Planning and Implementing NPIV For System z

Session 9479

Dr. Steve Guendert, Brocade Communications
sguender@brocade.com

Agenda

- Introduction
- NPIV background
- FCP Channels on the mainframe
- NPIV Planning
- NPIV implementation and configuration

Node Port ID Virtualization (NPIV)

Basics and background on NPIV and System z

zSeries/System z server virtualization

- zSeries/System z support of zLinux
  - Mainframe expanded to address open system applications
  - Linux promoted as alternative to Unix
  - Mainframe operating system virtualization benefits
    - Availability, serviceability, scalability, flexibility
- Initial zSeries limits
  - FCP requests are serialized by the operating system
    - FCP header does not provide image address
    - FICON SB2 header provides additional addressing
  - Channel ports are underutilized

FCP channel/device sharing-summary

- Different host operating systems sharing access to a single FCP channel may access the same fibre channel port via this channel.
- While multiple operating systems can concurrently access the same remote fibre channel port via a single FCP channel, fibre channel devices (identified by their LUNs) can only be serially re-used.
- In order for two or more unique operating system instances to share concurrent access to a single fibre channel or SCSI device (LUN), each of these operating systems must access this device through a different FCP channel.
- Should two or more unique operating system instances attempt to share concurrent access to a single fibre channel or SCSI device (LUN) over the same FCP channel, a LUN sharing conflict will occur, resulting in errors.

Consolidation

- Server consolidation
  - TCO compelling
  - IBM Project Big Green

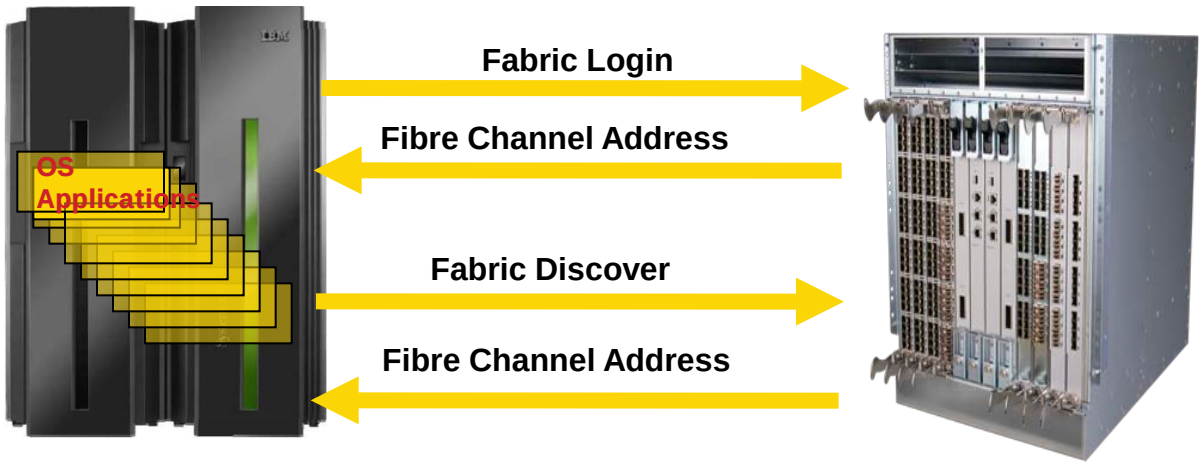
- I/O consolidation
  - Resulting cost/performance benefit is not competitive

Node Port ID Virtualization (NPIV)

- 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.

Server Consolidation-NPIV

- N_Port Identifier Virtualization (NPIV)
  - Mainframe world: unique to System z9 and later
  - zLinux on System z9/10/196 in an LPAR
  - Guest of z/VM v 4.4, 5.1 and later
  - 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



Using ELS to assign N_Port IDs

- FC standard defines a set of services used to establish communications parameters, each of which is called an *extended link service* (ELS).
- An ELS consists of a request sent by an N_Port and a response returned by a recipient port.
- One ELS, called a *fabric login* (FLOGI), is sent from an N_Port to its attached fabric port (F_Port) in the adjacent switch to request the assignment of an N_Port ID.

Using ELS to assign N_Port IDs: FLOGI

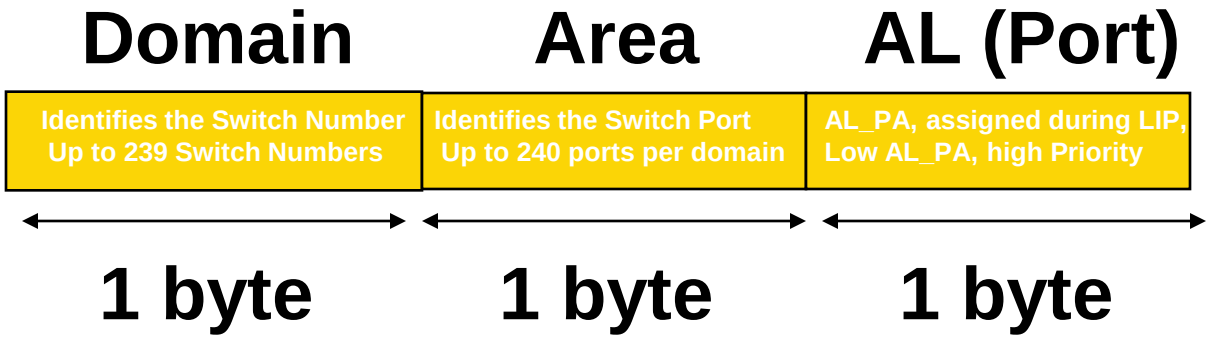
- The FLOGI request is the first frame sent from an N_Port to its adjacent switch.
- The purpose of the FLOGI ELS is to enable the switch and the N_Port to exchange initialization parameters
  - Includes unique identifiers known as *worldwide port names* (WWPNs)
  - Allows the fabric to assign an N_Port ID to the N_Port.
- The switch responds with the N_Port assigned to the requesting N_Port.
- Because the N_Port that sends the FLOGI request does not yet have an N_Port ID, it sets the S_ID in the FLOGI request to zero.
  - The switch responds with a FLOGI-accept response that contains the assigned N_Port ID.
  - The “HBA” uses this assigned N_Port ID as the S_ID when sending subsequent frames.

Need to request multiple N_Port IDs: FDISC

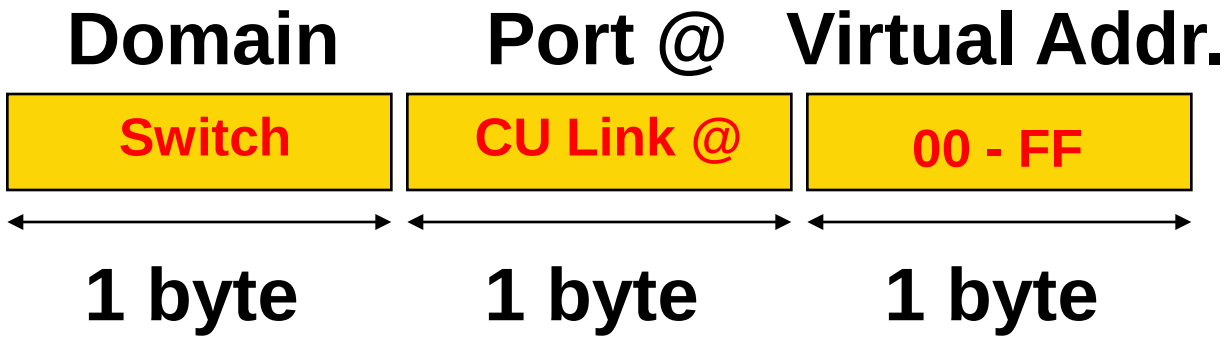
- Fabric Discovery (FDISC) is another ELS
- Original purpose is to verify an existing login with the fabric is still valid.
- The FDISC was always sent with a non-zero S_ID (the presumed S_ID of the sender).
- This made it possible to obtain additional N_Port IDs by an extension of the FDISC ELS.
- An unlimited number of additional N_Port IDs could be obtained

System z N-port ID Virtualization

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



FICON Express2, Express4 and Express 8 adapters now support NPIV

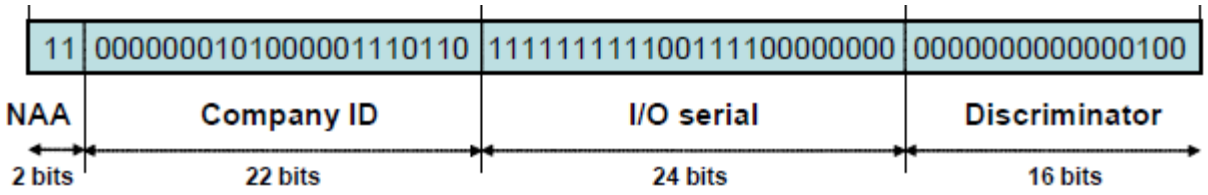


Without NPIV…..

- Each operating system image that has an FCP port is identified to the fabric by the permanent WWPN of the port.
  - All OS images then have the same access rights within the fabric.
- The permanent WWPN of the port determines:
  - Zone membership for all images sharing the port.
  - Logical Unit Number (LUN) access rights for all images sharing the port.

