



International Technical Support Organization

## IBM System z9 I/O and Networking Enhancements

[ibm.com/redbooks](http://ibm.com/redbooks)



IBM System z9 Workshop

© 2005 IBM Corporation

[ibm.com/redbooks](http://ibm.com/redbooks)

International Technical Support Organization



## Notices

This information was developed for products and services offered in the U.S.A.

Note to U.S. Government Users Restricted Rights — Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to: IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

### COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.



IBM System z9 Workshop

© 2005 IBM Corporation

2

## Agenda



## z9-109 I/O & Network Enhancements

### I/O Enhancements

- Up to 16 x 2.7GB STIs per book\* with Redundant I/O Interconnect
- Up to 80% increase in I/O bandwidth per book
- Multiple Subchannel Sets (MSS)
  - Second subchannel set definable per LCSS
  - Up to 63.75K Subchannels for Set-0
- Modified Indirect-Data-Address Word (MIDAW) facility
- FICON Express2 – 64 Open Exchanges
- Up to 28 FICON Express2/Express, Crypto Express2, OSA-Express2/Express features per I/O cage
  - Up to 336 FICON Express2 channels per machine (Up to 256 for the Model S08)
- Storage Area Network (SAN) enhancements:
  - N\_Port ID Virtualization for FCP
  - Program Directed re-IPL for FCP
  - Remote Node ID Reporting for FICON FC



### Networking enhancements

- HiperSockets IPv6
- OSA-Express2 1000BASE-T Ethernet
- OSA-Express2 OSN (OSA for NCP support)
- GARP VLAN management (GRVP)

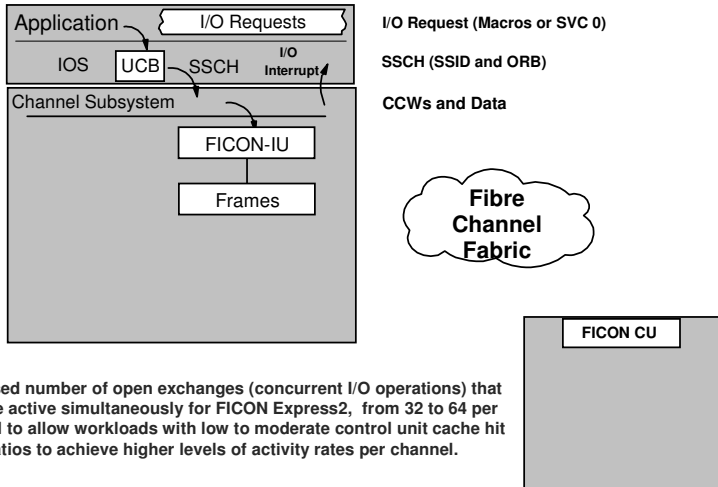
## I/O and Networking Enhancements



## z9-109 Modified Indirect-Data-Address Word (MIDAW) facility

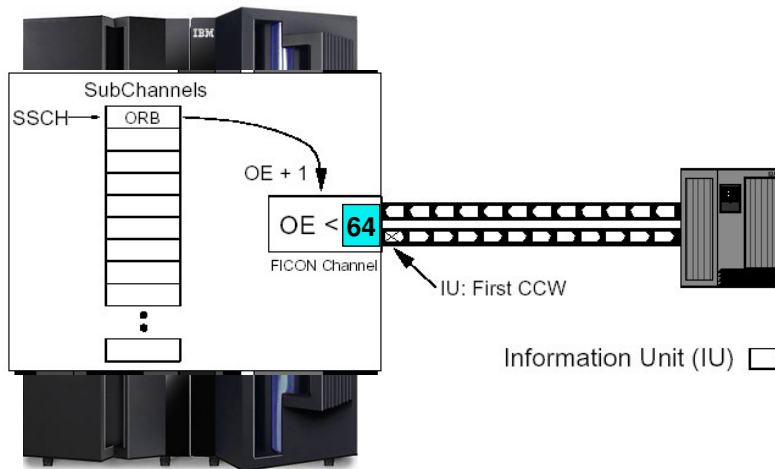
- **Modified-Indirect-Data-Address Word facility:**
  - The MIDAW facility is a new CCW-indirect-data-address word facility being added to z/Architecture to coexist with the current IDAW facility
  - Both MIDAW and IDAW facilities offer, for FICON and ESCON channels, alternatives to using CCW data chaining in channel programs
  - Both facilities are designed to **reduce channel, fabric, and control unit overhead by reducing the number of CCWs and frames processed**
  - The MIDAW facility is usable in certain cases where the IDAW facility is not because it does not have IDAW boundary and data length restrictions
- **Designed to:**
  - **Be compatible with existing disk control units**
  - **Increase the number of I/O operations per second** that can be processed and thus **move more data per second**, especially on faster FICON channels
- **Applications that may benefit include:**
  - DB2, VSAM, Partitioned Data Set Extended (PDSF), Hierarchical File System (HFS), z/OS File System (ZFS), and other datasets exploiting striping and compression

