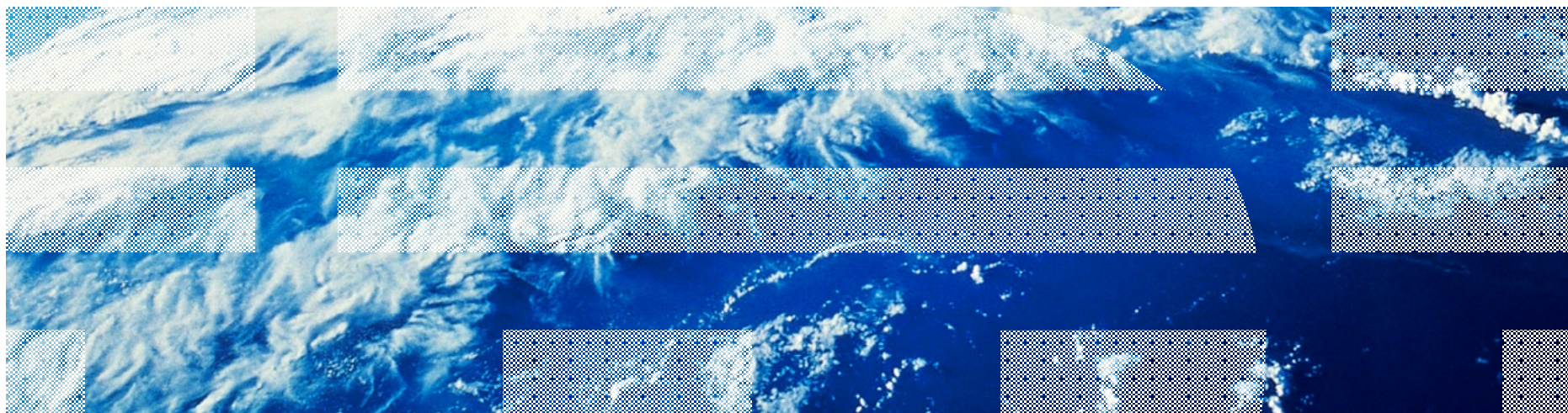# New Networking Options with z/VSE

Ingo Franzki

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the 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.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Agenda

§ Networking Overview

§ TCP/IP Products
- IPv6/VSE
- TCP/IP for VSE/ESA

§ IPv6 basics

§ Attachments
- OSA Express
- HiperSockets

§ Laver 2 & Layer 3 Support

§ VLAN Support

§ IEDN Support

§ Fast Path to Linux on System z

§ Tuning Tips

Networking with z/VSE - Overview

IBM System z
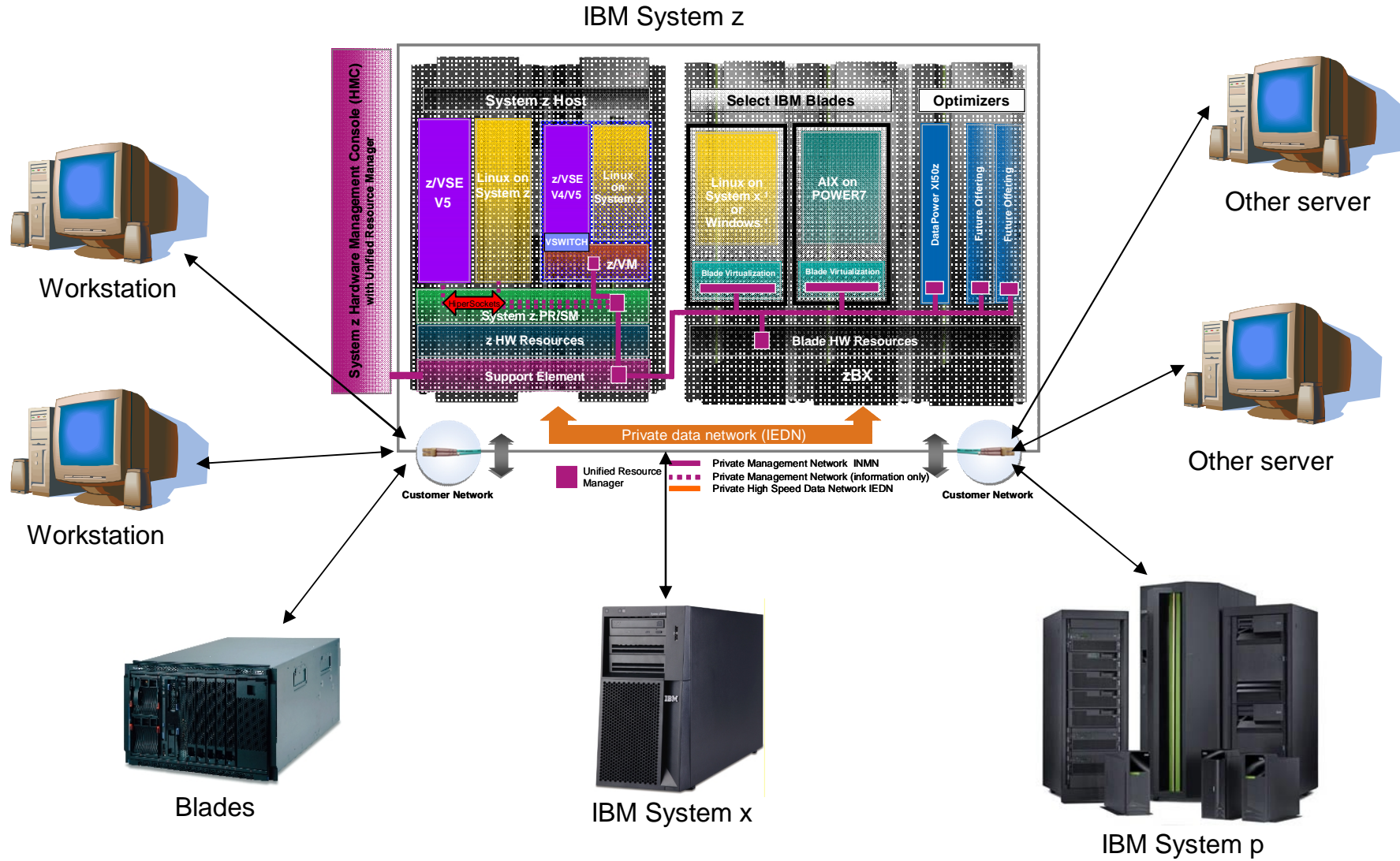
TCP/IP Products

§ **IPv6/VSE V1.1** (licensed from Barnard Software, Inc)
  – IPv6/VSE provides:
    • An **IPv6 TCP/IP stack**
    • IPv6 application programming interfaces (APIs)
    • IPv6-enabled applications
  – The IPv6 TCP/IP stack of IPv6/VSE can be run concurrently with an IPv4 TCP/IP stack within one z/VSE system
  – The IPv6/VSE product also includes
    • A **full-function IPv4 TCP/IP stack**
    • IPv4 application programming interfaces
    • IPv4 applications.
  – The IPv4 TCP/IP stack does not require the IPv6 TCP/IP stack to be active.
  – With **z/VSE V5.1** IPv6/VSE became a **base product**. With z/VSE V4.3 it is an optional product
  – Supports Layer 2 and 3 mode (z/VSE V5.1)
  – Supports Virtual LAN (VLAN) (z/VSE V5.1)

§ **TCP/IP for VSE/ESA V1.5** (licensed from CSI International)
  – Supports IPv4 only
  – Layer 3 mode only
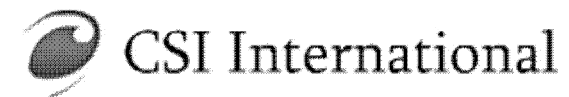
§ **Fast Path to Linux on System z** (part of z/VSE V4.3 or later)

IPv6 Basics

§ IPv6 Addresses

- 128 Bits in length (16 bytes)
  - 4 times larger than a IPv4 address
- Up to 2^{128} (about $3.4{\times}10^{38}$) unique addresses
  - That's approximately $5{\times}10^{28}$ (roughly 2^{95}) addresses for <u>each</u> of the roughly 6.8 billion ($6.8{\times}10^{9}$) people alive in 2010.
  - In another perspective, this is the same number of IP addresses per person as the number of atoms in a metric ton of carbon!