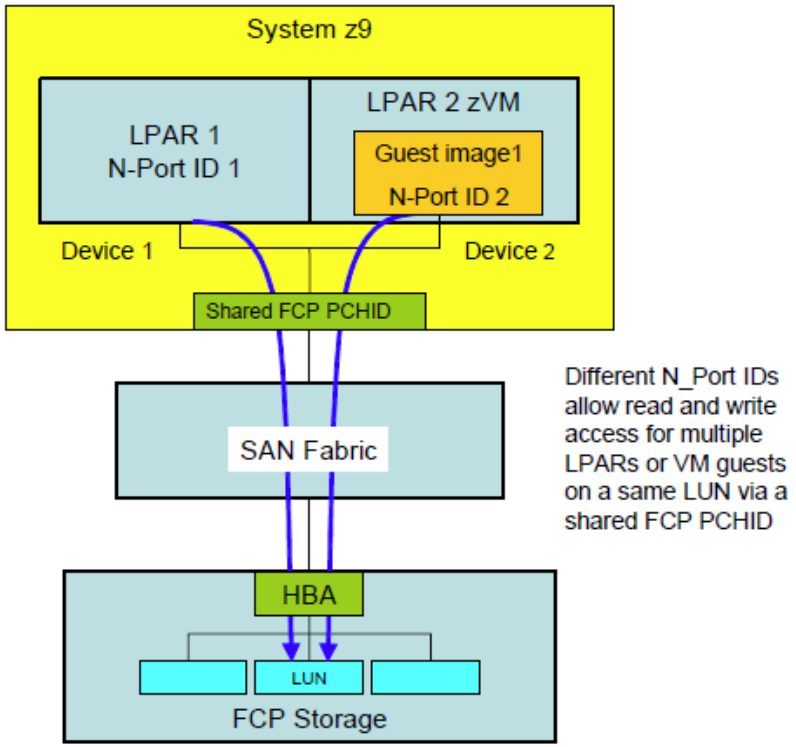
With NPIV…..

- The Service Element (SE) creates new WWPNs for the FCP port at FLOGI.
- A unique WWPN is assigned to each OS image sharing the port.
- The generated NPIV WWPN is registered with the fabric switch.
  - This uniquely identifies each image for fabric zoning and LUN masking.
- For increased address space with System z, the low order 7 bits of the I/O serial field are combined with the 16 bit discriminator field.
  - Increases address space to over 8000000 unique WWPNs for a single FCP port.



Graphic: IBM

Example illustration



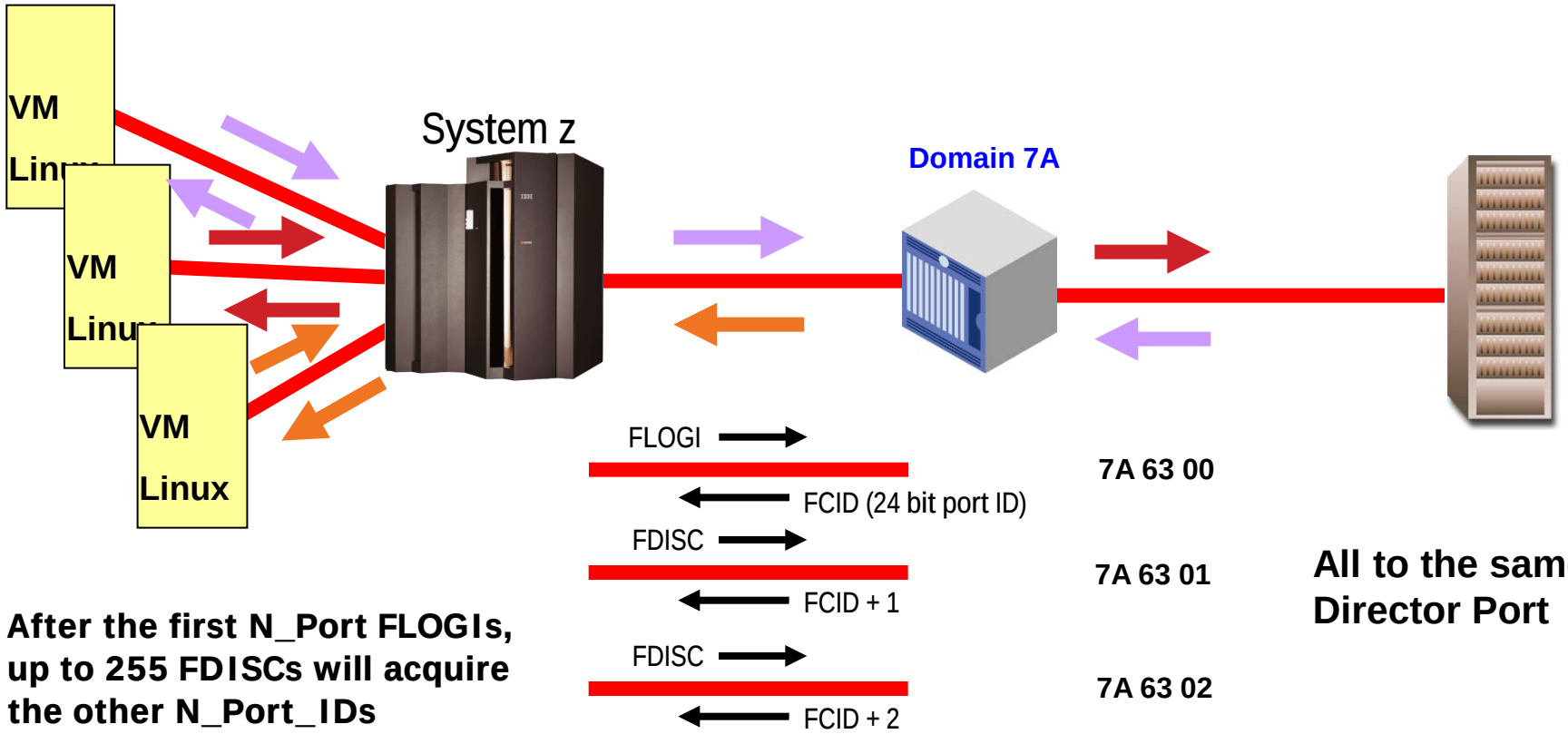
- In this figure, the two LPARs share a single physical FCP port.
- Each instance registers with the fabric's name server.
- The NPIV WWPN is supported in the FDISC process.
- Each LPAR receives a different N_Port ID to allow multiple LPARs or VM guests to read and write to the same LUN using the same physical port.
  - Without NPIV, writing to the same LUN over a shared port is not allowed.

Graphic: IBM

System z N-port ID Virtualization-summary

One System z server port can have up to 255 NP-IDs
•IBM has told us it wants this expandable to thousands

- NPIV on the System z
  - FCP Driver for System z
  - Same CHPIDs as used for FICON



Etc. for up to 256 FCIDs per channel path

FCP channels on the mainframe

FICON and FCP Mode

- A FICON channel in Fibre Channel Protocol mode (CHPID type FCP) can access FCP devices through a single Fibre Channel switch or multiple switches to a SCSI device

- The FCP support enables z/VM, z/VSE, and Linux on System z to access industry-standard SCSI devices. For disk applications, these FCP storage devices use Fixed Block (512-byte) sectors instead of Extended Count Key Data (ECKD) format.

- FICON Express8, FICON Express4, FICON Express2, and FICON Express channels in FCP mode provide full fabric attachment of SCSI devices to the operating system images, using the Fibre Channel Protocol, and provide point-to-point attachment of SCSI devices.

FICON and FCP Mode (Continued)

- The FCP channel full fabric support enables switches and directors to be supported between the System z server and SCSI device, which means many “hops” through a storage area network (SAN).

- FICON channels in FCP mode use the Queued Direct Input/Output (QDIO) architecture for communication with the operating system.

FCP channel and device sharing

- An FCP channel can be shared between multiple Linux operating systems, each running in a logical partition or as a guest operating system under z/VM.
- To access the FCP channel, each operating system needs its own QDIO queue pair defined as a data device on an FCP channel in the HCD/IOCP.
- These devices are internal software constructs and have no relation to physical devices outside of the adapter.
- These QDIO devices are also referred to as subchannels.

FCP channel and device sharing

- An FCP channel can be shared between multiple Linux operating systems, each running in a logical partition or as a guest operating system under z/VM.
- To access the FCP channel, each operating system needs its own QDIO queue pair defined as a data device on an FCP channel in the HCD/IOCP.
- These devices are internal software constructs and have no relation to physical devices outside of the adapter.
- These QDIO devices are also referred to as subchannels.
- The host operating system uses these subchannels as vehicles to establish conduits to the FCP environment.
- Each subchannel represents a virtual FCP adapter that can be assigned to an operating system running either natively in an LPAR or as a guest OS under z/VM.

NPIV PLANNING AND IMPLEMENTATION

Best practices and good ideas

Planning guide

- Good reference material available from IBM Redbooks
- 2005 Redbook
  - Z9
  - 6064
  - 2 Gbps

NEWS FLASH !!!

- An update to the 2005 Redpaper is going to happen
- Joint paper (IBM-Brocade)
- Reflection of:
  - New mainframe technology (zEnterprise)
  - Improved channel technology (FICON Express 8S)
  - Improved switching technology

General planning considerations

- Do not use more than 32 subchannels per physical channel in NPIV mode.
- Do not use more than 128 total target logins per physical channel in NPIV mode.
  - Example: in a configuration with 32 subchannels limit the target logins to no more than an average of 4.
- Using more subchannels, target logins, or both can cause timeout errors.

Using NPIV in the Real World

- IBM's 2005guidelines recommend that a customer deploy no more than *32 virtual nodes per port-pair link* because of potential timeout issues
  - This is based on numbers that came with the 2005 2Gb environment
- There is a desire to move beyond 32 NPIV definitions per port, but the limiting factor is not the per port limit, but the fabric limit:
- Consequently, the 32 virtual links per physical link limitation is due to fabric and switch support limitations, which is a function of both memory resources and processing – which will get better over time
- The bigger problem is the name server size explosion, which is something that we will have to address before NPIV can be fully exploited to its maximum potential.

General planning considerations (2)

- Zone each NPIV WWPN individually.
  - Reduces fabric traffic since all participants in a zone would be notified when another N_Port joins/leaves the fabric.
- Consider using multipathing.
  - For details, please refer to the Redbook *Linux on zSeries: Fibre Channel Protocol Implementation Guide* (SG24-6344).
- Enable NPIV on the SAN switch/director prior to enabling it on the mainframe.
  - If NPIV is not enabled on the switching device first, the attempt to establish a connection to the fabric will fail for all subchannels that are operated in NPIV mode.

