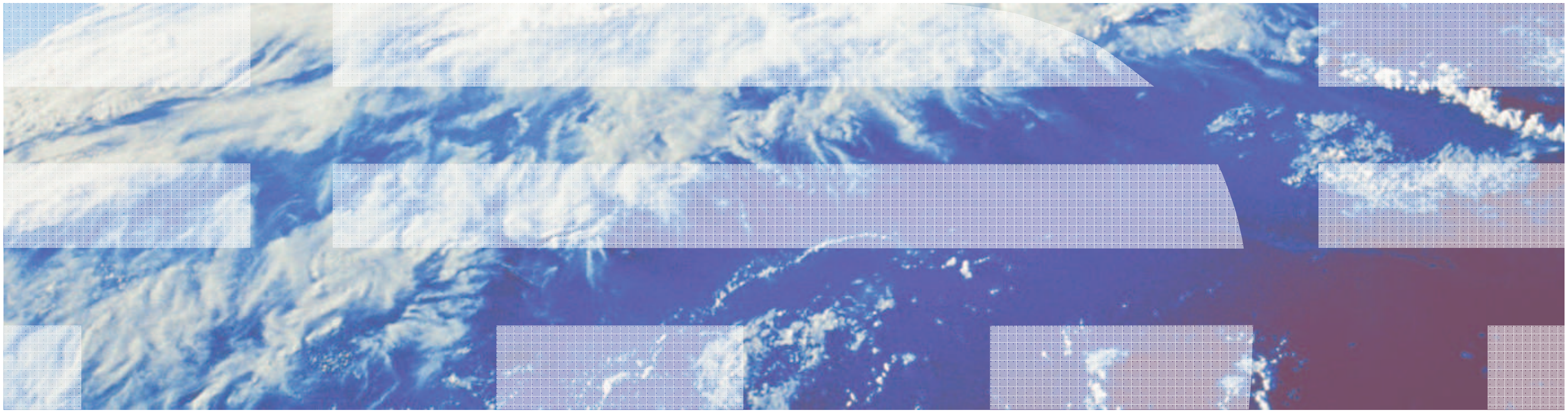
Alan Altmark
z/VM and Linux IT Consultant, IBM Lab Services
March 2011

The z/VM Virtual Switch
Advancing the Art of Virtual Networking

Session 8441



Note

References to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe on any of the intellectual property rights of IBM may be used instead. The evaluation and verification of operation in conjunction with other products, except those expressly designed by IBM, are the responsibility of the user.

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

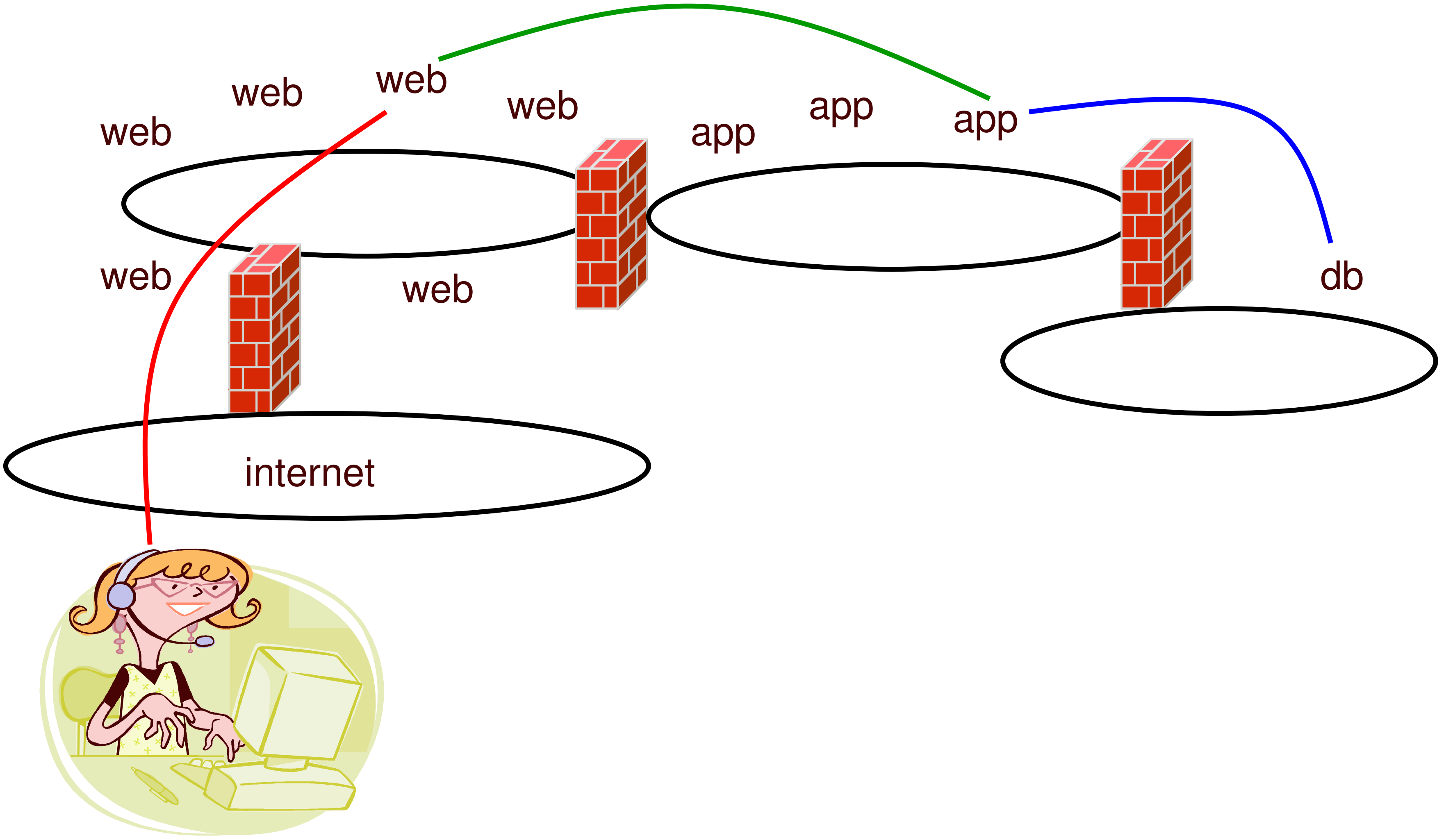
IBM IBM logo DB2 z/OS z/VM

Other company, product, and service names may be trademarks or service marks of others.

