NPIV Technology Introduction

Session 3814

Stephen Guendert
Brocade

stephen.guendert@brocade.com
Topics

• Problem Description
• WIIFM?
• Technology Summary
• Next SHARE
Problem Description and Solution
Problem: Providing M/F I/O Virtualization

- Linux for S/390 and zSeries is a "pure" Linux from a user point of view
  - Supports the S/390 and zSeries processor architecture and devices that are specific to S/390 and zSeries environments
  - Linux for S/390 and Linux for zSeries automatically inherits important strengths and reliability features of the S/390 and zSeries hardware.

- z-types now supports zLinux (usually as a VM guest)
  - Mainframe expanded to address open system applications
  - Linux promoted as alternative to Unix
  - Mainframe OS virtualization benefits
    - Availability, serviceability, scalability, flexibility

- Benefit trade-offs
  - FCP I/O requests become serialized by the OS subsystem
    - FCP header doesn’t provide image addresses
  - Any one Linux image usually does not push much I/O
  - Mainframe Channel ports are often underutilized
Solution: N_Port Identifier Virtualization

- N_Port becomes virtualized
  - Supports multiple images behind a single N_Port

- N_Port requests more than one FCID
  - FLOGI provides first address
  - FDISC provides additional addresses

- All FCID’s associated with one physical port
Market Analysis

A.K.A.: Why should we even care about NPIV? WIIFM?
Linux is Big Business

• IBM, Sony and Phillips have created a company for sharing Linux patents, royalty free
  • This is a break-through in protecting vendors and customers from patent royalty disputes resulting from freely shared Linux code
  • Linux can now challenge Microsoft who has argued that relying on “open source” software, like Linux, poses legal risks

• If the patent risk is lifted, Linux use could accelerate very fast
  • Might be used in medical devices, electronics and traffic lights
  • Definitely more use in the data center

• IDC estimates that worldwide Linux business will grow 25.9% annually, doubling from $20 billion in 2005 to $40 billion by 2008
Other reasons to care

- Learning new technology is a career enhancing move.
- The world is moving to a converged data center where the data is what matters (data centricity) and not the operating system type or server platform type.
- If you can show management cost savings, that generally is a career enhancing move.
  - Linux and NPIV may save your company operating cost $$.
Technology Summary
Standards and NPIV

- **FC-FS**
  - Describes FDISC use to allocate additional N_Port_IDs
    - Section 12.3.2.41
  - NV_Ports are treated like any other port
    - Exception is they use FDISC instead of FLOGI

- **FC-GS-4**
  - Describes
    - Permanent Port Name and Get Permanent Port Name command
      - Based on the N_Port ID (G_PPN_ID)
    - The PPN may be the F_Port Name

- **FC-LS**
  - Documents the responses to NV_Port related ELSs
    - FDISC, FLOGI and FLOGO
    - Reference 03-338v1
More Standards on NPIV

- **FC-DA**
  - Profiles the process of acquiring additional N_Port_IDS
    - Clause 4.9

- **FC-MI-2**
  - Profiles how the fabric handles NPIV requests
    - New Service Parameters are defined in 03-323v1
    - Name Server Objects in 7.3.2.2 and 7.3.2.3
FCP Channels on the mainframe

- FICON Express2/4 channel cards FCP microcode.
  - Allows support for Linux SCSI data payload
  - Allows Linux running on host to access industry standard SCSI devices.

- Queued direct input/output (QDIO) architecture is used by FICON channels in FCP mode to communicate with the OS.
  - Derived from the same QDIO architecture defined for Hipersockets communications and for OSA Express.

- Define channel type using HCD/IOCP and the QDIO data devices
  - No requirement for defining fibre channel storage devices/controllers, or directors/switches.
    - Addressed by OS using WWNs, fibrechannel IDs and Logical Unit Numbers (LUNs).
FCP LUN Access Control

- Predecessor to NPIV
- Z990, 890, z9
- Provides host based control of access to storage controllers and their devices as identified by LUNs.
- Allows read only sharing of FCP SCSI devices among multiple operating system images.
- Gives end user the ability to define individual access rights to storage controller ports as well as devices for each operating system image.
FCP LUN Access Control (2)

- Can significantly reduce the number of FCP channels needed to provide controlled access to data on FCP SCSI devices.
  - Without it, FCP channels will prevent logical units from being opened by multiple Linux images simultaneously. (first come first served)
  - Without it, one Linux image can block other images from accessing the data on one or more logical units.

- MIF is not exploited by the FCP industry standard architecture.
Node Port ID Virtualization (NPIV)

- Unique to System z9
- Allows each operating system sharing an FCP channel to be assigned a unique virtual world wide port name (WWPN).
  - Used for both device level access control in a storage controller (LUN masking) and for switch level access control on a fibre channel director/switch (zoning).
- A single, physical FCP channel can be assigned to multiple WWPNs and appear as multiple channels to the external storage network.
- The virtualized FC Node Port IDs allow a physical fibre channel port to appear as multiple, distinct ports.
  - IO transactions are separately identified, managed, transmitted, and processed just as if each OS image had its own unique physical N port.
NPIV

- Available on System z9
  - In a logical partition
  - z/VM 4.4 or later.

- IBM designed NPIV so that multiple users can have their own Fibre Channel ID on the same N_Port.

- It doesn’t work with loop and basically behaves like loop with multiple N_Port IDs on a single N_Port

- NPIV has significant scaling implications that will be demonstrated.