General planning considerations (3)

- Switching devices typically limit the number of supported N_Port IDs.
  - Because each NPIV WWPN is assigned an N_Port ID at login, this limit can be exceeded.
- Some switching devices also will limit the number of N_Port IDs that can be assigned to a physical port.
  - This limit varies by vendor/model
  - This limit also may be configurable on some switching devices.

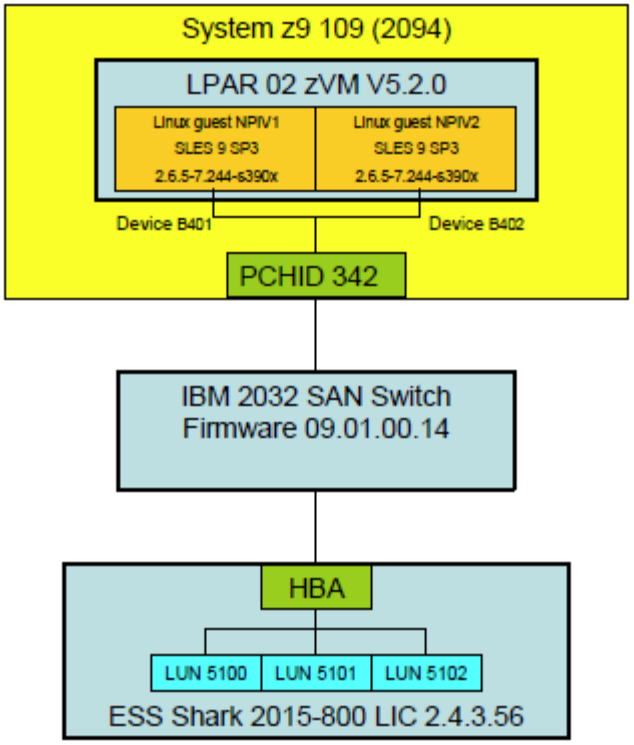
General planning considerations (4)

- Each login from an NPIV mode subchannel into a storage subsystem counts as a separate host login.
- Consult with your storage vendor for the specific limits in your configuration.
- Example:
  - IBM ESS model 800 supports up to 16 Host Bus Adapters (HBAs).
    - Each HBA supports up to 124 host logins
    - ESS 800 itself supports up to 512 host logins

Configuration section

- Section has many screen shots/captures
- 20 slides, but they will go by fairly quickly
- Some of the graphics used are from IBM software, Brocade management software, and/or illustrations from some of the references cited at the end of the presentation.

Configuration of NPIV



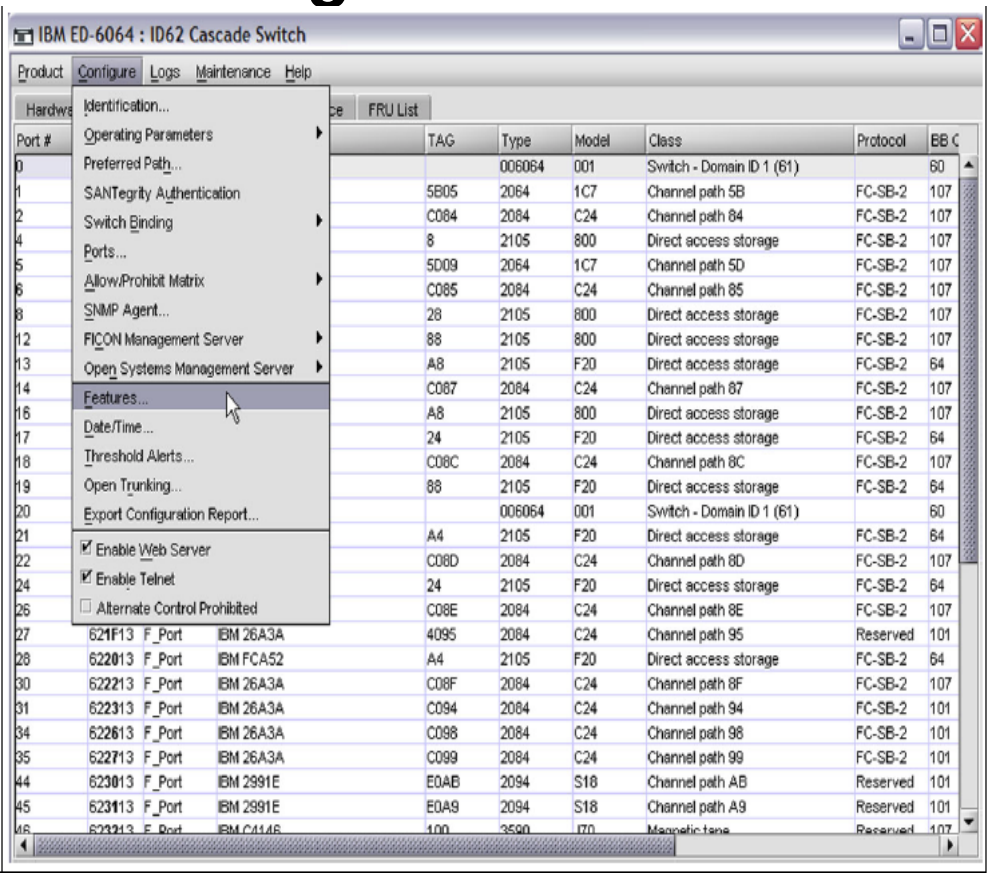
- Next several slides will walk through configuration steps.
- Configure:
  - NPIV on the switching device
  - NPIV on the mainframe
  - Fabric Security
  - Linux server
- For more specific details, please see references listed at end.

Graphic: IBM

Configuring NPIV on the switching device

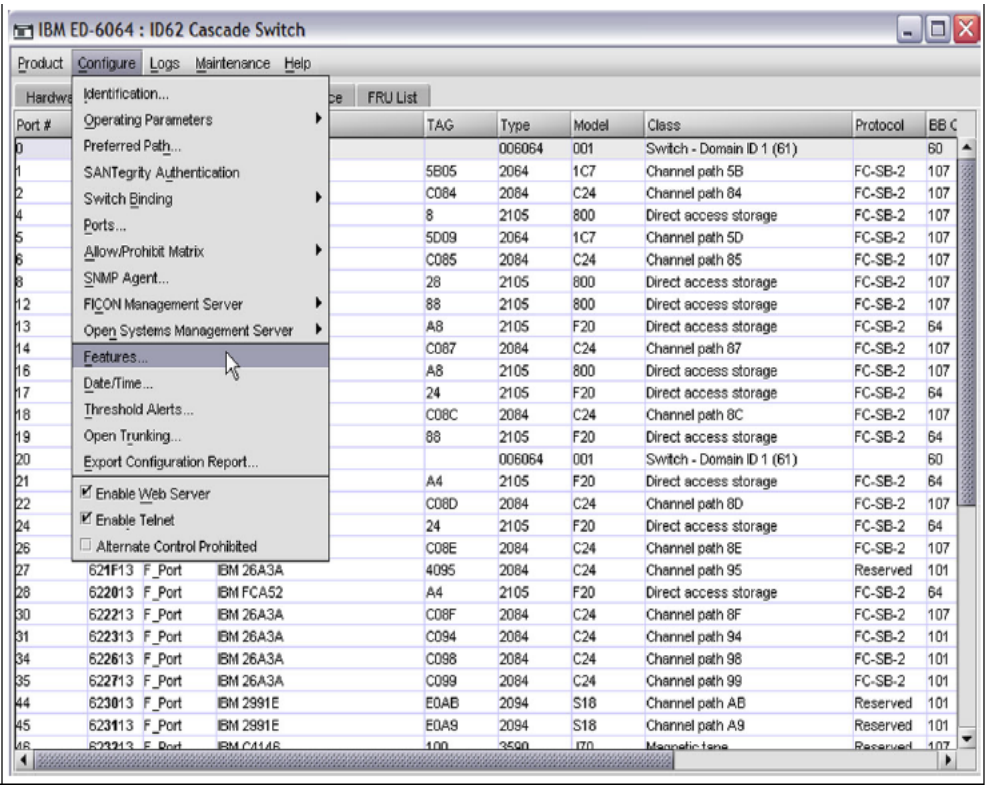
- Example of a specific switch, check your model's documentation for details on your switches.
- First, enable NPIV on the switch.
  - Add the NPIV feature key (if required)
  - Activate the NPIV feature
  - Configure NPIV on a specific switch port

Adding NPIV feature to switch

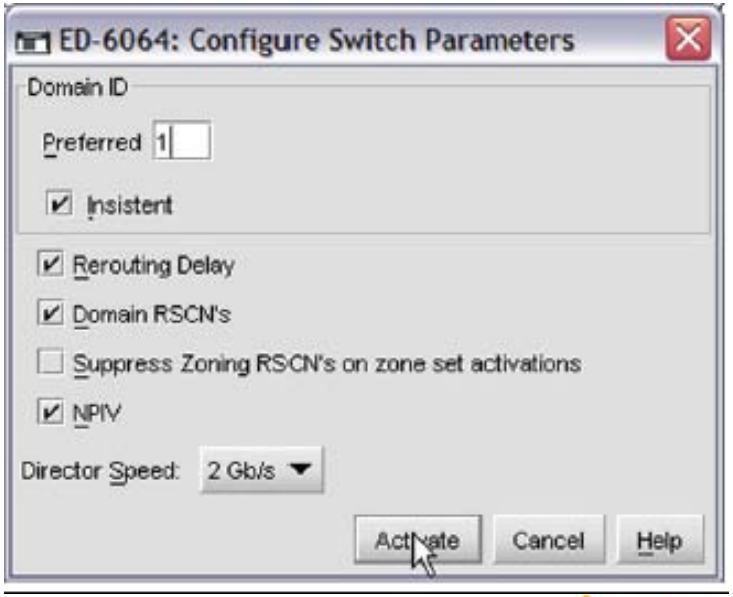


- Start the switch's management application and log onto switch.
- In this example, select the **Configure** drop down menu, and go to **Features**.
- In the **Features** configuration menu, click **New** and enter the feature key provided by the manufacturer.

