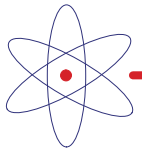
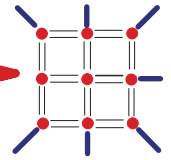
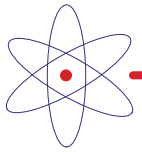
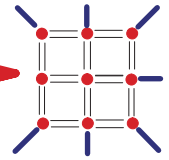




VISA: Netstation's Virtual Internet SCSI Adapter

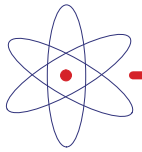
Rod Van Meter
rdv@isi.edu

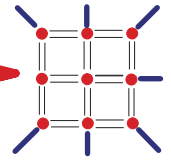
ASPLOS
October 5, 1998

Correction to Table 7

benchmark	15 secs	10%
file system	34	24%
network bcopies	38	26%
other networking	21	14%
gprof	16	11%
misc. kernel	21	14%
total CPU 1GB read	145 secs	

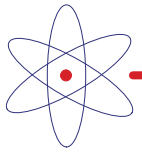


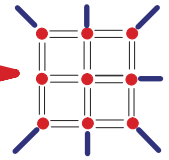


Conventional Wisdom

IP can't be used to communicate with peripherals.

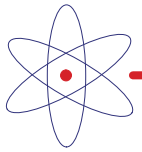
Wrong!

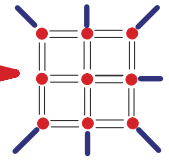




Goals of This Talk

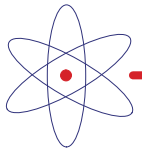
- Establish that IP offers desirable properties for communicating with peripherals.
- Demonstrate that IP can be used effectively *in the host* for communicating with peripherals.

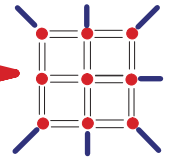




Non-Goals of This Talk

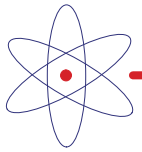
- Security
- IP in the peripherals
- Orthogonal & covered elsewhere
- http://www.isi.edu/netstation/

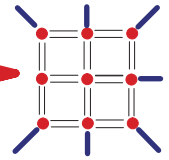




Outline

- Netstation
- Networking for peripherals
- Advantages of IP
- SCSI over IP in the host

