy*

Potential Solutions

- *Pipeline RPC in 1 process*
- *Simplify data translation*



Tuning the File System

File Server

Replace existing file data transfer mechanism

- *Pipelined RPC*
- *In-kernel protocols*
- *Lightweight data encoding*

Port server code directly to disk interface

- *Remove Sbus bottleneck on disk data (aggregate)*
- *RPC allows simple server code*

Initial implementation is modified NFS

ATOMIC-ATM Gateway

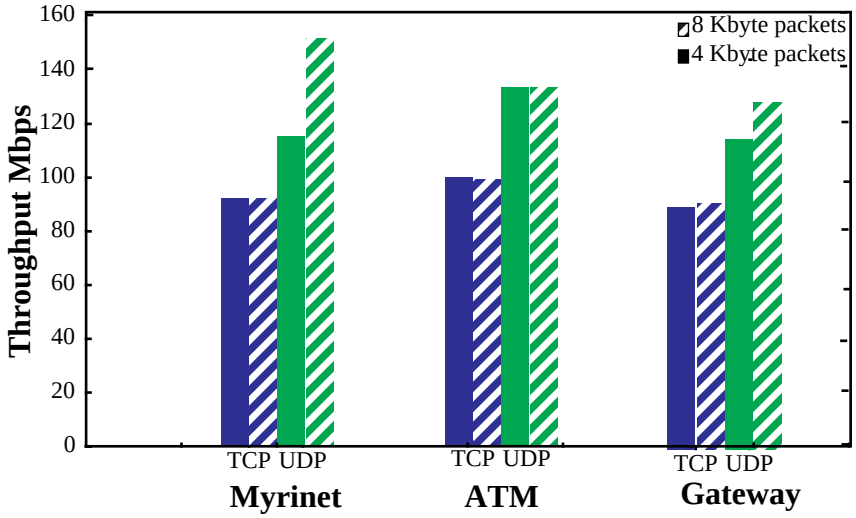
Gateway

Goals

- *Access other media at high speed*
- *Limited hardware*

Host-based solution

- *SPARC 20/71*
- *IP routing*
  - BSD kernel-based
  - Direct inter-interface

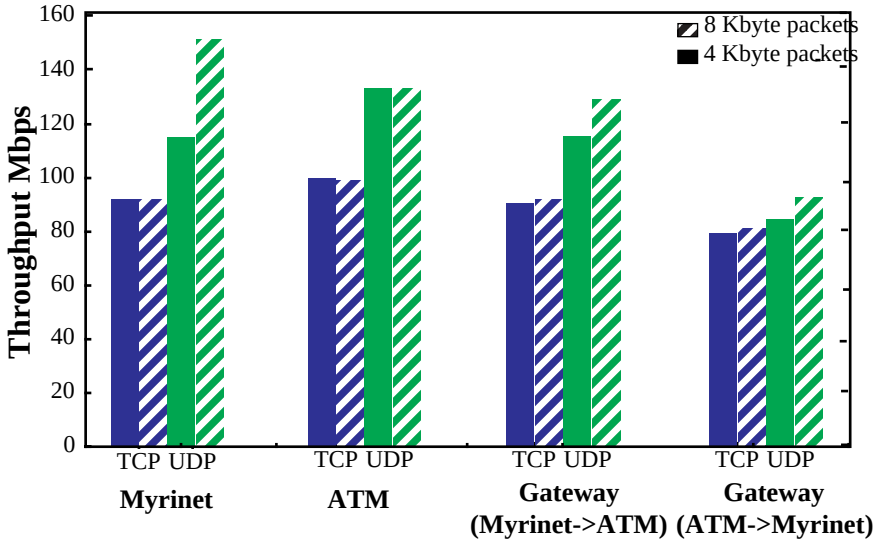


Host-based Gateway Issues

Gateway

Hardware Interactions

- *FORE ATM*
  - fly-by DMA
  - 155 Mbps limit
- *Myricom LANai*
  - High overhead DMA
  - 200 Mbps+ bandwidth



