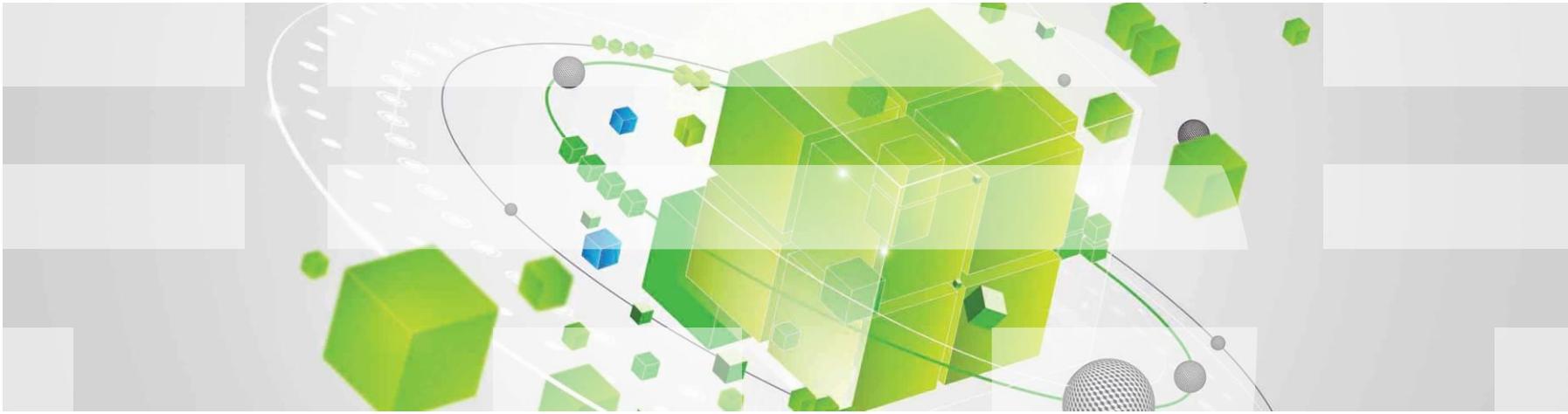
IBM z Systems z13™ and z13s™
SMC-D / ISM Introduction: z/OS Overview

Jerry Stevens (sjerry@us.ibm.com)

March 16, 2016



IBM z Systems™ z13 / z13s SMC-D and ISM Introduction: Topics

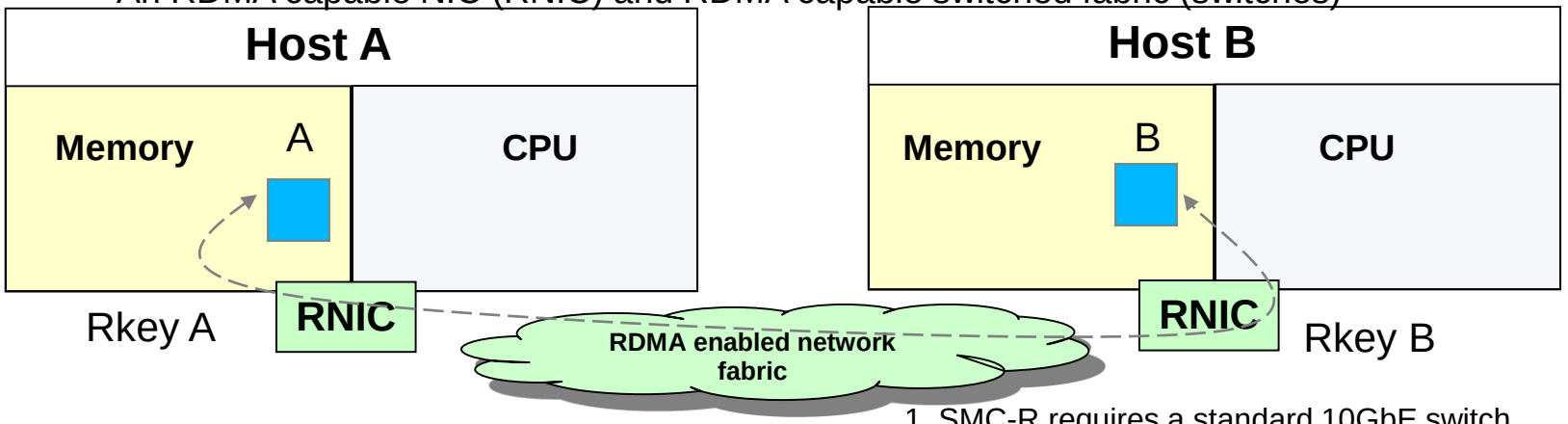
1. Brief review of SMC-R
2. Shared Memory Communications – Direct Memory Access (SMC-D):
   - Introduction: SMC-D: Summary of SMC-D and ISM functions
   - Objectives / Value of SMC-D and ISM (performance overview)
3. IBM z System z13 Internal Shared Memory (ISM) virtual PCI function
4. Getting started: Setup requirements for enabling SMC-D:
   - z13™ / z13s™ system firmware and software requirements
   - ISM System definitions (defining FIDs in HCD)
   - z/OS Communications Server configuration requirements (enabling SMC-D)
   - IP connectivity and VLANs
5. Testing / verification / feedback of SMC-D (scenarios)
6. SMC Applicability Tool (SMC-AT)

Topic 1. Review of SMC-R

Review: RDMA (Remote Direct Memory Access) Technology Overview

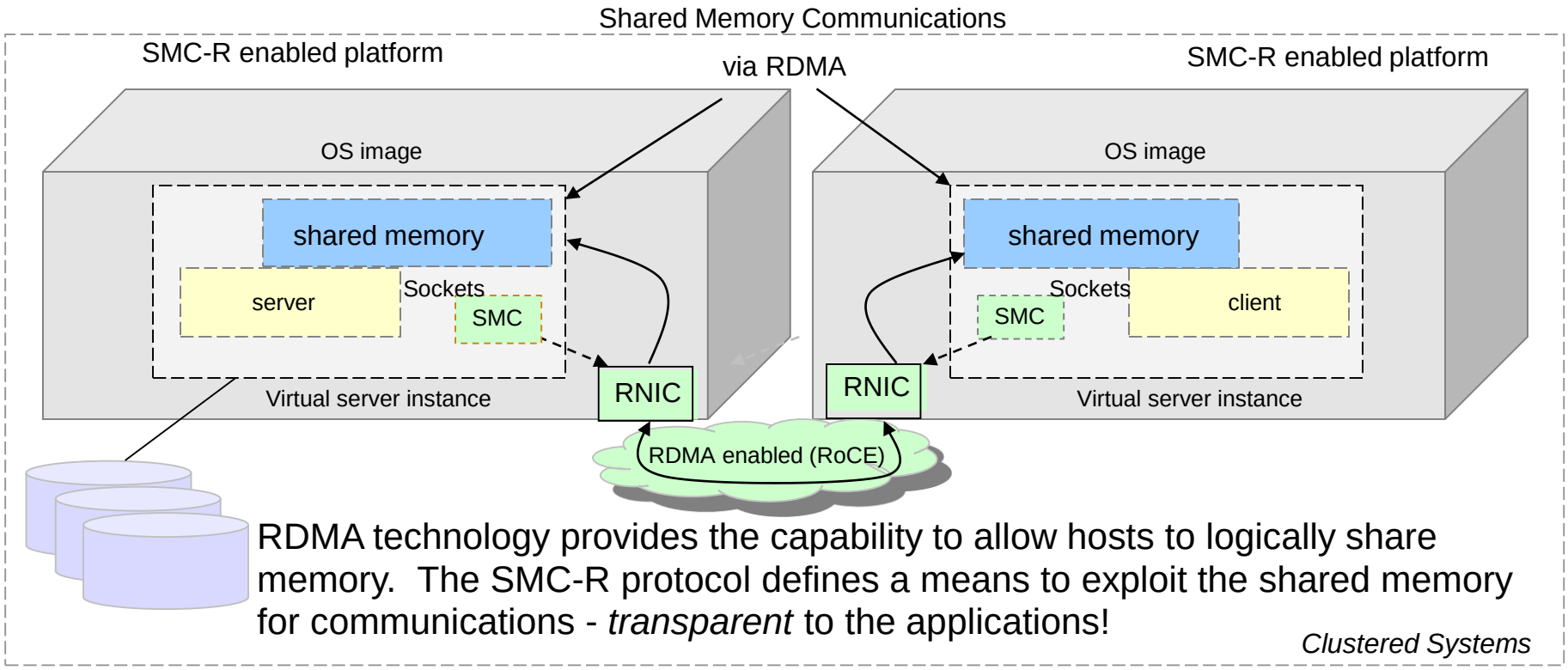
Key attributes of RDMA

- Enables a host to read or write directly from/to a remote host's memory ***without*** involving the remote host's CPU
  - By registering specific memory for RDMA partner use
  - Interrupts ***still required*** for notification (i.e. CPU cycles are not completely eliminated)
- Reduced networking stack overhead by using streamlined, low level, RMDA interfaces
  - Low level APIs such as uDAPL, MPI or RDMA verbs allow optimized exploitation
    - *For applications/middleware willing to exploit these interfaces*
- Key requirements:
  - A reliable "lossless" network fabric (LAN for layer 2 data center network distance)
  - An RDMA capable NIC (RNIC) and RDMA capable switched fabric (switches)[1]



1. SMC-R requires a standard 10GbE switch

Review: Shared Memory Communications over RDMA (SMC-R)

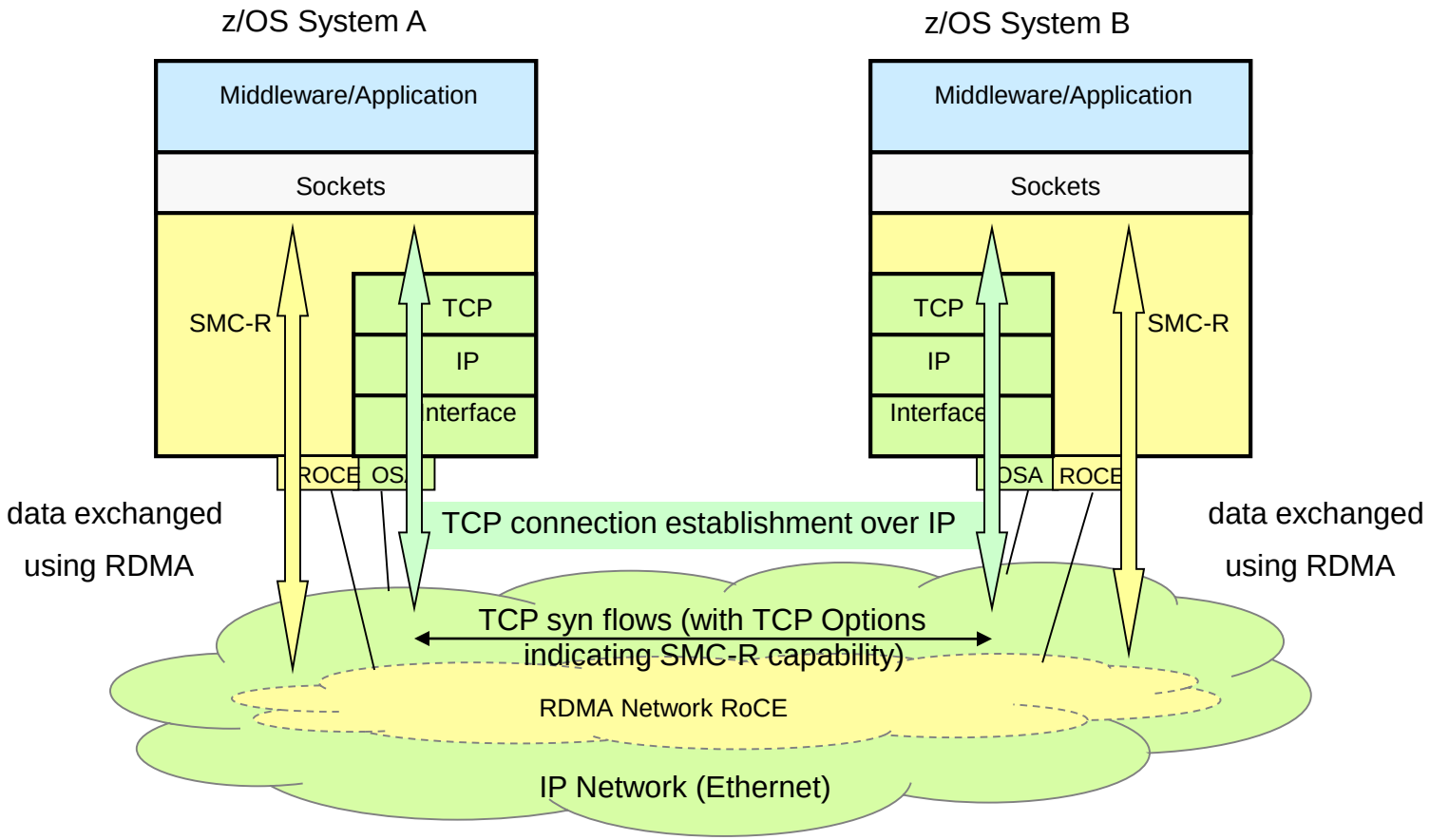


SMC-R is an *open* sockets over RDMA protocol that provides transparent exploitation of RDMA (for TCP based applications) while preserving key functions and qualities of service from the TCP/IP ecosystem that enterprise level servers/network depend on!

Draft IETF RFC for SMC-R:

http://www.rfc-editor.org/rfc/rfc7609.txt

Review: Dynamic Transition from TCP to SMC-R

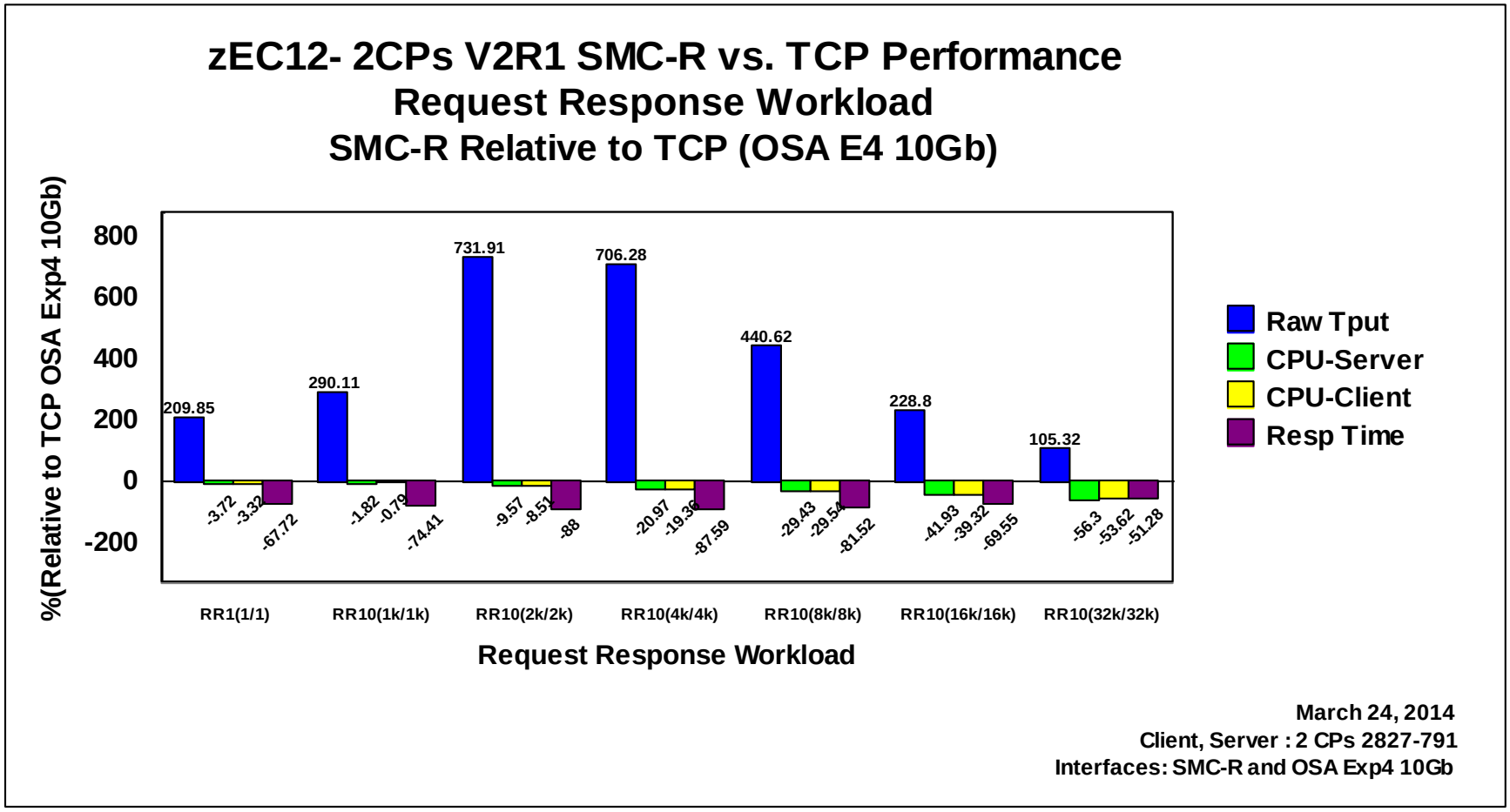


Dynamic (in-line) negotiation for SMC-R is initiated by presence of TCP Options

TCP connection transitions to SMC-R allowing application data to be exchanged using RDMA

Review: z/OS SMC-R Performance Relative to TCP (OSA Ex4 10Gb)

Request Response Workload with different payload. SMC-R provides significantly better performance compared to TCP (OSA Exp4 10Gb).