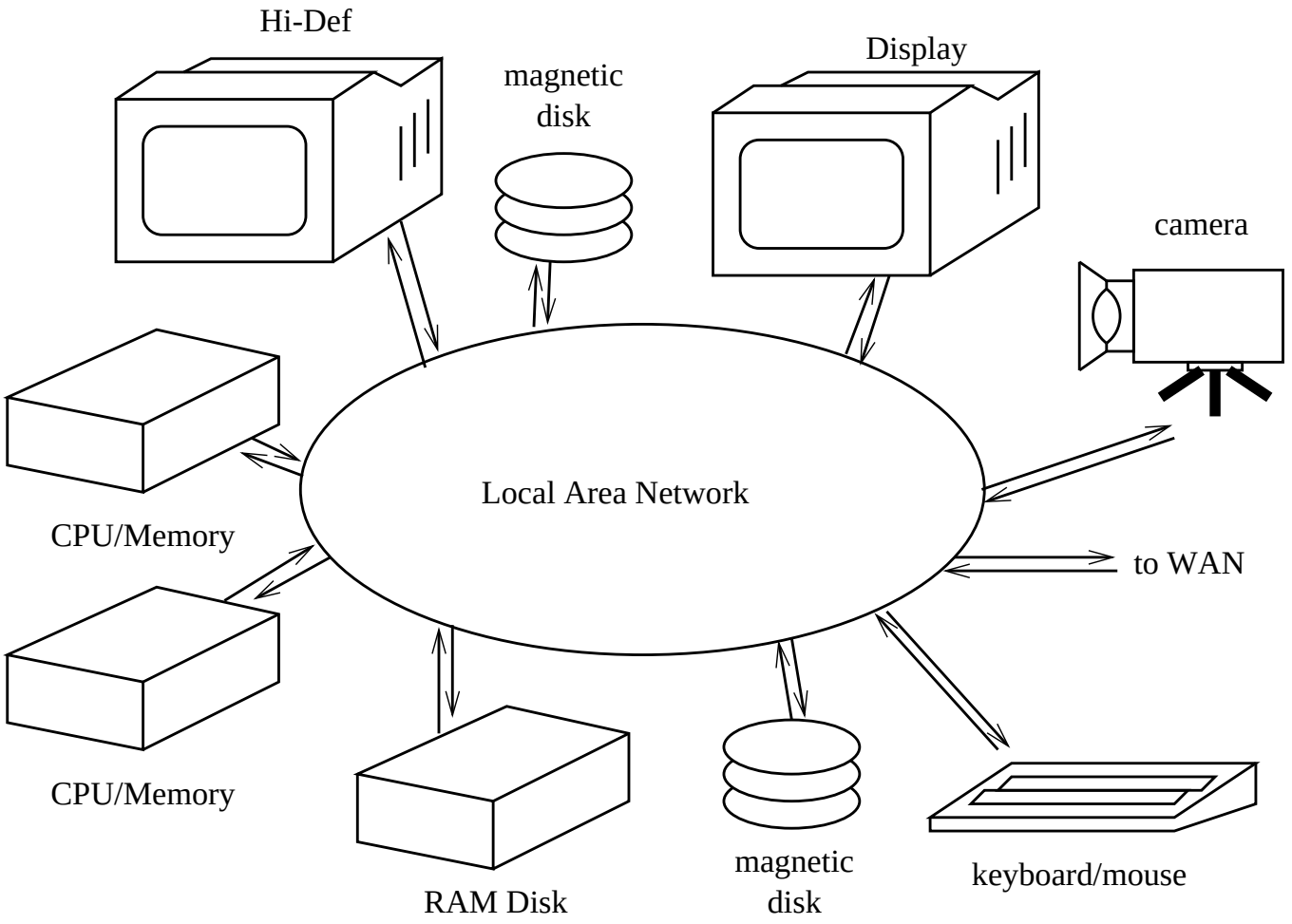


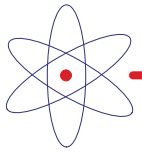


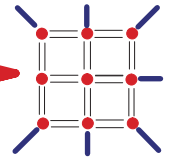
Netstation

- Replace I/O bus with a gigabit network
- Buses not scaling in:
  - # of devices connected
  - aggregate bandwidth
  - distance

A Netstation





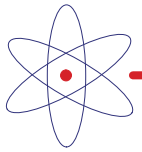


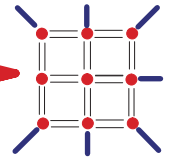
Netstation Research Areas

- Network-attached peripheral (NAP) HW & SW
- NAP-capable OS
- Network protocols
- Security and access models

Networking for Peripherals

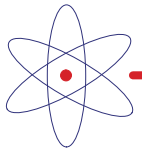
- Lots of choices
- Lots of R&D effort
- Spectrum from bus to network
- Poor interoperability

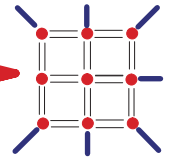




Networking Choices

- HiPPI
  - 800
  - 6400
- SSA
- Fibre Channel
  - fabrics
  - Arbitrated Loop
- ATM
- Firewire (1394)
- Myrinet
- ESCON
- switched SCSI
- gigabit ethernet
- ServerNet
- GigaNet
- WAN channel extenders
- VAXcluster CI



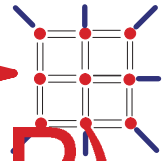


Scaling I/O Interconnects

as I/O nets get larger and more complex...

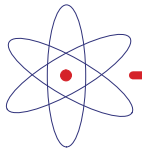
- media bridging
- congestion control
- flow control
- demultiplexing
- legacy systems
- latency
- security
- reliability
- heterogeneity

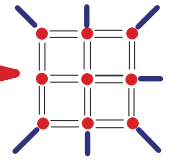
all become bigger problems!



