Datacenter Fabric Workshop
Windows IB

Windows ULPs
(IPoIB, SRP, WSD)

Fab Tillier
SilverStorm Technologies
ftillier@silverstorm.com

August 22, 2005

Agenda

- **IPoIB Miniport**
- SRP Miniport
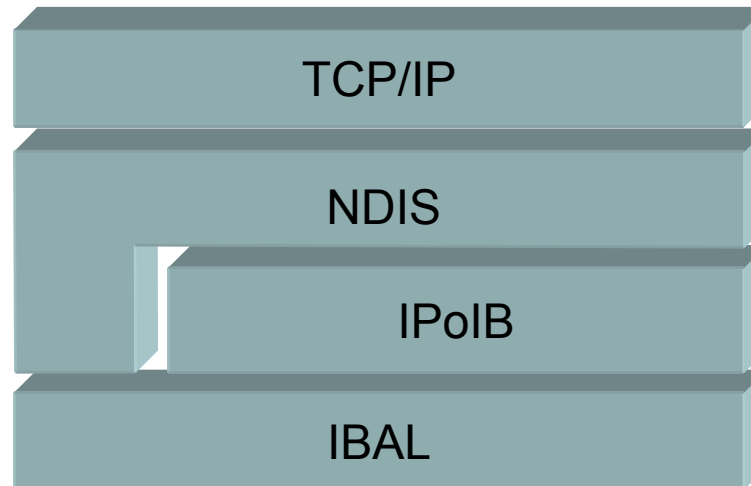- Winsock Direct Provider

IPoIB

- NDIS-WDM Miniport

- Appears as 802.3 device to the OS
  - Requires lookup from Ethernet to IPoIB MAC on every send and receive

- Performs DAPL ATS service registration for assigned IP addresses

- Single MR of all of physical memory
  - No registration in speed path

NDIS-WDM Driver Model

- ## NDIS Miniport Upper Edge
  - Looks like a standard NDIS Miniport driver
- ## WDM Lower Edge
  - Work outside the NDIS wrapper for device access

IPoIB To Do

- On-demand send-only multicast join
- ATS IPv6 support
- Common buffer for receive buffers
  - Proper DMA mappings

IPoIB Enhancements

- ## LBFO
  - Load balance between multiple instances
- ## IP to IB Path resolution
  - Access ARP cache
  - Eliminates need to perform SA queries for ATS records
- ## IPoIB RC
  - Report single MTU (~64K), perform IP fragmenting for UD MTU

Agenda

- IPoIB Miniport
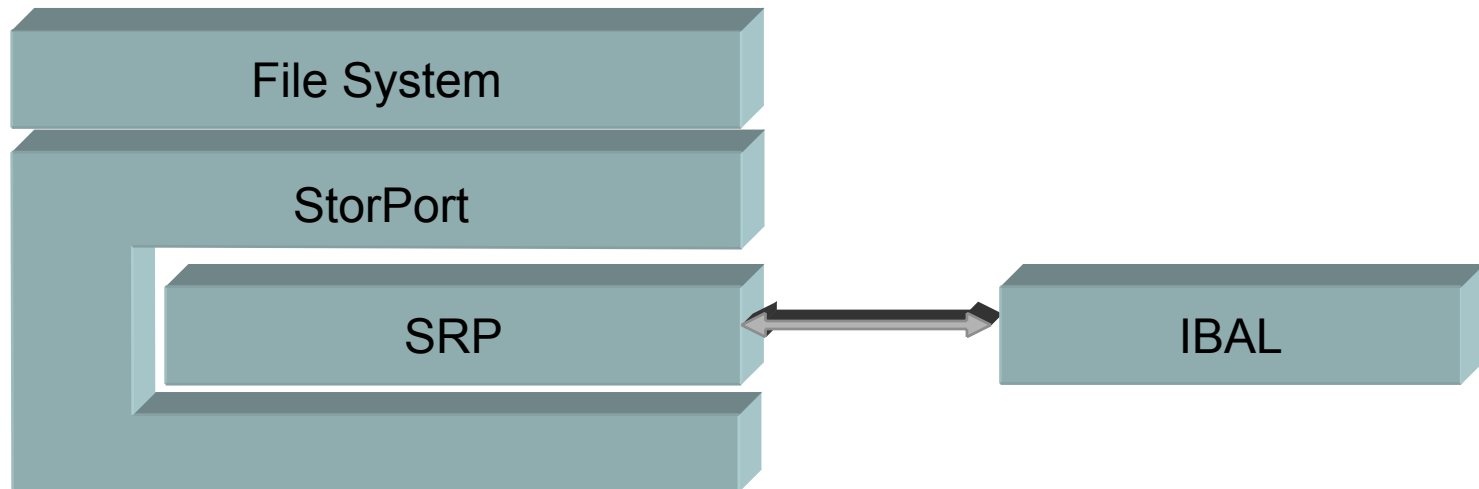- SRP Miniport
- Winsock Direct Provider

SRP

- Storport Miniport
  - Available only on Windows 2003
- Uses single MR of all of physical memory
  - Must trust SRP target
  - No registration in speed path
  - Requirement due to I/O at DISPATCH
- Failover support
  - IB path failures only
  - No failover between I/O controllers

