A Tunable Communications Library for Data Injection

Craig Ulmer and Sudhakar Yalamanchili

Center for Experimental Research in Computer Systems

Outline

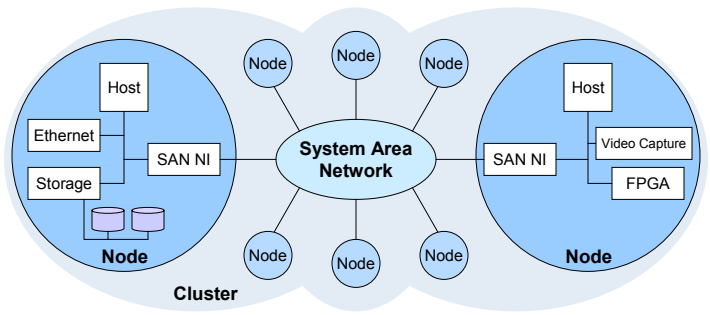
- Background
  - Large variety in cluster computer hardware
- Insulate users from hardware environment
  - Task: Data injection to peripheral devices
- TPIL: Tunable PCI Injection Library
- Performance measurements
- Conclusions

Background

- Cluster computers
  - Affordable parallel processing
  - Multiple clusters at an institution
  - Different hardware in each generation
- Variety of peripheral devices
  - System Area Network (SAN) NICs
  - Storage adaptors
  - Multimedia controllers

Example

- Georgia Tech x86 cluster computers
  - 5 Clusters (PPro to Pentium IV)
- Active System Area Network (ASAN)
  - 6 Types of PCI device

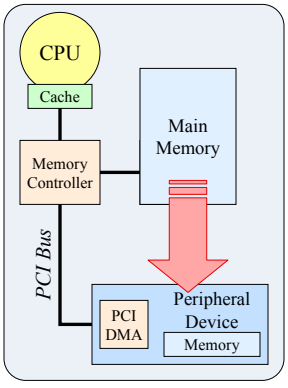


Research Question

- How do we insulate users from hardware?
  - Want programs to perform well everywhere
  - Self tuning software

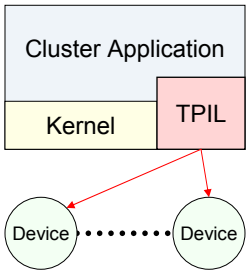
Focus: Data Injection

- Inject data from host to peripheral device
  - Common task
- Challenging
  - No host-level DMA engine
  - Virtual/Physical memory
  - Multiple mechanisms



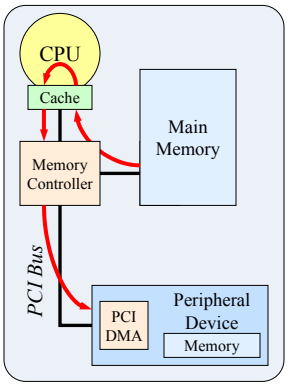
TPIL: Tunable PCI Injection Library

- Portable library for x86
  - Implements multiple injection mechanisms
  - Provides built in benchmarking
  - Selects best transfer method for injection size
- Transfer mechanisms
  - User space PIO transfers
  - Kernel space DMA transfers



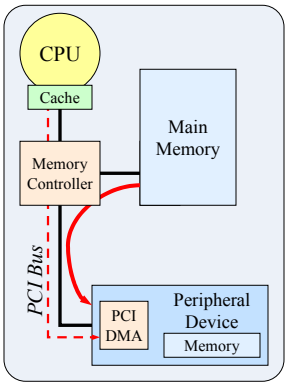
Programmed I/O Transfers

- CPU transfers data
  - Pull into CPU, Push to I/O
  - Good for small bursts
- Methods:
  - Plain memcpy()
  - MMX registers (8x64b)
  - SSE cache directives
    - ** Write-combining*



