The IBM System z9, FICON/FCP Intermix and Node Port ID Virtualization (NPIV)

By Stephen R. Guendert

IT HAS BEEN POSSIBLE TO COMBINE MAINFRAME FICON STORAGE NETWORKS and open systems fibre channel Storage Area Networks (SAN) onto a common storage network since early 2003. This is commonly known as FICON/FCP Intermix and/or protocol intermix mode (PIM). IBM has also been one of the industry's biggest proponents of Linux, particularly Linux on the mainframe. It has become possible to consolidate many stand alone open systems servers onto a mainframe running multiple Linux images on a single footprint. In July 2005 IBM announced System z9, and along with it Node Port ID Virtualization, more commonly known as NPIV. This article will briefly discuss the concepts behind FICON/FCP Intermix, FCP channels on the mainframe, and finally, NPIV.

INTERMIX BASIC CONCEPTS

There are several key concepts to understand when discussing FICON/FCP Intermix. First, the open systems SAN Fibre Channel Protocol (FCP-SCSI-3) and FICON (FC-SB-2/3) are merely different Upper Layer Protocols (ULP) in the overall Fibre Channel Standard. (see figure 1). As you can see, the difference between open systems SAN and FICON is at the FC-4 (data payload packet). So, essentially, your open systems SAN directors/switches and mainframe FICON directors/switches are identical hardware. FICON directors/switches will typically be bought with a software feature known as Control Unit Port, or CUP. CUP allows the operators to manage the FICON directors in-band using the same management tools they are used to with ESCON.

WHY INTERMIX?

Let's face it. Open systems and mainframe environments have not traditionally shared system and/or staff resources. They have completely different cultures, almost to the point of sometimes resembling the current two-party system in the United States political arena. So, why would anyone consider moving to an intermix environment? There are 3 primary reasons.

First, less than 40% of ESCON end users have migrated to FICON since native FICON became available in 2001. Many of these same end users already have a well established open systems SAN(s) that has (have) been delivering reliable, high performance connectivity for several years. These organizations, particularly the larger ones, will often have open ports available on some of their directors/switches. By looking at FICON/FCP intermix as an option, they might consider a short-term allocation of some of their unused SAN ports to test
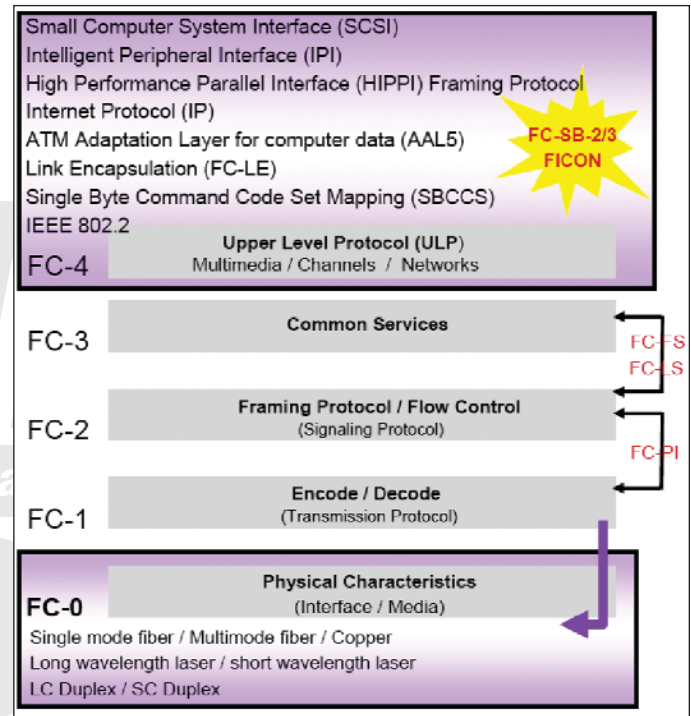

Figure 1: Fibre Channel Standard

FICON, or to complete an initial FICON deployment once testing is complete. This could be beneficial to an organization looking to delay the purchase of a dedicated FICON infrastructure until the connectivity requirements are sufficient enough to justify purchasing separate FICON directors.