SRP Driver Model

- StorPort designed for locally attached HW
  - Interacting outside of framework not "normal"
  - Miniport does not see important events
    - Device removal
    - Driver unload
  - Most entry points invoked at DISPATCH_LEVEL

SRP Enhancements

- DDBD Support
  - Provides better performance vs. IDBD
  - Uses a virtually contiguous MR to describe discontiguous SGLs
  - Requires registration/deregistration in speed path
    - I/O path invoked at DISPATCH_LEVEL
    - Requires IRP-based verb support (async)
- Multi-Path I/O
  - Allow failover and load balancing across multiple I/O controllers
- Migrate to iSCSI driver model
  - Support Windows XP
  - Designed to support working outside of the iscsiport framework

Agenda

- IPoIB Miniport
- SRP Miniport
- Winsock Direct Provider
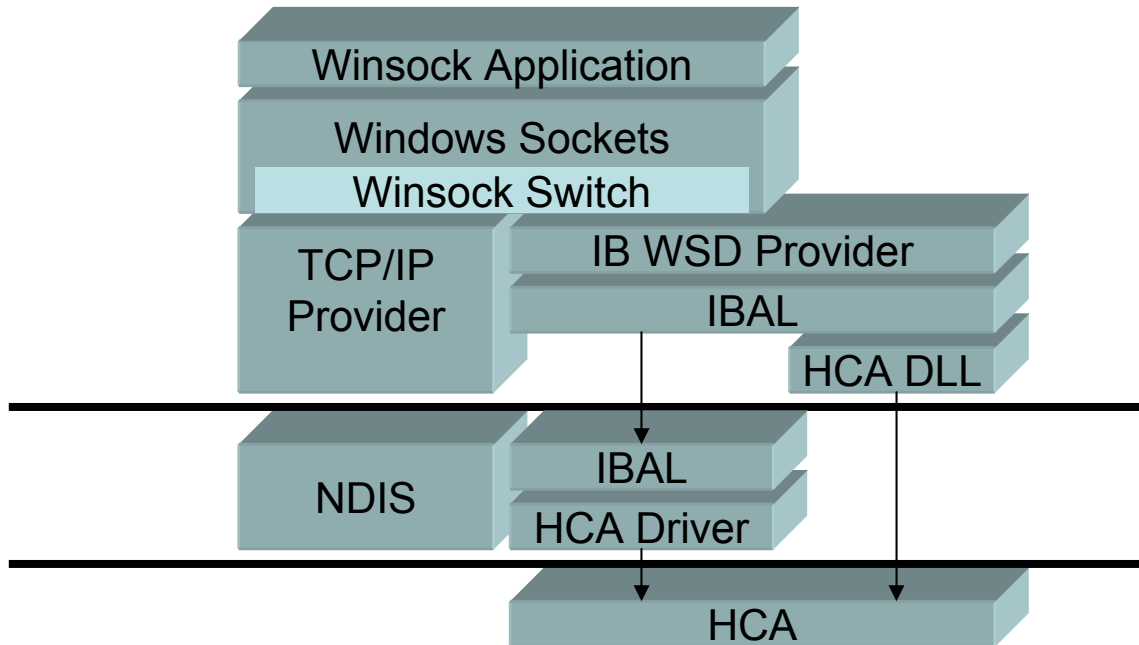
Winsock Direct Provider

- Transparent support for socket apps
  - Zero-copy
  - RDMA
- Framework provided by Microsoft
  - Part of Windows Server 2003
- Uses ATS to for address resolution
  - IP to GIDs (one to many)
  - GID to IPs (one to many)

WSD Driver Model

- Infrastructure provided by Microsoft
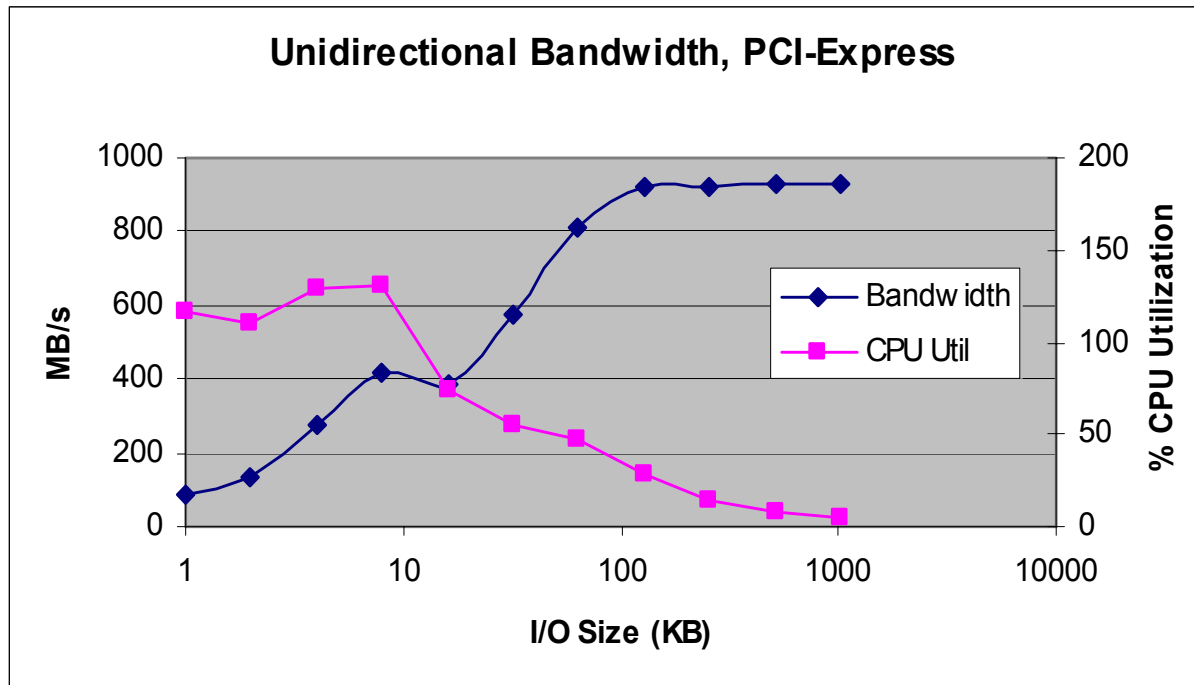- Available in Server SKUs only