- Hard Zoning works only at the N_Port Level
N_Port Identifier Virtualization

- The Mainframe N_Port becomes virtualized via NPIV
  - Supports multiple Linux guests behind a single N_Port
  - Allows hundreds or thousands of Linux guests in a VM machine to share a physical channel
  - Not a FICON implementation - native FCP

- The Mainframe N_Port requests more than one FCID from the FICON switching device F_Port
  - FLOGI provides first (physical) address
  - FDISC provides additional (virtual) addresses

- All FCID’s are associated with one FICON switching device physical port which creates the virtualized environments

- A physical port can represent up to 256 virtual ports
  - Virtual N_Port (NV_Port) retains full N_Port capabilities
  - IBM has announced support on wide variety of switching products
A Simplified Schematic of Linux390 on a z-mainframe without NPIV

One FCP Channel Per Linux

No parallelism so it is very difficult to drive I/O for lots of Linux images

Probably very little I/O bandwidth utilization

z9-109

Linux A
Linux B
Linux C
Linux D

Line Card

6064 or 6140 Chassis

200 MBps per port

Copyright 2007 Brocade
A Simplified Schematic of Linux390 on a z9-109 with NPIV

Lots of Parallelism

One FCP Channel for many Linux images

Much better I/O bandwidth Utilization per path

6064 or 6140 Chassis

Copyright 2007 Brocade
**z9-109 N-port ID Virtualization**

**FC-FS 24 bit fabric addressing – Destination ID (D_ID)**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Area</th>
<th>AL (Port)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies the Switch Number</td>
<td>Identifies the Switch Port</td>
<td>AL_PA, assigned during LIP, Low AL_PA, high Priority</td>
</tr>
<tr>
<td>Up to 239 Switch Numbers</td>
<td>Up to 240 ports per domain</td>
<td></td>
</tr>
</tbody>
</table>

1 byte 1 byte 1 byte

---

**Domain** | **Port @** | **Virtual Addr.**

Switch | CU Link @ | 00 - FF

1 byte 1 byte 1 byte

---

*Copyright 2007 Brocade*
FICON Express2 or Express4 Cards

FICON Express Channel Card w/o NPIV

- 1 WWN
- 1 WWPN
- 1 WWPN
- 1 WWPN

FICON Express Channel Card with NPIV

- 1 WWN
- <=255 WWPNs
- <=255 WWPNs
- <=255 WWPNs

 WWN - World Wide Name
 WWPN - World Wide Port Name
Example: Mainframe SAN with NPIV

- One z9 server port can have up to 255 N_Port_IDs
  - IBM has told us it wants this expandable to thousands - future
  - The switch or fabric probably limits the maximum number

- After the first NV_Port FLOGIs, up to 255 FDISCs will acquire the other N_Port_IDs
z9 N-port ID Virtualization

- NPIV on the z9 – NOW!
  - FCP Driver for z9
  - Same channel card as used for FICON

- Provides Managed Storage Services

An ability to configure how many Linux guests will run in a VM machine.
Up to 256 per F_Port and up to 1,024 per Director

Etc. for up to 256 FCIDs per channel path
**z9 N-port ID Virtualization**

Configure > Ports > NPIV

---

**This feature is not installed**

**NPIV State:** Disabled

<table>
<thead>
<tr>
<th>Port #</th>
<th>Name</th>
<th>Attached WWN</th>
<th>Port Type</th>
<th>Login Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>None</td>
<td>G Port</td>
<td>1</td>
</tr>
</tbody>
</table>

**Node Port Identifier Virtualization (NPIV)**

Method for assigning multiple Fibre Channel addresses (NPort IDs) to a single physical node port (NPort). Typically, the N Port associated with NPIV is a server device, but device type is not limited.

---

*Copyright 2007 Brocade*
Many new WWPNs will have to be handled by zoning
Zoning will probably have to be by WWN

- <= 510 active target port connects per CHPID
- Up to 255 logical subchannels per CHPID
  At 255 subchnls/CHPID, < 3 targets per WWPN
Fabric Management of NPIV

Typical FICON zoning is “port” zoning to allow any-to-any

- What if Linux (yellow 3) only does HDS and Tape?
  - Linux Y3 should not be “zoned” to get to all the rest
  - But “other” yellow Linux users might need other devices
  - Port Zoning would have to allow Y3 to everyone
  - So NPIV is probably going to require WWN zoning in the fabric!
  - And possibly much more zoning management

Copyright 2007 Brocade
Fabric Management of NPIV

Instead of 60, or 136 or 252 ports trying to use 4 ports for cascaded links, now there could be 1,024 virtual ports trying to use the 4 cascaded links. How will you ever assure that no ISL oversubscription will be taking place?

- Only Trunking has any hope to manage this!
- FSPF routing of cascaded links is static
- Prohibit paths to block cascaded links is static
- Preferred paths for user defined ISL routing is static
- Trunking to manage cascaded links is **dynamic**!
  - Dynamic workload removal from oversubscribed cascaded links
  - Provides for user defined oversubscription thresholds

Copyright 2007 Brocade
Future of Server Virtualization

• Virtualization
  • N_port ID Virtualization
    • FCP Driver for zSeries (January 2002)

• Managed Storage Services

• eBusiness “On Demand”
  • zLinux Performance
N_Port Identifier Virtualization Summary

- **Enabling Technology**
  - Supports Mainframe and open server virtualization
  - Enables “in fabric” virtualization

- **Vendor support**
  - IBM is primary motivator
  - zLinux is first application
    - Enables economic deployment of zLinux on z9 Series
    - Enhances FCP utilization of z9 channel link
    - Supports eBusiness “On Demand” initiative (worth $4B to IBM)
Next SHARE

- NPIV Presentation based on two part zJournal article appearing in Dec/Jan 2006/07 and Feb/March 2007 issues.
  - Virtual fabric tagging and other new advances
Questions?
Questions?

Thank you!