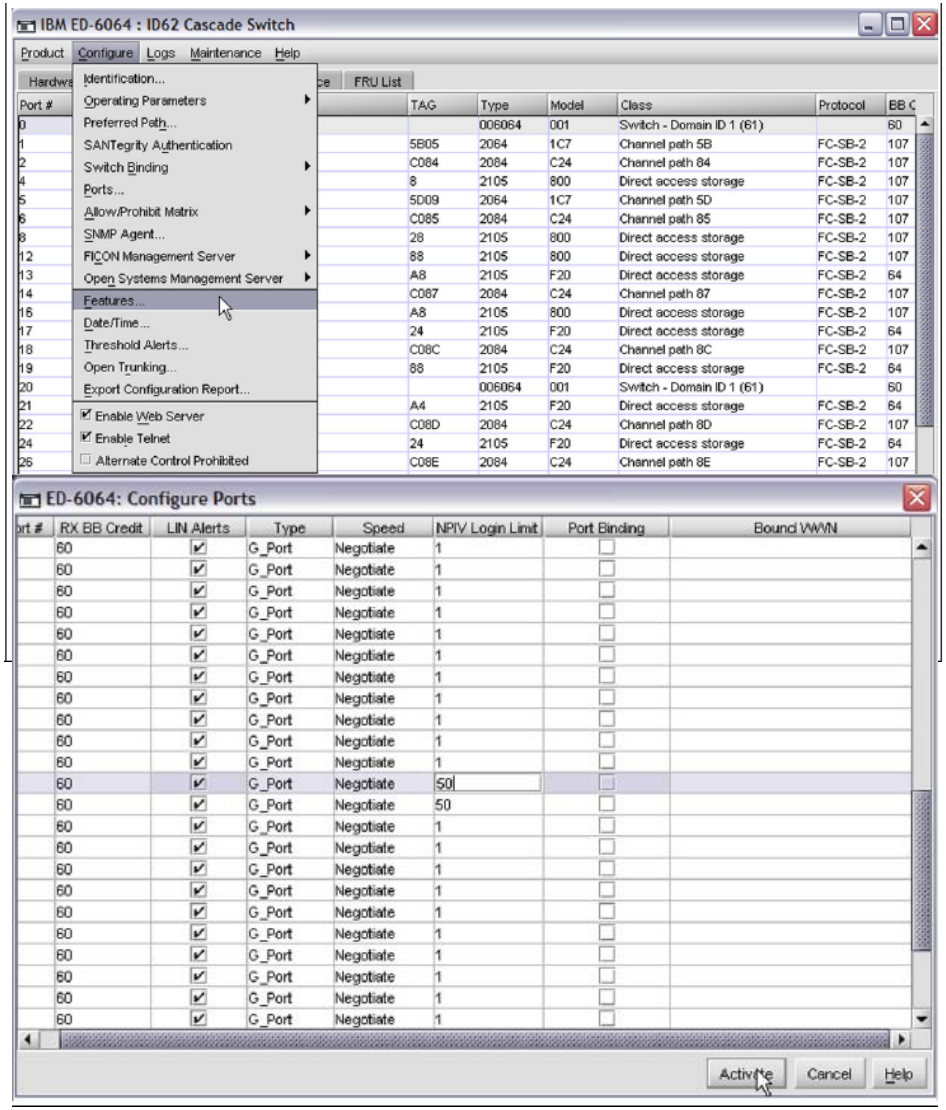
Activating the NPIV feature on switch



- Select the **Configure** dropdown, then **Operating Parameters,** then **Switch Parameters** to open the Configure Switch Parameters menu.
- Select the **NPIV** option
- Click **Activate**



Configuring NPIV on an individual port

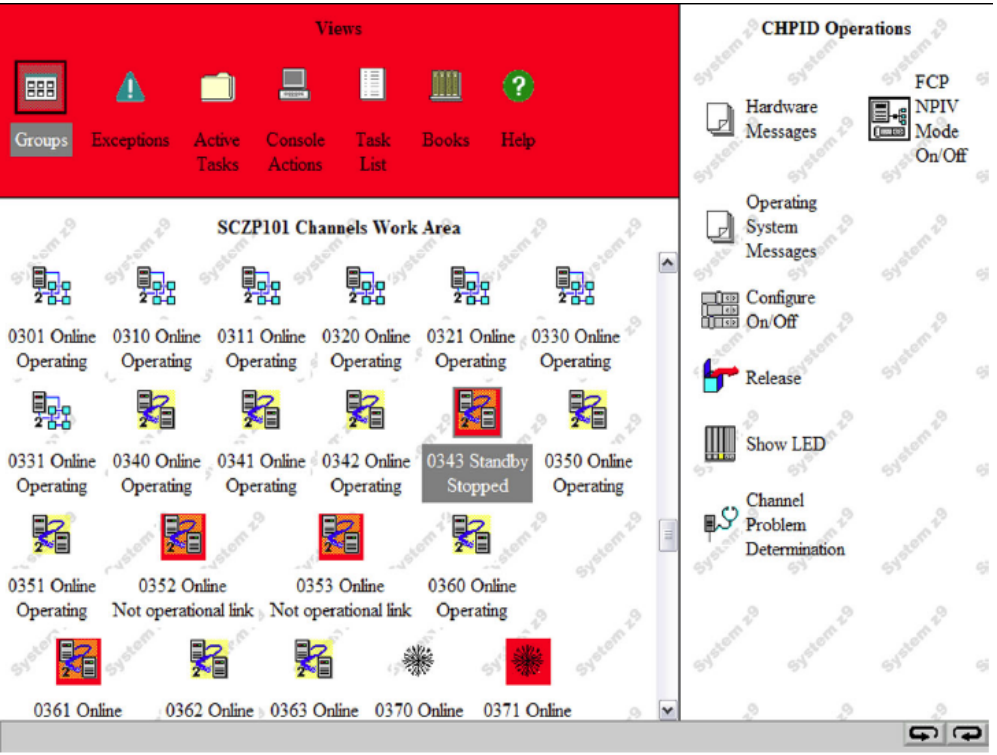


- Select **Configure,** drop down to **Ports**
- Double click on **the NPIV login limit** for the port and enter the desired login limit.
  - This will limit the total number of WWPN logins (both NPIV and the default WWPN logins)
  - If you set this value too small, an out of resource in fabric error message is reported on the Linux host.
- Click **Activate** to complete the configuration.

Configuring NPIV on the mainframe

- Example we use is a System z9
- Configure NPIV on the switch first
  - If NPIV is enabled on our System z9 but not on the switch, the FCP CHPID reverts to non-NPIV mode on Fabric Login.
  - A CHPID off/on will then be required to enable NPIV once the switch has NPIV configured/enabled. (Disruptive)
- To enable NPIV, we will
  - Enable the NPIV feature on the mainframe
  - Find the NPIV WWPNs for the FCP CHPID
  - Find the permanent WWPN for the FCP CHPID

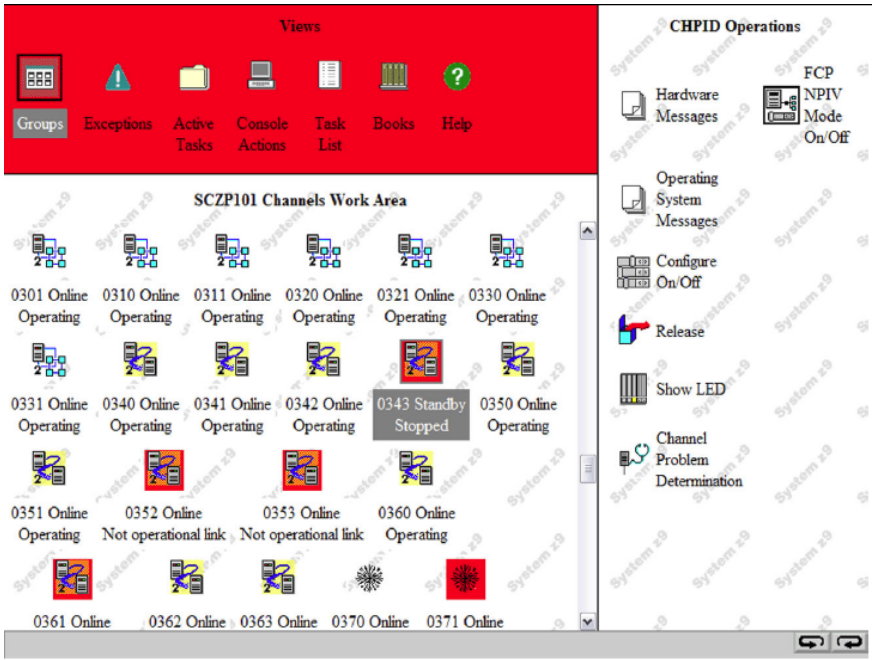
Enabling the NPIV feature on our z9

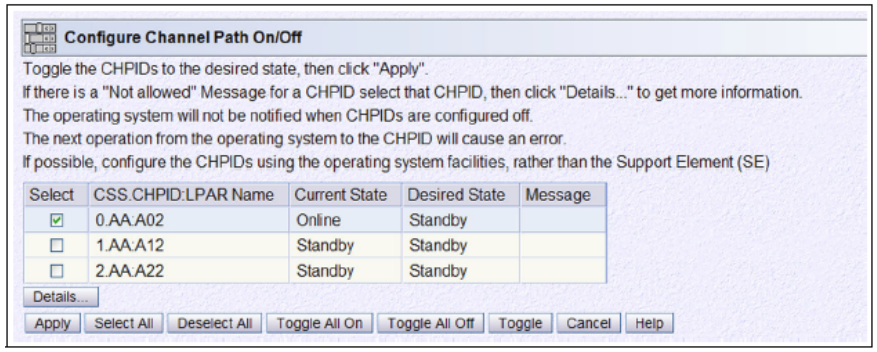


- Can be enabled from the CHPID operations menu in the SE
- From the HMC, select the Single Object to navigate to the SE:
  - Select your CPC, right click, and select the **Channels** option.
  - Scroll over to the **CHPID Operations** task on the right.
- The NPIV feature can be selectively enabled for individual LPARs.
- From the **CHPID Operations** menu……