WSD Performance

- Dual 3.2GHz EM64T Server
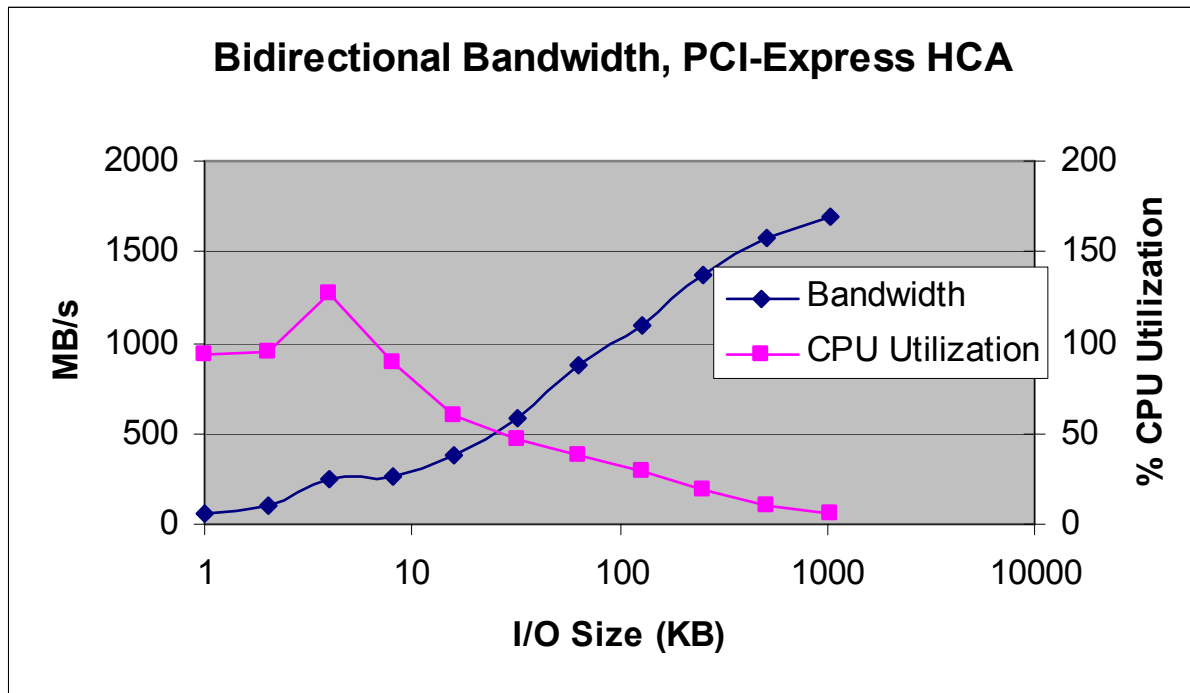- Windows 2003, x64 Edition
- PCI-Express HCA

Unidirectional Bandwidth, PCI-Express

WSD Performance

- Dual 3.2GHz EM64T Server
- Windows 2003, x64 Edition
- PCI-Express HCA

Bidirectional Bandwidth, PCI-Express HCA

WSD Testing

- WHQL tests for WSD
  - In HCT 12.101
  - Passing majority of tests, minor issues being investigated

WSD To Do

- Flush cached MR on VM changes
    - WSD Switch provides notification
    - Need to implement MR cache callback
- Use IPoIB IP→IB Address Translation
    - Eliminate SA queries
    - Access cached information
- Performance Tuning

Resources

- ## OpenIB WiKi
  - https://openib.org/tiki/tiki-index.php?page=OpenIB+Windows

- ## Openib-windows mailing list
  - http://openib.org/mailman/listinfo/openib-windows

- ## Sign up to contribute
  - http://windows.openib.org/openib/contribute.aspx

Q & A