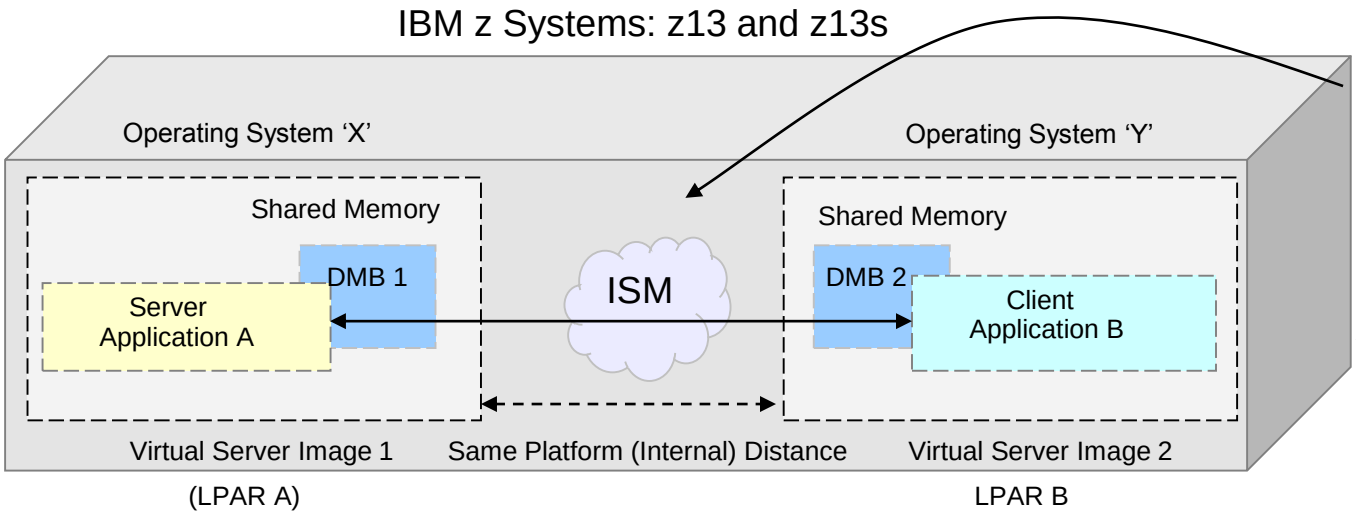


SMC-R Key Attributes - Summary

- ✓ Optimized Network Performance (leveraging RDMA technology)
- ✓ Transparent to (TCP socket based) application software
- ✓ Leverages existing Ethernet infrastructure (RoCE)
- ✓ Preserves existing network security model
- ✓ Resiliency (dynamic failover to redundant hardware)
- ✓ Transparent to Load Balancers
- ✓ Preserves existing IP topology and network administrative and operational model

Topic 2. Shared Memory Communications – Direct Memory Access (SMC-D Introduction)

Shared Memory Communications-Direct Memory Access (SMC-D) over Internal Shared Memory (ISM)

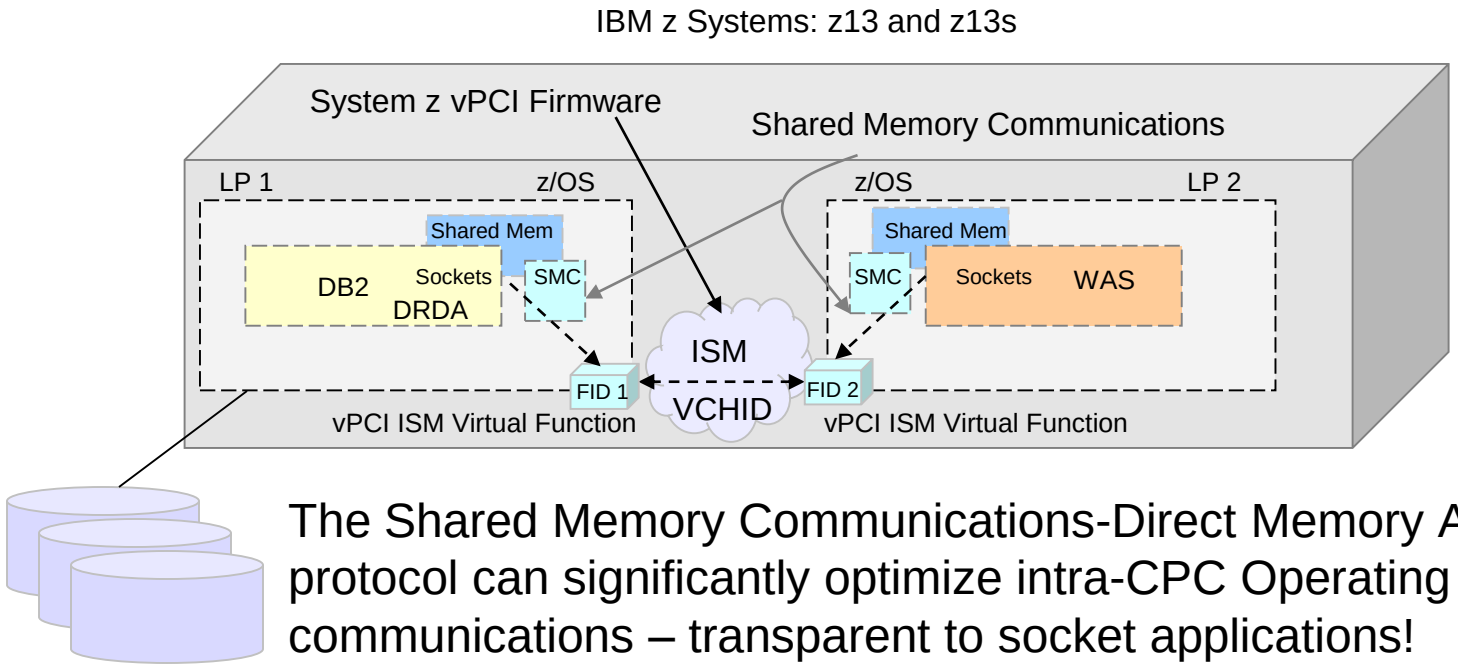


"Shared Memory" across unique OS instances within the same CPC

SMC-D (over ISM) extends the value of the Shared Memory Communications architecture by enabling SMC for direct LPAR to LPAR communications. SMC-D is very similar to SMC-R (over RoCE) extending the benefits of SMC-R to same CPC operating system instances without requiring physical resources (RoCE adapters, PCI bandwidth, NIC ports, I/O slots, network resources, 10GbE switches etc.).

Note 1. The performance benefits of SMC-R (cross CPC) and HiperSockets (within CPC) are similar to each other. SMC-D / ISM provides significantly improved performance benefits above both within the CPC.
Reference performance information: http://www-01.ibm.com/software/network/commserver/SMCR/

SMC-D over ISM:
Internal Shared Memory vPCI Function with ISM VCHIDs

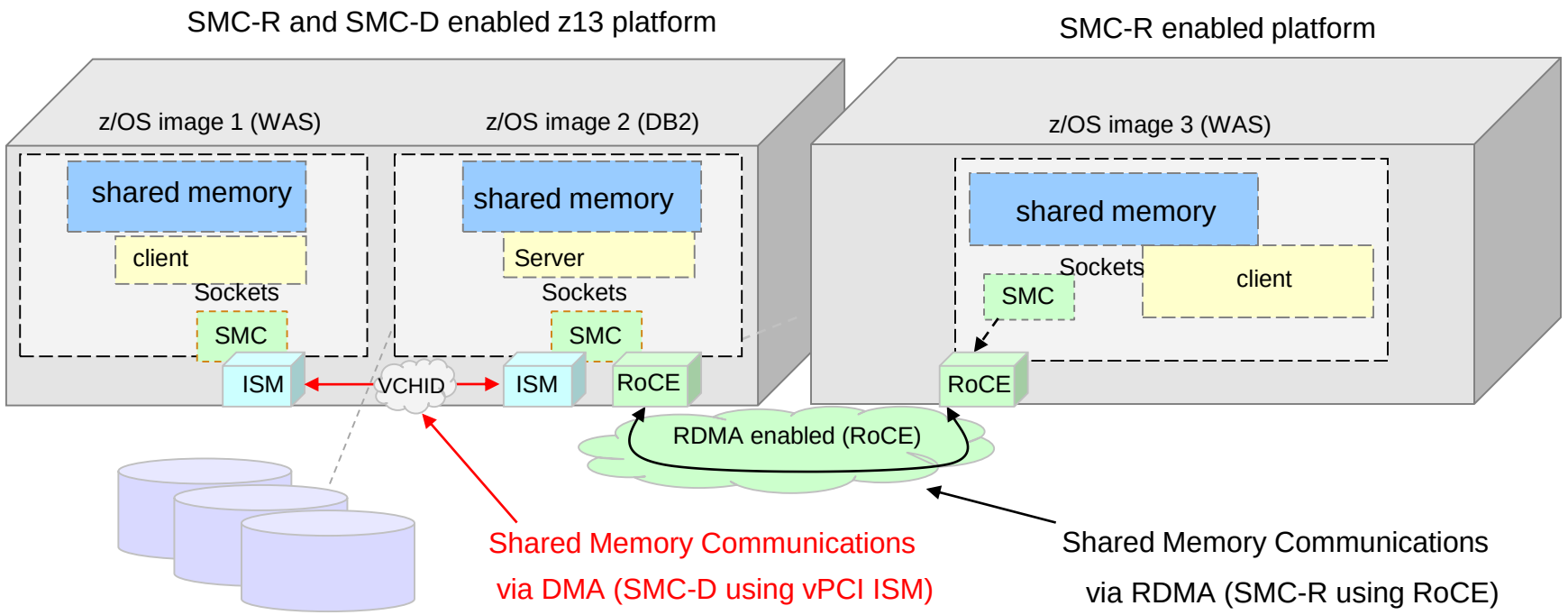


The Shared Memory Communications-Direct Memory Access (SMC-D) protocol can significantly optimize intra-CPC Operating Systems communications – transparent to socket applications!

- Tightly couples socket API communications / memory within the CPC.
- Eliminates TCP/IP processing in the data path.
- ISM is a z System firmware solution (leveraging existing OS virtual memory and does require additional hardware).

Shared Memory Communications within the enterprise data center (RoCE) and within System z (ISM)

Clustered Systems: Example: Local and Remote access to DB2 from WAS (JDBC using DRDA)



Both forms of SMC can be used concurrently combining to provide a highly optimized solution.

Shared Memory Communications: via System z PCI architecture:

1. RDMA (SMC-R for cross platforms via RoCE)
2. DMA (SMC-D for same CPC via ISM)

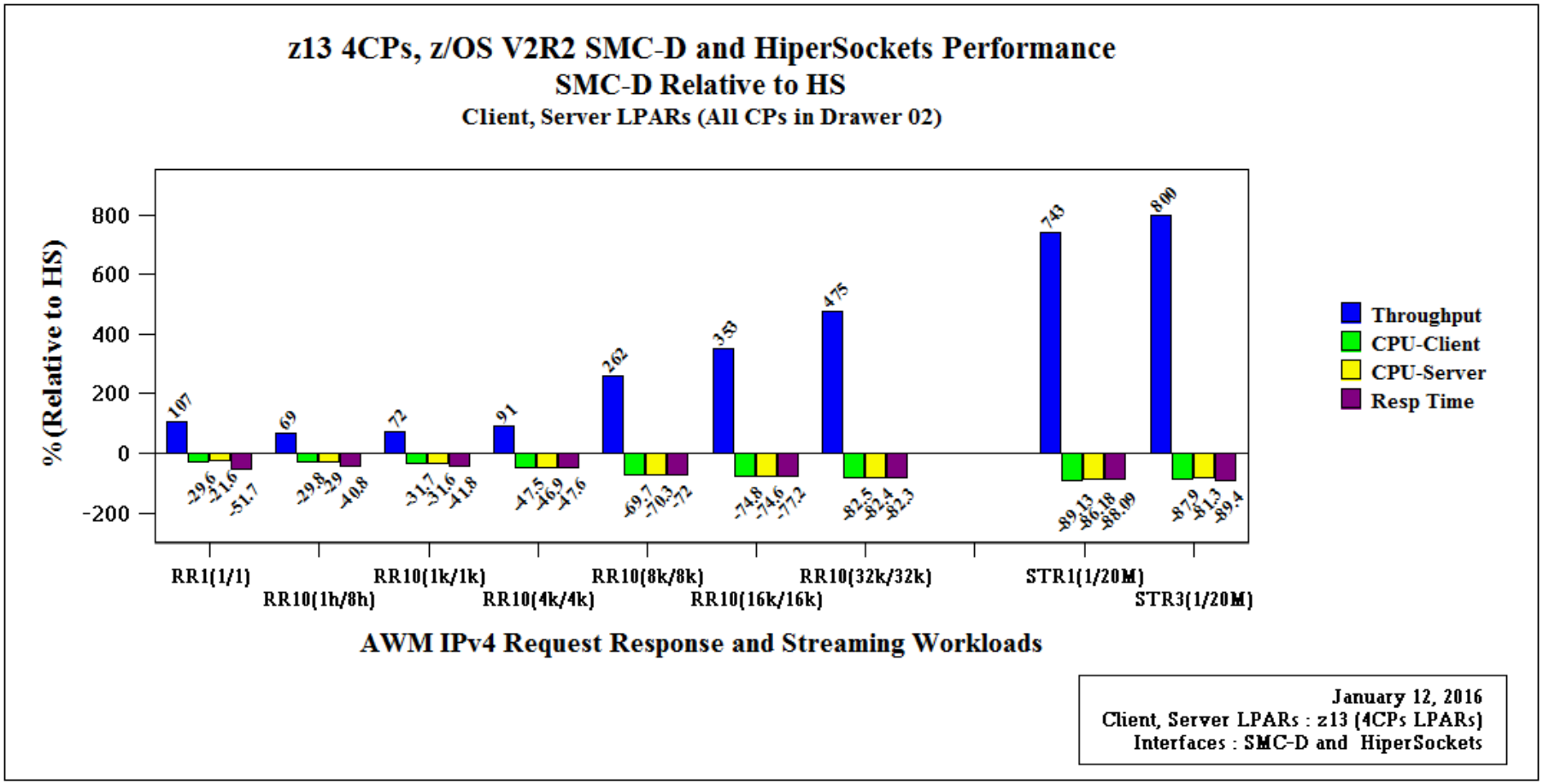
Note. To better understand the IP connectivity shown in this example see chart 35.

SMC-D Performance Benefits and Value (Performance Overview)

- The value of the next generation of highly optimized internal CPC communications is about **providing significantly improved network performance**[1] using tightly coupled socket API communications / memory within the CPC **without additional hardware**
- Network **improvement attributes** are typically described as **latency, throughput, CPU cost and scalability.** Improvements in network performance can potentially improve (increase) application workload transaction rates while **reducing CPU cost**.
- The network latency characteristics provided by SMC-D are compelling:
  - network latency is typically expressed as “network round trip time.” This latency attribute can translate to an improved overall application transaction rate for z/OS to z/OS workloads.
  - **Workloads that are network intensive and transaction oriented** (sometimes described as “request/response” workloads) -- that require multiple and even hundreds of network (“client/server”) flows to complete a single transaction **will realize the most benefit.**

1. Refer to SMC-R website (URL in backup) for SMC-D detailed performance information (additional benchmarks to be added)

HiperSockets Comparison



Up to 9x the throughput! See breakout summary on next chart.