Solution: the Internet Protocol (IP)

- Strengths:
  - heterogeneity
  - scalability
- Weaknesses:
  - complexity
  - performance
- Note: most of those LANs support IP!

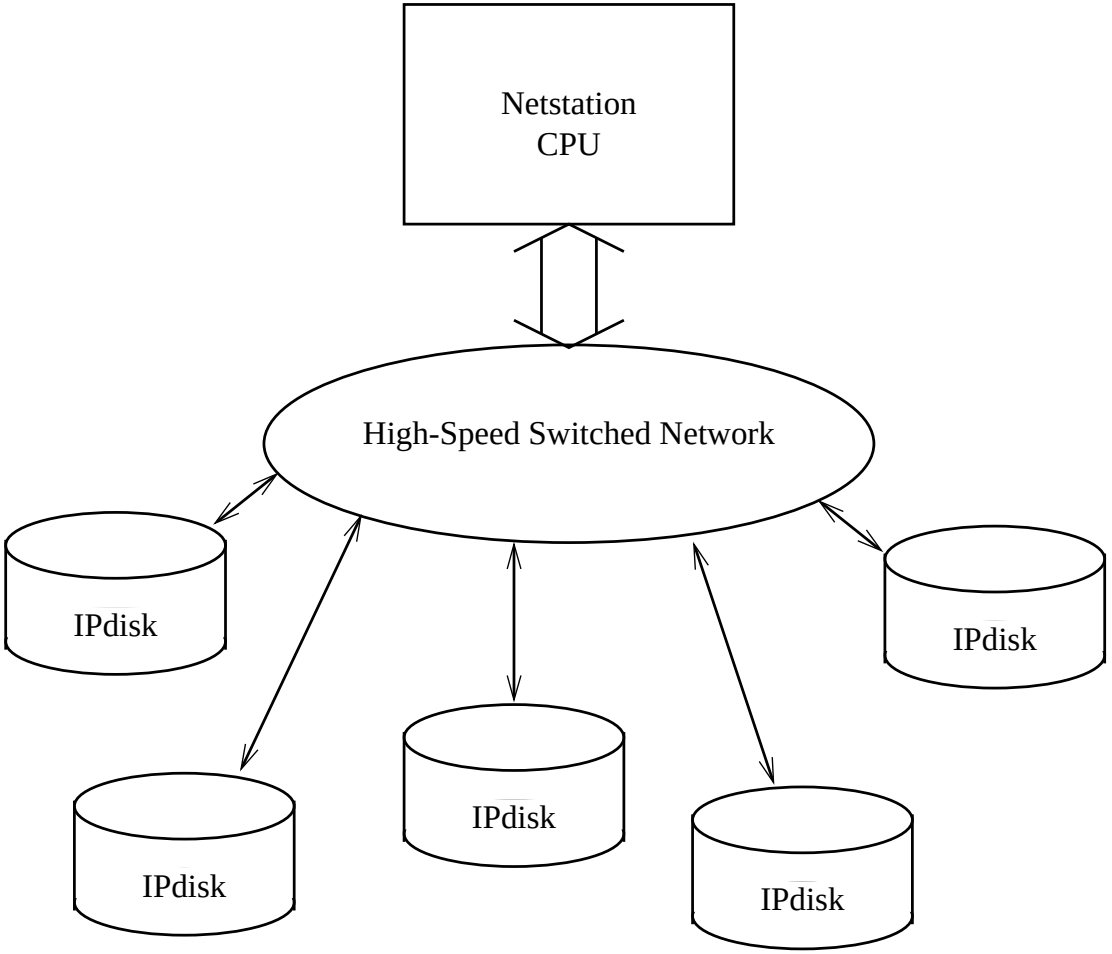


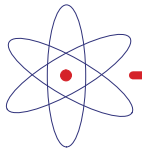


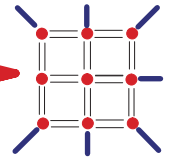
VISA: Virtual Internet SCSI Adapter

- Demonstrate IP acceptable for peripherals
  - device performance
  - CPU load
- Single host to device
- Proof of concept
- Primarily OS experiment
- Convince the storage community