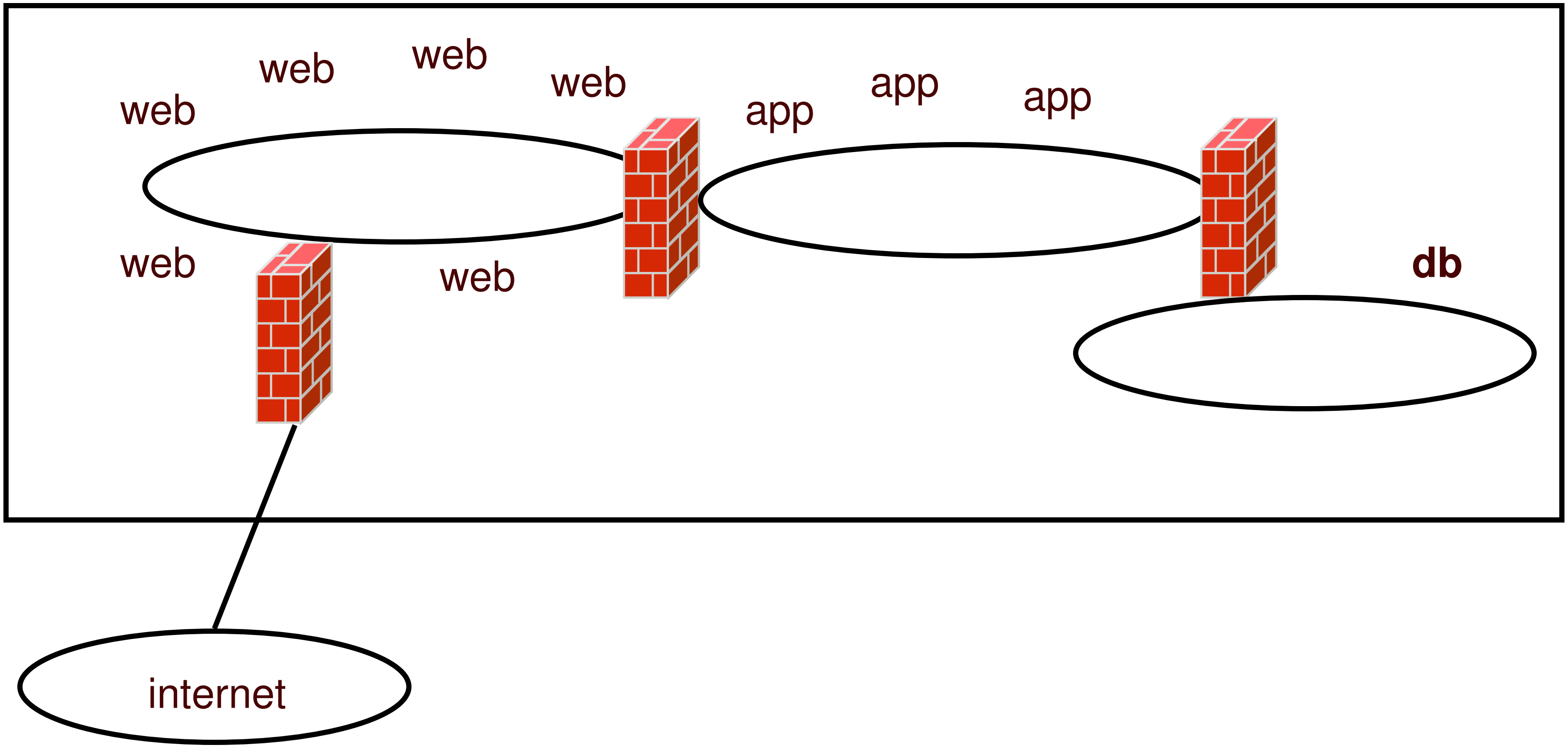
Topics

- Overview
- Multi-zone Networks
- Virtual Switch
- Virtual NIC

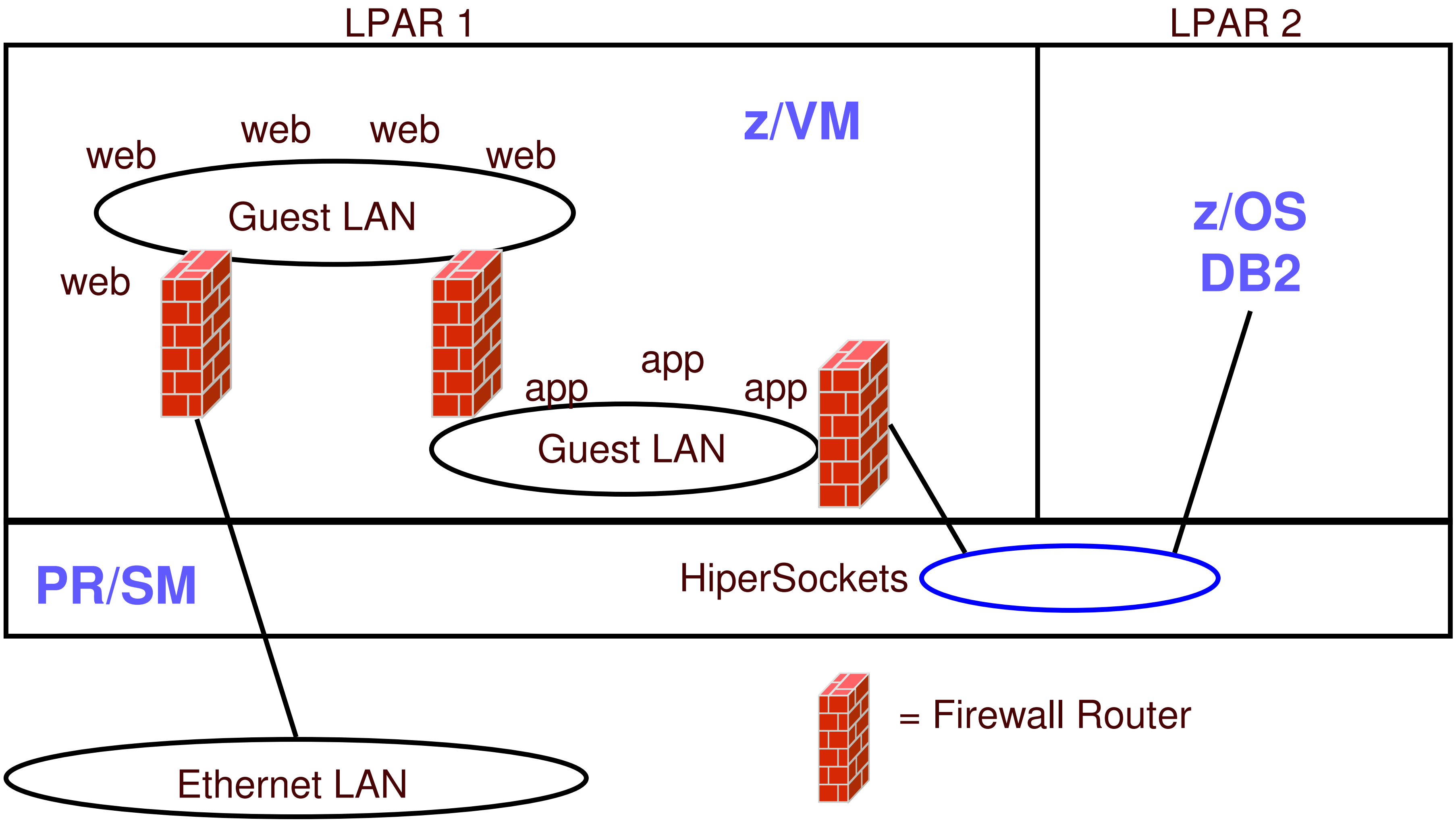