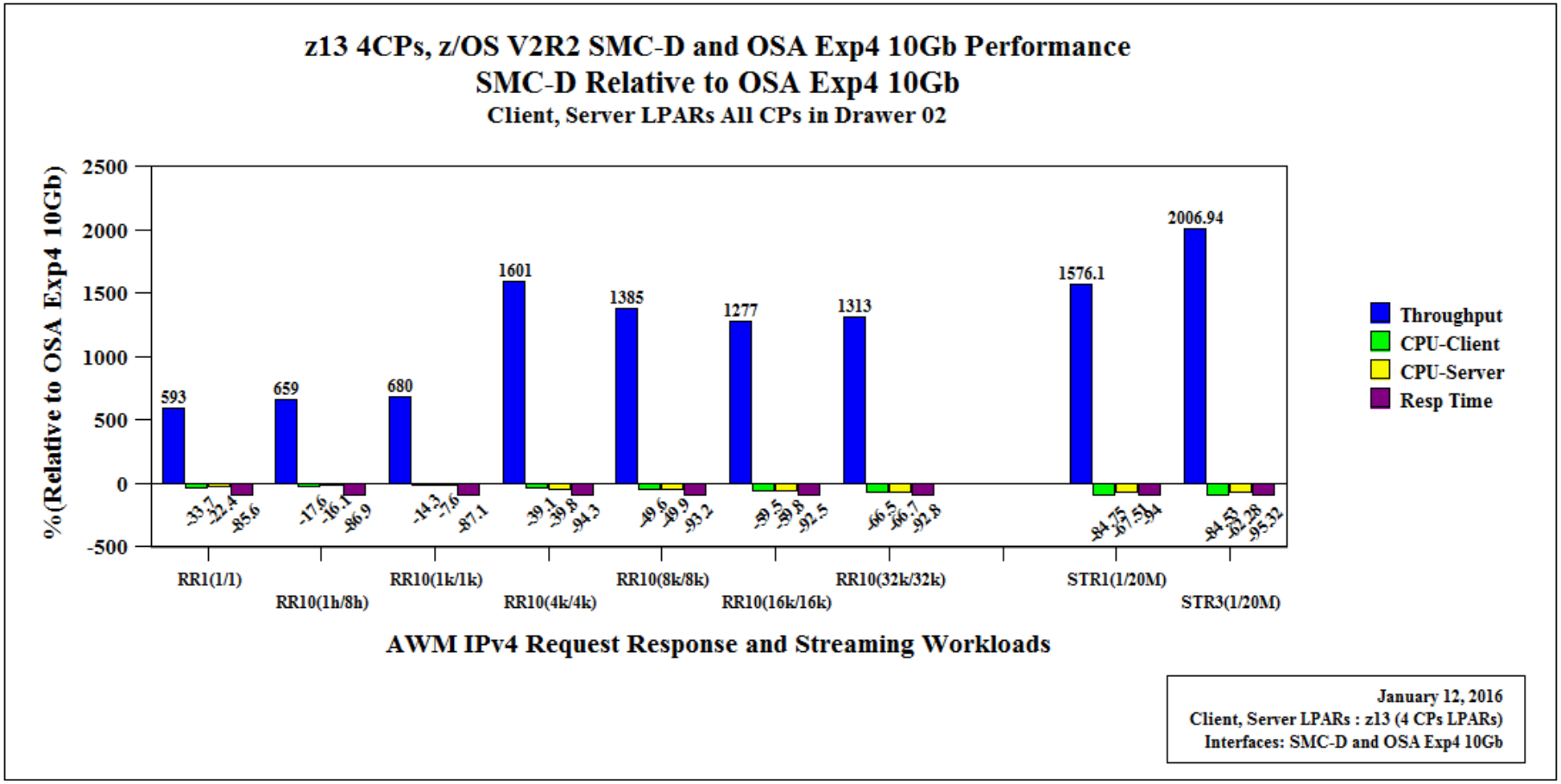
SMC-D / ISM to HiperSockets Summary Highlights

- **Request/Response Summary for Workloads with 1k/1k – 4k/4k Payloads:**
  - **Latency: Up to 48% reduction in latency**
  - **Throughput: Up to 91% increase in throughput**
  - **CPU cost: Up to 47% reduction in network related CPU cost**

- **Request/Response Summary for Workloads with 8k/8k – 32k/32k Payloads:**
  - **Latency: Up to 82% reduction in latency**
  - **Throughput: Up to 475% (~6x) increase in throughput**
  - **CPU cost: Up to 82% reduction in network related CPU cost**

- **Streaming Workload:**
  - **Latency: Up to 89% reduction in latency**
  - **Throughput: Up to 800% (~9x) increase in throughput**
  - **CPU cost: Up to 89% reduction in network related CPU cost**

OSA Comparison



Up to 21x the throughput! See breakout summary on next chart.

SMC-D / ISM to OSA Summary Highlights

- **Request/Response Summary for Workloads with 1k/1k – 4k/4k Payloads:**
  - **Latency: Up to 94% reduction in latency**
  - **Throughput: Up to 1601% (~17x) increase in throughput**
  - **CPU cost: Up to 40% reduction in network related CPU cost**

- **Request/Response Summary for Workloads with 8k/8k – 32k/32k Payloads:**
  - **Latency: Up to 93% reduction in latency**
  - **Throughput: Up to 1313% (~14x) increase in throughput**
  - **CPU cost: Up to 67% reduction in network related CPU cost**

- **Streaming Workload:**
  - **Latency: Up to 95% reduction in latency**
  - **Throughput: Up to 2001% (~21x) increase in throughput**
  - **CPU cost: Up to 85% reduction in network related CPU cost**
- **FTP:**
  - **For Binary Get and Put:**
    - **Up to 58% lower (receive side) CPU cost and**
    - **Up to 26% lower (send side) CPU cost and equivalent throughput**

Shared Memory Communications architecture

Faster communications that preserve TCP/IP qualities of service

- Shared Memory Communications – Direct Memory Access (SMC-D) optimizes z/OS for improved performance in '***within-the-box'*** communications versus standard TCP/IP over HiperSockets or Open System Adapter

Typical Client Use Cases:

- Valuable for multi-tiered work co-located onto a single z Systems server without requiring extra hardware
- Any z/OS TCP sockets based workload can seamlessly use SMC-D without requiring any application changes

SMC Applicability Tool (SMCAT) is available to assist in gaining additional insight into the applicability of SMC-D (and SMC-R) for your environment

*Up to **61%** CPU savings for FTP file transfers across z/OS systems versus HiperSockets**

*Up to **9x** improvement in throughput with more than a **88%** decrease in CPU consumption and a **90%** decrease in response time for streaming workloads versus using HiperSockets**

*Up to **91%** improvement in throughput and up to **48%** improvement in response time for interactive workloads versus using HiperSockets**

* All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM.

Shared Memory Communications architecture

Faster communications that preserve TCP/IP qualities of service

***Memory-to-memory communications** using high speed protocols and direct memory placement of data for faster communications*

Shared Memory Communications Remote Direct Memory Access (SMC-R)

- Use the RoCE Express hardware feature to enable shared memory communications between two servers
- Up to 50% CPU savings for FTP file transfers across z/OS systems versus standard TCP/IP *
- z/OS V2.2 Communications Server now automatically selects between TCP/IP and RoCE

Shared Memory Communications Direct Memory Access (SMC-D)

- Use firmware-based Internal Shared Memory to optimize inter-system operating system communications LPAR to LPAR
- Valuable for multi-tiered work co-located onto a single z Systems server without requiring extra hardware
- Up to **61%** CPU savings for FTP file transfers across z/OS systems versus HiperSockets **

Any z/OS TCP sockets-based workload can **seamlessly** use SMC-R or SMC-D without application changes

SMC Applicability Tool (SMCAT) helps assess benefit of SMC-R and SMC-D for your environment

Connection level security is preserved with SMC-R and SMC-D

* Based on internal IBM benchmarks in a controlled environment using z/OS V2R1 Communications Server FTP client and FTP server, transferring a 1.2GB binary file using SMC-R (10GbE RoCE Express feature) vs. standard TCP/IP (10GbE OSA Express4 feature). The actual CPU savings any user will experience may vary.

** All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided “AS IS” and no warranties or guarantees are expressed or implied by IBM.

SMC-D and ISM (vPCI) Overall Value Points

*Provides **Highly optimized:** improved throughput, reduced latency and CPU cost for intra-CPC communications along with:*

- ✓ Provides the same list of key SMC-R value points:
  - ✓ Transparent to socket applications, no IP topology changes, preserves connection level security, VLAN isolation, transparent with load balancers, etc.
- ✓ …without requiring hardware (adapters, card slots, switches, PCI infrastructure, fabric management, etc.)… cost savings
- ✓ Provides superior resiliency / High Availability (no hardware failures)
- ✓ Provides high scalability, bandwidth and virtualization (i.e. 8k virtual functions)
- ✓ Preserves security (connection level security + secure internal communications)
- ✓ Preserves value of z Systems co-location of workloads (e.g. highly optimized internal communications)
- ✓ Enabled in z/OS with a single TCP/IP profile keyword [1]

Note 1. ISM VCHID and FIDs must be defined in HCD (or IOCDS)

Topic 3. IBM System z13™ Internal Shared Memory (ISM) (ISM Introduction)

IBM z Systems™ z13™ and z13s™ SMC-D with ISM Introduction

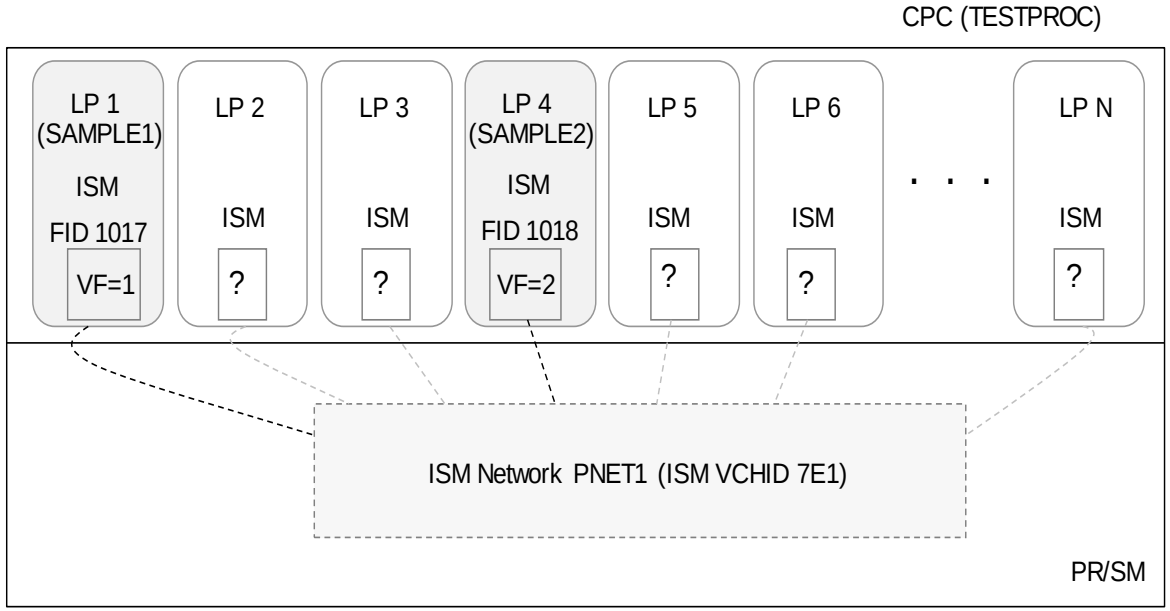
Description

- The IBM z13 and z13s introduces Internal Shared Memory (ISM) virtual PCI function. ISM is a virtual PCI network adapter that enables direct access to shared virtual memory providing a highly optimized network interconnect for z Systems intra-CPC communications.
- ISM is supported by z/VM 6.3 (PTF) with pass thru guest support.
- IBM z/OS V2R2 (PTF) introduces the capability to exploit ISM with **Shared Memory Communications-Direct Memory Access** (SMC-D).

- For more information on new z13™ and z13s™

  http://www-03.ibm.com/systems/z/announcement.html

Introduction: IBM System z13 / z13s
Internal Shared Memory (ISM) virtual PCI Function



FUNCTION FID=1017,PCHID=7E1,VF=1,PART=((SAMPLE1),(SAMPLE1,SAMPLE2)),PNETID=(PNET1),TYPE=ISM
FUNCTION FID=1018,PCHID=7E1,VF=2,PART=((SAMPLE2),(SAMPLE1,SAMPLE2)),PNETID=(PNET1),TYPE=ISM

Internal Shared Memory (ISM) Overview

- ISM enables the ability for Operating Systems (LPARs) to share virtual memory (similar to RDMA)
- New “Internal Shared Memory” (ISM) VCHID Type
  (ISM VCHID concepts are similar to IQD (HiperSockets) VCHID)
- ISM is based on existing z System’s PCI architecture (i.e. virtual PCI Function / adapter)
- Introduces a new PCI Function type (ISM virtual PCI function)
- System admin / configuration / operations follows the same process (HCD/IOCDS) as existing PCI functions (e.g. RoCE Express, zEDC Express, etc.)
- ISM supports Dynamic I/O
- Supported by z/VM when z/OS is a guest on z/VM (PCI device support)
- Enables highly optimized next generation intra-CPC communications (SMC-D)

continued…

Internal Shared Memory (ISM) Overview (part 2)

- Provides adapter virtualization (Virtual Functions) with high scalability:
  - 32 ISM VCHIDs per CPC (each VCHID represents a unique internal shared memory network each with a unique Physical Network ID)
  - 255 VFs per VCHID (8k VFs per CPC)
    (i.e. the maximum no. of virtual servers that can communicate over the same ISM VCHID is 255)
- Each ISM VCHID represents a unique (isolated) internal network, each having a unique Physical Network ID (PNet IDs are configured in HCD/IOCDS)
- ISM VCHIDs support VLANs (i.e. can be sub-divided into VLANs)
- ISM provides a GID (“Global ID” internally generated by firmware) that corresponds with each ISM FID. The GID is used to locate / address a host on an ISM network (VCHID)
- MACs (VMACs), MTU, physical ports[1] and Frame size are all N/A
- ISM is supported by z/VM (for passthru guest access to support the new PCI function)

Note 1. ISM VCHIDs provide support for a single logical port (also see PNet ID topic)

Topic 4. Getting Started: Install / Configure / Enable: ISM and SMC-D Enablement Overview

Four steps:

1. Upgrade System z (firmware)
2. Install required software
3. Define ISM FIDs with PNet ID (HCD definitions)
4. Enable SMCD (TCP/IP profile)

Steps 1 and 2: ISM System z13 and z/OS SMC-D Requirements

1. IBM z Systems: z13 (driver level 27 (GA2)) or z13s

2. z/OS software (PTF) requirements:

    1. CommServer VTAM: OA48411 UA80711
    2. CommServer TCP/IP: PI45028 UI35411
    3. z/OS (IOS): OA47913 UA80812
    4. HCD: OA46010
    5. IOCP: OA47938 UA90986
    6. HCM: IO23612
    7. RMF: OA49113 UA80445

Note. For a complete / current list of PTFs refer to the PSP bucket.

Step 3. HCD - Defining ISM in HCD

- ISM is defined as a PCI device. VCHIDs and FIDs / VFs must be defined in HCD (or IOCDS)
  - A VCHID represents a virtual PCI Adapter, which also represents a unique (isolated) internal network
  - FIDs/VFs are assigned to LPARs as reconfigurable PCI functions
  - FIDs that are defined to the same VCHID are eligible to communicate with each other
- Some examples of Service Element (SE) panels are included in the backup (page 67)
- HCD Change processor steps are shown in backup (required before you can define ISM)
- HCD (and IOCDS) samples of ISM definition and a corresponding HCD IQD sample definition (with PNet ID) follow.