- IPv6 address are usually written as eight groups of four hexadecimal digits (each group representing 16 bits, or two bytes), where each group is separated by a colon (:).
  - Example: 2001:0db8:85a3:08d3:1319:8a2e:0370:7344
- Leading zeroes in a group may be omitted (but at least one digit per group must be left):
  - 2001:0db8:0000:08d3:0000:8a2e:0070:7344 is the same as
    2001:db8:0:8d3:0:8a2e:70:7344
- A string of consecutive all-zero groups may be replaced by two colons. In order to avoid ambiguity, this simplification may only be applied once:
  - 2001:db8:0:0:0:0:1428:57ab is the same as 2001:db8::1428:57ab
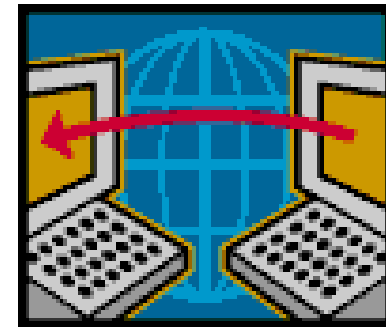
IPv6 Basics - addressing

§ IPv6 Addresses gets assigned to interfaces (network adapters)

§ One interface (network adapter) can have multiple IPv6 addresses
  – Assigned address
  – Link local address (FE80::/10)
    • typically built using the MAC address

§ Every IPv6 address has a "scope":
  – Link local
  – Site local
  – Global

§ IPv6 addresses are typically composed of two logical parts:
  – Routing prefix
    • The length of the prefix is specified with the address separated by a slash: /64
  – Interface identifier
    • Usually automatically determined from the MAC address of the interface
  – Internet service providers (ISPs) usually get assigned the first 32 bits (or less) as their network from a regional internet registry (RIR)

IPv6 Basics – auto configuration

Goal: Plug 'n' Play network

§ An IPv6 endpoint needs at least 3 pieces of information to be able to communicate:
  – IPv6 address
  – IPv6 network
  – IPv6 gateway

§ Right after the start, an endpoint only knows its link local address
  – E.g. determined from the MAC address of the interface
 – With that, it can only communicate within its local network segment

§ The interface then uses Neighbor Discovery Protocols to search for routes in its local network segment
  – It sends requests to the multicast address FF02::2, which all routers are reachable at (Router Solicitation)
  – Available routes then reply with information about the network

§ Router also send Router Advertisements in regular intervals to all hosts in the network(s) segment they are responsible for

§ ICMPv6 provides essential functions in an IPv6 network
  – Address Resolution Protocol (ARP) is replaced by Neighbor Discovery Protocol (NDP)
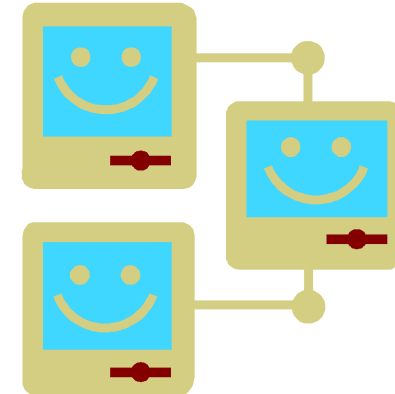
Migration from IPv4 to IPv6

§ Contrary to popular belief, IPv6 is not backward compatible !

§ But: IPv4 and IPv6 networks can be used concurrently over the same cable and with the same endpoint

Transition methods:

§ **Dual IP Stacks**

– That's the easiest possibility

– The IP stack supports both protocols concurrently

• Examples: Linux since Kernel 2.6, Windows since XP SP1

– Existing IPv4 applications can continue to run unchanged

• Applications can be IPv6-enabled over time, one after the other
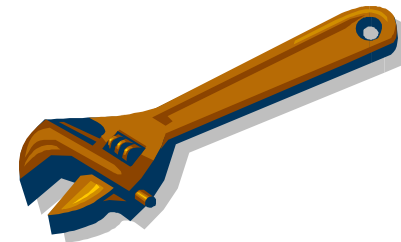
§ **Tunneling**

– IPv6 packets are sent as payload of other protocols (usually IPv4) to a tunneling broker, which is located in an IPv6 network. The broker extracts the IPv6 packet from the payload and sends it as IPv6 packet through IPv6 routing to the final destination.

• Example: 6in4 using Tunneling-Broker

Migration from IPv4 to IPv6

Which infrastructure parts needs to be migrated?

§ **Layer 1 devices (e.g. hubs)**

– Those are completely transparent for IPv6

§ **Layer 2 devices (switches)**

– Devices which have been purchased within the last 10 years most likely support IPv6 already

§ **Layer 3 devices (routers)**

– Usually not required for local LANs

– Today most router manufacturer provide IPv6 capable routers

– Routers that use Multiprotocol Label Switching (MPLS) are protocol independent

§ **Endpoints (PCs, Server, etc.)**

– Most modern operating systems support IPv6

§ **Applications**

– May have to be adapted (IPv6-enabled) to be able to work with IPv6 addresses

Why should a z/VSE customer care about IPv6?

Independent on your concrete benefits

à You will have to care about IPv6,
  sooner or later!

Why?

- Your internet service provider (ISP) migrates to IPv6

- On 3 February 2011, the Number Resource Organization (NRO) announced that the free pool of available IPv4 addresses is now fully depleted.

- Your customers or partners are only reachable via IPv6 (e.g. China)

- Governmental organizations may only allow manufacturers of IPv6 capable products and applications to participate in advertised biddings

  - Example: The US Department of Defense (DoD) only allows products that are on the "Unified Capabilities Approved Products List" (UC APL) for its advertised biddings.

    - "This list is used by procurement offices in the DoD and the U.S. Federal agencies for ongoing purchases and acquisitions of IT equipment"

IPv6 enabled applications

The following applications and tools are part of the IPv6/VSE product:

§ FTP Server (POWER queues, VSAM catalogs, SAM file, z/VSE libraries, …)

§ Batch FTP Client

§ TN3270E server (TN3270/TN3270E Terminal & TN3270E Printer Sessions)

§ Network Time Protocol Server (NTP server)

§ Network Time Protocol Client (NTP client)

§ System Logger Client

§ Batch Email Client

§ Batch LPR

§ Batch Remote Execution client (REXEC)

§ Batch PING

§ GZIP data compression

§ REXX automation

Home grown applications may need to get adapted (IPv6 enabled)

© 2012 IBM Corporation

OSA Express

OSA Express 4s, OSA Express 3, OSA Express 2

§ OSA Express supports various features such as:

- 10 Gigabit Ethernet
- Gigabit Ethernet
- 1000BASE-T Ethernet

§ CHPID types

- **OSC**   OSA-ICC (for emulation of TN3270E and non-SNA DFT 3270)
- **OSD**   Queue Direct Input/Output (QDIO) architecture
- **OSE**   non-QDIO Mode (OSA-2, for SNA/APPN connections)
- **OSN**   OSA-Express for NCP: Appears to z/VSE as a device-supporting channel data link control (CDLC) protocol.
- **OSX**   OSA-Express for zBX. Provides connectivity and access control to the Intra-Ensemble Data Network (IEDN) from z196 and z114 to Unified Resource Manager functions.

OSA Express in QDIO Mode

§ **For an OSA Express adapter in QDIO mode, you need 3 devices**
   - A read device
   - A write device
   - A datapath device

§ **Add the devices in the IPL procedure as device type OSAX:**
   - `ADD cuu1-cuu3,OSAX`

§ **In TCP/IP for VSE define a LINK:**
   - `DEFINE LINK,ID=...,TYPE=OSAX,`
       `DEV=cuu1 (or DEV=(cuu1,cuu2)),`
       `DATAPATH=cuu3,`
       `IPADDR=addr,`
       `...`

§ **In IPv6/VSE define a DEVICE:**
   - `DEVICE device_name OSAX cuu1 portname cuu3`

§ **For each LINK of an OSAX device, the TCP/IP partition requires 1050K partition GETVIS (ANY) space and 1050K for SETPFIX (ANY)**

CSI International

BSI BARNARD SOFTWARE, INC.