DMA Transfers

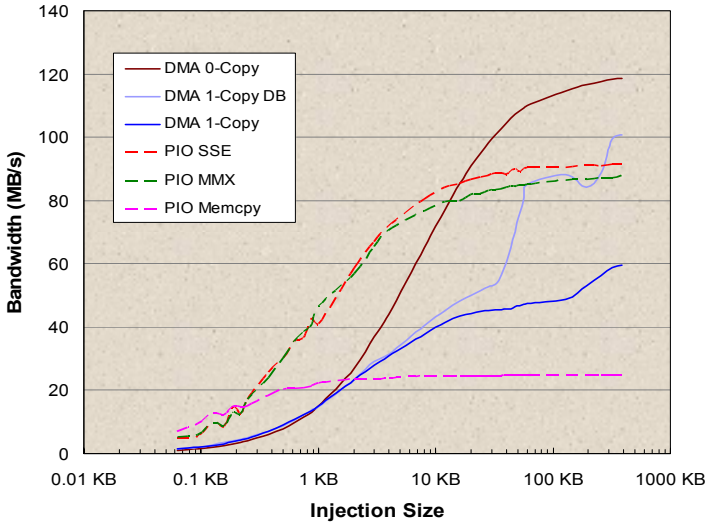
- Card DMA engine transfers data
  - DMA from physical memory
  - Good for long bursts
- Methods:
  - 1-Copy
  - 1-Copy double-buffered
  - 0-Copy



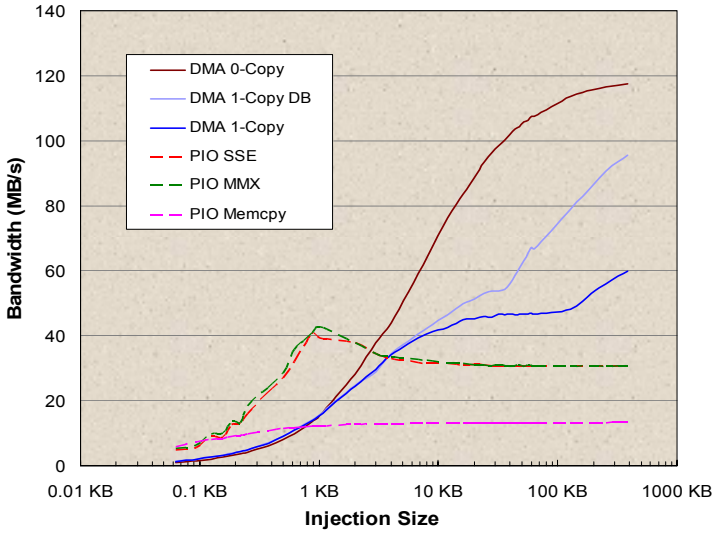
Performance Results

- TPIL driver extensions:
  - Myrinet LANai 4 NI
  - Myrinet LANai 9 NI (64b/66MHz PCI)
  - Celoxica RC-1000 FPGA
- Three clusters
  - Pentium Pro-200MHz
  - Pentium III-550MHz
  - Pentium IV-1.7GHz
- Linux 2.4 Kernel

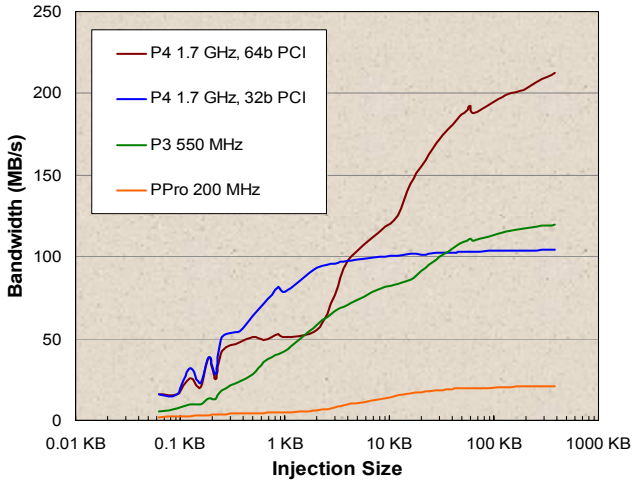
Test 1:Myrinet LANai 4, PIII-550



Test 2: Celoxica RC-1000, PIII-550



Test 3: Myrinet LANai 9, All Hosts



Conclusions

- Large variety in hardware performance
  - Different cutoffs for each mechanism
- TPIL a library for data injections
  - Selects best transfer method based on size
  - Eases task of moving between clusters
- Future work
  - Add mechanisms, support other cards
  - Allow customized interactions
  - Perform similar work for other tasks