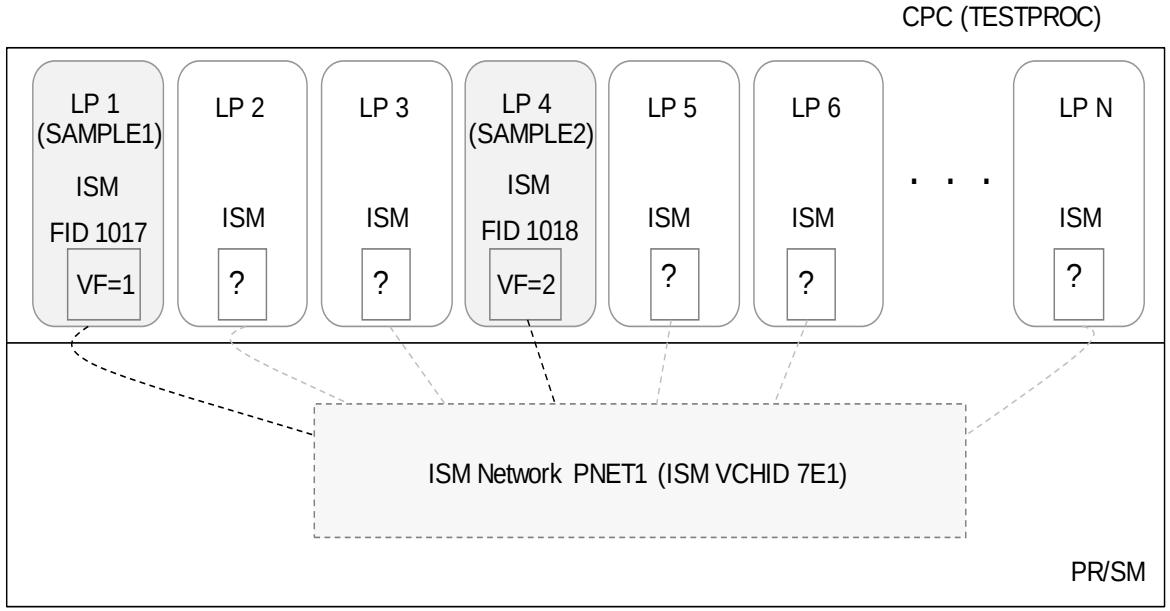
Associating ISM with your IP Network devices (OSA or HS)

- ISM Functions must also be associated with another Channel (CHID), either:
  1. IQD (a single IQD / HiperSockets) channel **or**…
  2. OSD channels
     A single **ISM VCHID can not be associated with both** (IQD and OSD)
- The association of an ISM VCHID (Function IDs) to the channel(s) is created by defining (HCD) matching **Physical Network IDs** (PNet IDs)
- The channel devices (OSD or IQD) provide IP connectivity and are associated with ISM based on having matching PNet IDs
- ISM (like IQD) supports a single PNet ID per ISM VCHID (a single “logical port”)
- PNet IDs are dynamically discovered by z/OS (from HCD config)

Associating ISM with your IP Network devices (part 2)

- ISM PNet IDs must:
  - be unique among other ISM VCHIDs for this System
  - match a corresponding IQD VCHID **or** OSD Channel(s)
- Additional PNet ID information is illustrated in the following charts:
  - "IP Connectivity" (topology) examples (concepts) of matching PNet IDs
  - New PNet ID Netstat displays

ISM Configuration Example (see the following HCD charts)



FUNCTION FID=1017,PCHID=7E1,VF=1,PART=((SAMPLE1),(SAMPLE1,SAMPLE2)),PNETID=(PNET1),TYPE=ISM
FUNCTION FID=1018,PCHID=7E1,VF=2,PART=((SAMPLE2),(SAMPLE1,SAMPLE2)),PNETID=(PNET1),TYPE=ISM

Add PCIe Function

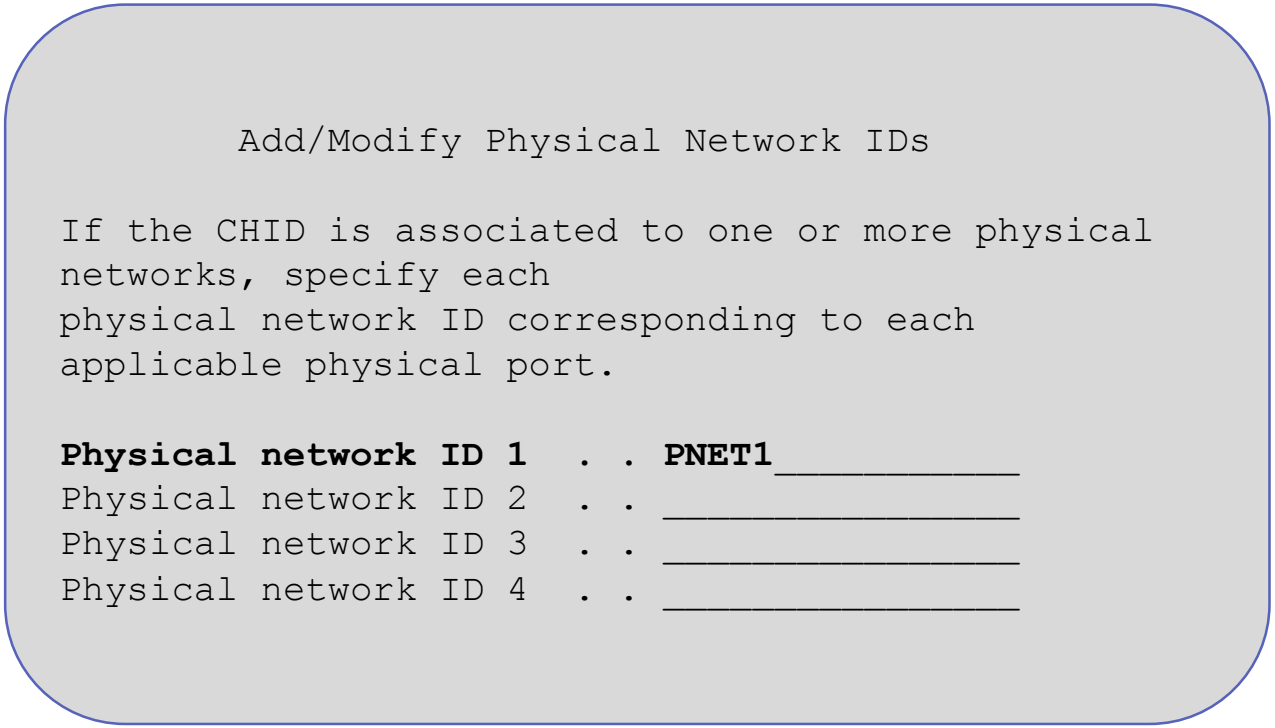
Define the ISM function:

1. action f on processor to see the PCIe function list
2. action add on function list (PF11 or line command add like)
   Note the ISM VCHID 7E1

```
                Add PCIe Function

  Specify or revise the following values.

  Processor ID  . . . . : TESTPROC     testprocessor

  Function ID . . . . . . 1017
  Type  . . . . . . . . . ISM            +

  CHID  . . . . . . . . . 7E1  +

  Virtual Function ID . . 1    +

  Description . . . . . . test scenario
```

Press Enter

Add/Modify ISM PNet ID



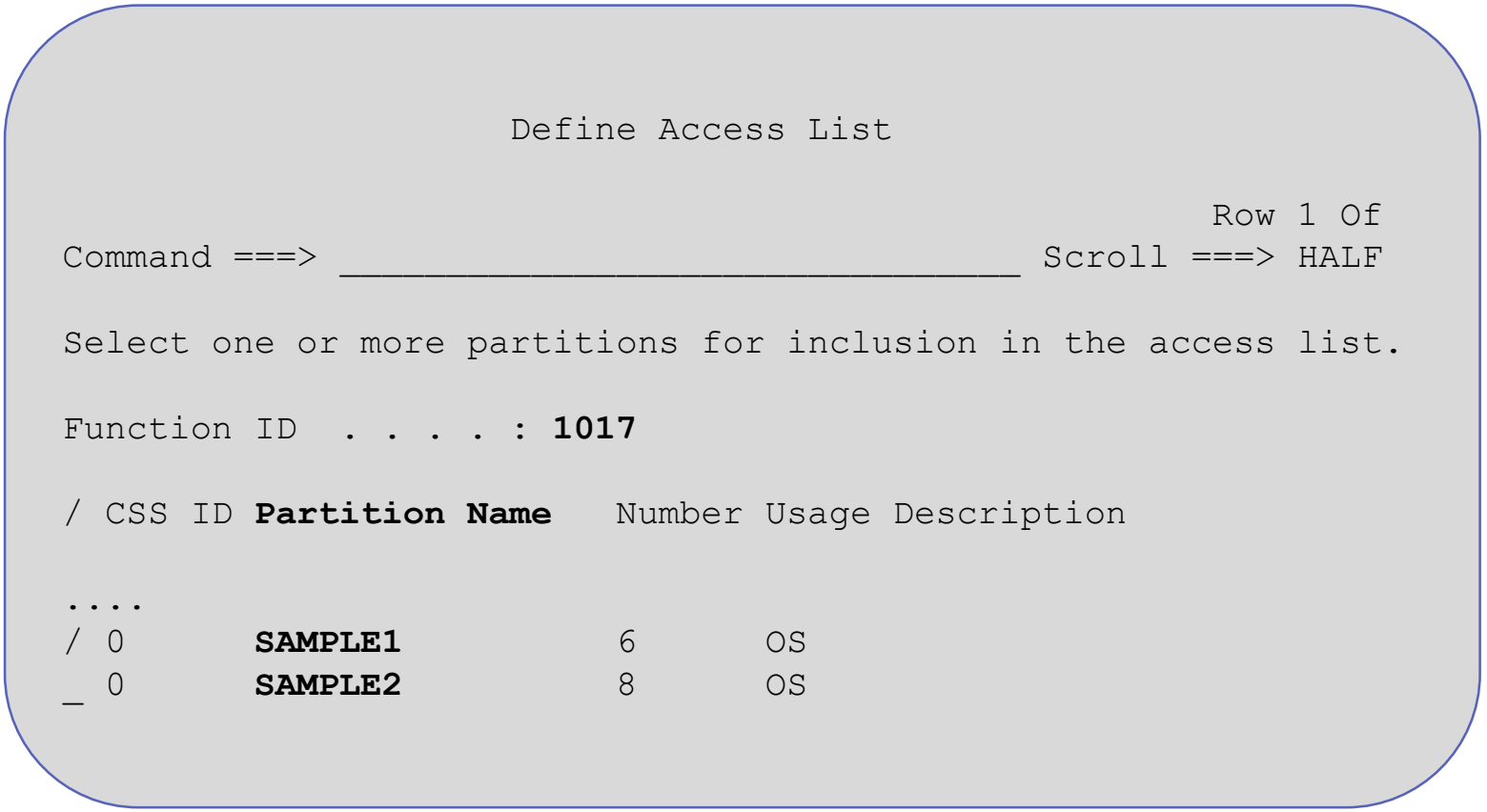
Press Enter

PNet ID Notes.

1. ISM supports a single PNet ID per ISM VCHID
2. ISM PNet IDs must be unique among other ISM VCHIDs for this System
3. ISM PNet IDs must match a corresponding IQD VCHID or OSD Channel(s)

Define Access List

Allows access to this ISM Function (FID) from specific partitions.



Press Enter

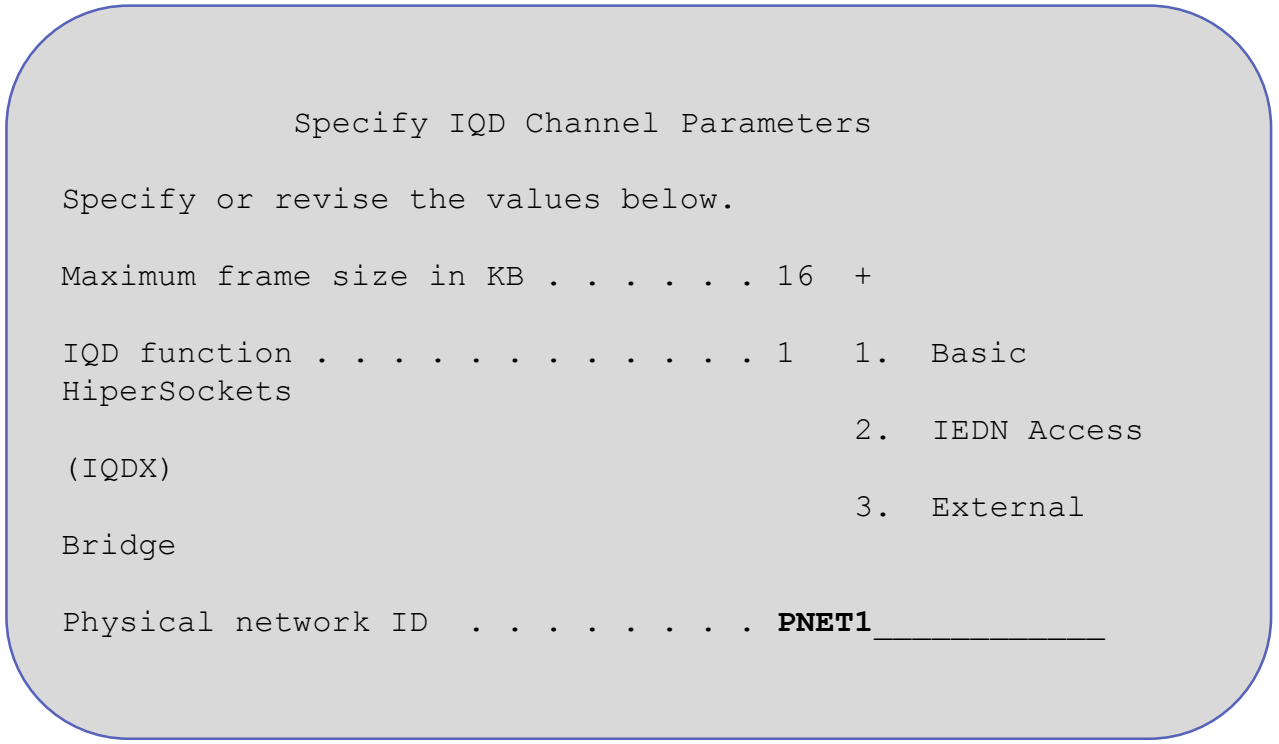
Note. The selected partition (SAMPLE1 in this example) must also be in the Access List for the corresponding IQD or OSD Channel

Define HiperSockets (IQD) Channel (to be associated with ISM VCHID)

```
                    Add Channel Path

 Specify or revise the following values.

 Processor ID . . . . : TESTPROC     testprocessor
 Configuration mode . : LPAR
 Channel Subsystem ID : 0

 Channel path ID  . . . . 11    +            Channel ID
7E0  +
 Number of CHPIDs . . . . 1
 Channel path type  . . . IQD   +
 Operation mode . . . . . SHR   +
 Managed  . . . . . . . . No   (Yes or No)   I/O Cluster
________  +
 Description  . . . . . . sample IQD__________________

 Specify the following values only if connected to a
switch:
 Dynamic entry switch ID  __   + (00 - FF)
 Entry switch ID  . . . . __   +
 Entry port . . . . . . . __   +
```

Press Enter

Define IQD Parameters



Press Enter

Define IQD Access List

```
                        Define Access List

Row 1 of
 Command ===> ____________________________________________
Scroll ===> HAL

 Select one or more partitions for inclusion in the access
list.

 Channel subsystem ID : 0
 Channel path ID  . . : 11     Channel path type  . : IQD
 Operation mode . . . : DED    Number of CHPIDs . . : 1

 / CSS ID Partition Name   Number Usage Description
...
 / 0      SAMPLE1          6      OS
 / 0      SAMPLE2          8      OS
```

Press Enter

Enter on the candidate list as well, and you are back on the chpid list.

Press F3 twice to go back to the processor list

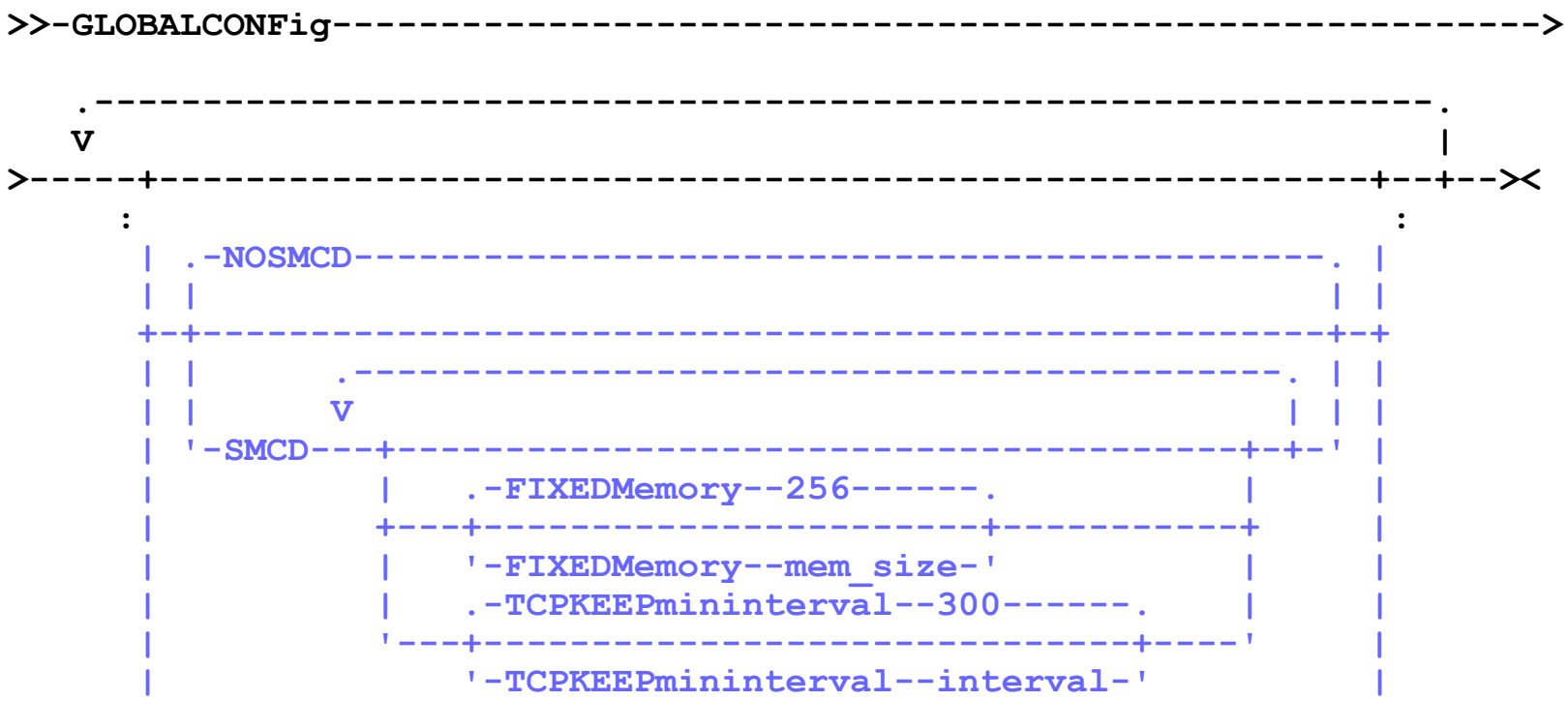
Result: Chpid 11 on PCHID 7E0 is now defined and has partition SAMPLE1 and SAMPLE2 of CSS 0 in its access list. PNETID is PNET1

z/OS CommServer Exploitation of Internal Shared Memory (ISM)