# FICON 64 Open Exchanges



Increased number of open exchanges (concurrent I/O operations) that may be active simultaneously for FICON Express2, from 32 to 64 per channel to allow workloads with low to moderate control unit cache hit ratios to achieve higher levels of activity rates per channel.

# FICON 64 Open Exchanges....



## z9-109 Availability Enhancements

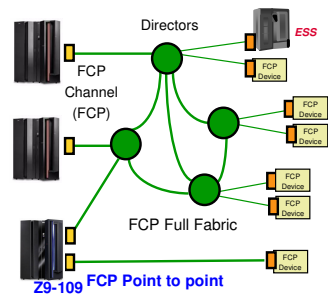
### ■ Request Node Identification Data (RNID)

- Designed to assist with the isolation of FICON cabling-detected errors:
  - In a fiber optic environment, with extended distances, resolution of fiber optic cabling problems can be a challenge. To help facilitate resolution, the operating system can now request the RNID data for each device or control unit attached to native FICON channels and display the RNID data using an operator command.
- This RNID data is now:
  - Formatted and displayed on the SE "Analyze Control Unit Header" IOPD panel.
  - Provided to z/OS so it can display it on the "D M=DEV" (Display Device Matrix) command to help debug configuration/cabling problems.

### ■ FICON Link Incident Reporting

- Designed to allow an operating system image to register for link incident reports improving the ability to capture data for link error analysis without operator intervention. z/OS will display this information on consoles and will also save it in the system log and in LOGREC.

## z9-109 FCP Attachment Options for z/VM and Linux



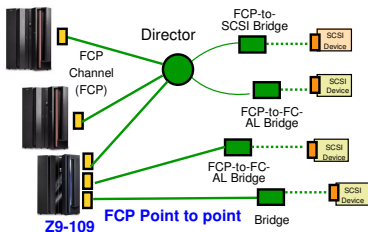
### ■ FCP Full Fabric Connectivity

- Homogeneous, single vendor fabric
- Fibre channel directors, switches
- **NPIV support**



### ■ FCP point to point – Designed to support all FICON features supported on z9-109

- **Direct attachment to FCP CU port**
- NPIV support not applicable



### ■ FCP switched to SCSI Bridge

- FCP-to-SCSI Bridges
- FCP-to-FC-AL bridge



### ■ FCP point to point – Designed to support all FICON features supported on z9-109

- FCP-to-SCSI bridge
- FCP-to-FC-AL bridge

Supported devices: [www.ibm.com/servers/eserver/zseries/connectivity/#fcp](http://www.ibm.com/servers/eserver/zseries/connectivity/#fcp)

## FCP Channel N\_Port ID Virtualization (NPIV)

- Multiple operating system images can concurrently access the same or different SAN attached devices (LUNs) via a single, shared FCP Channel.
- N\_Port ID Virtualization (NPIV) provides each OS instance a set of virtual FCP Channel ports.
  - Each virtual FCP Channel port has its own unique SAN identity (WWPN)
  - A virtual FCP Channel port behaves like a physical FCP Channel port
- Each virtual FCP Channel port identifier (WWPN) can be used in device-level access control in storage controllers (LUN masking), as well as switch-level zoning.
- NPIV allows FCP channel sharing among operating system instances, as has been previously available for ESCON and FICON channels.
- Advantages:
  - Improved channel utilization
  - Less hardware required
  - Reduced complexity of physical I/O connectivity.

