

INFORMATION SCIENCES INSTITUTE

Production Use of a Gigabit LAN

ATOMIC-2



Joe Touch, Ted Faber, Annette DeSchon, Avneesh Sachdev USC/ISI HPCC Division

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.









INFORMATION SCIENCES INSTITUTE

ISI Installation Status



April 24, 1996



Areas of Research

Transport to the application

- Faster stock protocols, dual-stack implementations

File service

- NFS for small files

Gateway via host interfaces

- Router line-cards aren't available

Security protocols - Performance vs. strength trade-offs

Transport Protocols

Atomic Transport Protocol (ATP)

- user-space reliable transfer
- tuned down to device DMA
- used for dual-stack PVM
 - user ATP data, kernel TCP control
 - 65 Mbps w/TCP only, 164 dual-stack
 - useful for 64KB+ data blocks



ATPng

- full duplex, reliable transport inside LAN

- dual-stack implementation
- compatibility library for TCP applications
 - substitute select, read, write, etc. calls via library that calls system version
 - connect-time choice of TCP or ATP for WAN interoperation



Networked File Systems

Measured protocol/system bottlenecks

- synchronous RPC
- multi-process pipelining
- xdr copy

Optimizations

- single-process RPC pipelining
- off-load RPC / NFS into RAM-disk CPU
- omit xdr copy intra-LAN
- aggregate requests and data (like Nagle's alg.)







Gateway Design

Kernel IP limitations

- bus limited to 160 Mbps
 320 Mbps total, 2x crossing
- DMA has high latency
- asymmetric performance

Planned optimizations

- peer DMA / PCI for higher BW
- sub-MTU transfers / pipelining for low latency
- routing coordination of host and peer DMA







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





INFORMATION SCIENCES INSTITUTE





http://www.isi.edu/atomic2/

Recent Results

Installation

- 33 hosts via in-wall wiring; production since 2/95
- Calren ARC / GTE SPAN OC-3c up

Gateway

- production Sun SPARC gateway installed

File Server

- NFS RPC optimization testing - 30% improved raw RPC

Security

- IPv4 AH / MD5 completed