OSA Express Multi-Port support

§ **OSA Express 3 or later provides 2 ports per CHPID for selected features**

- Default is port 0
- To use port 1, you must specify this at the DEFINE LINK or DEVICE/LINK statement:
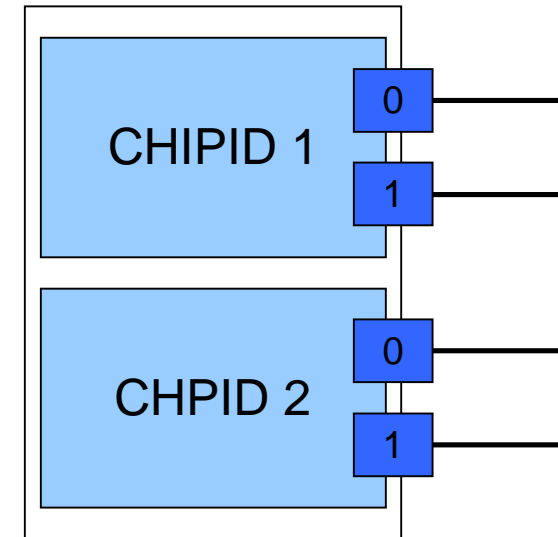 - **TCP/IP for VSE:**
    ```
    DEFINE LINK,ID=...,TYPE=OSAX,
            DEV=cuu1 (or DEV=(cuu1,cuu2)),
            DATAPATH=cuu3,
            OSAPORT=1,
            ...
    ```
  - **IPv6/VSE:**
    ```
    DEVICE device_name OSAX cuu1 portname cuu3
    LINK device_name adapter_no IPv6_addr netmask mtu
    ```
- For CHIPID type OSE (non-QDIO mode) you must use OSA/SF to select the OSA port

CHIPID 1

0

1

CHPID 2

0

1

HiperSockets

§ **"Network within the box" functionality**

   – allows high speed any-to-any connectivity among operating systems

   – without requiring any physical cabling

§ **CHPID type IQD**

   – Uses the QDIO (Queue Direct I/O) architecture

   – For an HiperSockets adapter, you need 3 devices

- A read device
- A write device
- A datapath device

   – Add the devices in the IPL procedure as device type OSAX with mode 01:

- `ADD cuu1-cuu3,OSAX,01`

   – Frame size is defined via CHPARM parameter (formerly OS=nn):

- CHPARM=00 (default): 16K (MTU=8K)
- CHPARM=40   24K (MTU=16K)
- CHPARM=80   40K (MTU=32K)
- CHPARM=C0   64K (MTU=56K)
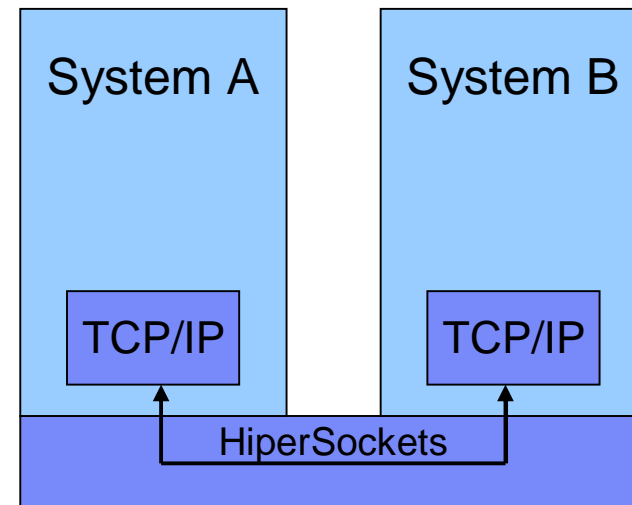
Layer 2 vs. Layer 3 Mode

§ **Layer 2:**

- TCP/IP stack passes a **frame** to the network card
- Addressing uses **MAC addresses**
- TCP/IP stack must perform ARP to translate IP to MAC

§ **Layer 3:**

- TCP/IP Stack passes an (IP) **packet** or **datagram** to the network card
- Addressing uses IP addresses (IPv4 or IPv6)
- The network card performs ARP to translate IPv4 to MAC

OSI Model:

Data	**7. Application Layer**	Application
Data	**6. Presentation Layer**	representation encryption
Data	**5. Session Layer**	Inter host comm.
Segment	**4. Transport Layer**	Flow control
Packet/ Datagram	**3. Network Layer**	Logical addressing
Frame	**2. Data Link Layer**	Physical addressing
Bit	**1. Physical Layer**	Media

Layer 2 vs. Layer 3 Mode (continued)

§ **Layer 2:**

– Supported by **IPv6/VSE** product (BSI) with **IPv6**
  OSA Express adapter (OSD, OSX) only, no HiperSockets

§ **Layer 3:**

– Supported by **IPv6/VSE product** (BSI) with **IPv4 and IPv6**
– Supported by **TCP/IP for VSE** product (CSI) with **IPv4**

§ **VSWITCH:**

– z/VM allows to define VSWITCH in Layer 2 or layer 3 mode
– z/VSE V4.2 and 4.3:
  • Supports Layer 3 VSWITCH (IPv4 only)
– z/VSE V5.1:
  • Supports Layer 2 VSWITCH (IPv4 and IPv6)
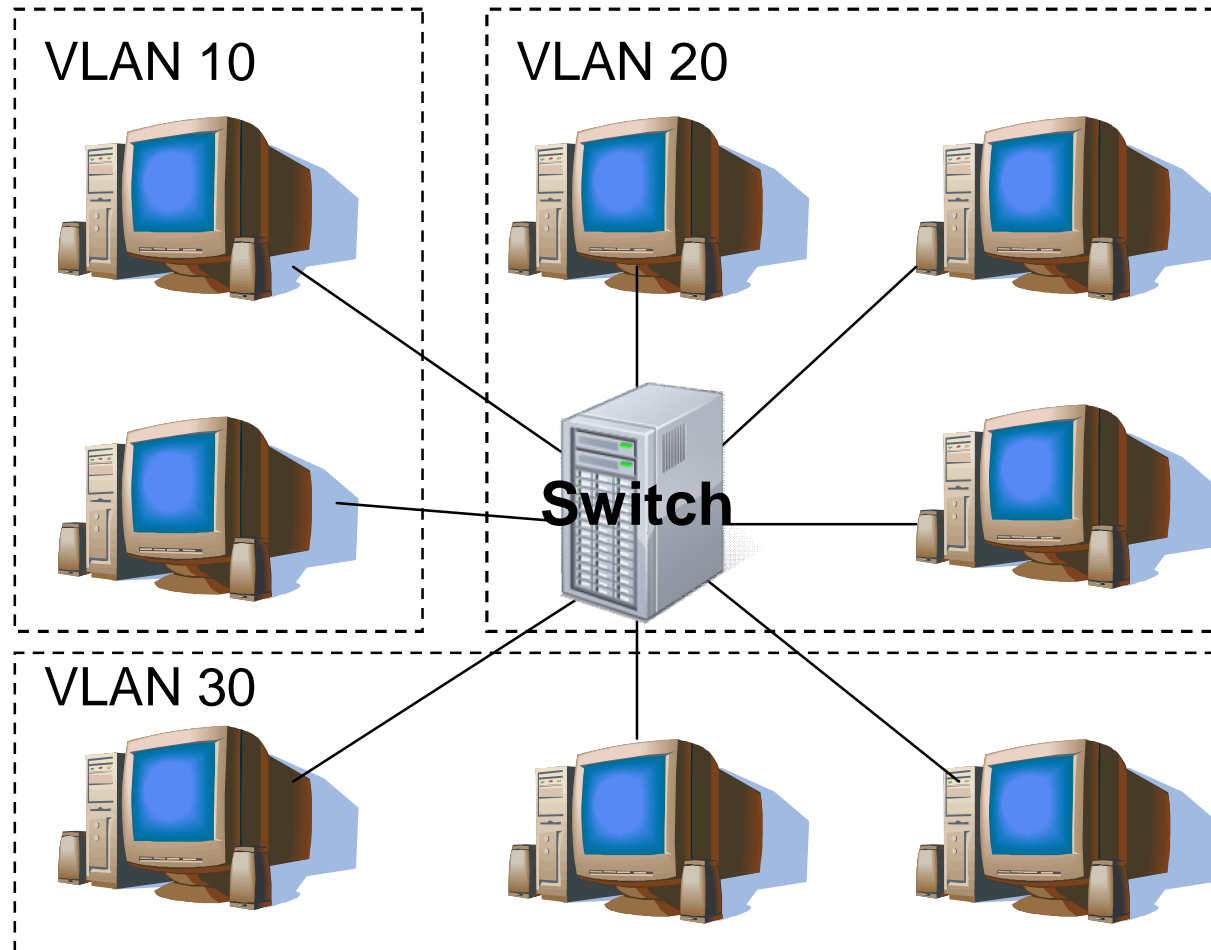 • Supports Layer 3 VSWITCH (IPv4 only)