Multi-Zone Network



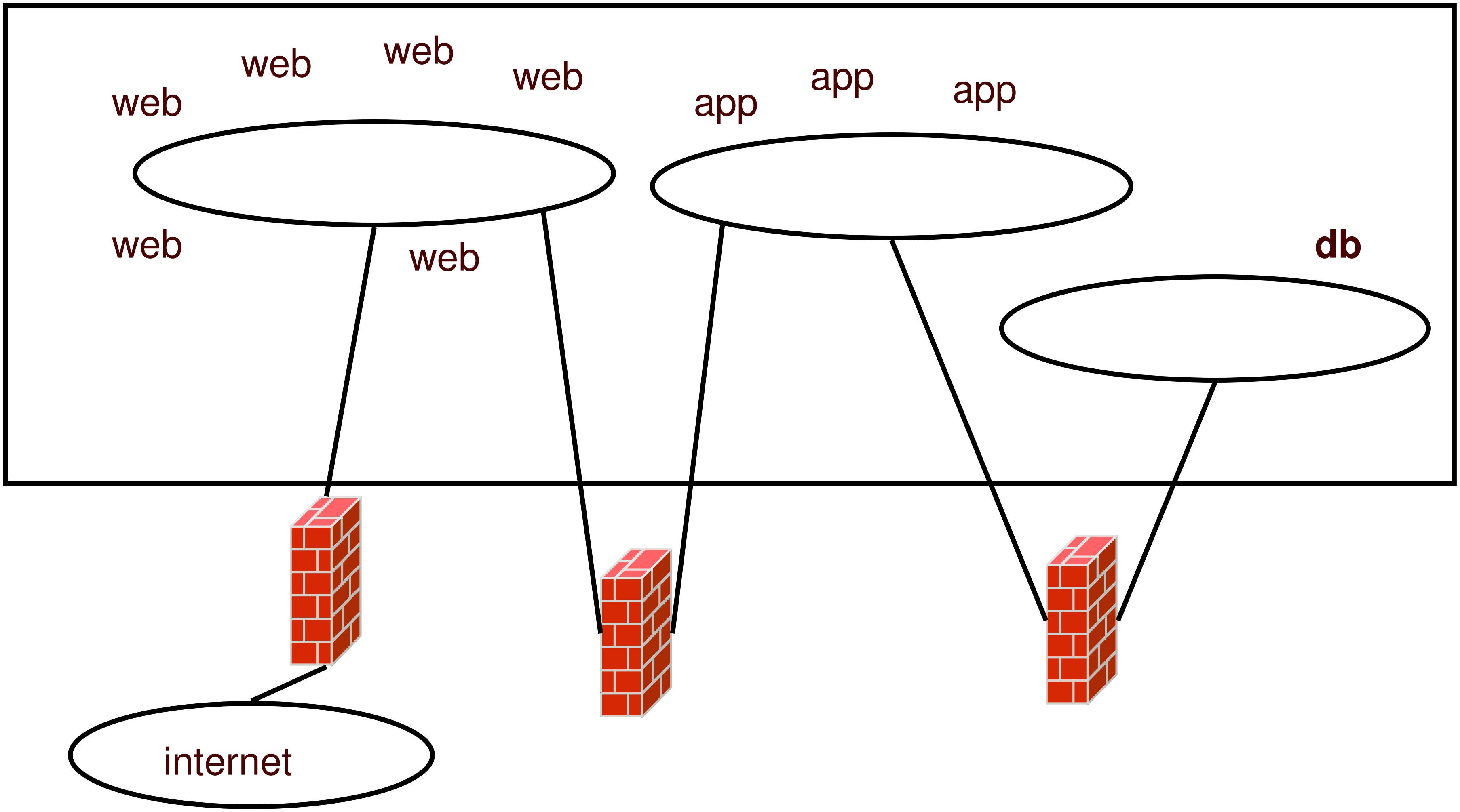
Multi-zone Network on System z



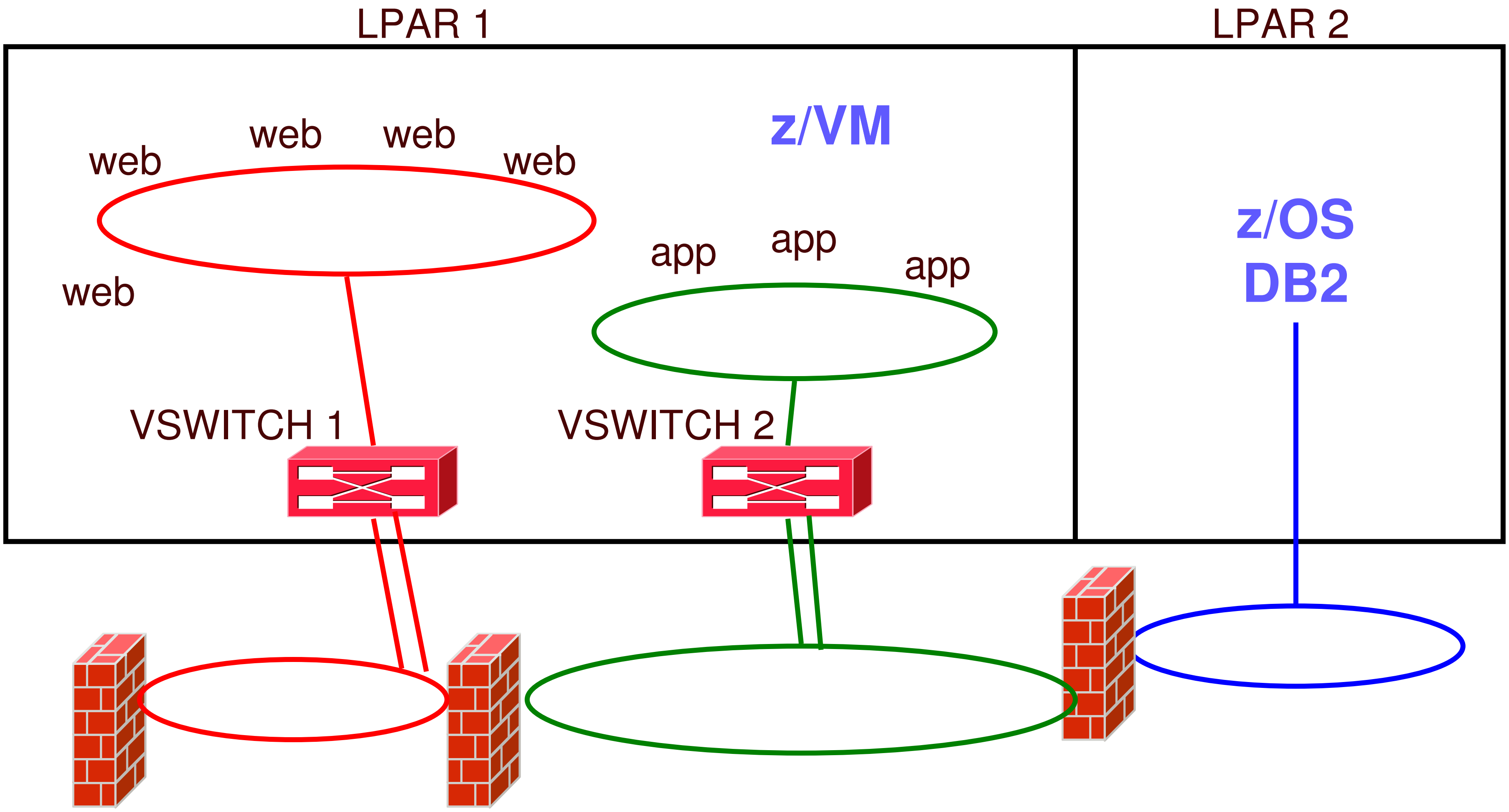