Secondly, the consolidation in switching infrastructure made possible by FICON/FCP intermix implementation can reduce total cost of ownership. With some exceptions, the latest generation of high port count directors from all vendors are designed with non-blocking architectures and full throughput to all ports, even when the director port count is fully populated. Many large IT organizations that currently have separate open systems and FICON storage networks are large enough to fully utilize their director ports, even when running segregated storage networks. However, there are many organizations that segregate open systems and mainframe environments that are not as large. Oftentimes these organizations are not fully utilizing their directors and could realize a significant cost savings by consolidating these separate, under-utilized directors onto a common storage network. This most likely would result in fewer "serial numbers" on the

floor, leading to lower maintenance, electrical, cooling, and other operational costs. Also, it allows an organization to have one common cable plant/fiber infrastructure rather than separate ones for mainframe and open systems. This would likely lead to lower physical infrastructure costs, as well as reduced management costs for the cable plant/infrastructure.

Third, intermix makes perfectly good sense for specialized environments requiring both flexibility and resource sharing. For example, quality assurance, test, development and disaster recovery environments/data centers would all fit this description. Already established FICON implementations will consider intermix with the likely evolution of storage subsystem based DASD mirroring applications from ESCON via channel extension technology to FCP for increased performance between sites.

Finally, if you are running, or are considering running Linux on your mainframe you will essentially be running an intermix environment because channels for supporting Linux images will need to be defined as FCP (even though you'll be using FICON/FICON Express/FICON Express 2 channel cards on the mainframe, the channels need to be defined as FCP).

FCP CHANNELS ON THE MAINFRAME

The IBM Systems z9 FICON Express2 channel cards provide support for Fibre Channel and Small Computer System Interface (SCSI) devices in Linux environments. The channel card microcode will be for FCP to support the SCSI data payload of Linux. The z9 FCP support allows Linux running on the host to access industry-standard SCSI devices. For disk applications, these FCP storage devices utilize Fixed Block (512 byte) sectors rather than the Extended Count Key Data (ECKD) format. The SCSI and/or FCP controllers and devices can be accessed by Linux on the System z9 as long as the appropriate I/O driver support is there. There are currently two supported methods of running Linux on the System z9. In can be run as a guest operating system under z/VM (version 4 release 3 and later releases). It can also be run natively in a logical partition.

The International Committee of Information Technology Standards (INCITS) developed the Fibre Channel Protocol Standard (FC-FCP) and subsequently it has been published as an ANSI standard. The z9

(and zSeries) FCP I/O architecture fully conforms to the Fibre Channel standards specified by INCITS. To review our earlier discussion, FCP is an upper layer fibre channel mapping of SCSI on a common stack of Fibre Channel physical and logical communication layers. FC-FCP and SCSI are supported by a wide range of controllers and devices which complement the storage attachment capability through FICON and ESCON channels.

The Queued Direct Input/Output (QDIO) architecture is used by FICON channels in FCP mode to communicate with the operating system. This architecture is derived from the same QDIO architecture defined for Hipersockets communications and for OSA Express. Rather than using control devices, FCP channels use/define data devices that represent QDIO queue pairs. This consists of a request queue and a response queue. Each of these queue pairs represents a communication path between the FCP channel and the operating system. An operating system can send FCP requests to the FCP channel via the request queue, and the response queue can be used by the FCP channel for passing completion indications and unsolicited status indications back to the operating system.

The FCP channel type still needs to be defined using HCD/IOCP, as do the QDIO data devices. However, there is no requirement for defining the fibre channel storage devices/controllers, or directors and switches. All of these devices will be configured on an operating system level using the parameters outlined in the industry standard fibre channel architecture. They will be addressed using World Wide Names (WWNs), Fibre Channel Identifiers (IDs), and Logical Unit Numbers (LUNs). After the addresses are configured on the operating system; they are passed to the FCP channel along with the corresponding fibre channel I/O request via a queue.

NODE PORT ID VIRTUALIZATION (NPIV)

The FCP industry standard architecture does not exploit the security and data access control functions of Multiple Image Facility (MIF). The z990 and z890 have a feature known as FCP LUN access control. This has been carried forward to the Systems z9. LUN access control provides host-based control of access to storage controllers and their devices as identified by LUNs. LUN
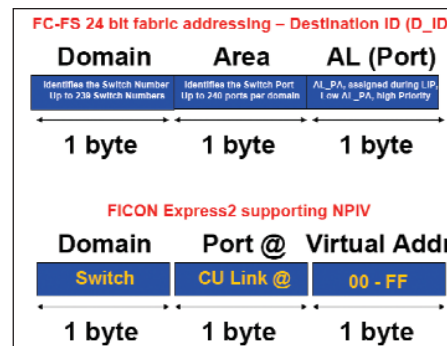