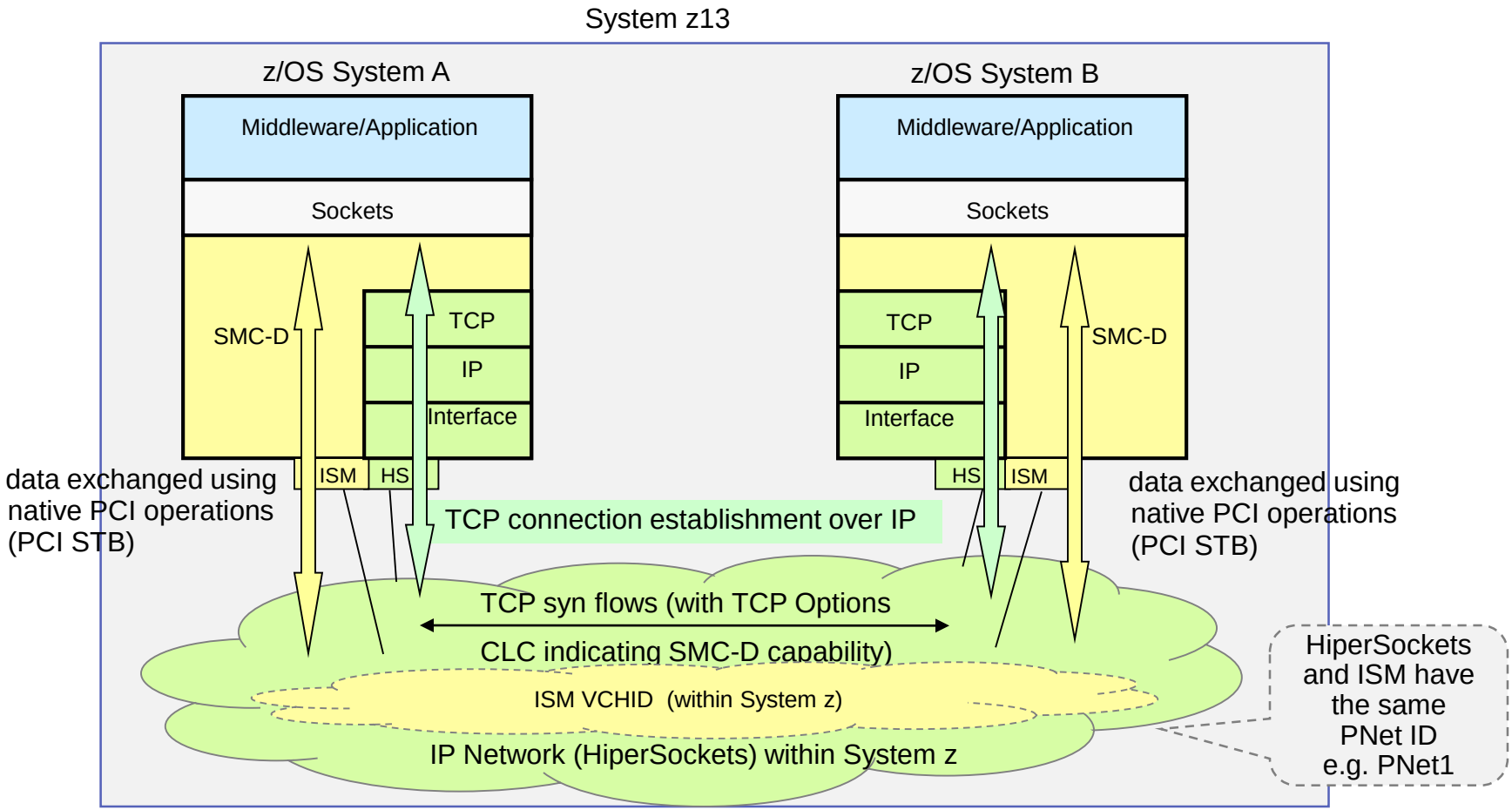
- ISM enables Shared Memory Communications-Direct Memory Access (SMC-D)
- Once the ISM HCD configuration is complete, SMC-D can be enabled in z/OS with a single TCP/IP parameter (**GLOBALCONFIG SMCD**).
- Notes:
  - ISM FIDs **are not** defined in the TCP/IP profile. ISM FIDs must be Configured On to z/OS and then the FIDs are dynamically discovered by TCP/IP.
  - An OS can be enabled for both SMC-R and SMC-D. SMC-D is used when both peers are within the same CPC (and using the ISM VCHID and VLAN).
  - ISM FIDs (VCHIDs) must be associated with an IP network. The association is accomplished by defining matching PNet IDs (e.g. HS and ISM).
    Notes:
    - Your OSA (or IQD channel) must have a PNet ID defined (and must match your ISM FID)
    - The OSA or IQD INTERFACE statement must have IPSubnet defined
  - Host virtual memory is managed by each OS (similar to SMC-R, logically shared memory) following existing z System's PCI I/O translation architecture (i.e. only minor changes required for z/VM guests). There are no required configuration changes.

Step 4. TCPIP Profile GlobalConfig SMCD

- SMCD parameter on GLOBALCONFIG (similar to SMCR)
- **Single Keyword! SMCD** is the only required setting to enable SMC-D
- Key difference from SMCR parameter: ISM PFIDs are not defined in TCPIP (ISM FIDs are auto discovered based on matching PNETID associated with the OSD or HiperSockets)



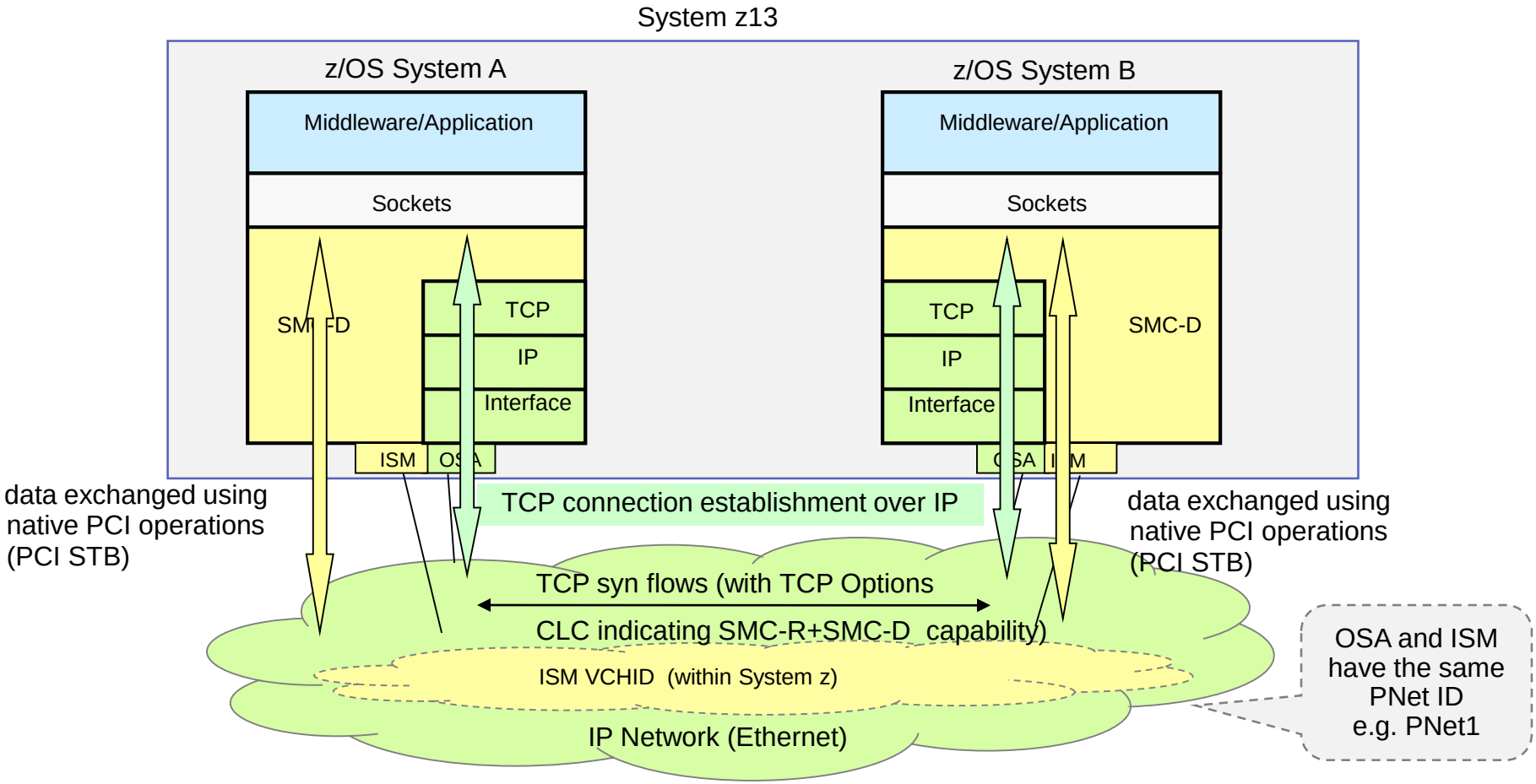
Dynamic Transition from TCP to SMC-D – (HiperSockets IP Network)



Dynamic (in-line) negotiation for SMC-D is initiated by presence of TCP Options

TCP connection transitions to SMC-D allowing application data to be exchanged using Direct Memory Access (LPAR to LPAR)

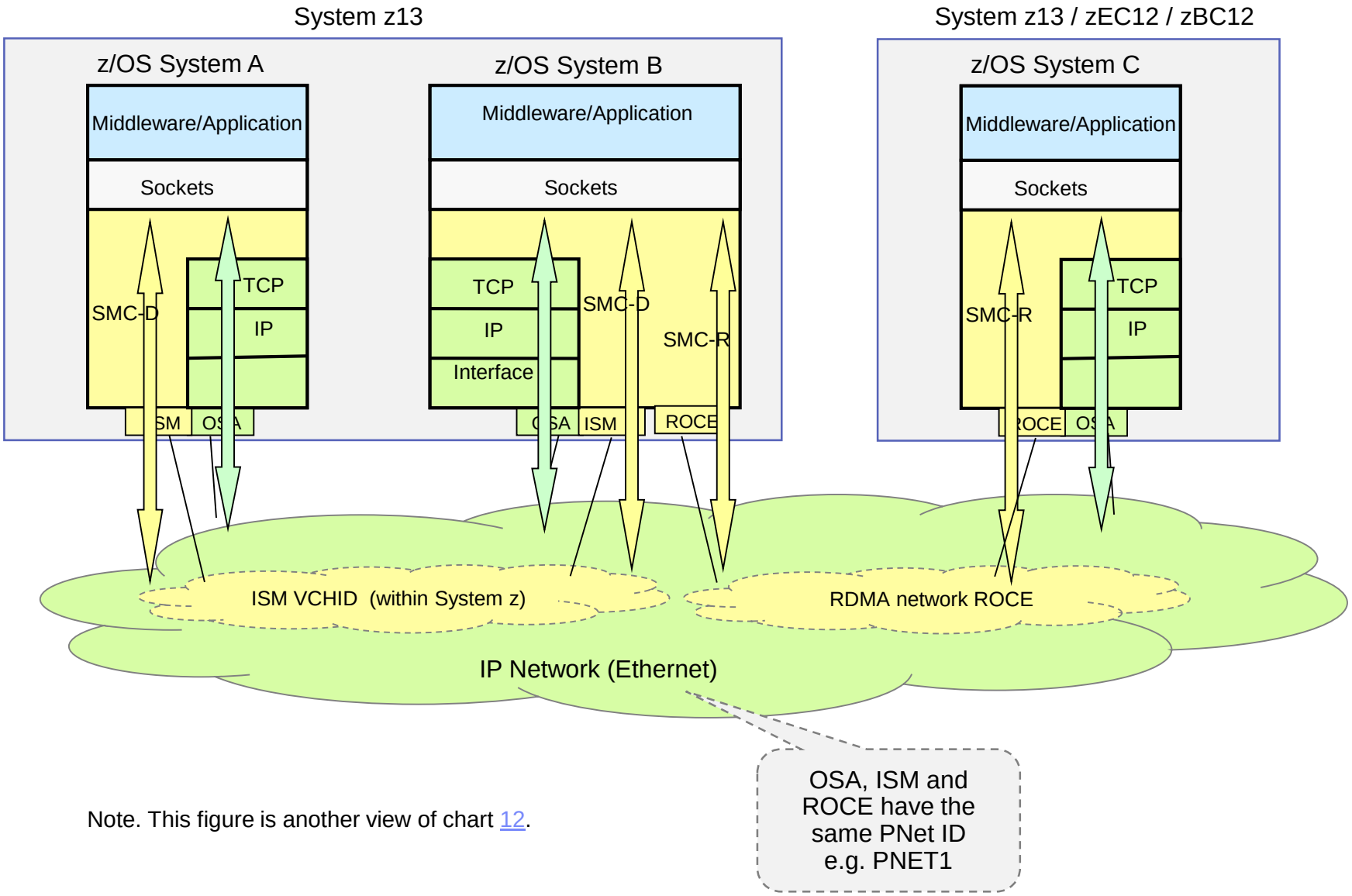
Dynamic Transition from TCP to SMC-D (OSA/LAN IP network)



Dynamic (in-line) negotiation for SMC-R&SMC-D is initiated by presence of TCP Options

TCP connection transitions to SMC-D allowing application data to be exchanged using Direct Memory Access (LPAR to LPAR)

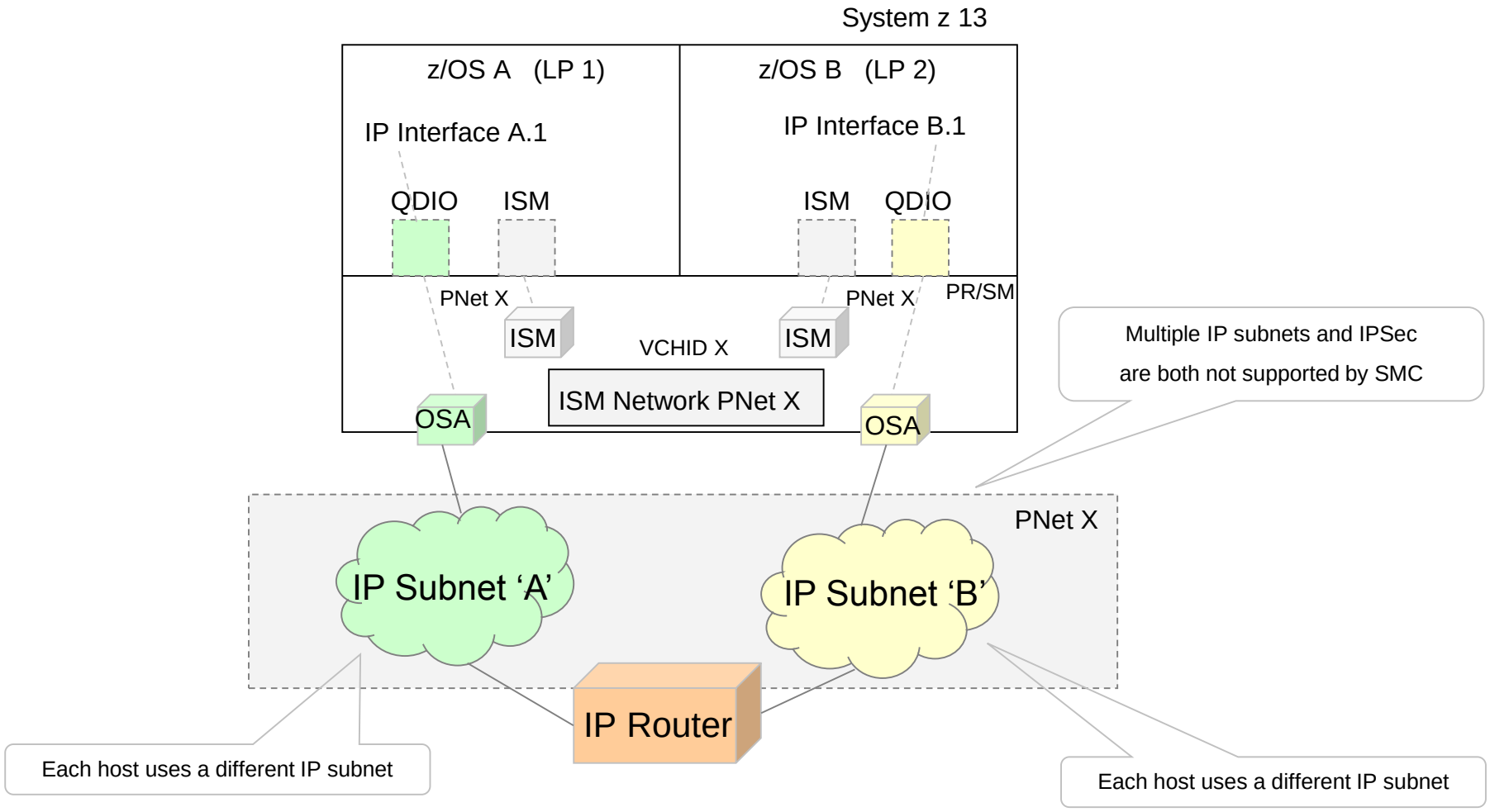
OSA/LAN IP network with SMC-D and SMC-R



Note. This figure is another view of chart 12.