Unusual gateway characteristics

- *Asymmetric performance*
- *May be solved by another bus*

MD5 and High Speed IP

Authentication and Security

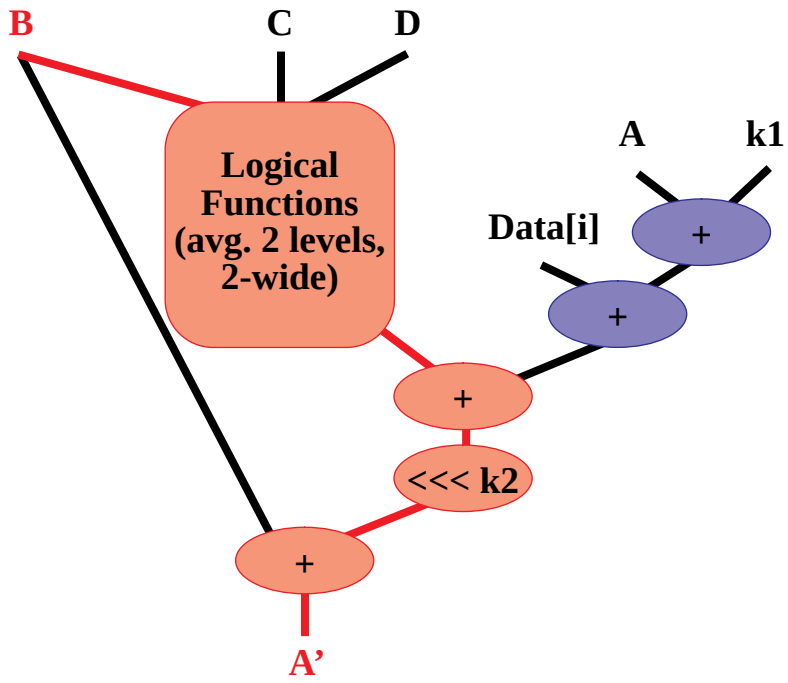
MD5 is the default required option for authentication in IP version 6

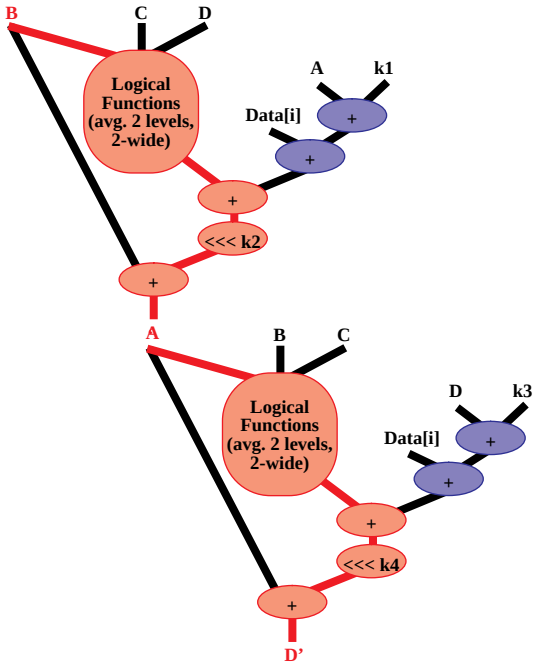
MD5 digest added to authenticated packets

- *Proves that the packet is from another MD5 entity*
- *Provides more than data integrity (not a checksum)*
- *Touches every byte of an authenticated packet*