Figure 2: Fibre Channel addressing

access control also allows read-only sharing of FCP SCSI devices among multiple operating system images. When a host channel is shared among multiple operating system images, the access control mechanism is capable of providing for either none, or all images to have access to a particular logical unit (device) or storage controller. FCP LUN access control gives the end user the ability to define individual access rights to storage controller ports as well as devices for each operating system image. LUN access control can significantly reduce the number of FCP channels that are needed to provide controlled access to data on FCP SCSI devices. Without LUN access control, FCP channels prevent logical units from being opened by multiple Linux images at the same time. In other words, access is granted on a first-come, first-served basis. This prevents problems with concurrent access from Linux images that are sharing the same FCP channel (which is also sharing the same world-wide port name). In effect, this means that one Linux image can block other Linux images from accessing the data on one or more logical units.

If the FCP channel is used by multiple independent operating systems (under z/VM or in multiple LPARs), the SAN is not aware of this fact, and it cannot distinguish among the multiple independent users. The end user can partially avoid this by requiring separate FCP channels for each LPAR. However, this does not solve the problem with z/VM.

Another method to control access to devices is the implementation of Node Port (N_Port) ID Virtualization (NPIV). NPIV is unique to the System z9 and is a major improvement. NPIV allows each operating system that is sharing an FCP channel to be assigned a unique virtual world-wide port name (WWPN). The virtual WWPN can be used for both device-level access control in a storage controller (LUN masking), and

for switch-level access control on a fibre channel switch and/or director (zoning). What NPIV is doing is allowing a single physical FCP channel to be assigned multiple WWPNs and appear as multiple channels to the external SAN environment. These virtualized FC N_Port IDs allow a physical fibre channel port to appear as multiple, distinct ports, providing separate port identification, and security within the fabric for each operating system image. The I/O transactions of each operating system image are separately identified, managed, transmitted, and are processed the same as if each operating system image had its own unique physical N_port.

NPIV is based on a recent extension to the Fibre Channel standards that allows a host bus adapter (HBA) to perform multiple logins to the SAN fabric via a single physical port. The switch/director to which the FCP channel is directly connected (i.e. the "entry switch") must support multiple N_Port logins. No changes are required for the downstream switches, devices, or control units. The switch and not the FCP channel itself provides the multiple WWPNs used for the virtualization. Therefore, NPIV is not supported with FCP point-to-point attachments. NPIV requires switched FCP attachment.

NPIV is available with Linux on System z9 in a logical partition or as a guest of z/VM V 4.4, 5.1 and later for SCSI disks accessed via dedicated subchannels and for guest IPL. For guest use of NPIV, z/VM 4.4 and z/VM 5.1, and z/VM 5.2 provide support transparently; i.e., to PTF is required. z/VM 5.1 and later provides NPIV support for VM-system use of SCSI disks (including emulated-FBA minidisks for guests). z/VM 5.1 requires a PTF to properly handle the case of a fibre channel switch/director not being able to assign a new N_Port ID when one is requested (due to the switch's capacity being exceeded). z/VM 5.2 will provide support at a later date allowing VM users and VM guest operating systems to obtain the worldwide port name(s) (WWPNs) being used in a virtual machine. The QUERY command will be enhanced for VM users, and virtualization of a machine function will be enhanced for VM guests.

Figure 2 illustrates the fibre channel addressing for NPIV.

CONCLUSION

The enhancements made to the IBM System z9 have made FICON/FCP intermix

a more attractive, viable, and realistic option for connectivity in the world's largest data centers. NPIV technology finally makes it realistic to run open systems and z/OS on a mainframe connecting everything via FICON Express channels into a common storage network. End users need to evaluate the cost of doing so, compared with the costs of operating large Windows environments. Perhaps the mainframe really is not dead after all! ↺

NaSPA member Steve Guendert is CNT Technical Solutions Group's Storage Principal Consultant and is the Director of their FICON practice. He is a PMI certified Project Management Professional (PMP), and he has an MBA, and a M.S. in Management Information Systems. He is a well regarded FICON consultant and is one of the industry's experts on the subject. He lectures and consults regularly on FICON technology as well as business case justifications for migrating from ESCON to FICON. He is currently a Ph D Candidate and is writing his doctoral dissertation on Enterprise I/O Subsystems.