à Be carefully when connecting z/VSE systems to already existing VSWITCHes

Virtual LAN (VLAN) - Overview

§ VLAN allows a physical network to be divided administratively into separate logical networks

§ These logical networks operate as if they are physically independent of each other

VLAN 10

VLAN 20

Switch

VLAN 30

Virtual LAN (VLAN) – Frame Tagging

§ **A VLAN tag is inserted into the Link Layer Header**

- **3 bit priority:** an be used to prioritize different classes of traffic (voice, video, data)

- **12 bit VLAN ID:** specifies the VLAN to which the frame belongs
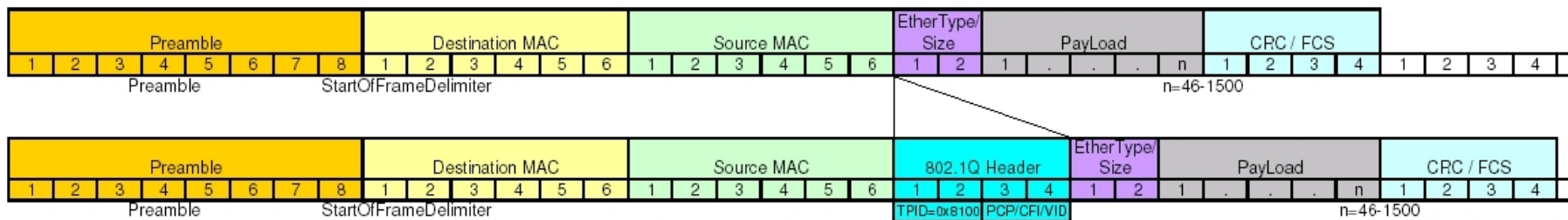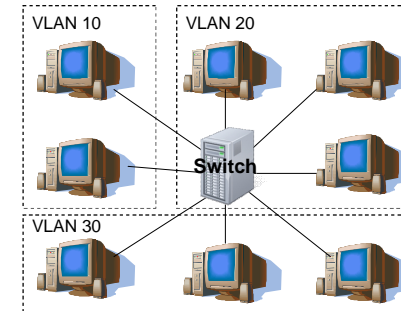
Source: Wikipedia: http://en.wikipedia.org/wiki/File:TCPIP_802.1Q.jpg

Virtual LAN (VLAN) – Trunc Port / Access Port

§ Switches have different types of ports

- **Access Port**
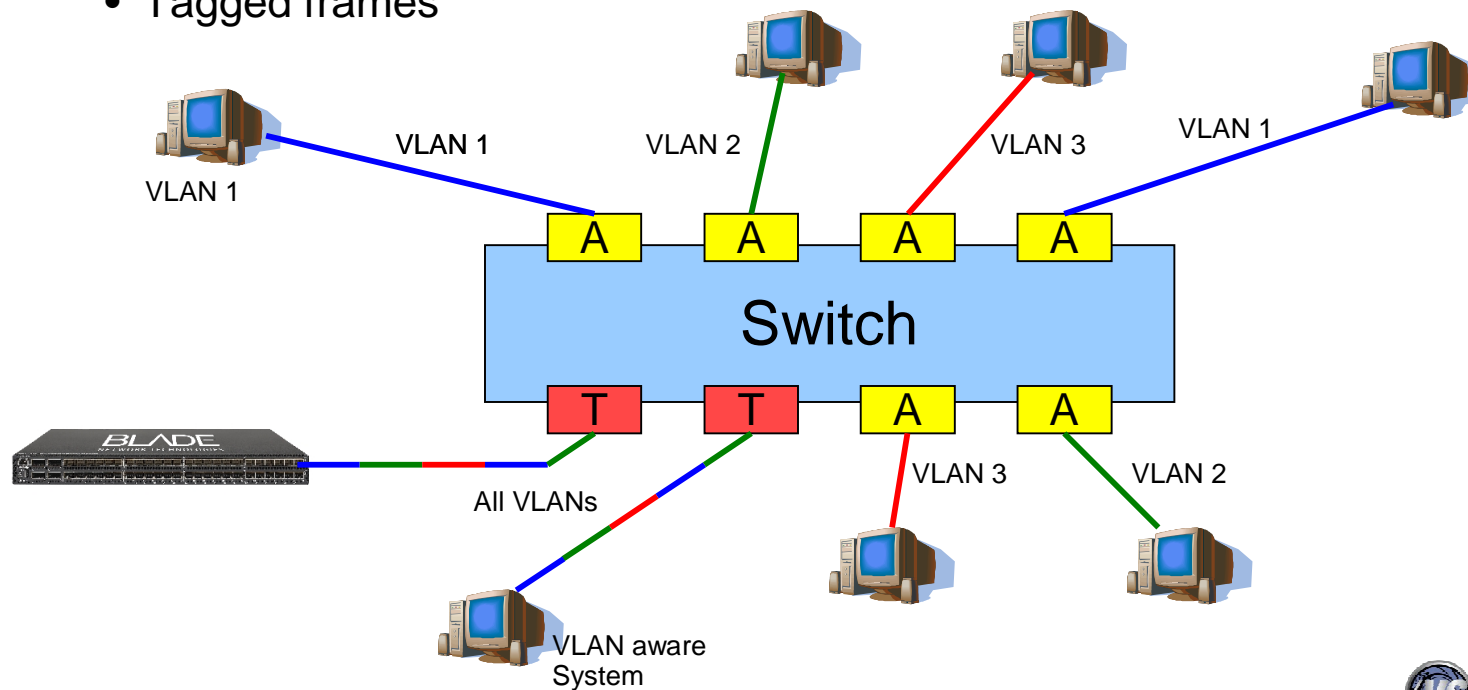 - Not VLAN-aware
  - Un-tagged frames
- **Trunc Port**
  - VLAN-aware
  - Tagged frames

VLAN 1

VLAN 1

VLAN 2

VLAN 3

VLAN 1

| A | A | A | A |

Switch

| T | T | A | A |

All VLANs

VLAN 3

VLAN 2

VLAN aware
System

Virtual LAN (VLAN) – z/VSE support

§ **z/VSE provides VLAN support for OSA Express (CHPID type OSD and OSX) and HiperSockets devices**

- In a **Layer 3** configuration, VLANs can be **transparently** used by **IPv6/VSE** and **TCP/IP for VSE/ESA**

- If you wish to configure VLANs for OSA-Express (CHPID type OSD and OSX) devices in a **Layer 2** configuration that carries **IPv6 traffic**, you require the **IPv6/VSE** product

§ **You can use one of the following two ways to configure your system to use VLAN:**

1. **Configure** one or more VLANs **in the TCP/IP stack** of **IPv6/VSE**

   - For details of IPv6/VSE commands, refer to IPv6/VSE Installation Guide

2. **Generate** and catalog phase **IJBOCONF** containing the **Global VLANs** to be used with your OSAX devices

   - z/VSE provides skeleton SKOSACFG to generate phase IJBOCONF

   - The VLANs contained in IJBOCONF can be **transparently** used for **Layer 3** links by **IPv6/VSE** and **TCP/IP for VSE/ESA**
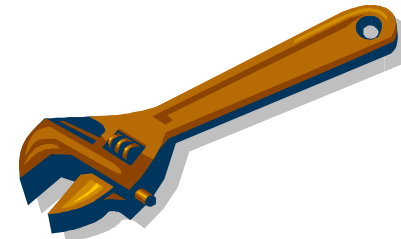
Intra-Ensemble Data Network (IEDN) support

§ **OSA-Express for zBX (CHPID type OSX)**

– Provides connectivity and access control to the Intra-Ensemble Data Network (IEDN) from zEnterprise 196 and 114 to Unified Resource Manager functions

§ **An Intra-Ensemble Data Network (IEDN) provides connectivity between:**

– A zEnterprise CEC (Central Electrical Complex) and System z Blade Center Extensions (zBXs)

– Two or more zEnterprise CECs

§ **z/VSE supports the IEDN network of a zEnterprise 196 or 114**

– **z/VSE V4.2, V4.3 and V5.1:**

  • z/VM VSWITCH and **OSDSIM** mode in a **z/VM 6.1** guest environment

– **z/VSE V5.1:**

  • **OSA Express for zBX** devices either in an **LPAR** or **z/VM** guest environment with **dedicated** OSAX devices