Experimental Configuration



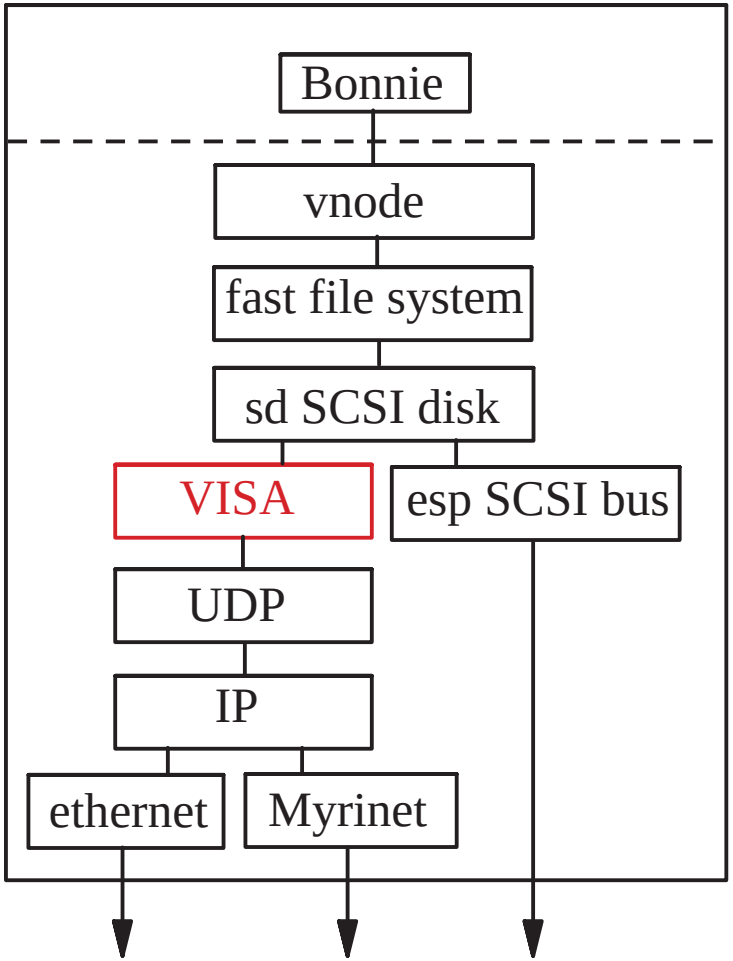


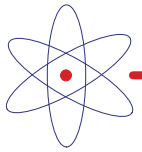


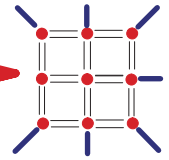
Architecture

- What
  - sends SCSI RPCs, receives data via network
  - accesses IPdisk
  - SunOS 4.1.3 scsi_transport layer
  - meshes with SCSI-3
- How
  - simple reliability over UDP
  - single-threaded pseudo-process
  - prefetch/consolidation in FS code above
  - single command per target -- no queueing

VISA Code Path

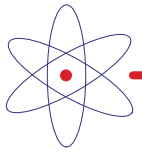


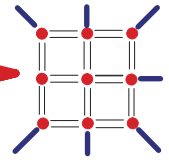




IPdisk

- Emulates disk NAP
- Full SCSI disk command set
- UDP or TCP
- Four types of store:
  - RAM
  - file
  - raw SCSI disk
  - solid state disk