Multi-zone Network with Guest LANs



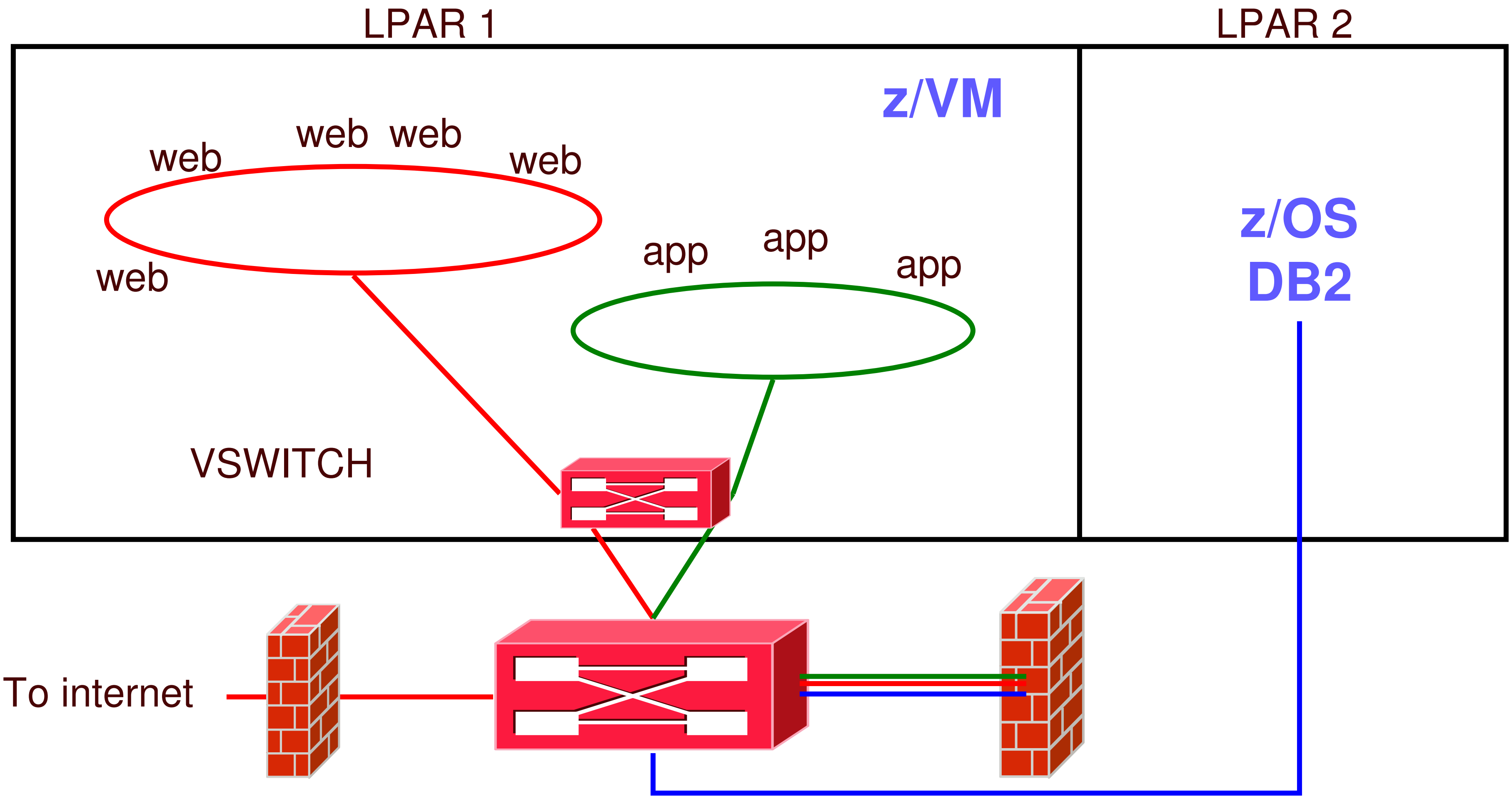
Multi-DMZ Network on zSeries with outboard firewall