MD5 Dataflow Analysis

Authentication and Security



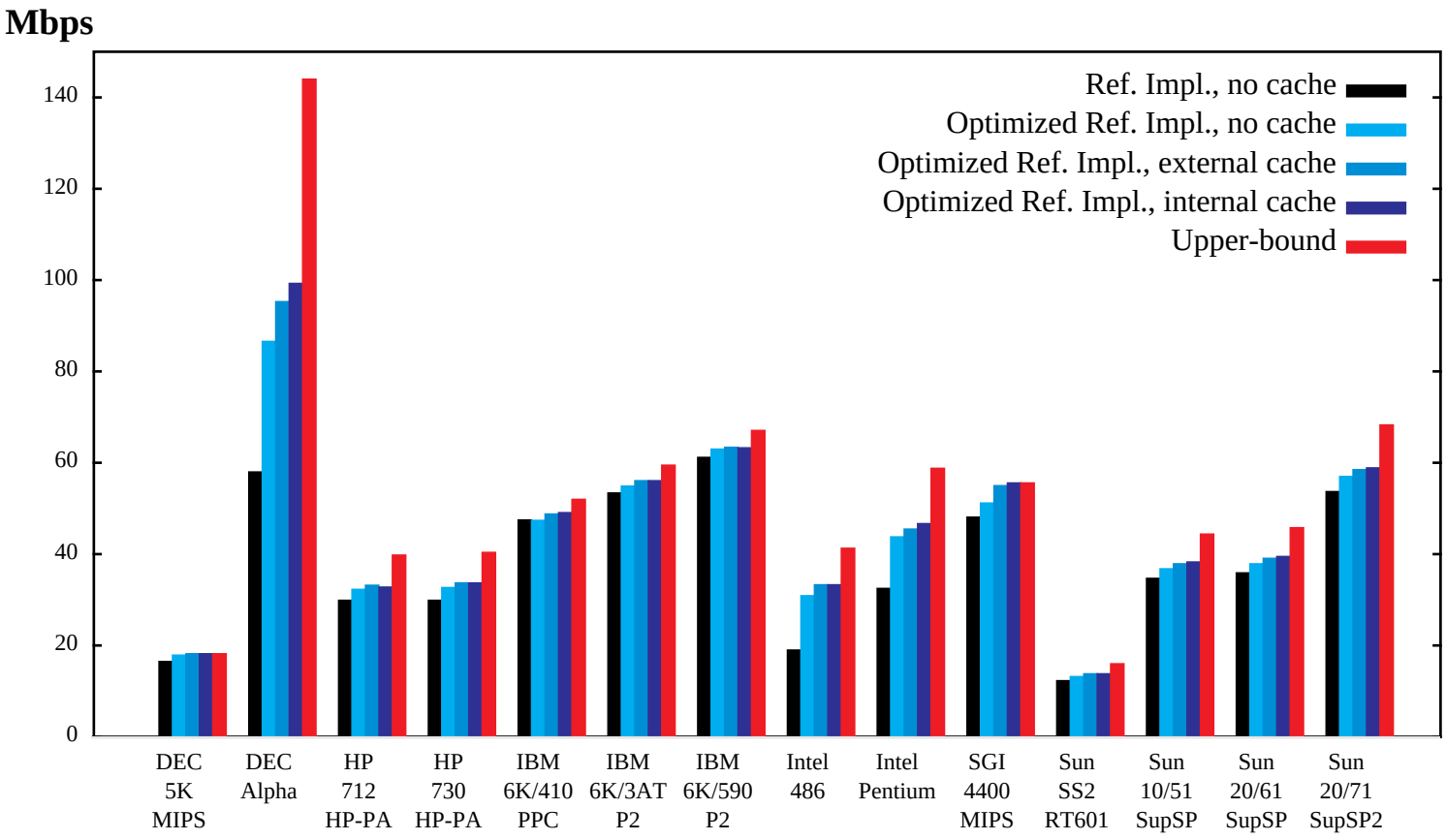


mostly serial
compute-bound
5-level critical path (2 adds, 3logicals)

sequentially dominated

MD5 Performance

Authentication and Security



Other options

Authentication and Security

Seeking fast algorithm

- *MD5 is 45 opcodes per word*
- *15 ops/wd OK / 2-4 for "low cost"*
- *Alt. Hash (AHA) currently at ~200 Mbps*

Conclusions

Current Status

- *File Server*
  - Analysis complete
  - Implementation underway
- *Protocol development*
  - ATP successful in improving PVM performance
  - New protocol in development
- *Gateway*
  - Tuning of current gateway continues
  - Searching for other hardware
- *Security*
  - MD5 analysis complete
  - Investigating MD5 in IP stack

http://www.isi.edu/atomic2