## z9-109 N\_Port ID Virtualization (NPIV)

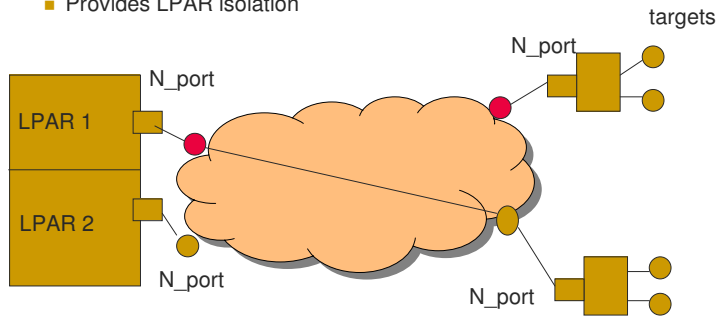


- NPIV provides the capability to define an FCP channel as multiple virtual FCP channels, each with its own unique virtual Fibre Channel port name and FC N\_Port ID.
- By assigning different virtual port names to different operating system images (running in logical partitions or as z/VM guests in virtual machines), operating systems can use the virtual FCP channels as if they were using dedicated physical FCP Channels
- The I/O transactions of each operating system image are separately identified, managed, and transmitted, and are processed by the fiber channel and SAN infrastructure as if each image had its own unique physical N\_Port. Each virtual FCP Channel port identifier (WWPN) can be used in device-level access control (LUN masking) in storage controllers and in switch-level zoning.
- Designed to permit multiple operating system images concurrently to access SAN attached devices (LUNs) via a single, shared FCP Channel.
  - Designed to work with all FICON features supported on z9-109
  - May reduce number of FCP channels needed
  - May simplify infrastructure by reducing ports and cabling
  - May improve channel utilization
- NPIV is an industry standard solution which supersedes FCP LUN Access Control
  - Requires FIBER Channel Switch support
  - Requires z/VM – LINUX support

## Zoning without FCP Channel Virtualization

- Zoning controls access to N\_Ports

- Provides LPAR isolation

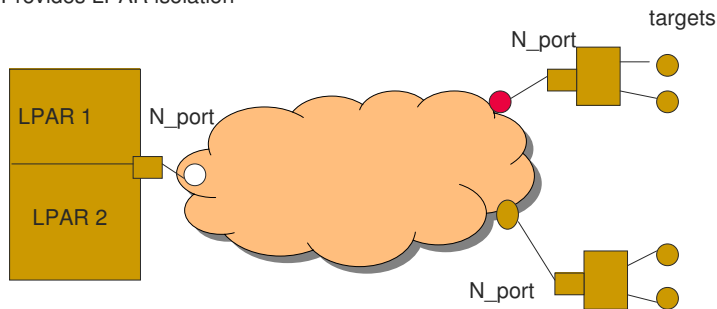


**Without NPIV**  
 -need separate N\_Ports  
 -policed by fabric

## Zoning with FCP Channel Virtualization

- Zoning controls access to N\_Ports

- Provides LPAR isolation



**With FCP Channel Virtualization**  
 -don't need separate N\_Ports  
 -policed by fabric

## VM Support of FCP Channel Virtualization

Although theoretically NPIV is "transparent" to operating systems, there are 3 things VM is doing to provide support:

1. **QUERY** command to obtain the default worldwide port name (WWPN) from the machine for a given FCP subchannel. V5.2 only
2. **Virtualization** of the underlying **CHSC command** that obtains the WWPN, such that guests can retrieve the WWPN. V5.2 (allows a Linux guest to do a query)
3. **VM System use of SCSI disks:** Change to **SCSI driver** such that, when a new N\_Port ID is requested from the switch and there are **no more currently available**, the driver will fail the request instead of retrying for an hour (and thus appearing to be "hung").

## LUN Masking without FCP Channel Virtualization and with FCP Channel Virtualization

- **LUN masking controls access to targets**
  - Without FCP Channel Virtualization
    - Server polices LUN masking as each LPAR needs its own mask
    - Intercepts LUN commands
  - With FCP Channel Virtualization
    - Server not involved
    - Each LPAR has an N\_Port ID and the target polices masking