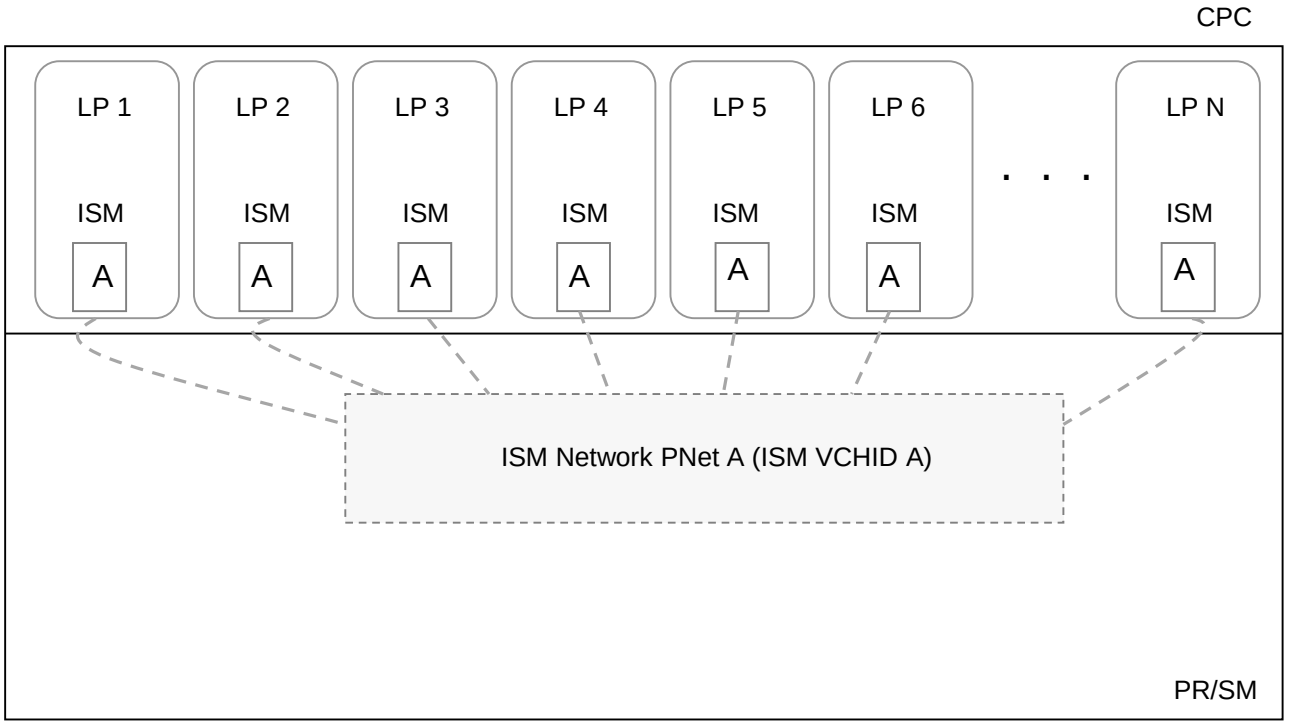
Multiple IP Subnets are not Supported by SMC (SMC-R or SMC-D)!

Peers must have direct connectivity over the same IP subnet to exploit SMC-R or SMC-D

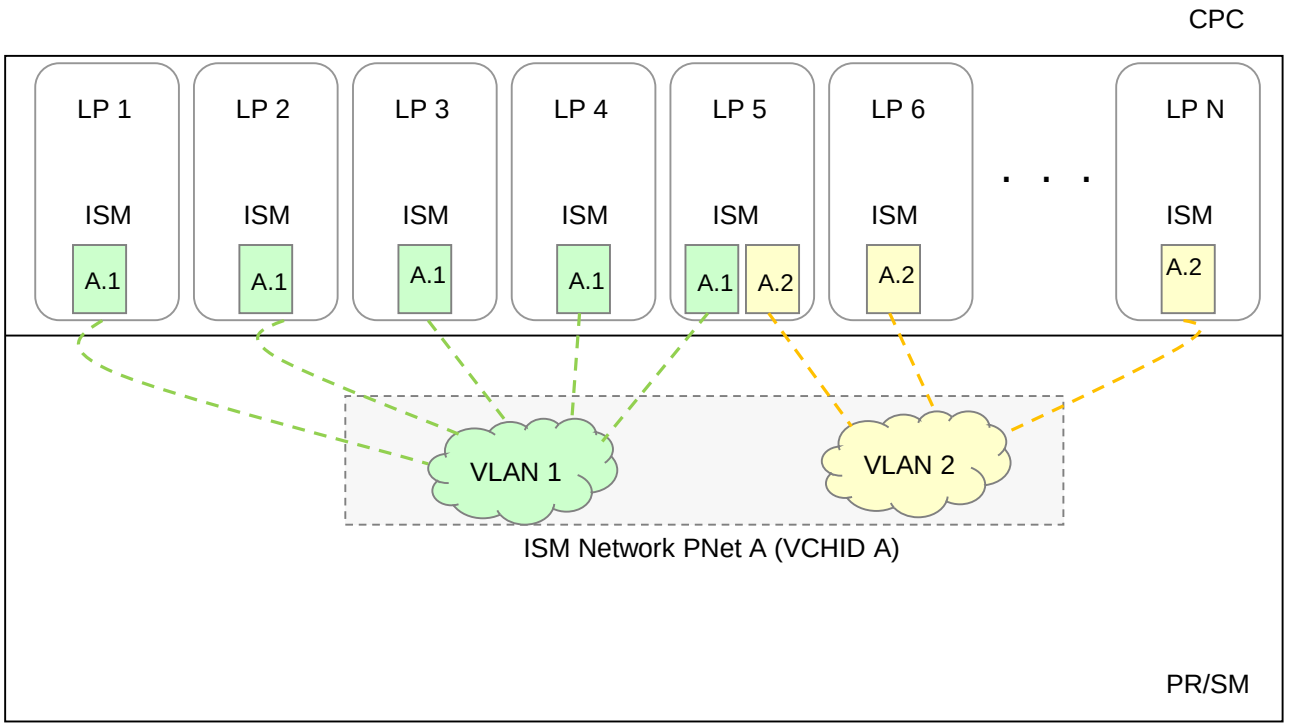


ISM VLAN Overview

ISM VCHID = Internal (ISM) Network (based on PNet ID)

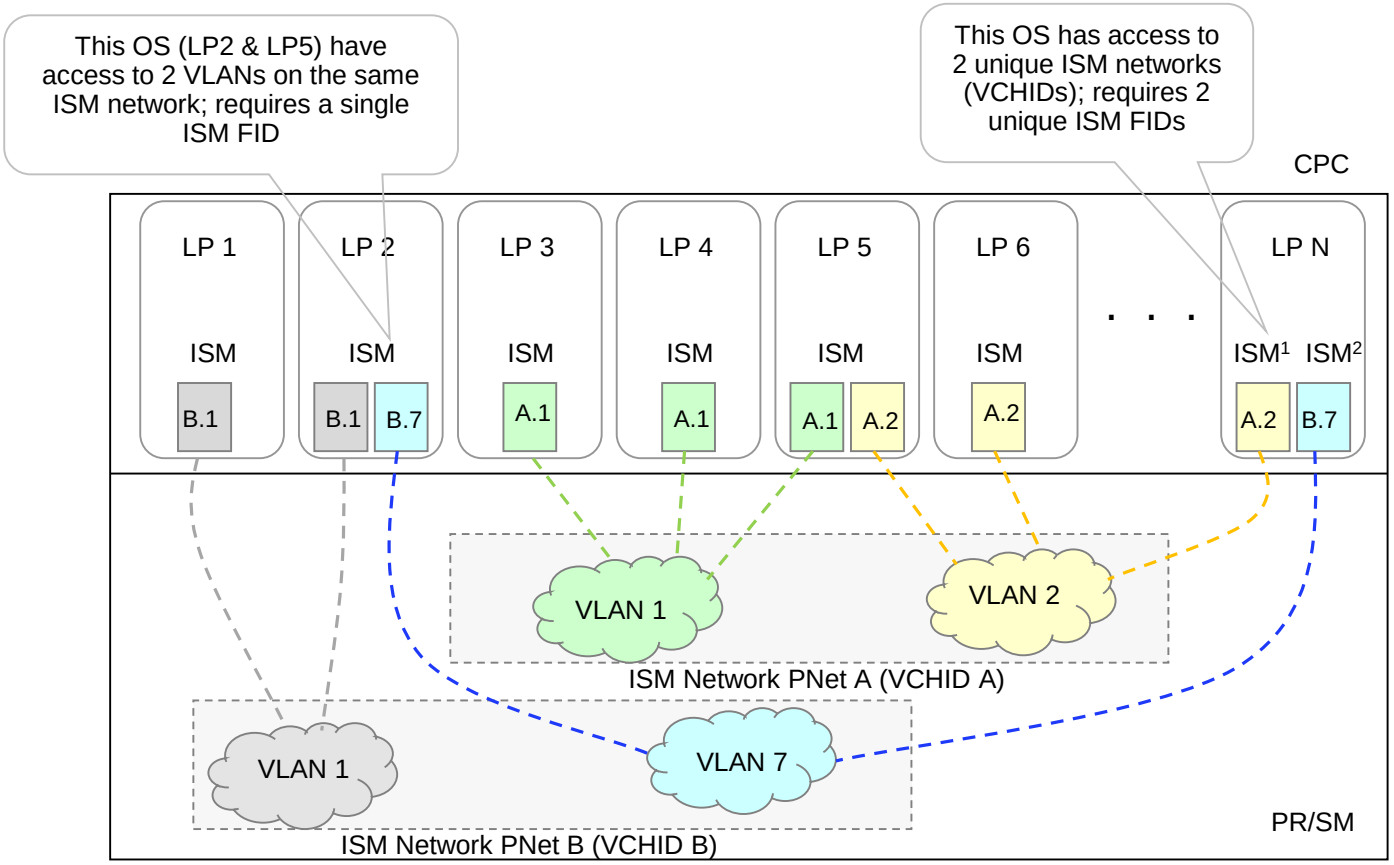


Subdividing ISM VCHIDs with VLANs (Isolating Workloads)



In z/OS ISM VLAN definitions are inherited from the associated IP interface

Isolating workloads with multiple VCHIDs and VLANs



Topic 5. Verification of ISM Usage

D PCIE (PFIDs that are in use)

Display PCIE shows the ISM PFIDs that are now allocated to VTAM (ALLC):

```
D PCIE
IQP022I 12.14.22 DISPLAY PCIE 691
PCIE     0010 ACTIVE
PFID      DEVICE TYPE NAME          STATUS  ASID  JOBNAME  PCHID VFN
00000005  10GbE RoCE Express       ALLC    0038  VTAMCS   0100  0005
00000001  10GbE RoCE Express       ALLC    0038  VTAMCS   0184  0001
00000500  ISM                      ALLC    0038  VTAMCS   07E0  0001
00000501  ISM                      ALLC    0038  VTAMCS   07E0  0002
00000502  ISM                      CNFG                   07E0  0003
00000503  ISM                      CNFG                   07E0  0004
00000504  ISM                      CNFG                   07E0  0005
00000600  ISM                      ALLC    0038  VTAMCS   07E1  0001
00000601  ISM                      ALLC    0038  VTAMCS   07E1  0002
00000602  ISM                      CNFG                   07E1  0003
00000603  ISM                      CNFG                   07E1  0004
00000604  ISM                      CNFG                   07E1  0005
```

Note. If you display a specific FID additional detail is provided including the configured PNET ID value.

Displaying the Configured PNet IDs for Channel Devices

OSD CHPID:

```
d m=chp(16)
IEE174I 11.20.44 DISPLAY M 612
CHPID 16:  TYPE=11, DESC=OSA DIRECT EXPRESS, ONLINE
DEVICE STATUS FOR CHANNEL PATH 16
     0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F
0096 +  +  +  +  +  +  +  +  +  +  +  +  .  .  .  +
SWITCH DEVICE NUMBER = NONE
PHYSICAL CHANNEL ID = 01C0
PNETID 1 = P1
********************** SYMBOL EXPLANATIONS ***********************
+ ONLINE    @ PATH NOT VALIDATED   - OFFLINE    . DOES NOT EXIST
* PHYSICALLY ONLINE   $ PATH NOT OPERATIONAL
```

IQD CHPID:

```
d m=chp(21)
IEE174I 11.21.19 DISPLAY M 615
CHPID 21:  TYPE=24, DESC=INTERNAL QUEUED DIRECT COMM, ONLINE
DEVICE STATUS FOR CHANNEL PATH 21
     0  1  2  3  4  5  6  7  8  9  A  B  C  D  E  F
0FD1 +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
SWITCH DEVICE NUMBER = NONE
ATTRIBUTES = MFS(24KB)
PNETID = P2
********************** SYMBOL EXPLANATIONS ***********************
+ ONLINE    @ PATH NOT VALIDATED   - OFFLINE    . DOES NOT EXIST
* PHYSICALLY ONLINE   $ PATH NOT OPERATIONAL
```

Displaying the Configured PNet IDs for PCI Devices

RoCE PFID:

```
d pcie,pfid=5
IQP024I 11.22.20 DISPLAY PCIE 618
PCIE     0010 ACTIVE
PFID      DEVICE TYPE NAME         STATUS  ASID  JOBNAME  CHID  VFN
00000005  10GbE RoCE Express       CNFG                   0100  0005
 CLIENT ASIDS: NONE
 PNetID 1: P1
 PNetID 2: P1
```

ISM PFID:

```
d pcie,pfid=500
IQP024I 11.22.30 DISPLAY PCIE 621
PCIE     0010 ACTIVE
PFID      DEVICE TYPE NAME         STATUS  ASID  JOBNAME  CHID  VFN
00000500  ISM                      CNFG                   07E0  0001
 CLIENT ASIDS: NONE
 PNetID 1: P1
```

Netstat DEvlinks/-d for a SMCD-enabled IQD interface

Shows the PNETID and the associated ISM interface:

```
D TCPIP,TCPIP2,NETSTAT,DEVLINKS,INTFNAME=IQD1
EZD0101I NETSTAT CS V2R3 TCPIP2 694
INTFNAME: IQD1              INTFTYPE: IPAQIDIO   INTFSTATUS: READY
    TRLE: IUTIQ421  DATAPATH: FD12     DATAPATHSTATUS: READY
    CHPID: 21
    PNETID: P2                 SMCD: YES
    IPBROADCASTCAPABILITY: NO
    ARPOFFLOAD: YES                  ARPOFFLOADINFO: YES
    CFGMTU: NONE                     ACTMTU: 16384
    IPADDR: 10.15.2.21/24
    VLANID: 200
    READSTORAGE: GLOBAL (3008K)
    SECCLASS: 255                    MONSYSPLEX: NO
    IQDMULTIWRITE: DISABLED
  MULTICAST SPECIFIC:
    MULTICAST CAPABILITY: YES
    GROUP             REFCNT         SRCFLTMD
    -----             ------         --------
    224.0.0.1         0000000001     EXCLUDE
      SRCADDR: NONE
  INTERFACE STATISTICS:
    BYTESIN                           = 0
    INBOUND PACKETS                   = 0
    INBOUND PACKETS IN ERROR          = 0
    INBOUND PACKETS DISCARDED         = 0
    INBOUND PACKETS WITH NO PROTOCOL  = 0
    BYTESOUT                          = 0
    OUTBOUND PACKETS                  = 0
    OUTBOUND PACKETS IN ERROR         = 0
    OUTBOUND PACKETS DISCARDED        = 0
  ASSOCIATED ISM  INTERFACE: EZAISM01
1 OF 1 RECORDS DISPLAYED
END OF THE REPORT
```