Multi-DMZ Network with two VSWITCHes

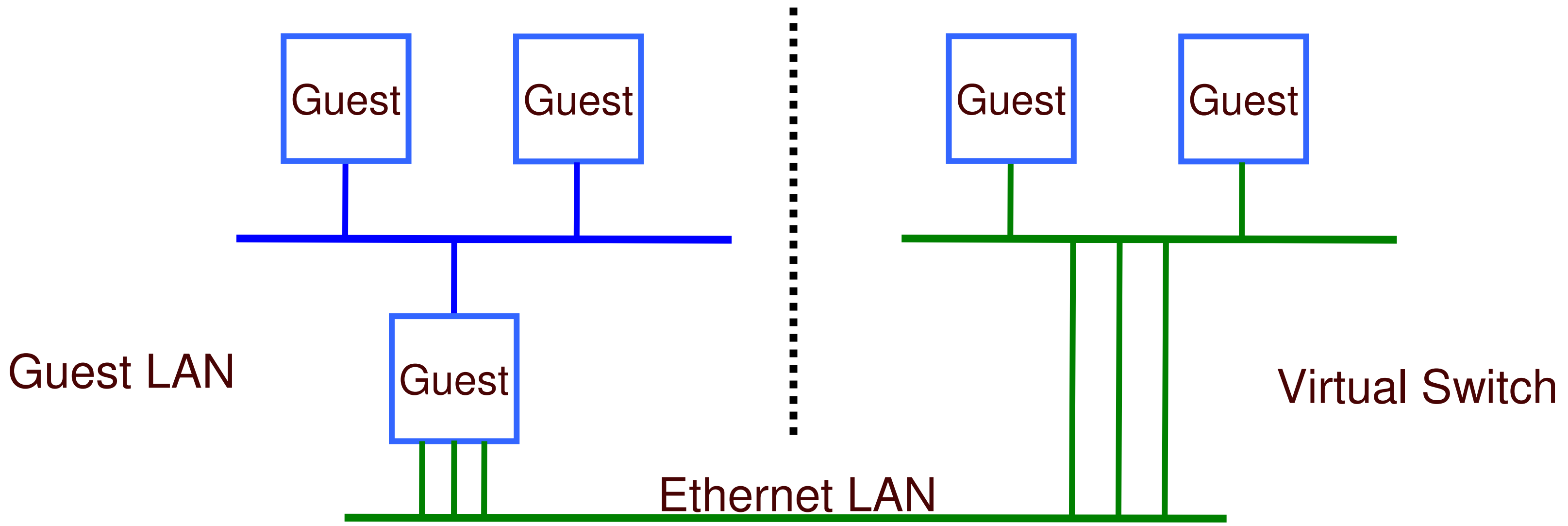


Multi-DMZ Network with VSWITCH (B)



With 1 VSWITCH, 3 VLANs, and a multi-domain firewall

Guest LAN vs. Virtual Switch



- Virtual router is required
- Different subnet
- External router awareness
- Guest-managed failover

- No virtual router
- Same subnet
- Transparent bridge
- CP-managed failover

Setting Guest LAN and VSWITCH defaults and limits

- Set global guest LAN attributes in the SYSTEM CONFIG file:

```
VMLAN LIMit PERSistent INFinite|maxcount
VMLAN LIMit TRANSient INFinite|maxcount
VMLAN ACNT|ACCOUNTing SYSTEM ON|OFF
VMLAN ACNT|ACCOUNTing USER ON|OFF
VMLAN MACPREFIX 020000-02FFFF
VMLAN MACIDRANGE SYSTEM x-y [USER a-b]
```

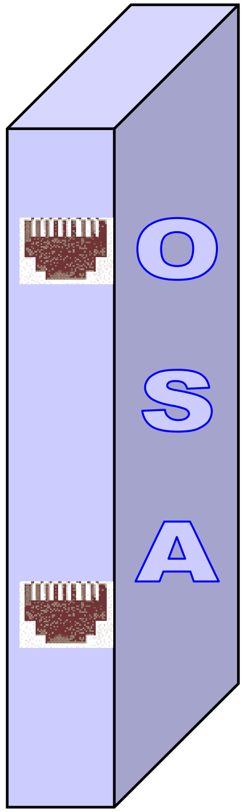
- `VMLAN LIMIT TRANSIENT 0` prevents dynamic definition of Guest LANs by class G users

Virtual MAC Addresses

- Each instance of CP should have a unique MACPREFIX
  - VMLAN MACPREFIX 020001
  - Reserve 020000 (the default) to recognize a misconfigured system
- Use MACIDRANGE to identify static vs. dynamic MAC addresses
  - VMLAN MACIDRANGE SYSTEM 000001-002FFF
    USER 002000-002FFF
  - USER range is a subset of SYSTEM range
  - Static MAC ids must come from USER range
- Virtual MAC = MACPREFIX || MACID
  - 020001 000123

What’s a ‘switch’ anyway?

© Cisco Corp



It creates LANs and routes traffic

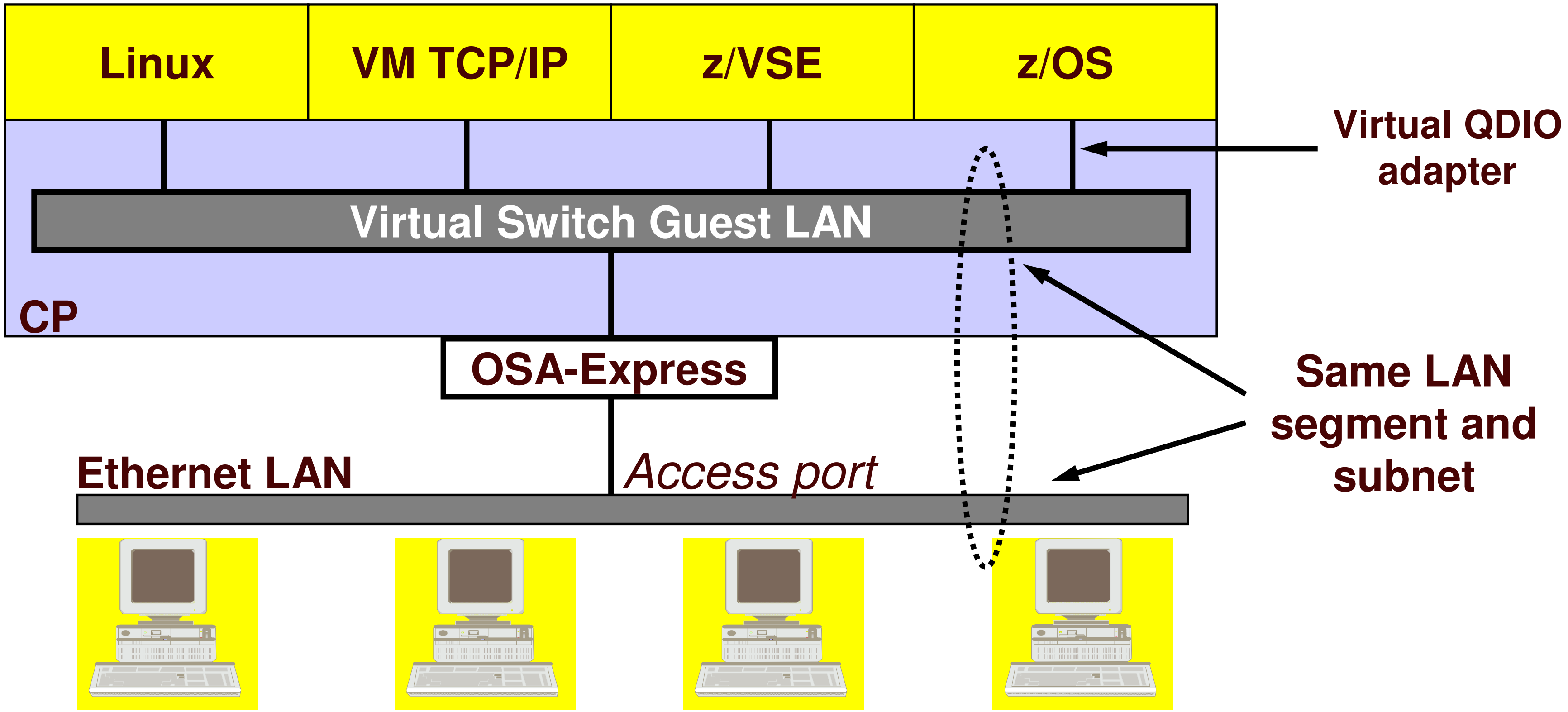
- Turn ports on and off
- Assign a port to a LAN segment
- Provides LAN sniffer ports