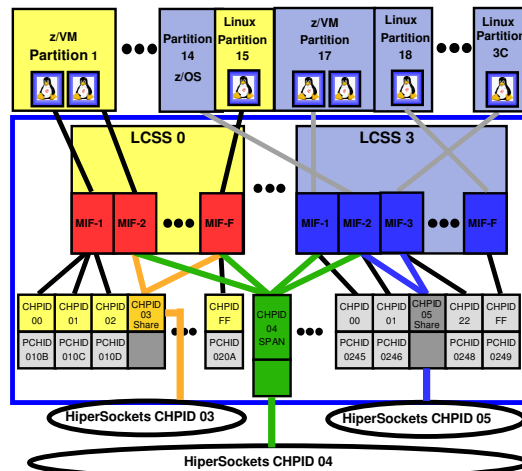
## Program-Directed Re-IPL

- **Program-Directed Re-IPL enables an operating system running natively in a logical partition to Re-IPL itself.**
- **This function is supported both for**
  - SCSI and ECKD devices
- **In order to initiate a Re-IPL, an operating system can determine**
  - **How** it was loaded (i.e., via Channel Control Words (CCWs) or via SCSI-type IPL, and
  - **From where** it was loaded in case of SCSI-type IPL (World Wide Port Name (WWPN) and Logical Unit Number (LUN) of the load device).
- **Using these parameters, the program can subsequently call a firmware function (SCLP call) to request that it be reloaded from the same load device, using the same load parameters.**
- **Software Support**
  - Linux on zSeries support for Program-Directed Re-IPL is expected to become available from IBM's Linux distribution partners at a later time.

**Note: z/VM already supports an interface that allows a program running as a guest under z/VM to re-IPL itself.**

## HiperSockets IPv6

- **Internet Protocol Version 6 (IPv6)**
- **More unique IP addresses**
  - Expands the IP address space
    - From 32 bits to 128 bits
  - Follow-on to IPv4
- **Minimum software:**
  - z/OS V1.7
  - z/VM V5.2 with PTFs in 1H2006



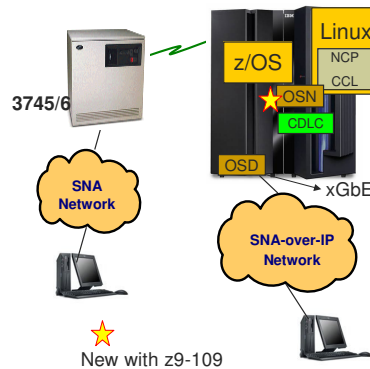
**Very High Speed Interconnection between programs running z/OS, z/VM or Linux**

## Communication Controller for Linux on zSystem

**New with z9-109:** OSA for NCP supports 3745/6 channel protocol for communication between VTAM® (or TPF) and the NCP running on the Communication Controller for Linux on zSeries

- Can ease customer migration by providing operational compatibility with existing 3745/6 environments
- Support for communication with SNA applications without anticipated changes to the operating system or application
- May help improve performance, scalability and usability of all CCL-based NCP environments via use of QDIO architecture

**Note:** z9-109 enhancements are planned to be exploited in a future CCL release



**Communications Controller for Linux on zSystem:** Helps remove customer dependency on IBM 3745/6 Communications Controller hardware for selected NCP environments.

CCL V1R1 available March 2005

## OSA-Express2 supporting CCL NCP



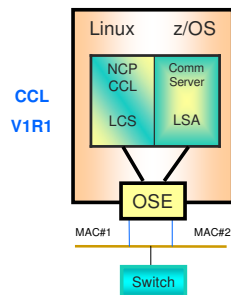
- Designed to help eliminate dependencies on hardware:
  - 3745/3746, ESCON, Token Ring
- OSA-Express2 (GbE, 1000BASE-T Ethernet)
  - New CHPID type – OSN
- Provides support for IBM Communication Controller for Linux on zSeries (CCL)
- The CCL image connects to OSA-Express2 OSN using extensions to the QDIO architecture. Linux support is in the QETH module.
- Allows TPF to exploit CCL
- Preserves mission critical SNA functions (such as SNI) and z/OS application workloads which depend upon these functions
- zSeries operating systems remain unchanged
- Helps collapse SNA inside server while exploiting and leveraging IP

## OSA-Express2 OSN key design advantages:

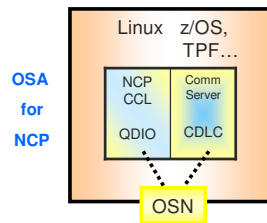
- Designed to appear to the operating systems as an ESCON channel connected to a 374x device type which exploits existing CDLC (Channel Data Link Control) protocols
- Allows system administrators to configure, manage, and operate their CCL NCPs just like their real 374x NCPs
- Does not require external hardware (cables or switches)
- Enables key NCP channel-related functions like loading and dumping of the NCP
- Allows multiple CCL images to communicate with multiple host OS images (up to 180 connections per CHPID type OSN)
- Can span multiple Logical Channel Subsystems