Graphic: IBM

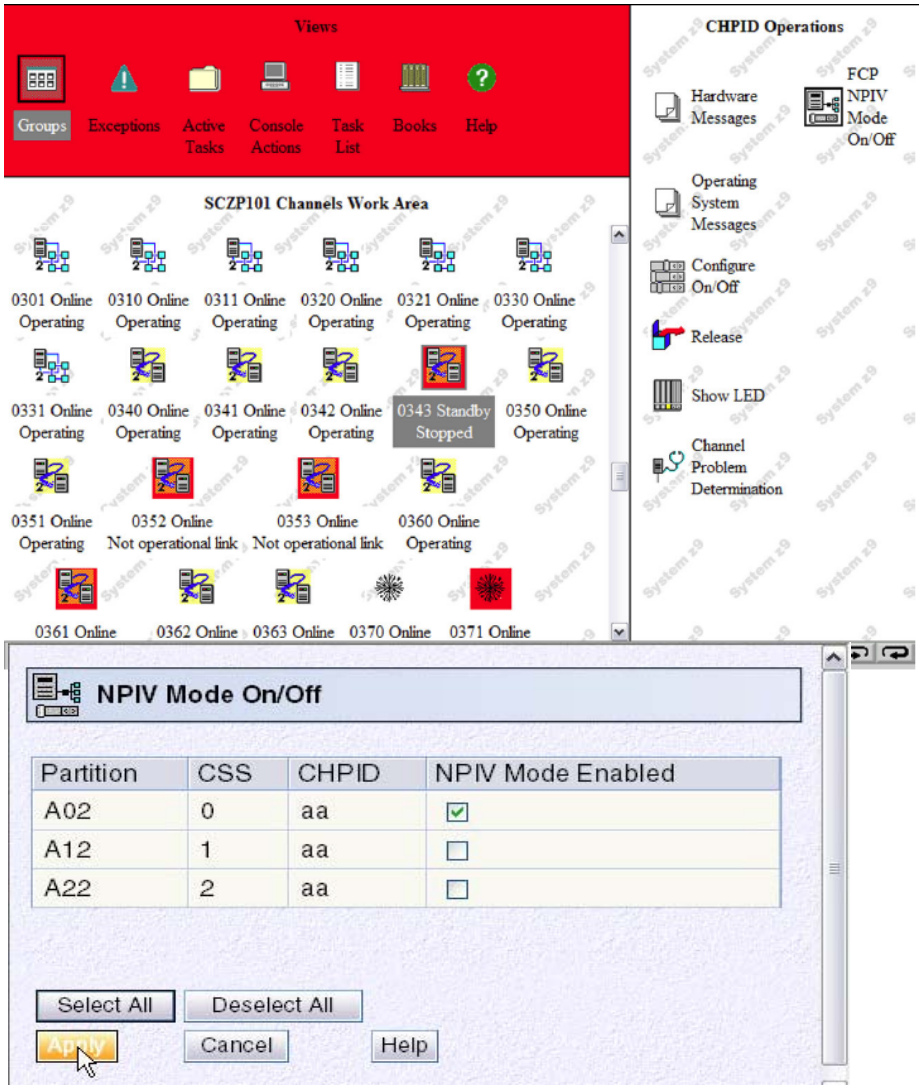
Enabling the NPIV feature on our z9 (2)





- Set the FCP CHPID to standby
  - Double click **Configure On/Off**
- Select the appropriate LPARs and click **Toggle** to change the **Desired State** option to standby.
- Click **Apply** to commit the changes

Enabling the NPIV feature on our z9 (3)

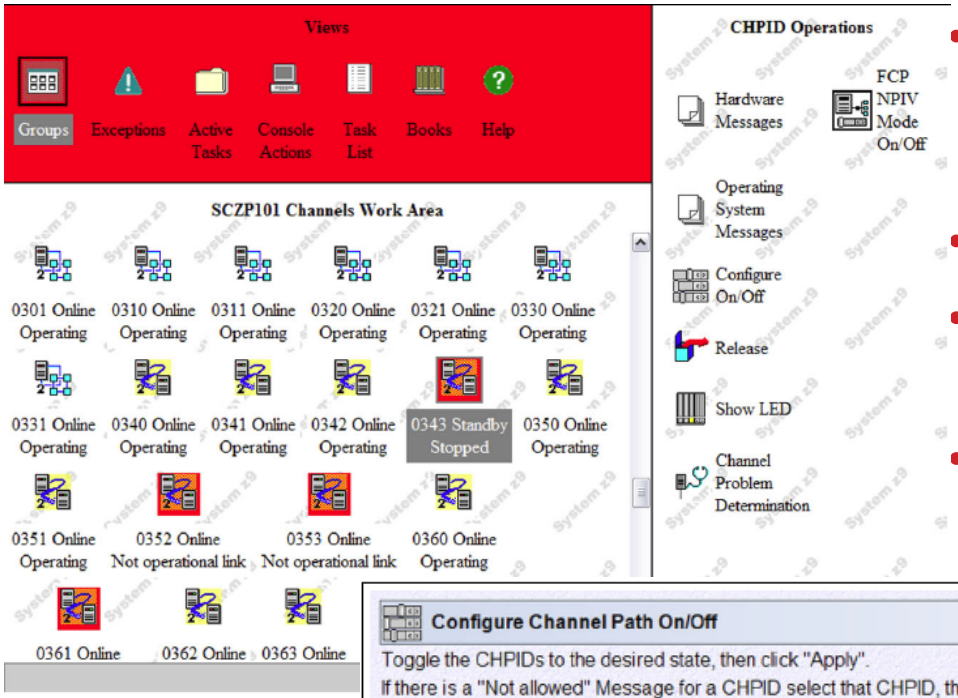


- Select your PCHID number in the Channel Work Area
- From the **CHPID operations** menu on the right, double click **FCP NPIV Mode On/Off** to open the **NPIV mode on/off menu**
- Select the **NPIV Mode Enabled** option for each LPAR.
- Click **Apply** to commit the changes

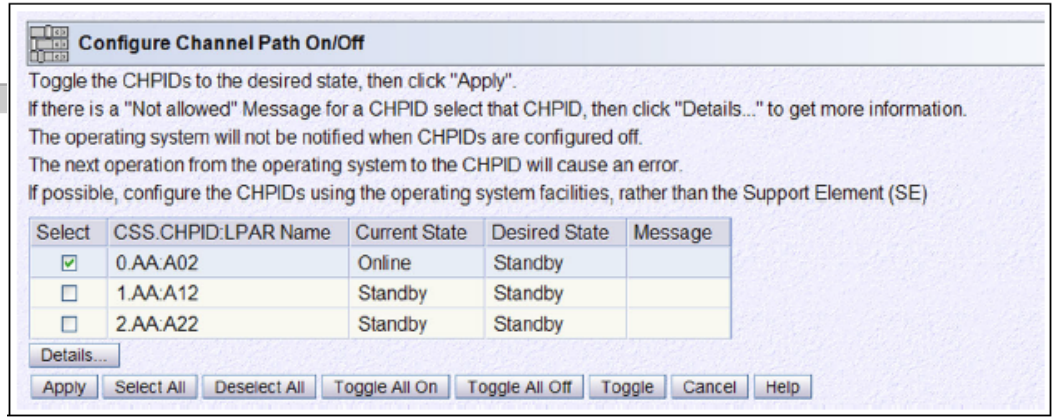
Important note

- To enable NPIV, the CHPID must be in standby state for the LPAR.
- If not, the **NPIV Mode On/Off** option is disabled to prevent any mode changes.

Setting the FCP CHPID online



- **CHPID Operations Menu**->double click **Configure Channel Path On/Off**
- Select the appropriate LPAR
- Click **Toggle** to change the **Desired State** option to **Online**
- Click **Apply** to commit the changes.