IEEE VLANs

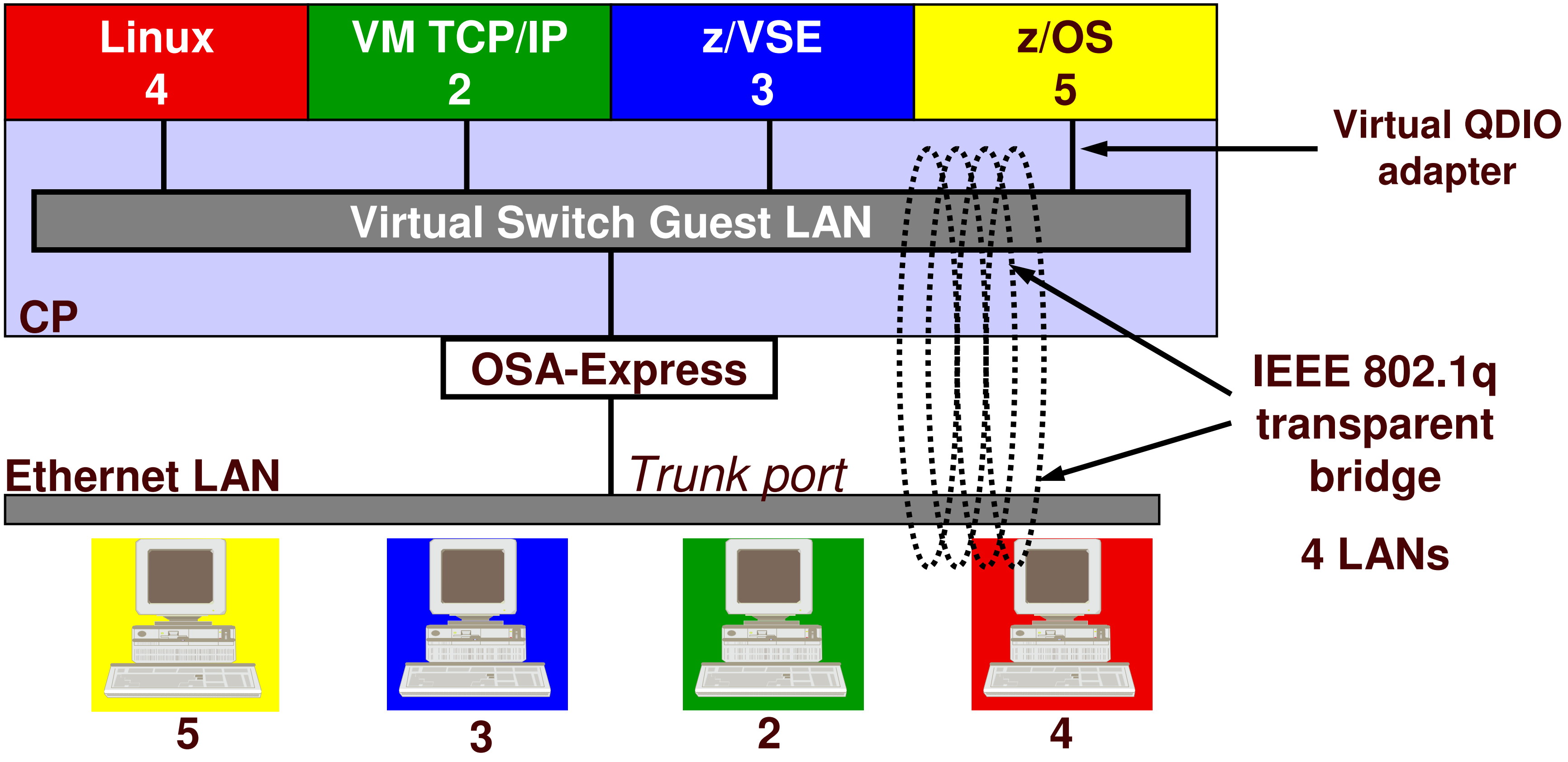
© Cisco Corp

- If you run out of ports, you don't throw it away, you daisy chain ("trunk") it to another switch.