  • This requires **VLAN** support

© 2012 IBM Corporation
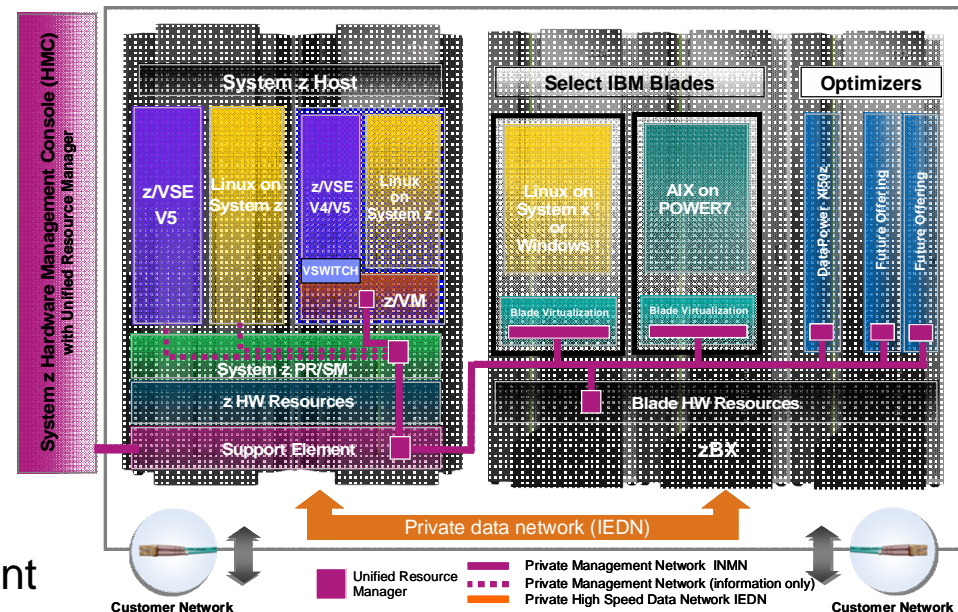
Linux Fast Path in a z/VM environment (z/VSE 4.3 or later)

Faster communication between z/VSE and Linux applications

Application

TCP/IP Stack

z/VSE Supervisor

z/VM

Application

TCP/IP Stack

Kernel

z/VM

HiperSockets, OSA, VSWITCH

network

Without LFP

Application

z/VSE Supervisor

IUCV

Application

LFP Daemon

TCP/IP Stack

Kernel

z/VM

System z10, z196, z114

network

network

With LFP under z/VM

24

© 2012 IBM Corporation
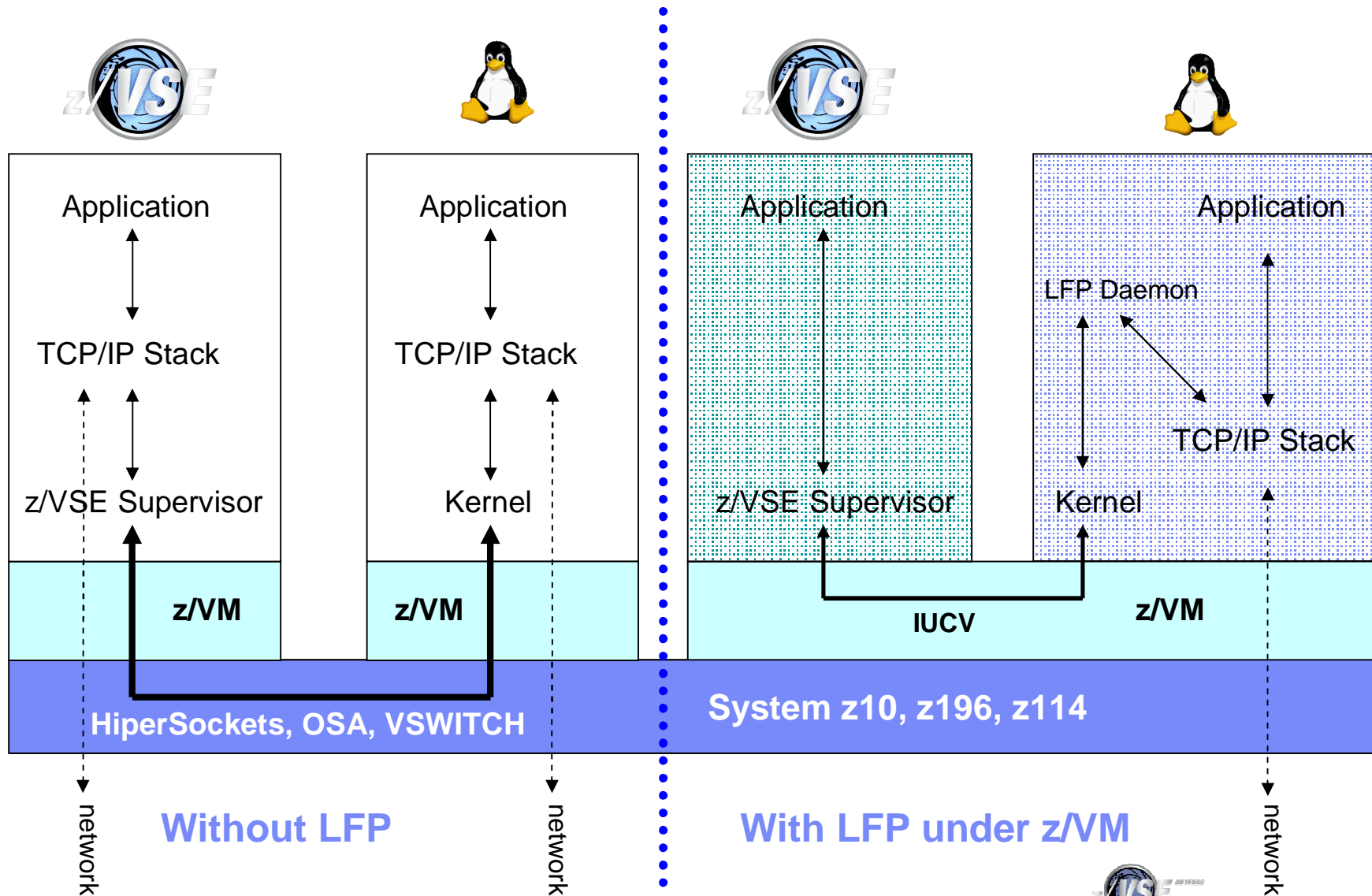
New: z/VSE z/VM IP Assist (VIA) (z/VSE 5.1 + z/VM 6.1)

With z/VM IP Assist (VIA), no Linux is needed to utilize the LFP advantage

NEW!

Application	Application	Application
↕	↕	↕
TCP/IP Stack	TCP/IP Stack	z/VSE Supervisor
↕	↕	
z/VSE Supervisor	Kernel	