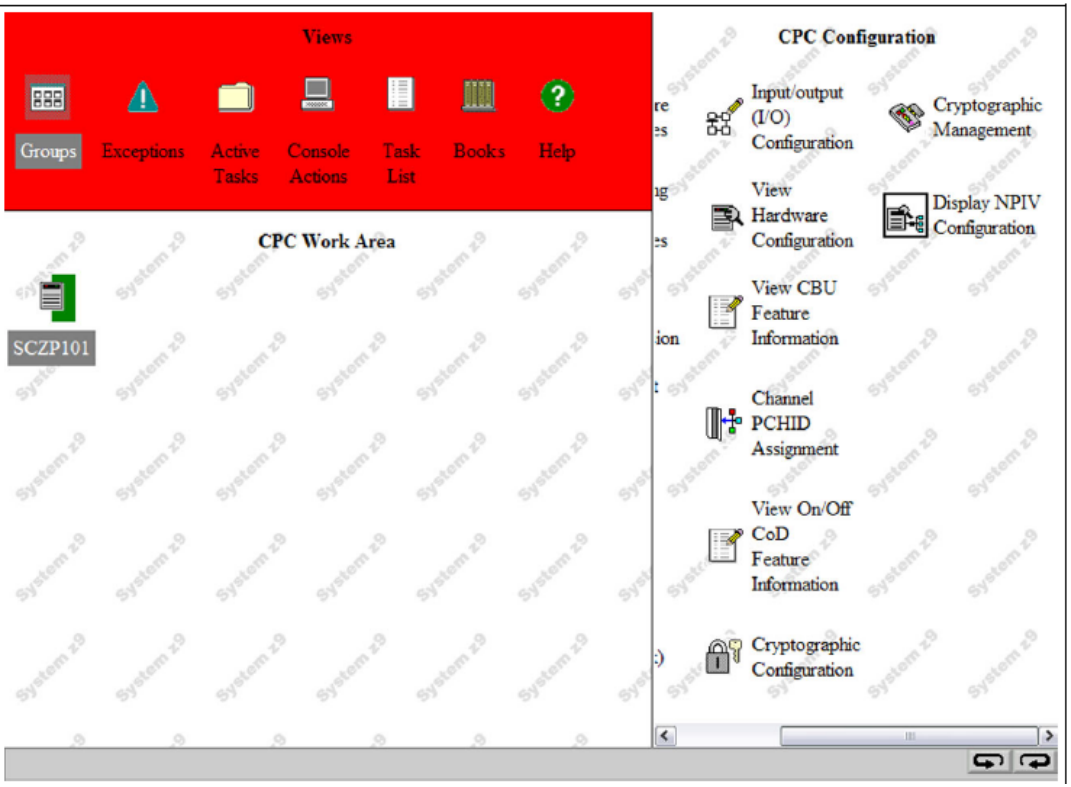


Finding the NPIV WWPNs for the FCP CHPID

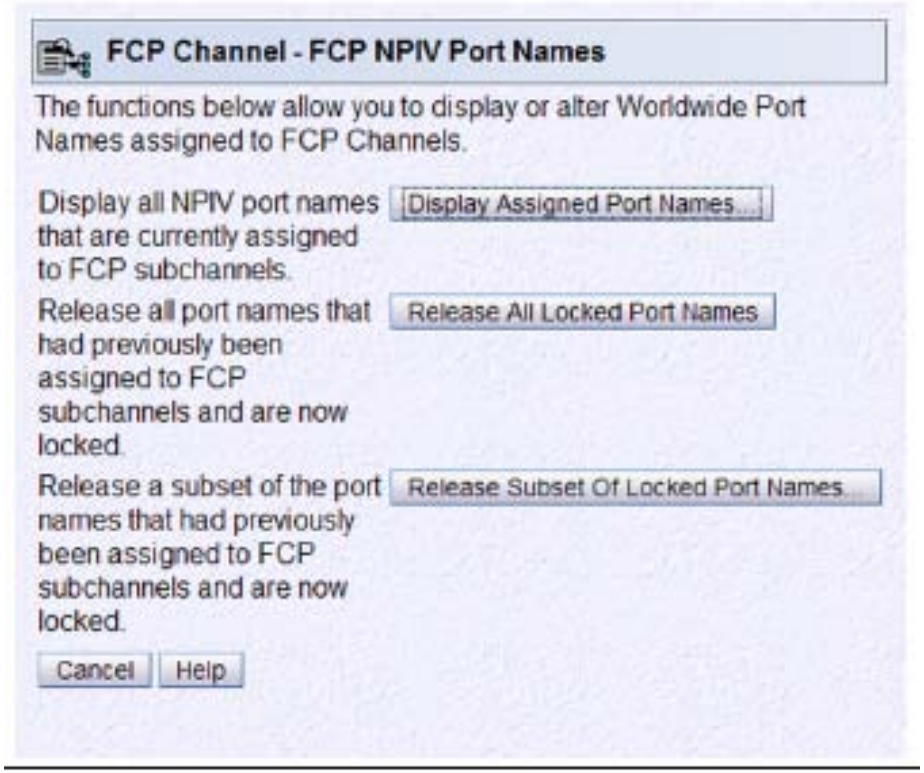
- Once we have enabled the NPIV mode, we can find the NPIV WWPNs assigned to an FCP CHPID.
- We need to know these WWPNs so we can:
  - Configure LUN masking
  - Configure zoning in the SAN
- We'll access the NPIV WWPNs from the CPC Configuration menu in the SE.

Finding the NPIV WWPNs for the FCP CHPID (2)

- Click **Display NPIV Configuration** to navigate to the **FCP channel –FCP NPIV Port Names** menu.



Finding the NPIV WWPNs for the FCP CHPID(3)



- Click **Display Assigned Port Name** to open the **Display Assigned Port Names Menu** (next slide).
- It's a good idea to restrict the number of WWPNs displayed by selecting the **Show NPIV=On** option.

Finding the NPIV WWPNs for the FCP CHPID(4)

Display Assigned Port Names

Partition	CSS	IID	CHPID	SSID	Device Number	WWPN	IOCDS	NPIV Mode
A02	00	02	a8	00	b400	c05076ffcf000000	A1	On
A02	00	02	a8	00	b401	c05076ffcf000004	A1	On
A02	00	02	a8	00	b402	c05076ffcf000008	A1	On
A02	00	02	a8	00	b403	c05076ffcf00000c	A1	On
A02	00	02	a8	00	b404	c05076ffcf000010	A1	On
A02	00	02	a8	00	b405	c05076ffcf000014	A1	On
A02	00	02	a8	00	b406	c05076ffcf000018	A1	On
A02	00	02	a8	00	b407	c05076ffcf00001c	A1	On
A02	00	02	a8	00	b408	c05076ffcf000020	A1	On
A02	00	02	a8	00	b409	c05076ffcf000024	A1	On
A02	00	02	a8	00	b40a	c05076ffcf000028	A1	On
A02	00	02	a8	00	b40b	c05076ffcf00002c	A1	On

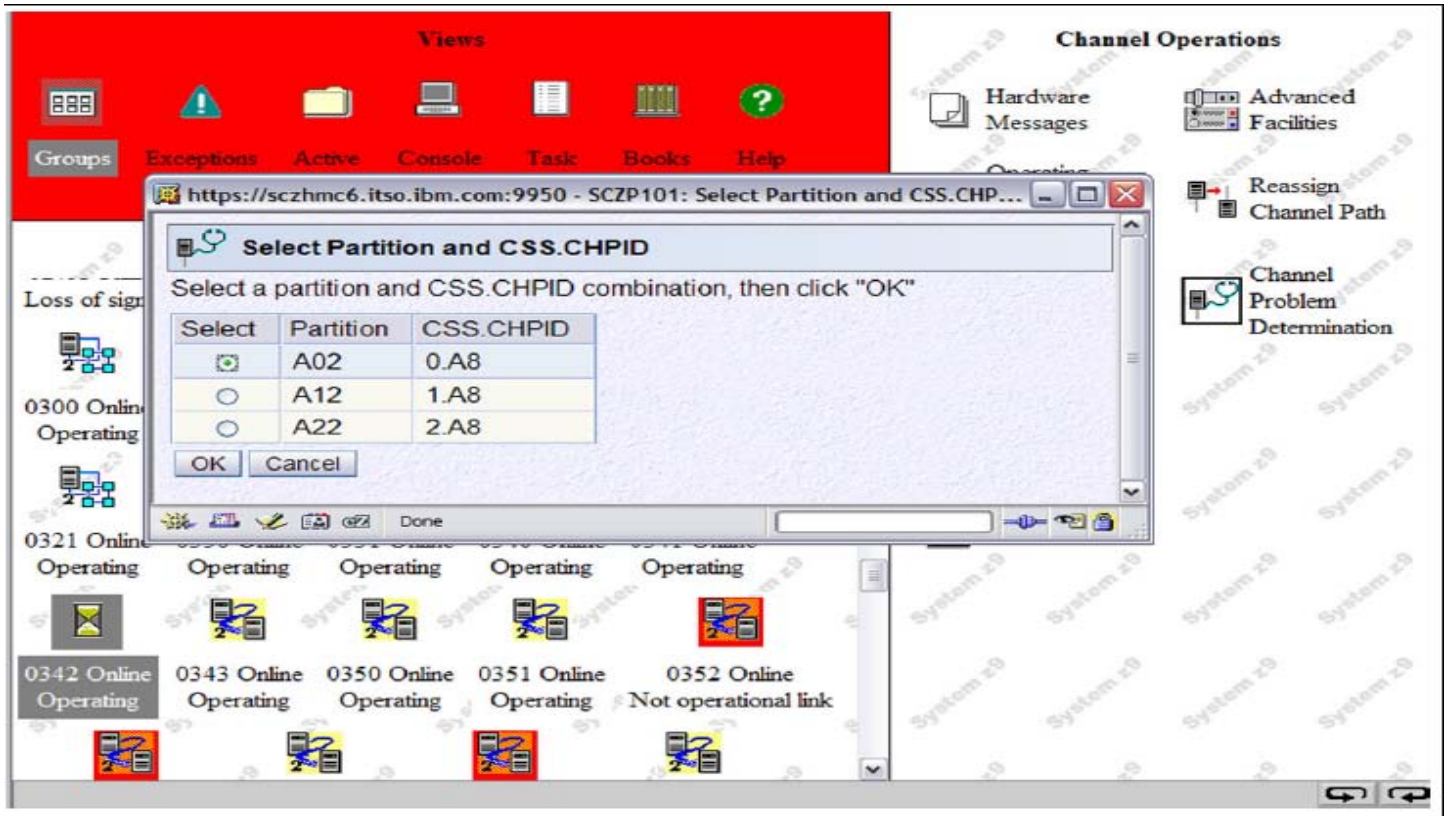
Transfer via FTP | Cancel | Help | Show all | Show NPIV=On

- Each device number in an LPAR is assigned a unique NPIV WWPN.
- Click **Transfer via FTP** to copy a text version of this menu to an FTP server.