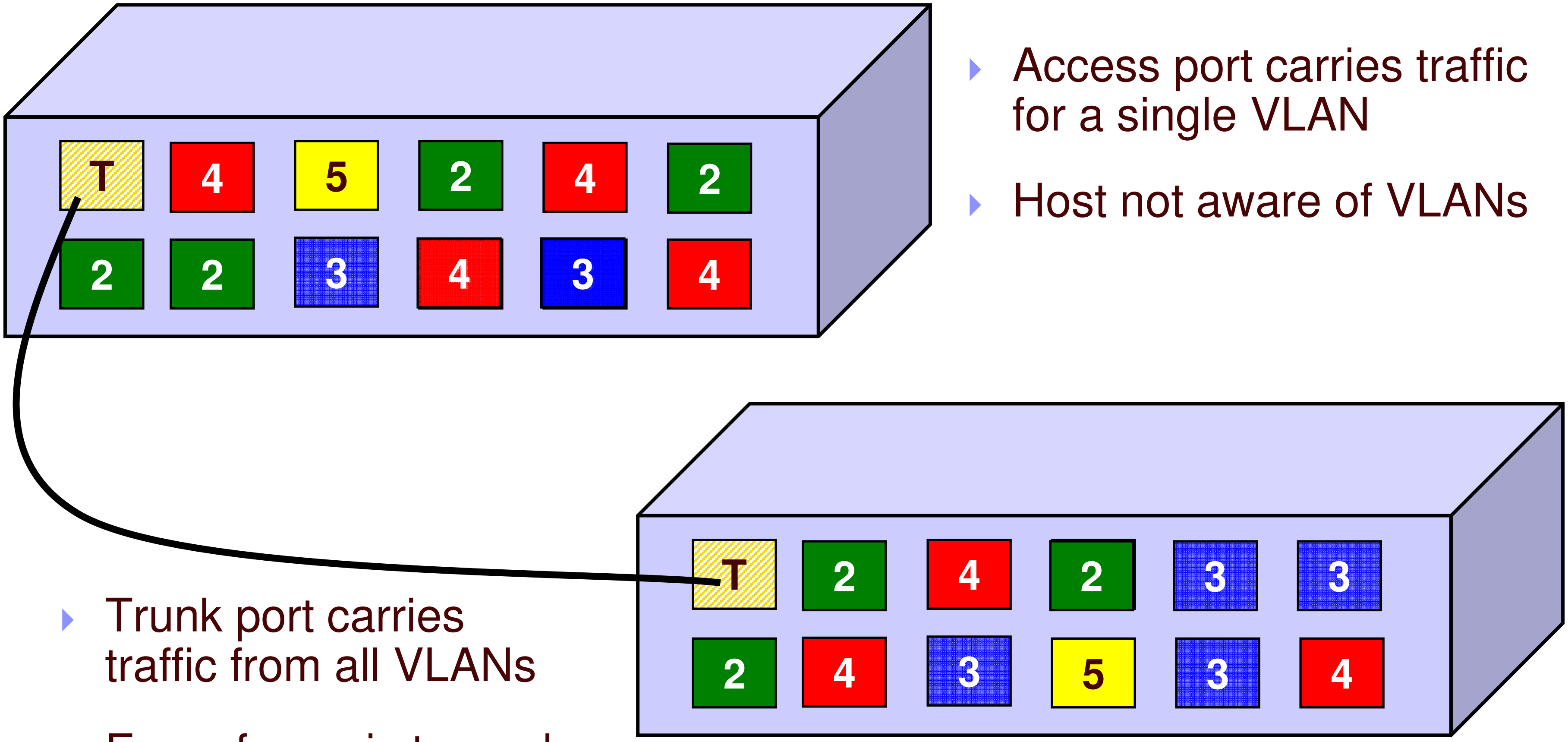
z/VM Virtual Switch – VLAN unaware
Sees only a single LAN segment