Netstat DEvlinks/-d for a SMCD-enabled OSD interface

Shows the PNETID and the associated RNIC and ISM interfaces:

```
D TCPIP,TCPIP2,NETSTAT,DEVLINKS,INTFNAME=OSD1
EZD0101I NETSTAT CS V2R3 TCPIP2 700
INTFNAME: OSD1                INTFTYPE: IPAQENET   INTFSTATUS: READY
    PORTNAME: HYDRA960  DATAPATH: 0962     DATAPATHSTATUS: READY
    CHPIDTYPE: OSD              SMCR: YES
    PNETID: P1                  SMCD: YES
    SPEED: 0000001000
    IPBROADCASTCAPABILITY: NO
    VMACADDR: 0200014860B0   VMACORIGIN: OSA    VMACROUTER: ALL
    ARPOFFLOAD: YES                  ARPOFFLOADINFO: YES
    CFGMTU: NONE                     ACTMTU: 8992
    IPADDR: 10.15.1.21/24
    VLANID: 100                      VLANPRIORITY: DISABLED
                  .
                  .
                  .
  ASSOCIATED RNIC INTERFACE: EZARIUT10001
  ASSOCIATED ISM  INTERFACE: EZAISM02
IPV4 LAN GROUP SUMMARY
LANGROUP: 00002
  NAME              STATUS      ARPOWNER          VIPAOWNER
  ----              ------      --------          ---------
  OSD1              ACTIVE      OSD1              YES
1 OF 1 RECORDS DISPLAYED
END OF THE REPORT
```

Netstat Devlinks all PNetIDs (new)

Netstat DEvlinks/-d PNETID * shows a summary of all the active interfaces that have a PNetID configured organized by PNetID value:

```
D TCPIP,TCPIP2,NETSTAT,DEVLINKS,PNETID=*
EZD0101I NETSTAT CS V2R3 TCPIP2 881
PNETID: P2
  INTFNAME: IQDIOINTF6         INTFTYPE: IPAQIDIO6
  INTFNAME: IQDIOLNK0A0F0217   INTFTYPE: IPAQIDIO
  INTFNAME: EZAISM01           INTFTYPE: ISM        ASSOCIATED: YES
PNETID: P1
  INTFNAME: V6OSD1             INTFTYPE: IPAQENET6
  INTFNAME: OSD1               INTFTYPE: IPAQENET
  INTFNAME: EZAISM02           INTFTYPE: ISM        ASSOCIATED: YES
  INTFNAME: EZARIUT10003       INTFTYPE: RNIC       ASSOCIATED: YES
7 OF 7 RECORDS DISPLAYED
END OF THE REPORT
```

Netstat ALL/-A for a connection using SMCD

Shows the SMCD status and a SMCD reason code if SMCD could not be used

```
D TCPIP,TCPIP2,NETSTAT,ALL,IPPORT=10.15.2.31+21
EZD0101I NETSTAT CS V2R3 TCPIP2 791
CLIENT NAME: OSASUP13                  CLIENT ID: 00000032
  LOCAL SOCKET: 10.15.2.21..1024
  FOREIGN SOCKET: 10.15.2.31..21
    BYTESIN:            0000000000000000174
    BYTESOUT:           0000000000000000029
    SEGMENTSIN:         0000000000000000007
    SEGMENTSOUT:        0000000000000000007
    STARTDATE:          08/19/2015       STARTTIME:          16:16:38
    LAST TOUCHED:       16:16:38         STATE:              ESTABLSH
                  .
                  .
                  .
  RECEIVEBUFFERSIZE:  0000245760       SENDBUFFERSIZE:     0000184320
    RECEIVEDATAQUEUED:  0000000000
    SENDDATAQUEUED:     0000000000
    SENDSTALLED:        NO
    SMC INFORMATION:
      SMCDSTATUS:       ACTIVE
        LOCALSMCDLINKID:4B020000         REMOTESMCDLINKID:   4B030000
        LOCALSMCRCVBUF: 64K              REMOTESMCRCVBUF:    64K
    ANCILLARY INPUT QUEUE: N/A
    APPLICATION DATA:   EZAFTP0C C OSASUP1   C      D
```

Netstat DEvlinks/-d SMC

Shows all ISM and RNIC interfaces and associated SMC link information

```
D TCPIP,TCPIP2,NETSTAT,DEVLINKS,SMC
EZD0101I NETSTAT CS V2R3 TCPIP2 833
INTFNAME: EZAISM01          INTFTYPE: ISM        INTFSTATUS: READY
  PFID: 0600  TRLE: IUT00600  PFIDSTATUS: READY
  PNETID: P2
  GIDADDR:  02008581C9172964
  INTERFACE STATISTICS:
    BYTESIN                          = 6567
    INBOUND OPERATIONS               = 17
    BYTESOUT                         = 41
    OUTBOUND OPERATIONS              = 4
    SMC LINKS                        = 1
    TCP CONNECTIONS                  = 1
    INTF RECEIVE BUFFER INUSE        = 64K
  SMCD LINK INFORMATION:
    LOCALSMCDLINKID: 4B020000  REMOTESMCDLINKID: 4B030000
      VLANID: 200
      LOCALGID:  02008581C9172964
      REMOTEGID: 01008582C9172964
      SMCDLINKBYTESIN:           6567
      SMCDLINKINOPERATIONS:      17
      SMCDLINKBYTESOUT:          41
      SMCDLINKOUTOPERATIONS:     4
      TCP CONNECTIONS:           1
      LINK RECEIVE BUFFER INUSE: 64K
INTFNAME: EZAISM02          INTFTYPE: ISM        INTFSTATUS: READY
                .
                .
                .
3 OF 3 RECORDS DISPLAYED
END OF THE REPORT
```

Summary: Verification of SMC-D

- Requires (at least) two z/OS instances (LPARs or z/VM guests) executing on the same z13 CPC (GA2 or z13s)
- Both z/OS instances must:
  - be defined to use the same ISM VCHID
  - have their ISM FIDs Configured On to each z/OS (LPAR)
  - have direct access to the same IP network (IP subnet) via OSA or HiperSockets (i.e. hosts can communicate directly over the same IP subnet without traversing an IP Router).
  - define an IP interface with the same VLAN ID (if VLANs are used)
- Enable both ISM and SMC-D (see backup for Netstat examples)
  - Verify both ISM and SMC-D are enabled
  - start your test application (TCP sockets) workloads
  - Verify TCP connections dynamically exploit SMC-D
- Optional: Measure / compare your performance:
  - Working with your performance analyst consider comparing your TCP/IP (OSA or HS) performance benchmarks with SMC-D (ISM) benchmarks for the sample workloads you are most interested in evaluating

Topic 6. SMC Applicability Tool (SMC-AT)

Evaluating SMC Applicability and Benefits

As customers express interest in SMC-R and RoCE Express one of the initial questions asked is:

- "What benefit will SMC-R provide in my environment?"
  - Some users are well aware of significant traffic patterns that can benefit from SMC-R
  - But others are unsure of how much of their TCP traffic (in their environment) is:
    - z/OS to z/OS and
    - how much of that traffic is well suited to SMC-R
- This same set of customer questions will also apply to SMC-D
- RYO evaluation processes can be a time consuming activity that requires significant expertise.

SMC Applicability Tool Introduction

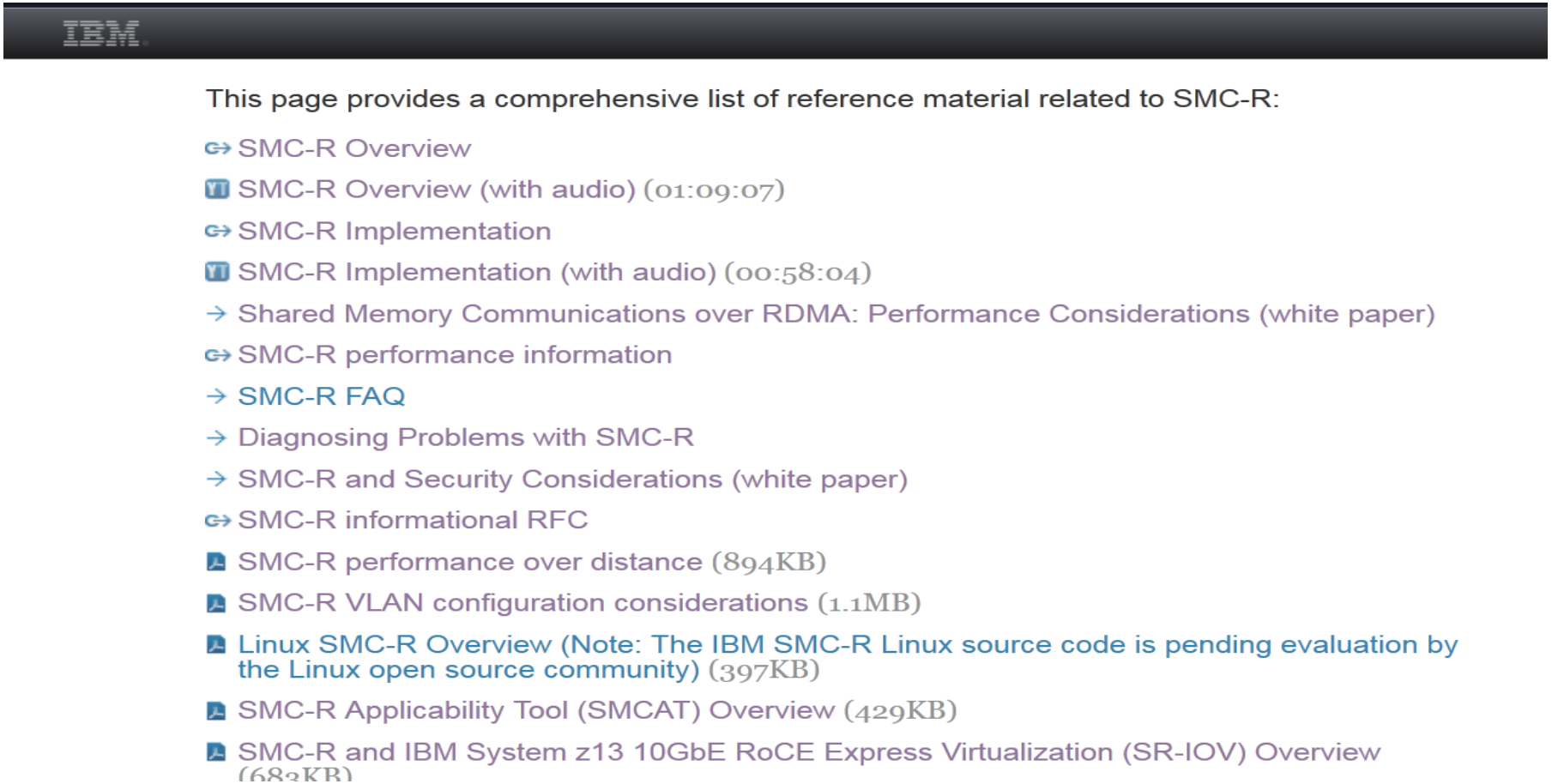
A new tool called SMC Applicability Tool (SMCAT) has been created that will help customers determine the value of SMC-R and SMC-D in their environment with minimal effort and minimal impact

- SMCAT is integrated within the TCP/IP stack:
  Gathers new statistics that are used to project SMC-R and SMC-D applicability and benefits for the current system
  - Minimal system overhead, no changes in TCP/IP network flows
  - Produces reports on potential benefits of enabling SMC-R / SMC-D
  - Does not require RoCE or ISM hardware or the SMC-R/D function. No IP configuration changes are required (measures your existing TCP/IP traffic).
- Available via the service stream on existing z/OS releases:
  - z/OS V1R13 - APAR PI48309 PTF UI31050
  - z/OS V2R1, V2R2 - APAR PI48155, PTFs UI31054 (2.1) and UI31055 (2.2)
- For additional SMC-AT information refer to:
  http://www.ibm.com/software/network/commserver/SMCR

SMC References

- ***SMC One Stop Shopping Web Page (Includes latest links to ALL other SMC References):***

  http://www.ibm.com/software/network/commserver/SMCR



IBM

This page provides a comprehensive list of reference material related to SMC-R:

- SMC-R Overview
- SMC-R Overview (with audio) (01:09:07)
- SMC-R Implementation
- SMC-R Implementation (with audio) (00:58:04)
- Shared Memory Communications over RDMA: Performance Considerations (white paper)
- SMC-R performance information
- SMC-R FAQ
- Diagnosing Problems with SMC-R
- SMC-R and Security Considerations (white paper)
- SMC-R informational RFC
- SMC-R performance over distance (894KB)
- SMC-R VLAN configuration considerations (1.1MB)
- Linux SMC-R Overview (Note: The IBM SMC-R Linux source code is pending evaluation by the Linux open source community) (397KB)
- SMC-R Applicability Tool (SMCAT) Overview (429KB)
- SMC-R and IBM System z13 10GbE RoCE Express Virtualization (SR-IOV) Overview (683KB)

Backup

Feedback, comments and questions are welcome.

Backup Topics (additional details):

1. HCD Change Processor
2. Service Element Screenshot Examples
3. Sample TRLE (display) Information

Backup Topic 1.1 Configuring ISM in HCD

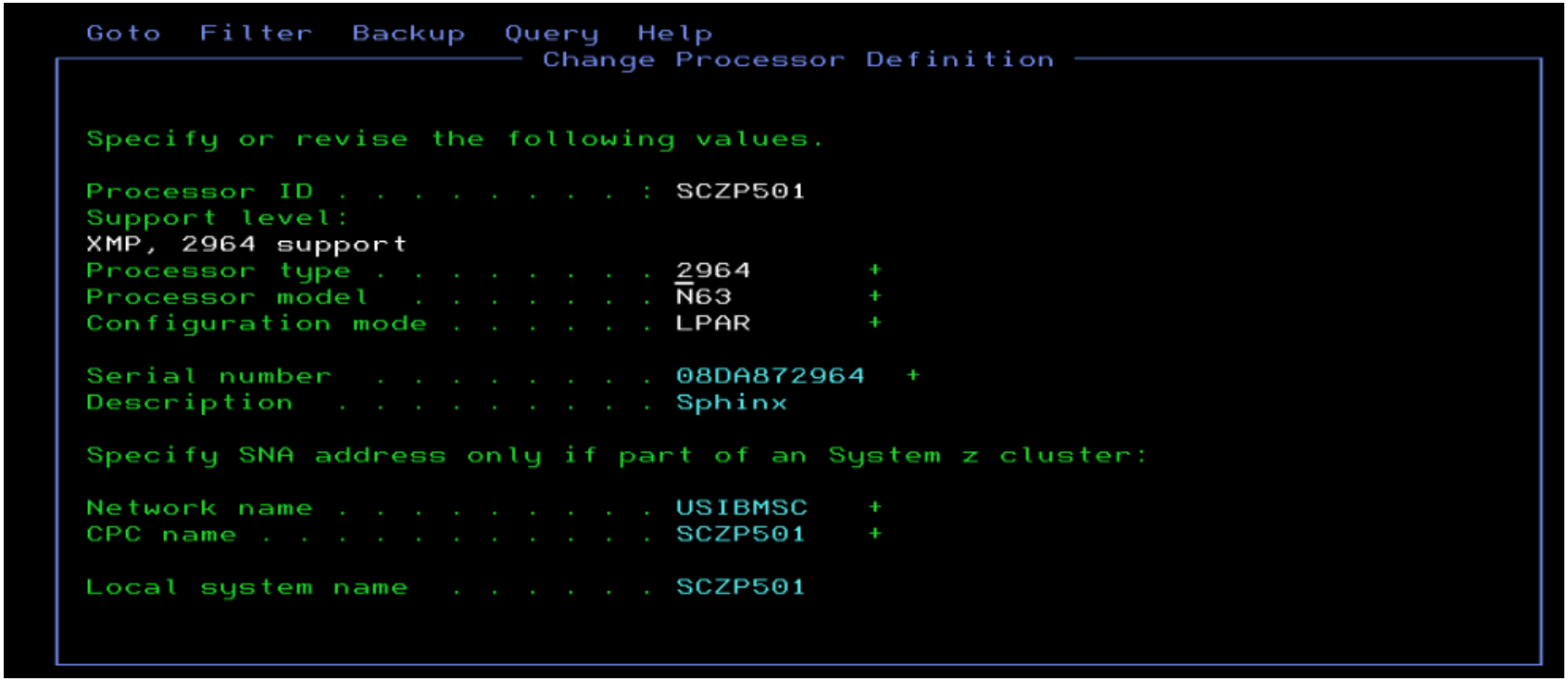
- HCD prereqs:
  - HCD APAR / PTF (OA46010)
  - Before you can define FID Type ISM you must first update your processor definition (see example in next charts)
- Notes:
  1. The maximum value that can be configured for a PCI FID (any FID type) is x0FFF
  2. The maximum number of VFs (FIDs) that can be configured for the same ISM PCHID = 255.