Finding the permanent WWPN for the FCP CHPID

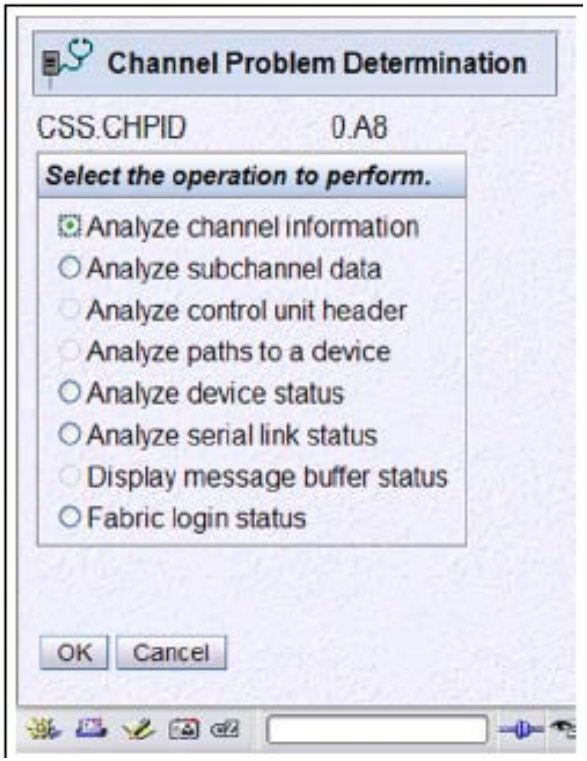
- The permanent WWPN is needed for LUN masking and zoning.
- We will find it by clicking **Channel Problem Determination** from the **Channel Operations** menu in the SE.

Finding the permanent WWPN for the FCP CHPID(2)



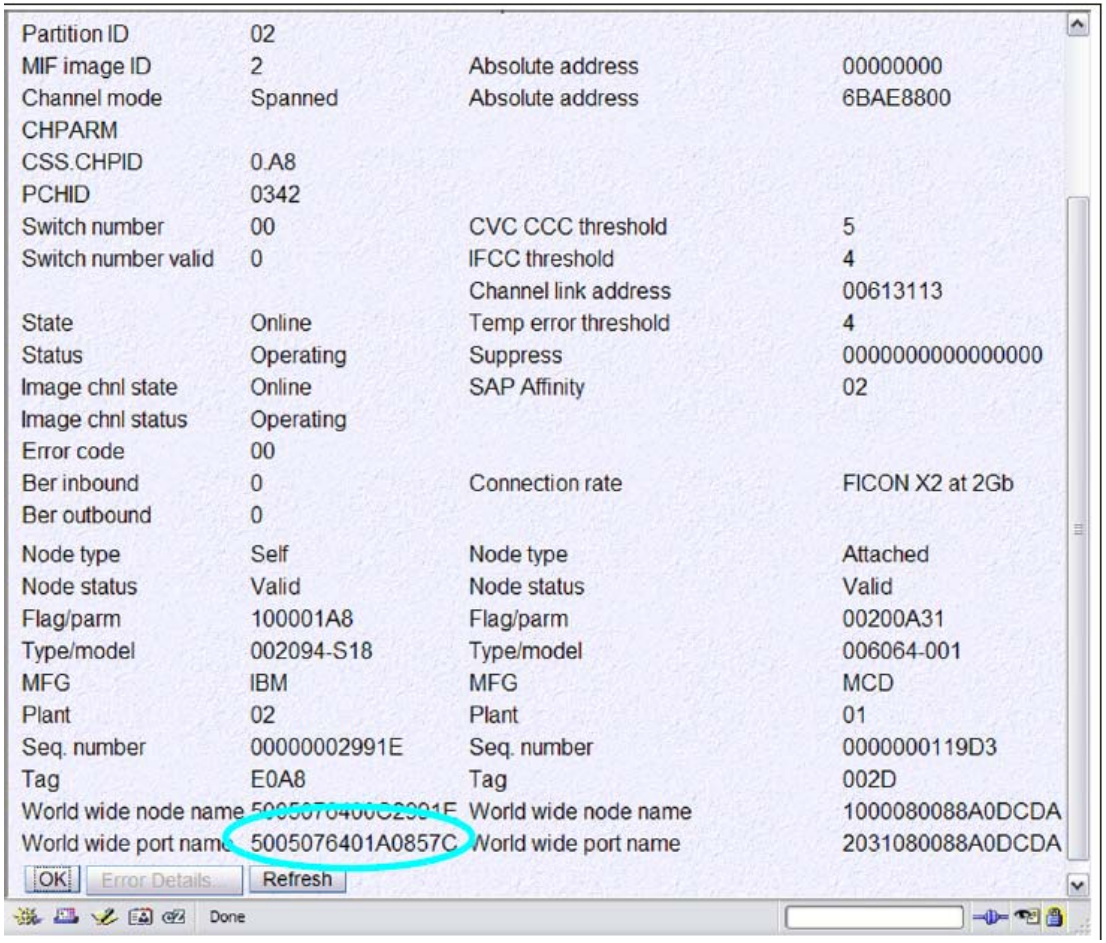
- Select the desired LPAR
- Click **OK** to open the **Channel Problem Determination** menu

Finding the permanent WWPN for the FCP CHPID(3)



- Select the **Analyze Channel Information** option
- Click **OK** to open the **Analyze Channel Information** menu shown on the next slide

Finding the permanent WWPN for the FCP CHPID(4)



Partition ID	02		
MIF image ID	2	Absolute address	00000000
Channel mode	Spanned	Absolute address	6BAE8800
CHPARM			
CSS.CHPID	0.A8		
PCHID	0342		
Switch number	00	CVC CCC threshold	5
Switch number valid	0	IFCC threshold	4
		Channel link address	00613113
State	Online	Temp error threshold	4
Status	Operating	Suppress	0000000000000000
Image chnl state	Online	SAP Affinity	02
Image chnl status	Operating		
Error code	00		
Ber inbound	0	Connection rate	FICON X2 at 2Gb
Ber outbound	0		
Node type	Self	Node type	Attached
Node status	Valid	Node status	Valid
Flag/parm	100001A8	Flag/parm	00200A31
Type/model	002094-S18	Type/model	006064-001
MFG	IBM	MFG	MCD
Plant	02	Plant	01
Seq. number	00000002991E	Seq. number	0000000119D3
Tag	E0A8	Tag	002D
World wide node name	[illegible]	World wide node name	1000080088A0DCDA
World wide port name	5005076401A0857C	World wide port name	2031080088A0DCDA

OK | Error Details... | Refresh

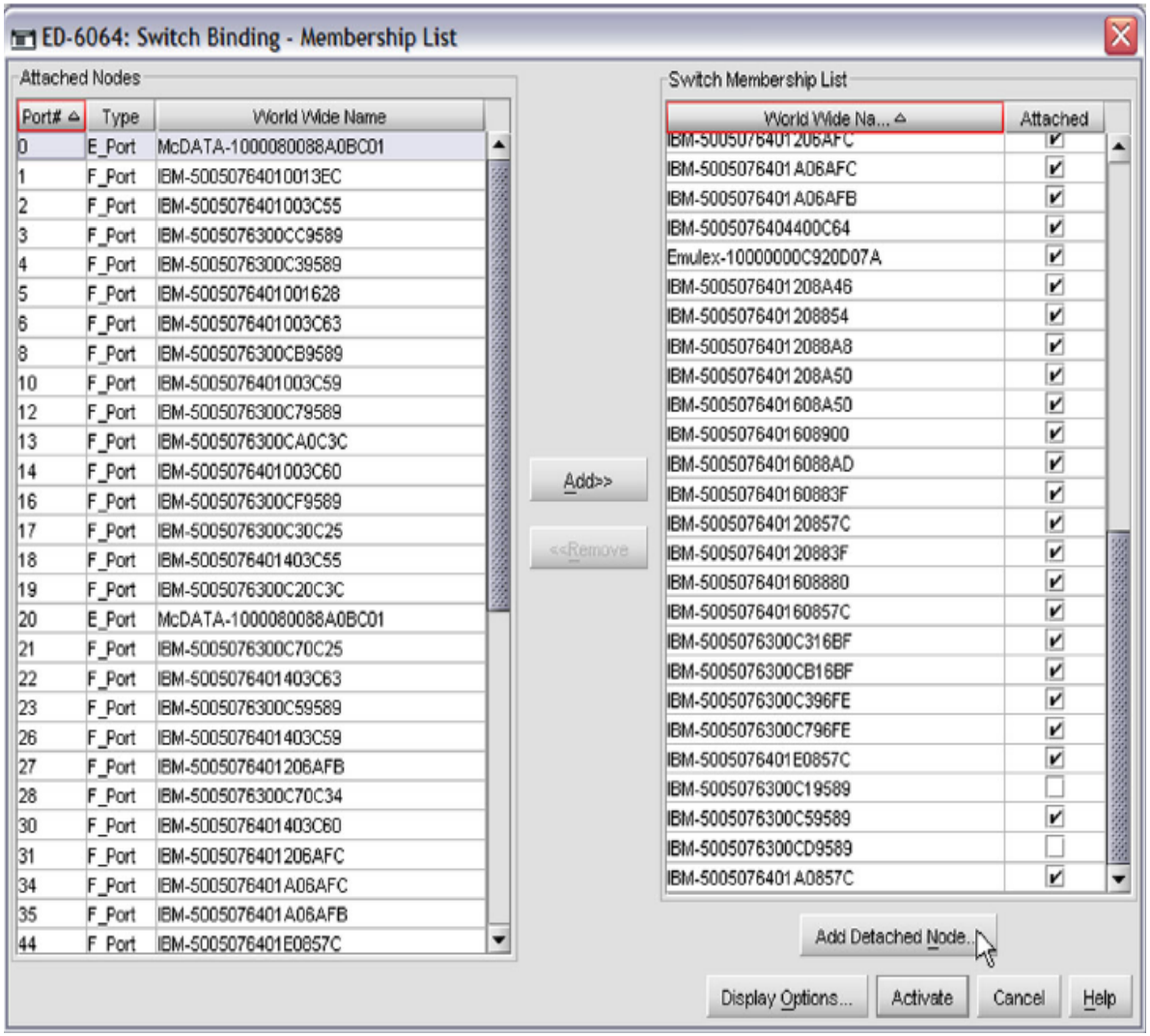
Done

Record the default WWPN (circled above in the screen shot)