z/VM | **z/VM** | IUCV → **VIA** | **z/VM 6.1**

HiperSockets, OSA, VSWITCH | **IBM zEnterprise (z196, z114)**

network | network | network

Without LFP | **With LFP + VIA**
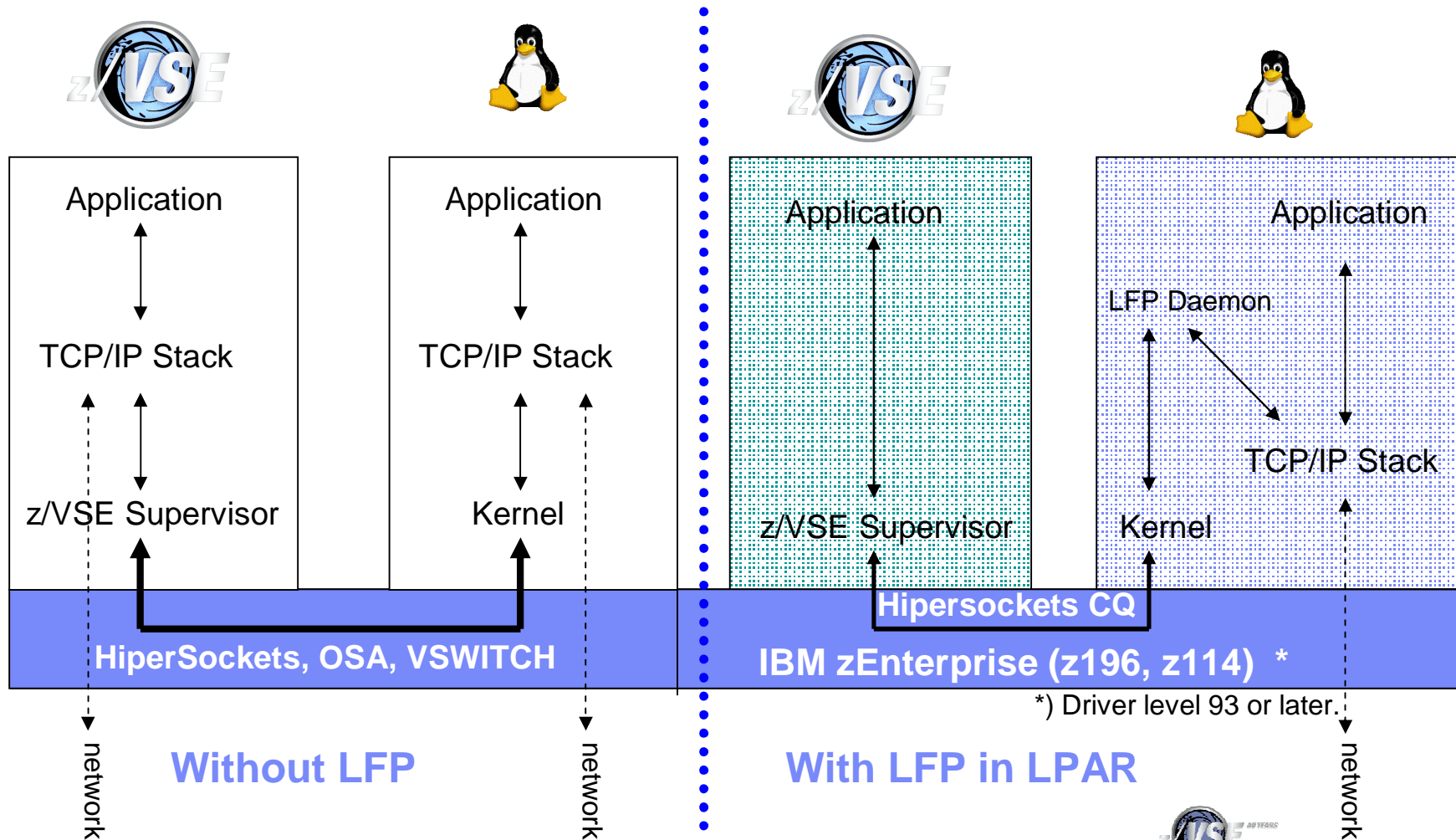
© 2012 IBM Corporation
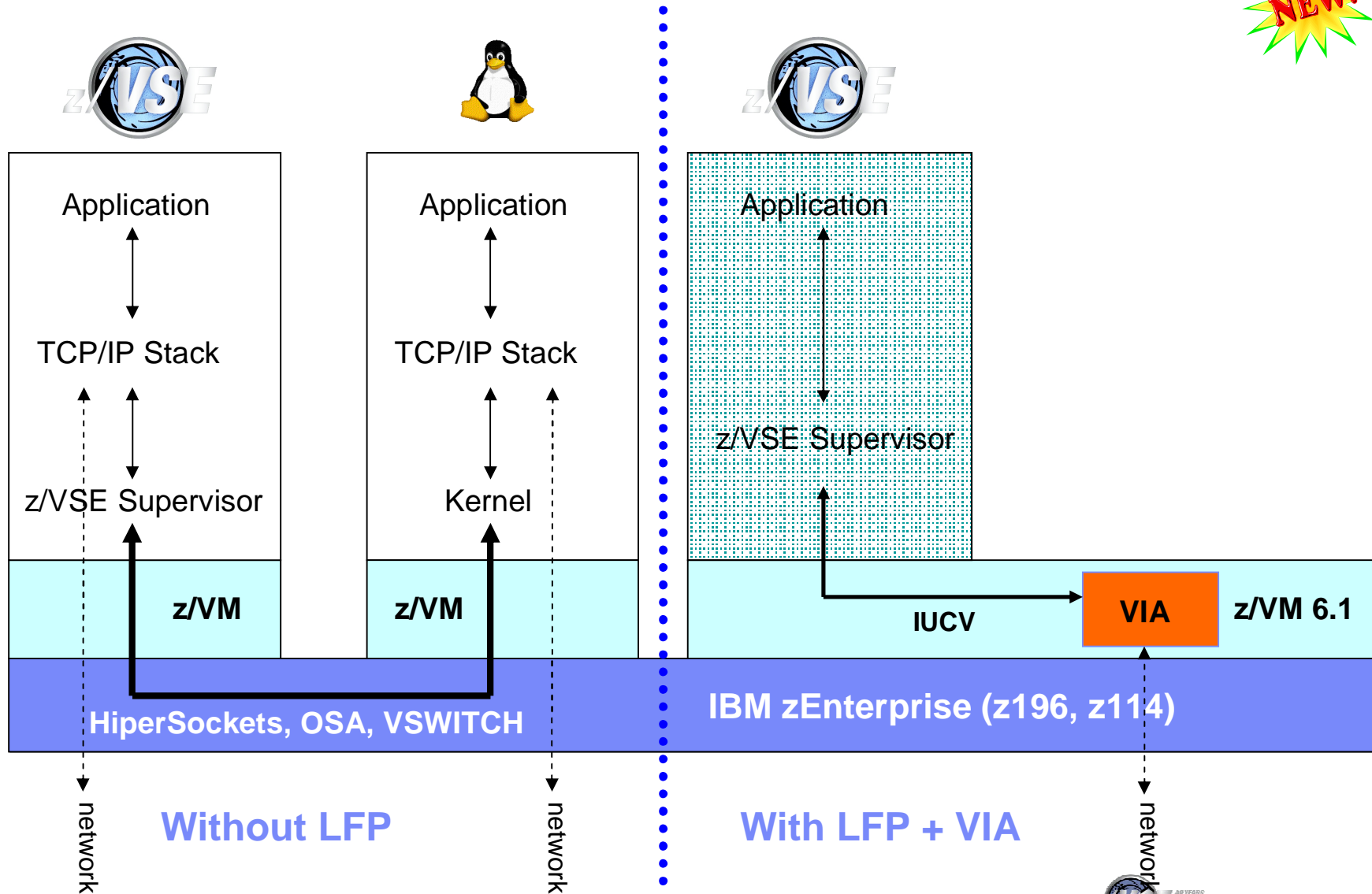
New: Linux Fast Path in an LPAR environment (z/VSE 5.1 + PTFs)

Faster communication between z/VSE and Linux applications

à Exploits the HiperSockets Completion-Queue support of IBM zEnterprise (z196, z114)

NEW!

Application	Application
↕	↕
TCP/IP Stack	TCP/IP Stack
↕	↕
z/VSE Supervisor	Kernel

HiperSockets, OSA, VSWITCH

Without LFP

network

Application	Application
↕	LFP Daemon
	↕ ↔ TCP/IP Stack
z/VSE Supervisor	Kernel

Hipersockets CQ

IBM zEnterprise (z196, z114) *

*) Driver level 93 or later.

With LFP in LPAR

network

Fast Path to Linux on System z (LFP)

è Linux Fast Path in a z/VM environment
- Both z/VSE and Linux on System z run as z/VM Guests in the same z/VM-mode LPAR on IBM z10, z114 or z196 servers
- Uses an IUCV connection between z/VSE and Linux

è Linux Fast Path in an LPAR environment
- Both z/VSE and Linux on System z run in their own LPARs on a zEnterprise server
- A HiperSockets connection is used between z/VSE and Linux on System z
- LFP requires the HiperSockets Completion Queue function that is available with a zEnterprise server (z196, z114)

§ The fast path to Linux on System z provides standard TCP/IP socket APIs for programs running on z/VSE
- Other than the basic socket API, no other tools are provided
- Since z/VSE V5.1: LFP supports IPv6

§ Allows selected TCP/IP applications to communicate with the TCP/IP stack on Linux without using a TCP/IP stack on z/VSE
- **All socket requests are transparently forwarded to a Linux on System z system**
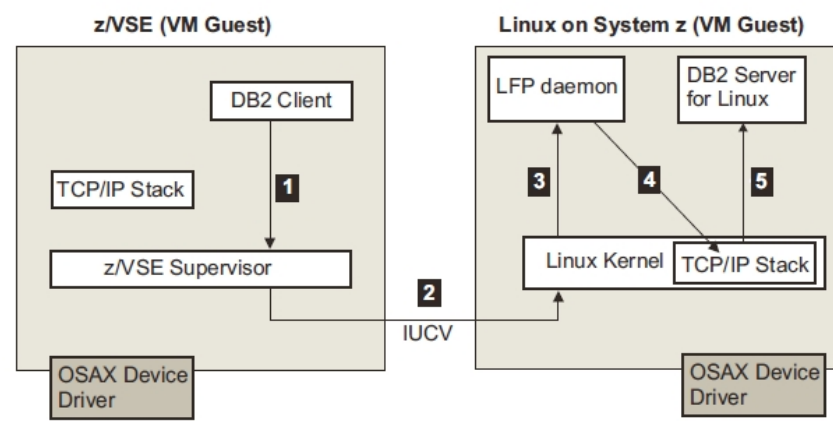
§ Possible performance increase due to:
- Less overhead for TCP/IP processing on z/VSE (TCP, sequence numbers and acknowledging, checksums, resends, etc)
- More reliable communication method (IUCV) compared to HiperSockets, which is a network device, with all its packet drops, resends, etc.