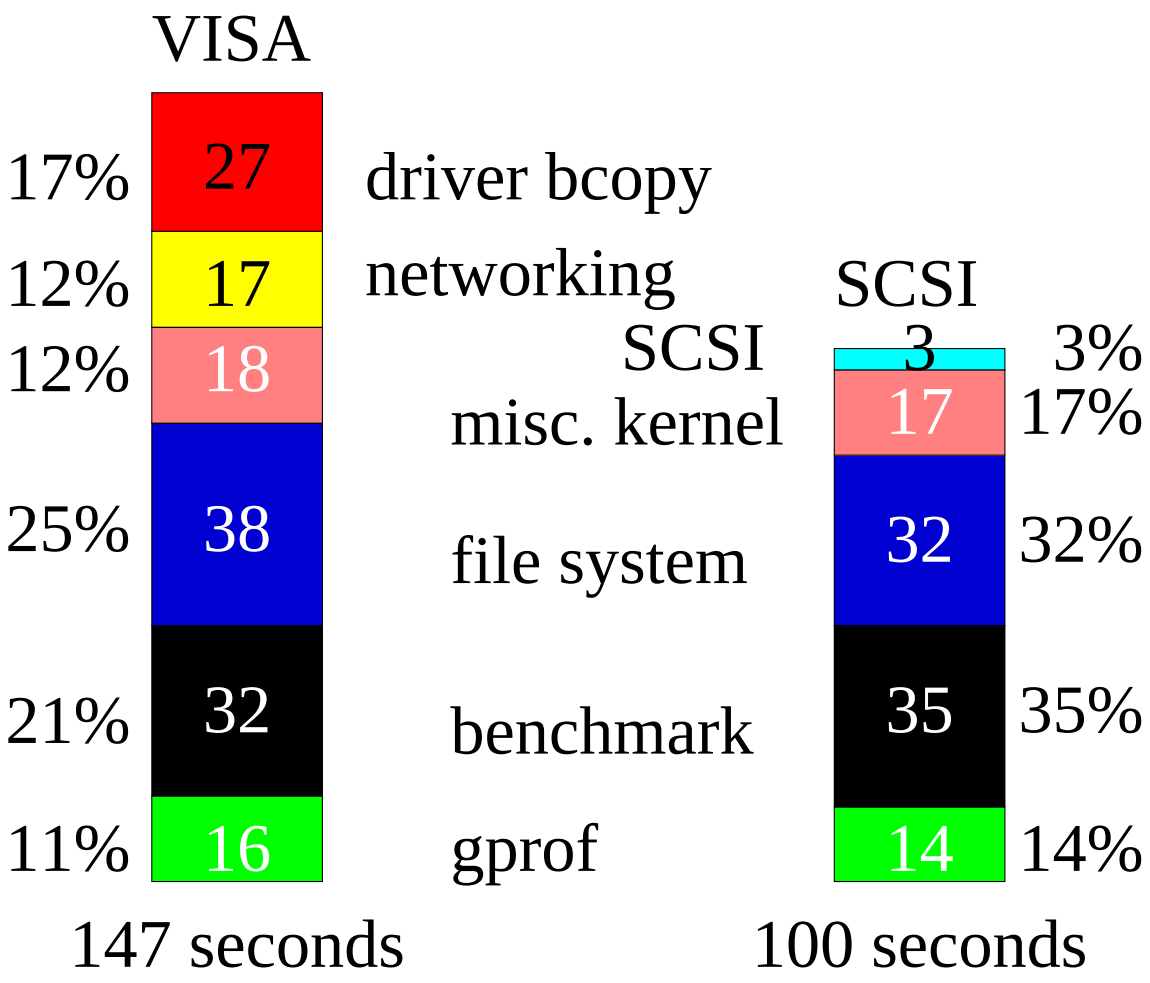




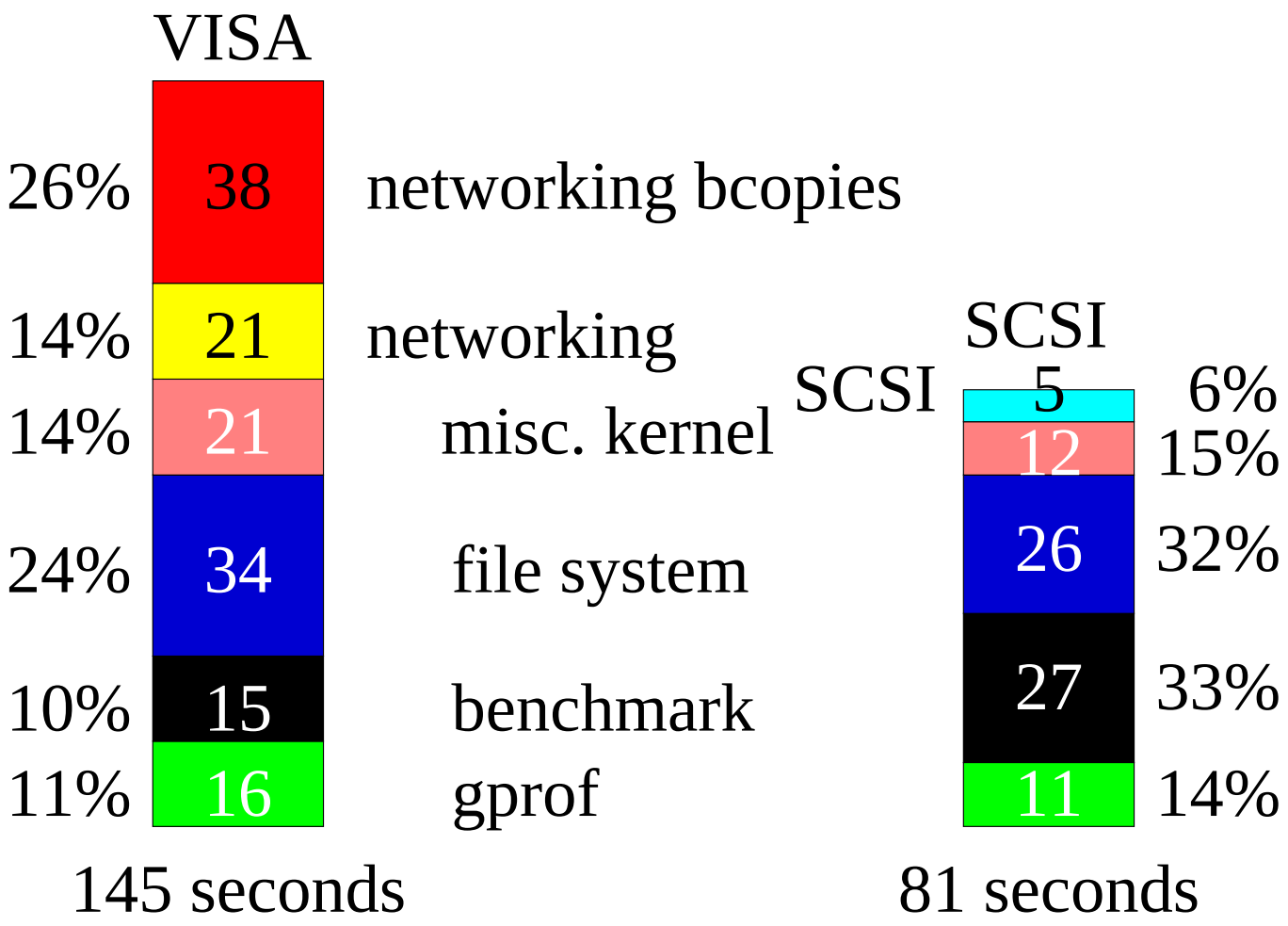
Performance

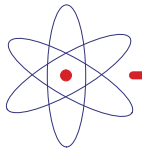
- Up to 72 Mbps write, 58 Mbps read through file system
- CPU limited at client (host) (75 MHz Sparc 20/71)
- Compares to:
  - 60 Mbps NFS
  - 107 Mbps TCP blast
  - 135 Mbps UDP blast
  - 115 Mbps FS write to buffer cache

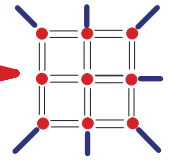
Write CPU Utilization (1GB)



Read CPU Utilization (1GB)

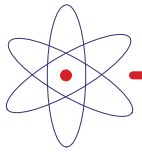


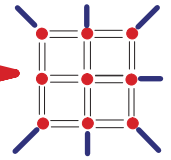




Results

- Demonstrated performance using IP at 2/3 SCSI speed (CPU bound)
- 80% possible w/ known SW optimizations
- Networking overhead comparable to file system
- Hardware support for demux and checksum will cut cost further
- Full TCP/IP in NIC doable, cuts cost to near zero





Conclusions

- Using IP brings big benefits
  - leverage existing R&D and infrastructure
  - heterogeneity
  - scalability
- Implementation cost is reasonable
- http://www.isi.edu/netstation/