## OSA-Express2 OSN (Open System Adapter for NCP)

- Supporting Channel Data Link Control (CDLC)
  - Channel protocol for communication from a host operating system to 374x
  - Communications Controller for Linux on zSeries (CCL)
  - Emulates 374x control unit



- ✓ CCL
- ✓ LAN connectivity – external communications
- ✓ Traditional connectivity
- ✓ OSA and non-QDIO (CHPID type OSE)
- ✓ Traffic flows on the LAN



- ✓ Traffic flows LPAR-to-LPAR
- ✓ OSA-Express2 (1000BASE-T, GbE) is the bridge
  - Internal communications – CCL to CDLC
  - SNA PU type 5 and PU type 2.1
- ✓ QDIO extensions for CHPID type OSN
- ✓ Unlike-to-unlike with "glue logic"

## VLAN management enhancement – support of GVRP

### IEEE 802.1

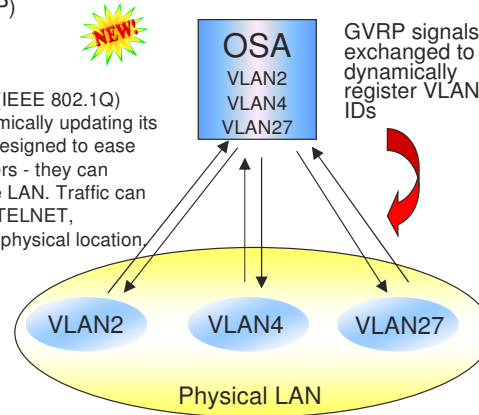
- o GARP VLAN Registration Protocol (GVRP)

- Propagate VLAN information

- o OSA communicates with GVRP-capable (IEEE 802.1Q) switches, registering its VLANs and dynamically updating its table as the VLANs change. VLANs are designed to ease the administration of logical groups of users - they can communicate as if they were on the same LAN. Traffic can be organized by traffic patterns (such as TELNET, FTP, banking transactions) rather than by physical location.

- Pre-requisites

- o OSA-Express2
    - o QDIO mode (CHPID type OSD)
    - o z/OS, z/VM, Linux on zSeries



Simplified network administration and management

## OSA technology refresh

Feature	Feature Name	z900	z800	z990	z890	Z9-109	CHPIDs
5201	OSA-2 Token Ring	X	N / A	N / A	N / A	N / A	OSA
5202	OSA-2 FDDI	X	N / A	N / A	N / A	N / A	OSA
2362	OSA-E 155 ATM SM	X	X	RPQ	N / A	N / A	OSD, OSE
2363	OSA-E 155 ATM MM	X	X	RPQ	N / A	N / A	OSD, OSE
2364	OSA-E GbE LX	X	X	C	C	C	OSD
2365	OSA-E GbE SX	X	X	C	C	C	OSD
2366	OSA-E Fast Ethernet	X	X	C	C	C	OSD, OSE
2367	OSA-E Token Ring	X	X	X	X	N / A	OSD, OSE
1364	OSA-E GbE LX	09/04	09/04	06/03	05/04	C	OSD
1365	OSA-E GbE SX	09/04	09/04	06/03	05/04	C	OSD
1366	OSA-E 1000BASE-T Ethernet	N / A	N / A	06/03	05/04	C	OSC, OSD, OSE
3364	OSA-E2 GbE LX	N / A	N / A	01/05	01/05	X	OSD, OSN
3365	OSA-E2 GbE SX	N / A	N / A	01/05	01/05	X	OSD, OSN
3366	OSA-E2 1000BASE-T Ethernet	N / A	N / A	N / A	N / A	X	OSC, OSD, OSE, OSN
3368	OSA-E2 10 GbE LR	N / A	N / A	01/05	01/05	X	OSD

LX = Long wavelength transceiver, SX = Short wavelength transceiver, LR - Long Reach transceiver  
 C = Carry forward on an upgrade from z900 to z990, z800 to z890. Replacements available.