Communication flows when using Linux Fast Path

Using a TCP/IP stack (CSI/BSI):
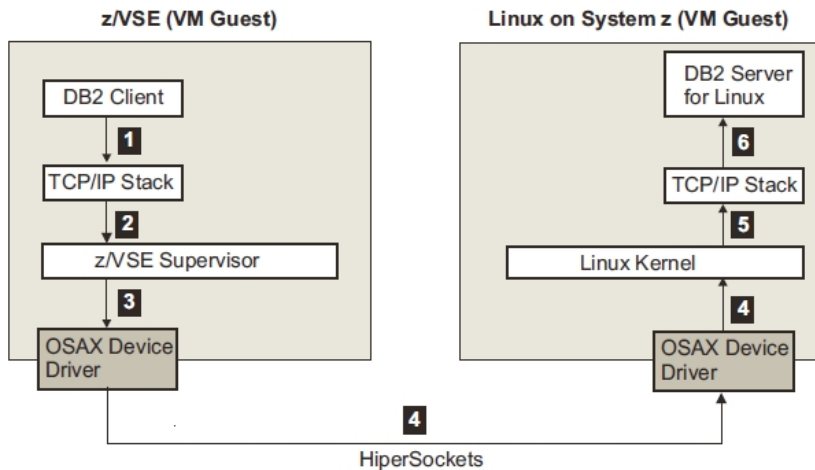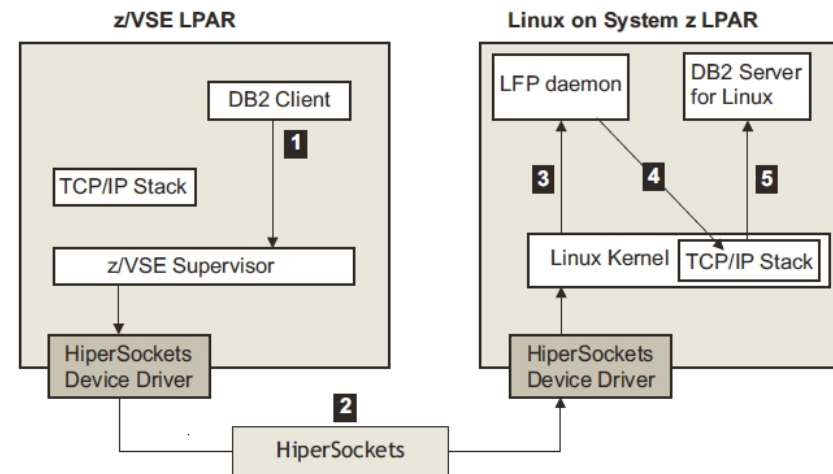

Using Linux Fast Path in a z/VM environment:

§ **Less overhead for TCP/IP processing on z/VSE**
  – Building of IP and TCP packets
  – Sequence numbers and acknowledging
  – Checksums
  – Retransmission of lost packets

§ **More reliable communication method compared to a traditional network device**
  – IUCV is a reliable communication method (z/VM)
  – HiperSockets Completion Queue support allows to build a reliable communication path (LPAR)

Using Linux Fast Path in an LPAR environment:

© 2012 IBM Corporation

Performance measurements using Linux Fast Path

Comparison TCP/IP for VSE versus Linux Fast Path (z/VM Environment):

Workload	TCP/IP for VSE	Linux Fast Path (LFP)	Difference

FTP (BSI FTP server)
§VSE à Linux (1GB)
    (NULL file, no I/O)
 	
§Linux à VSE (1GB)
    (NULL file, no I/O)
	19 MB/sec
	29% CPU (5% App + 24% TCPIP)
	
	21 MB/sec
	55% CPU (11% App + 44% TCPIP)
		72 MB/sec
		20% CPU (App)
		
		70 MB/sec
		20% CPU (App)
			3.7 times faster
			9% less CPU
			
			3.3 times faster
			35% less CPU

Socket Application (running 3 times) §VSE à Linux (100MB) §Linux à VSE (100MB)	4.6 MB/sec (*3 = 13.8 MB/sec) 9.7 MB/sec (*3 = 29.1 MB/sec) 26% CPU (3*1% App + 23% TCP/IP)	14.6 MB/sec (*3 = 43.8 MB/sec) 16.2 MB/sec (*3 = 48.6 MB/sec) 9 % CPU (3*3% App)	3.2 times faster 1,7 times faster 17% less CPU

Environment: IBM System z10 EC (2097-722). TCP/IP connection via shared OSA adapter.

à Significant benefits in transfer rate as well as CPU usage
à Reduced Sub Capacity Cost

z/VSE Fast Path to Linux on System z (LFP)

§ Most existing applications run unchanged with Linux Fast Path

– Provided they use one of the supported Socket API (LE/C, EZA or ASM SOCKET)

- And they do not use any CSI or BSI specific interface, features or functions
- Since z/VSE V5.1: LFP supports IPv6

§ IBM Applications supporting Linux Fast Path

– VSE Connector Server

– CICS Web Support

– VSE Web Services (SOAP) support (client and server)

– CICS Listener

– DB2/VSE Server and Client

– WebSphere MQ Server and Client

– VSAM Redirector

– VSE VTAPE

– VSE LDAP Support

– VSE Script Client

– POWER PNET

– TCP/IP-TOOLS included in IPv6/VSE product (e.g. FTP Server/Client)

§ Customer applications should run unchanged:

– Provided they use one of the supported Socket API (LE/C, EZA or ASM SOCKET)

TCP/IP Tuning: A simple picture might not be that simple in reality

© 2012 IBM Corporation
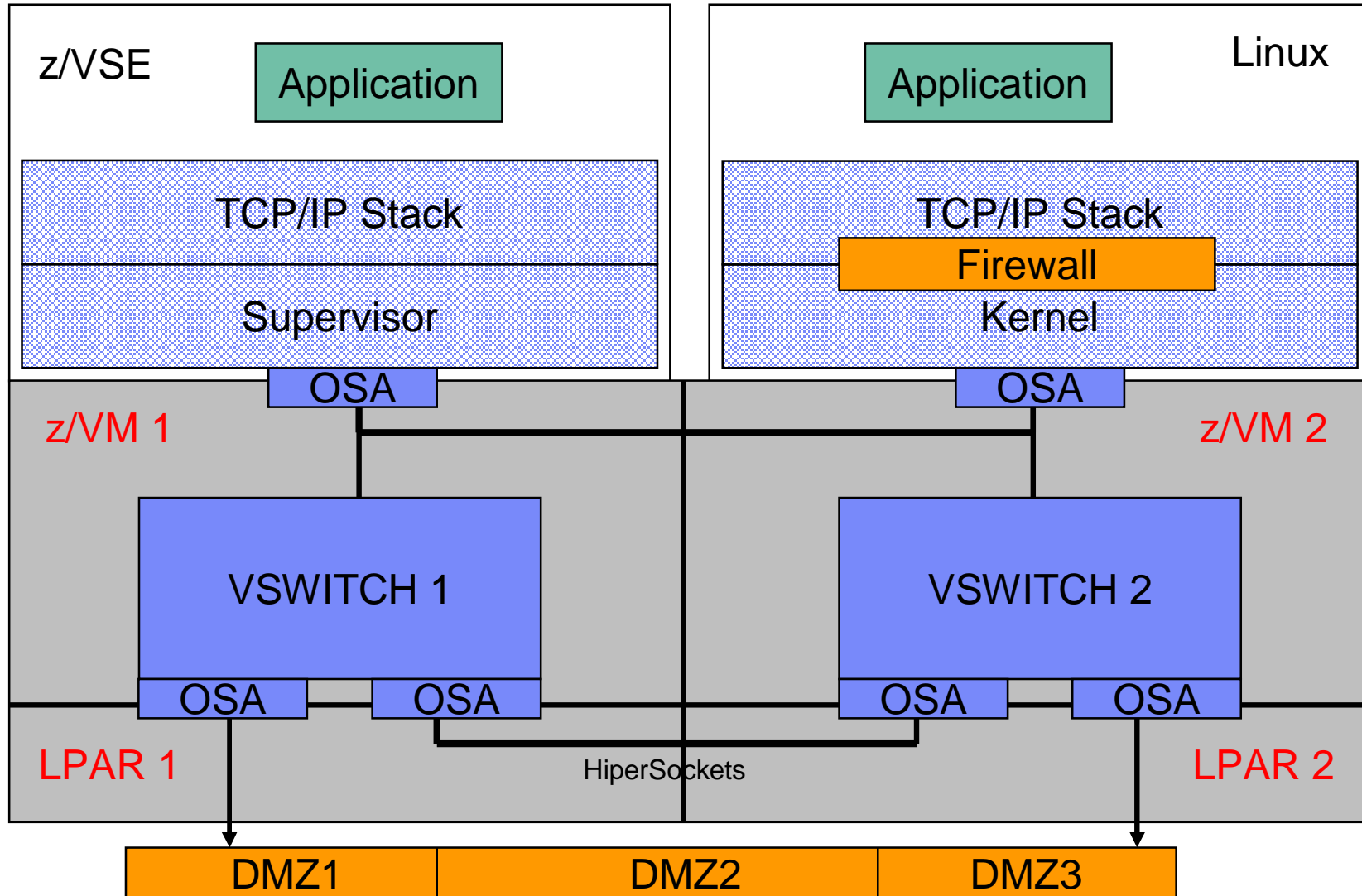
Shared OSA Adapter versus HiperSockets