Change Processor (HCD Processor List)

```
  Goto  Filter  Backup  Query  Help
--------------------------------------------------------------------------------
                              Processor List          Row 1 of 3 More:        >
Command ===> ____________________________________________________ Scroll ===> CSR

Select one or more processors, then press Enter. To add, use F11.

/ Proc. ID Type +    Model +   Mode+ Serial-# + Description
_ SCZP401  2827      H43       LPAR  00B8D72827 Helix
_ SCZP402  2827      H89       LPAR  0194D72827 Helix
c SCZP501  2964      N63       LPAR  08DA872964 Sphinx
```

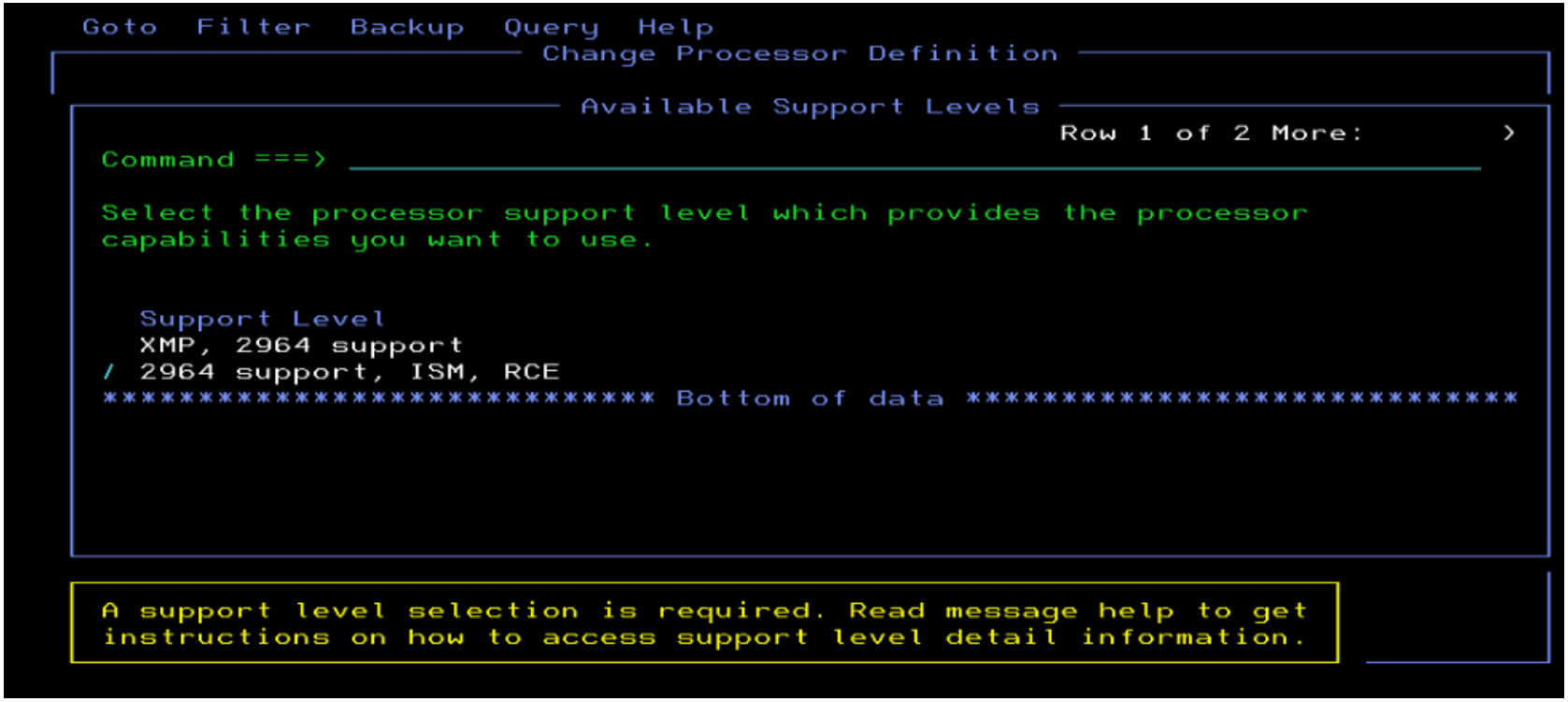
HCD option 1.3. C (Change) then press Enter

Change Processor Definition



Press enter

Change Processor Definition

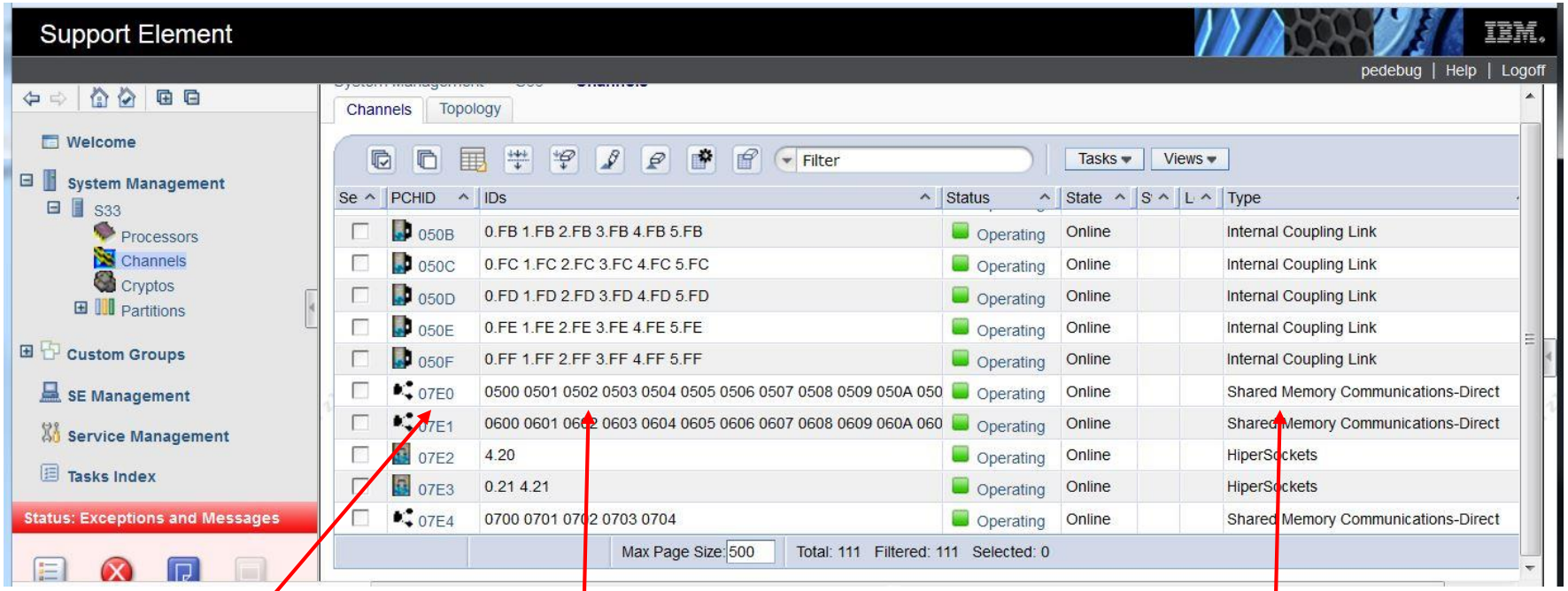


Select the new support level (slash) then Enter)

Backup Topic 2. Service Element (SE) Screenshot Examples

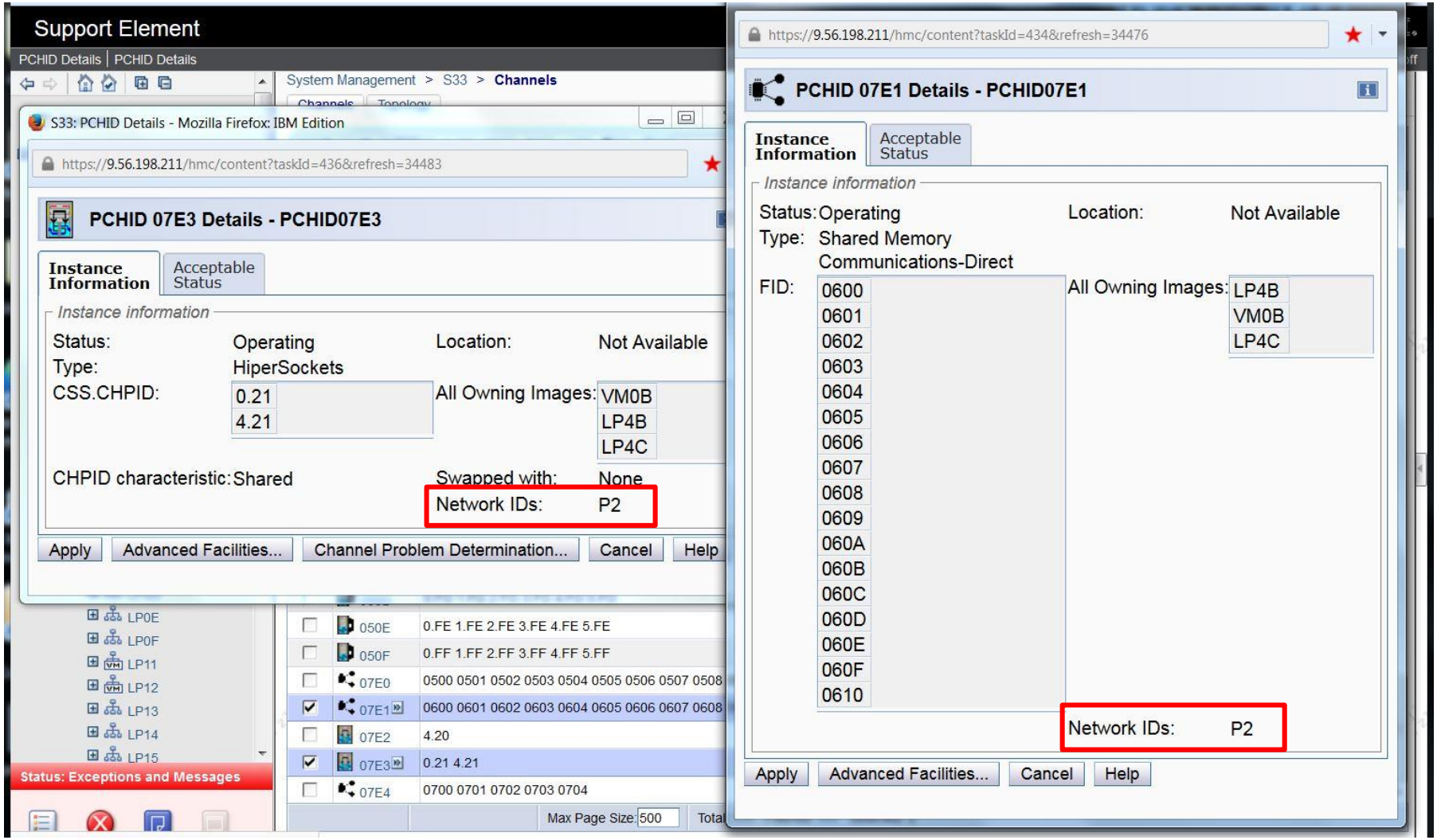
Note: This is not the same configuration as in the previous HCD example

Channel View



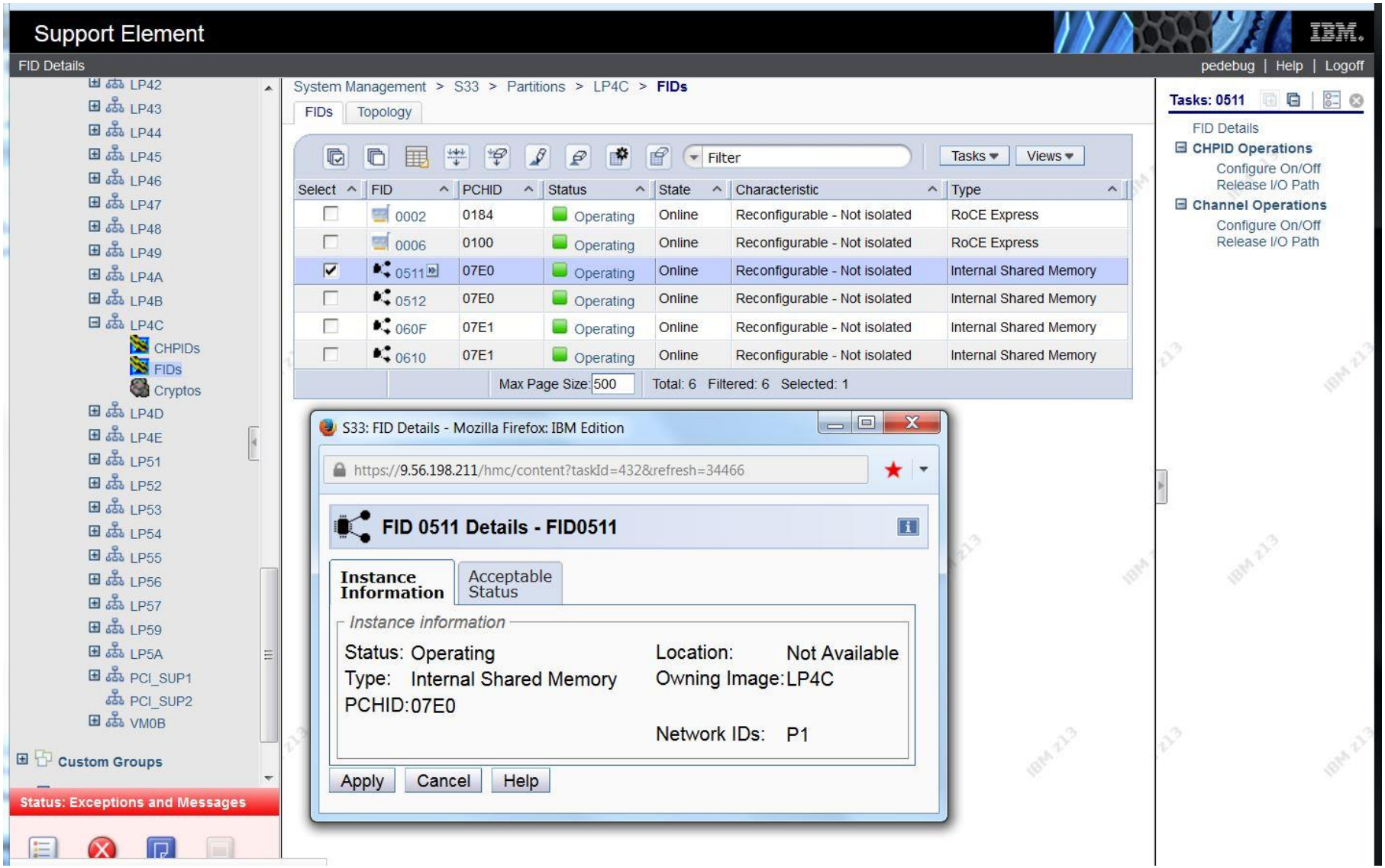
- VCHID range 07C0 -07FF is shared with HiperSockets VCHIDs
- Up to 255 FIDs per VCHID
- FIDs are unique per System

Channel Details



Network ID per VCHID,
associates an ISM VCHID with an HiperSockets VCHID (or OSA CHID)

FID View



Display TRL with CONTROL=ISM will show all ISM TRLEs

```
D NET,TRL,CONTROL=ISM
IST097I DISPLAY ACCEPTED
IST350I DISPLAY TYPE = TRL 725
IST924I -------------------------------------------------------------
IST1954I TRL MAJOR NODE = ISTTRL
IST1314I TRLE = IUT00501  STATUS = ACTIV       CONTROL = ISM
IST1314I TRLE = IUT00500  STATUS = ACTIV       CONTROL = ISM
IST1314I TRLE = IUT00601  STATUS = ACTIV       CONTROL = ISM
IST1314I TRLE = IUT00600  STATUS = ACTIV       CONTROL = ISM
IST1454I 4 TRLE(S) DISPLAYED
IST924I -------------------------------------------------------------
IST1954I TRL MAJOR NODE = HUBTRLES
IST1454I 0 TRLE(S) DISPLAYED
IST924I -------------------------------------------------------------
IST1954I TRL MAJOR NODE = VTMTRLES
IST172I NO TRLES EXIST
IST1454I 0 TRLE(S) DISPLAYED
IST924I -------------------------------------------------------------
IST1954I TRL MAJOR NODE = LOCTRLES
IST1454I 0 TRLE(S) DISPLAYED
IST924I -------------------------------------------------------------
IST1954I TRL MAJOR NODE = NETMTRLS
IST1454I 0 TRLE(S) DISPLAYED
IST314I END
```

Display TRL for an ISM TRLE

Shows the detailed TRLE information

```
D NET,TRL,TRLE=IUT00600
IST097I DISPLAY ACCEPTED
IST075I NAME = IUT00600, TYPE = TRLE 729
IST1954I TRL MAJOR NODE = ISTTRL
IST486I STATUS= ACTIV, DESIRED STATE= ACTIV
IST087I TYPE = *NA*                , CONTROL = ISM , HPDT =
*NA*
IST2418I SMCD PFID = 0600  VCHID = 07E1  PNETID = P2
IST2417I VFN = 0001
IST924I -----------------------------------------------
-----------
IST1717I ULPID = TCPIP2 ULP INTERFACE = EZAISM01
IST1724I I/O TRACE = OFF  TRACE LENGTH = *NA*
IST314I END
```

End of Material