Configuring Fabric Security

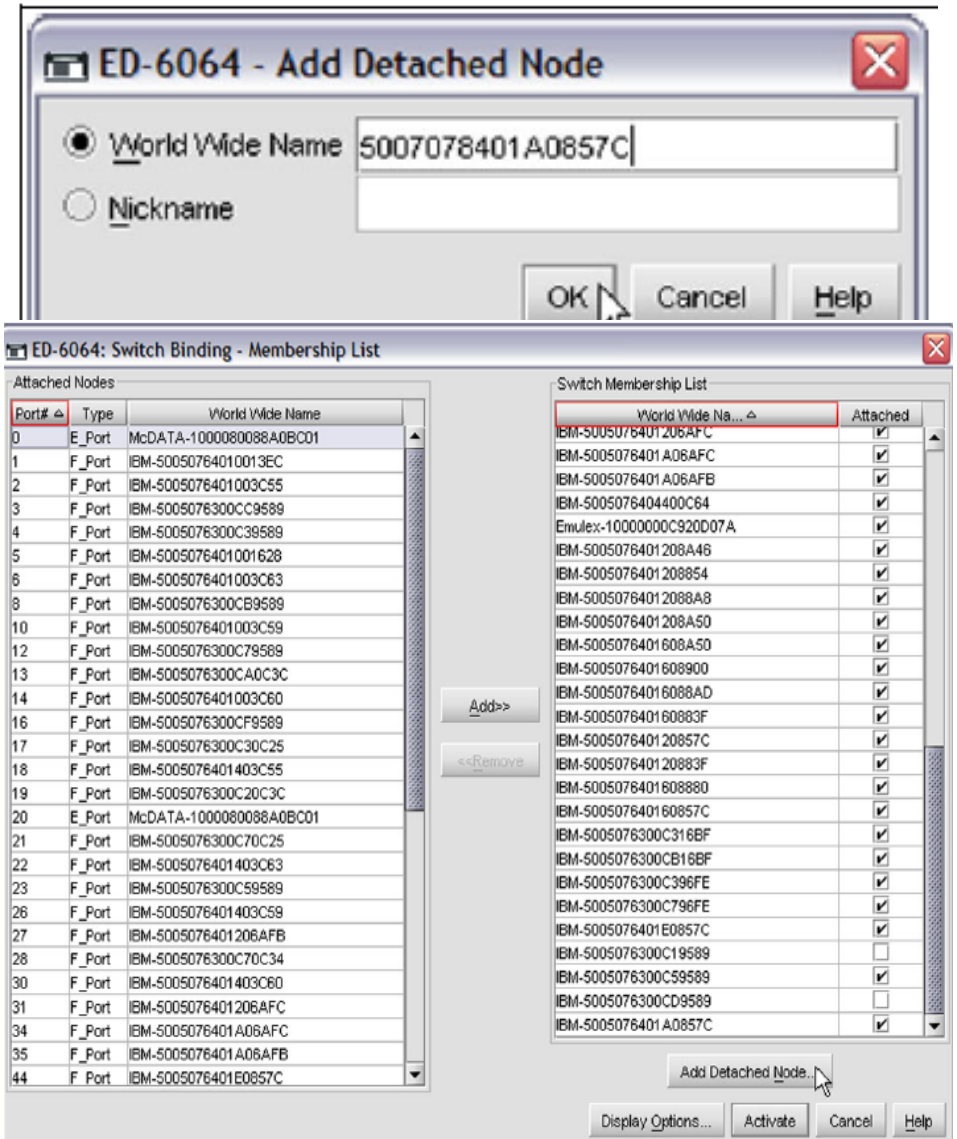
- The next step in the process is to define the NPIV and permanent WWPNs in the SAN fabric zoning and to the LUN masking on the DASD array.
- This configuration will be specific to the switch and array vendor/model.
  - Consult your vendor/model documentation for details on your specific requirements and steps.
  - *Linux on zSeries: Fibre Channel Protocol Implementation Guide* (SG24-6344) is an IBM Redbook that has a great general discussion on the subject of zoning and LUN masking.

Configuring Fabric Security-fabric zoning on the switching device



- **Switch binding membership list**
- Select **Add Detached Node**

Configuring Fabric Security-fabric zoning on the switching device(2)



- In the **Add Detached Node** menu/screen that opens, add the WWPN to the named zone.
- Click **OK**
- On the **Switch Binding Membership List** click **Activate** to commit the changes made.

Configuring Fabric Security-LUN Masking

- LUN masking grants a WWPN or group of WWPNs access to a specific LUN.
- This is configured on the disk array management tool(s).
- Perform this task following configuration of zoning
- Wide variety of models, each is somewhat different.
- Consult vendor documentation

Configuring the Linux server

- We are now ready to configure the Linux server to use the NPIV WWPN.
- Each device on the FCP CHPID is assigned a unique WWPN.
- The WWPN used by a Linux server is determined by the FCP device of the server.
- 3 steps:

Configuring the Linux Server Step 1

- When running under z/VM, the FCP device must be attached to the virtual machine of the Linux guest.
- To attach dynamically, use the **CP Attach** command.

```
Q CHPID A8
Path A8 online to devices B400 B401 B402 B403 B404 B405 B406 B407
Path A8 online to devices B408 B409 B40A B40B B40C B40D B40E B40F
Path A8 online to devices B410 B411 B412 B413 B414 B415 B416 B417
Path A8 online to devices B418 B419 B41A B41B B41C B41D B41E B41F
Path A8 online to devices B4FC B4FD
Ready; T=0.01/0.01 14:16:27
ATTACH B401 TO NPIV1
FCP B401 ATTACHED TO NPIV1 B401
Ready; T=0.01/0.06 14:16:46
ATTACH B402 TO NPIV2
FCP B402 ATTACHED TO NPIV2 B402
Ready; T=0.01/0.06 14:16:53
Q FCP
FCP  B401 ATTACHED TO NPIV1    B401 CHPID A8
FCP  B402 ATTACHED TO NPIV2    B402 CHPID A8
Ready; T=0.01/0.01 14:16:57
```

Configuring the Linux Server Step 1 (2)

- If the NPIV WWPN is unable to successfully login to the fabric, its WWPN is assigned the value 0x0000000000000000. Check zoning on the switching device.
- If an **“Out of resource in fabric”** message is shown, the reason might be that the **NPIV login limit** was set too small on the switch port.

```
FCP B401 ATTACHED TO NPIV1 B401
zfcp: adapter 0.0.b401: operational again
Feb 22 14:22:51 npiv1 kernel: crw_info : CRW reports slct=0, oflw=0, chn=0, rsc=
3, anc=1, erc=4, rsid=12
Feb 22 14:22:51 npiv1 kernel: zfcp: adapter 0.0.b401: operational again
zfcp: The adapter 0.0.b401 reported the following characteristics:
WWNN 0x5005076400c2991e, WWPN 0xc05076ffcf000004, S_ID 0x00613128,
adapter version 0x3, LIC version 0x600, FC link speed 2 Gb/s
zfcp: Switched fabric fibrechannel network detected at adapter 0.0.b401.
Feb 22 14:22:52 npiv1 kernel: zfcp: The adapter 0.0.b401 reported the following
characteristics:
Feb 22 14:22:52 npiv1 kernel: WWNN 0x5005076400c2991e, WWPN 0xc05076ffcf000004,
S_ID 0x00613128,
Feb 22 14:22:52 npiv1 kernel: adapter version 0x3, LIC version 0x600, FC link sp
eed 2 Gb/s
Feb 22 14:22:52 npiv1 kernel: zfcp: Switched fabric fibrechannel network detecte
d at adapter 0.0.b401.
```

Configuring the Linux Server Step 2

- Setup/Add the FCP disk to the Linux server using **YaST**

Configuring the Linux Server Step 3

- Verify the NPIV WWPN login and operation

Summary/questions

- NPIV background/introduction
- FCP Channels on the mainframe
- NPIV Planning
- NPIV implementation and configuration

- Feel free to contact me:
  - sguender@brocade.com

Key References

- S. Guendert. *Understanding the Performance and Management Implications of FICON/FCP Protocol Intermix Mode (PIM).* Proceedings of the 2008 CMG. Dec 2008.
- I. Adlung, G. Banzhaf et al. "FCP For the IBM eServer zSeries Systems: Access To Distributed Storage". *IBM Journal of Research and Development.* 46 No.4/5, 487-502 (2002).
- J. Srikrishnan, S. Amann, et al. "Sharing FCP Adapters Through Virtualization." *IBM Journal of Research and Development.* 51 No. ½, 103-117 (2007).
- J. Entwistle. "IBM System z10 FICON Express8 FCP Channel Performance Report". IBM 2009.
- American National Standards Institute. "Information Technology-Fibre Channel Framing and Signaling (FC-FS)." ANSI INCITS 373-2003.
- G. Bahnzhaf, R. Friedrich, et al. "Host Based Access Control for zSeries FCP Channels", *z/Journal* 3 No.4, 99-103 (2005)
- S. Guendert. "The IBM System z9, FICON/FCP Intermix, and Node Port ID Virtualization (NPIV). *NASPA Technical Support.* July 2006, 13-16.
- G. Schulz. *Resilient Storage Networks.* pp78-83. Elsevier Digital Press. Burlington, MA 2004.
- S. Kipp, H. Johnson, and S. Guendert. "Consolidation Drives Virtualization in Storage Networks". *z/Journal.* December 2006, 40-44.
- S. Kipp. H. Johnson, and S. Guendert. "New Virtualization Techniques in Storage Networking: Fibre Channel Improves Utilization and Scalability." *z/Journal*, February 2007, 40-46
- **Wong, Frederick. "Introducing N_Port Identifier Virtualization for IBM System z9". IBM 2005**

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

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

SHARE
Technology · Connections · Results

THANK YOU

SHARE
in Orlando
2011