To connect a z/VSE system with a Linux on System z you have 2 options:

1. Using a shared OSA Adapter

§ All traffic is passed through the OSA Adapter

§ The OSA Adapter has its own processor

§ Processing occurs asynchronous

§ Processing in OSA Adapter does not affect host processors

2. Using HiperSockets

§ Direct memory copy from one LPAR/Guest to the other

§ Memory copy is handled by the host processors

§ Processing occur synchronous

§ Consider mixed speed processors (full speed IFLs and throttled CPs)

à Memory copy performed by throttled CP is slower than memory copy performed by full speed IFL

TCP/IP Tuning: Performance tuning for HiperSockets

§ **When using HiperSockets to communicate between z/VSE and Linux, you may run into a "Target Buffer Full" condition**
  – This happens when z/VSE sends faster/more than Linux can receive
  – Per default Linux has 16 inbound buffers (64K per buffer = 1M per link)
  – To increase the number of buffers on Linux, use QETH option "buffer_count=128"
    • Use YAST to configure, or sysconfig scripts
    • Maximum of 128 buffers require 8MB of storage per link

§ **When TCP/IP for VSE encounters this situation (BUSY), it waits 500 msec until it retries to send the packet (Note: This does not apply to IPv6/VSE)**
  – Any additional packets to be sent are queued up
  – Problem can become dramatic, if more than 16 packets are queued up to be sent after BUSY situation
    • The resend will immediately flood the Linux buffers again, leading to the next BUSY situation, and so on....

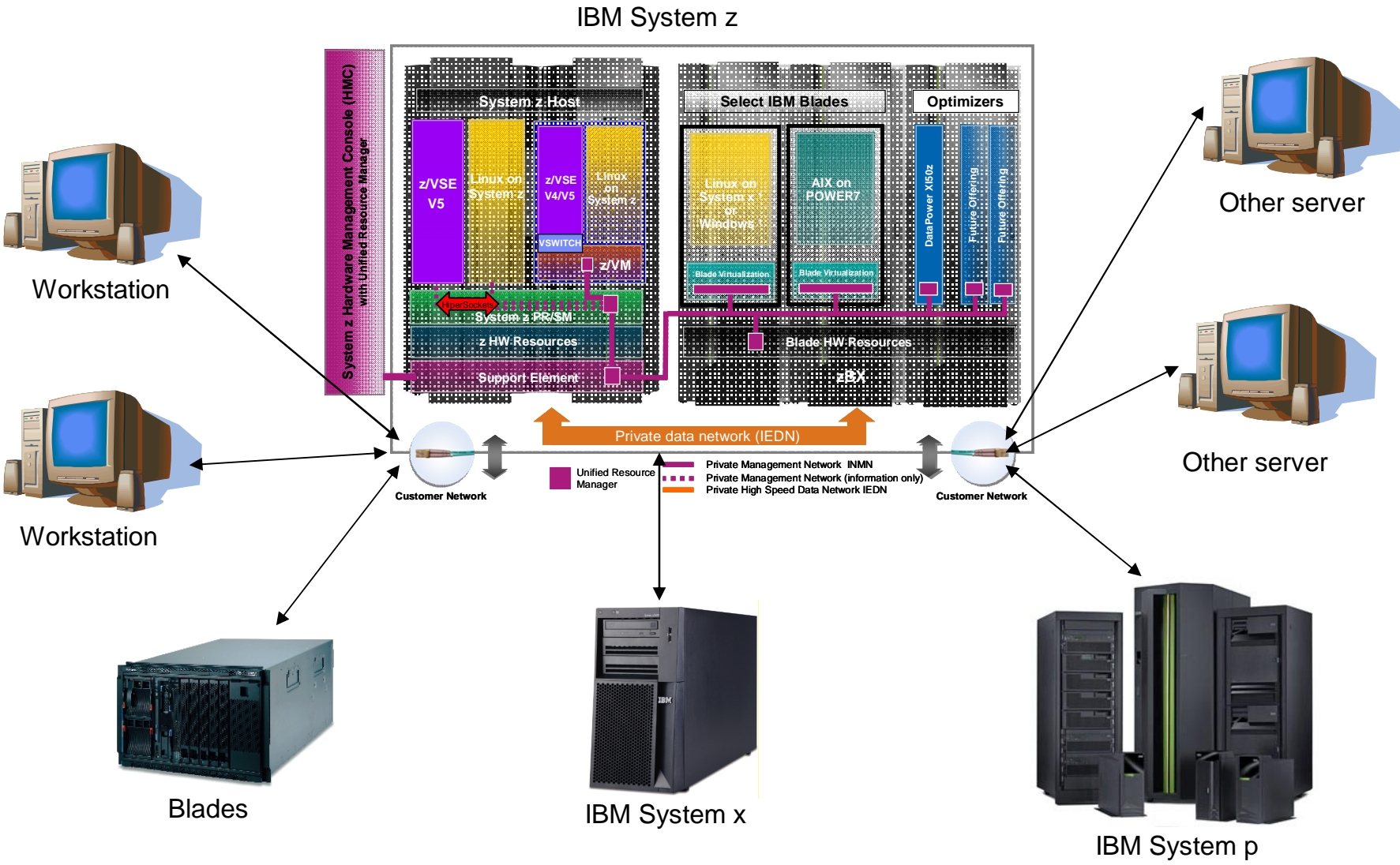§ **You can check via QUERY STATS,LINKID=xxxx [,RESET] if you have ever run into the BUSY situation** (RESET resets the counters)

```
C1 0065 0004: IPL615I  Busy mode..............................0   ß see here
C1 0065 0004: IPL615I  Busy mode, longest....................0
```

§ **You can configure a shorter BUSY wait time via DEFINE LINK command**
  – BUSY=nnn (shortest possible wait time is 100 msec)

Networking with z/VSE - Summary

IBM System z

System z Hardware Management Console (HMC) with Unified Resource Manager

System z Host

z/VSE V5

Linux on System z

z/VSE V4/V5

Linux on System z

VSWITCH

z/VM

HyperSockets

System z PR/SM

z HW Resources

Support Element

Select IBM Blades

Linux on System x or Windows

AIX on POWER7

Blade Virtualization

Blade Virtualization

Blade HW Resources

Optimizers

DataPower XI50z

Future Offering

Future Offering

Private data network (IEDN)

Customer Network

Customer Network

Unified Resource Manager

Private Management Network INMN
Private Management Network (information only)
Private High Speed Data Network IEDN

Workstation

Workstation

Blades

IBM System x

Other server

Other server

IBM System p

Questions ?