z/VM Virtual Switch – VLAN aware
Sees all authorized LAN segments



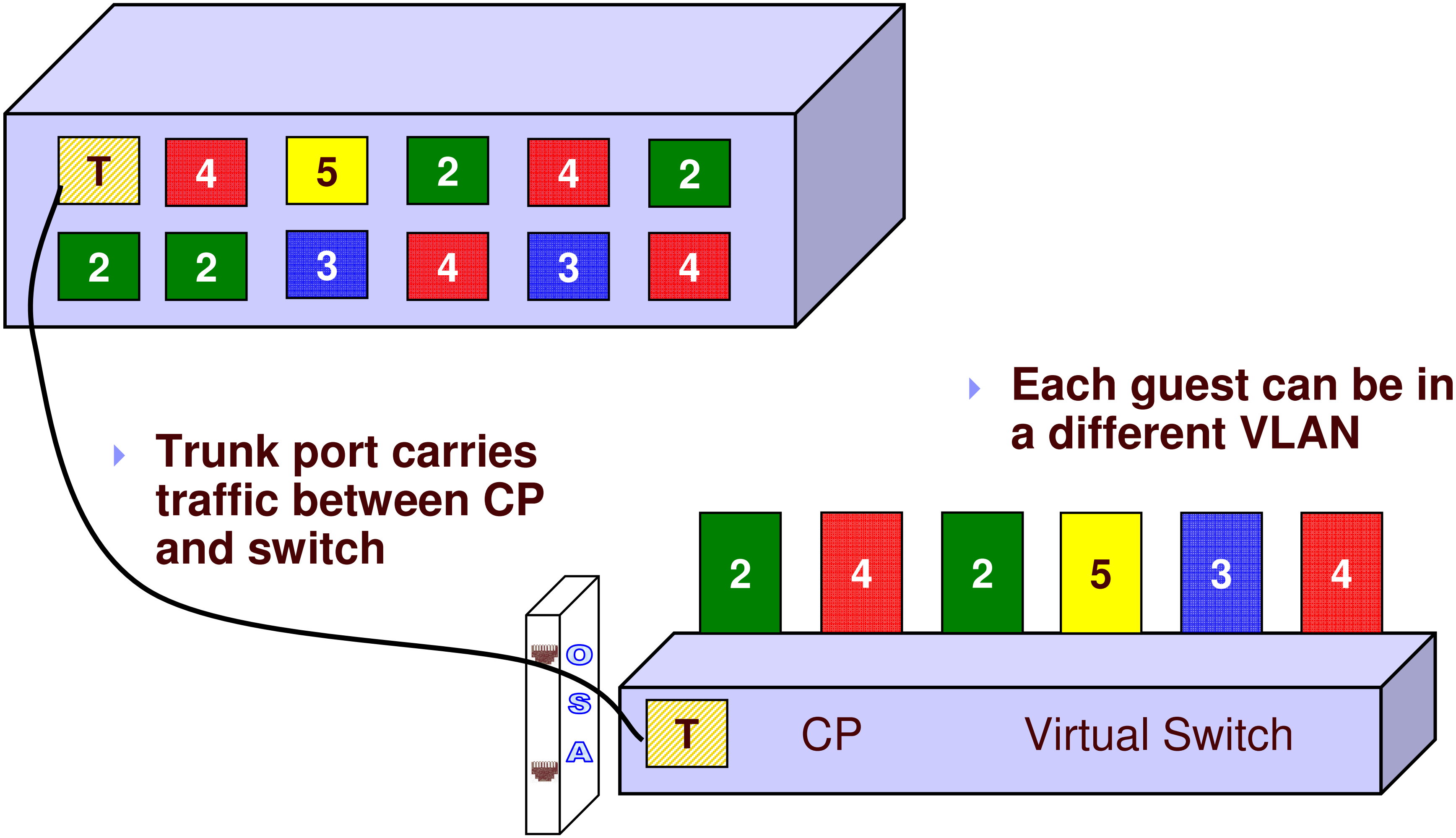
Trunk Port vs. Access Port



- Access port carries traffic for a single VLAN
- Host not aware of VLANs

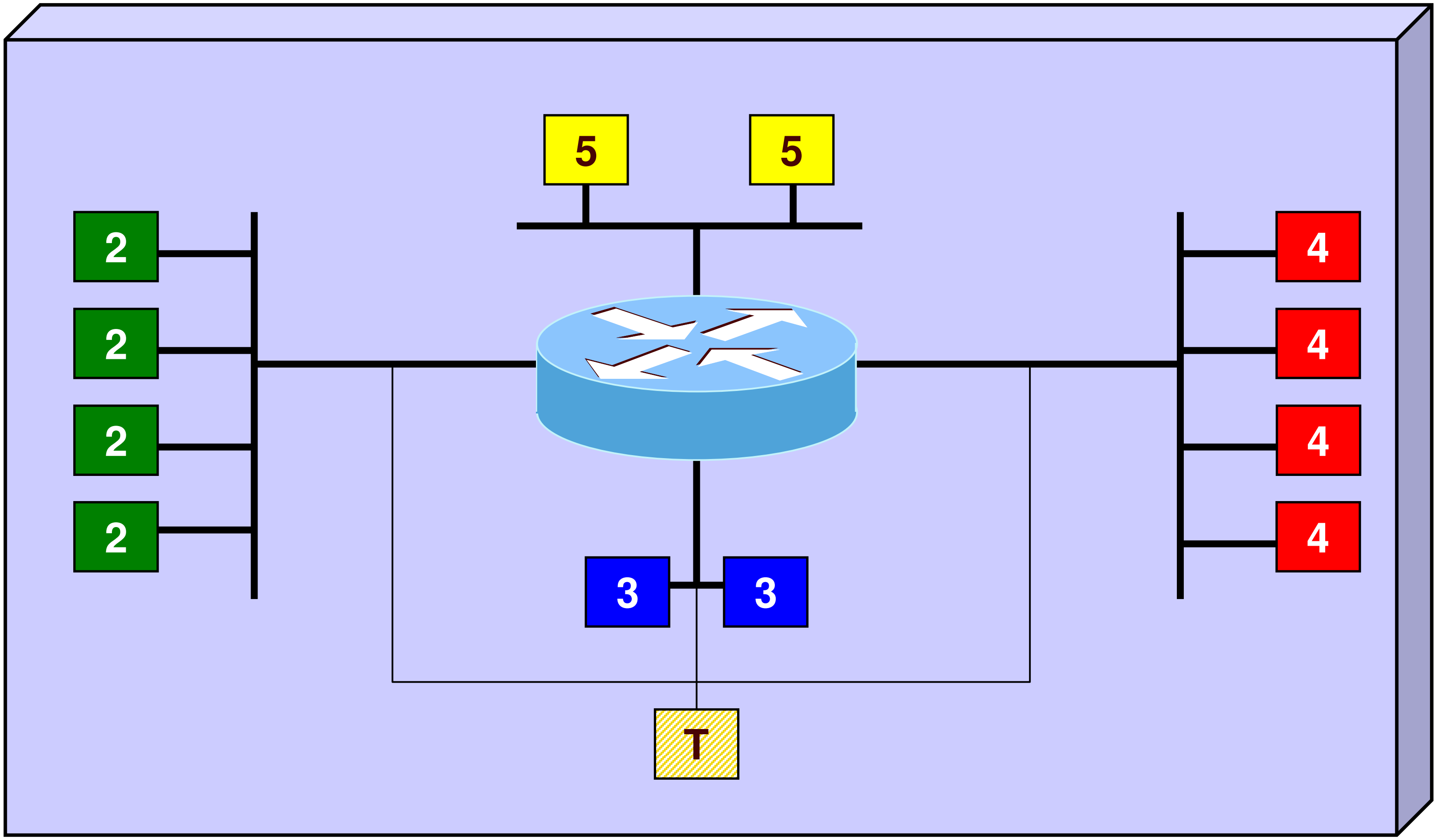
- Trunk port carries traffic from all VLANs
- Every frame is tagged with the VLAN id

Physical Switch to Virtual Switch



- **Trunk port carries traffic between CP and switch**
- **Each guest can be in a different VLAN**

A VLAN-aware switch: An inside look



Virtual Switch Attributes

- 1-8 character name
- Associated OSAs or Port group
- A controller virtual machine
  - DTCVSW1 and DTCVSW2
  - Starts, stops, and monitors OSAs
  - Not involved in data transfer
  - Do not ATTACH or DEDICATE devices
- Access list

Create a Virtual Switch

- SYSTEM CONFIG or CP command:

```
DEFINE VSWITCH name
            [RDEV NONE | cuu [cuu [cuu]] ]
            [NONROUTER | PRIROUTER]

            [VLAN UNAWARE | VLAN default_vid]
            [NATIVE 1 | native_vid]
            [GROUP group_name]

            [IP | ETHERNET]

            [CONNECT | DISCONNECT]
            [PORTTYPE ACCESS | PORTTYPE TRUNK]
            [CONTROLLER * | CONTROLLER userid]
Example:

DEFINE VSWITCH SWITCH12 RDEV 1E00 1F04
```

ETHERNET vs. IP

- ETHERNET = “Layer 2”
  - Each guest has a unique MAC address
  - Guest sends ethernet frame to NIC
  - OSA and CP have MAC address awareness
- IP = “Layer 3”
  - All guests have the same MAC address
  - Guest sends IP packets to NIC
  - OSA adds frame
  - OSA and CP have IP address awareness

Access list

- Only users in the access list can connect (couple) to this LAN or VSWITCH
- CP SET LAN or SET VSWITCH to GRANT or REVOKE access
  - RACF can control and audit access
- CP QUERY LAN or VSWITCH can show you the current access list and who is connected
  - Look at the DETAILS option

Vs. Guest LAN

- DEFINE LAN, SET LAN, QUERY LAN

- Owned by users or SYSTEM
- Class G can create (by default)
- Persistent vs. Transient
- Standalone LAN segment
- No connection to external network
  - Virtual router
  - Each Guest LAN needs its own subnet

Change the Virtual Switch access list

- Specify after DEFINE VSWITCH statement in SYSTEM CONFIG to add users to access list

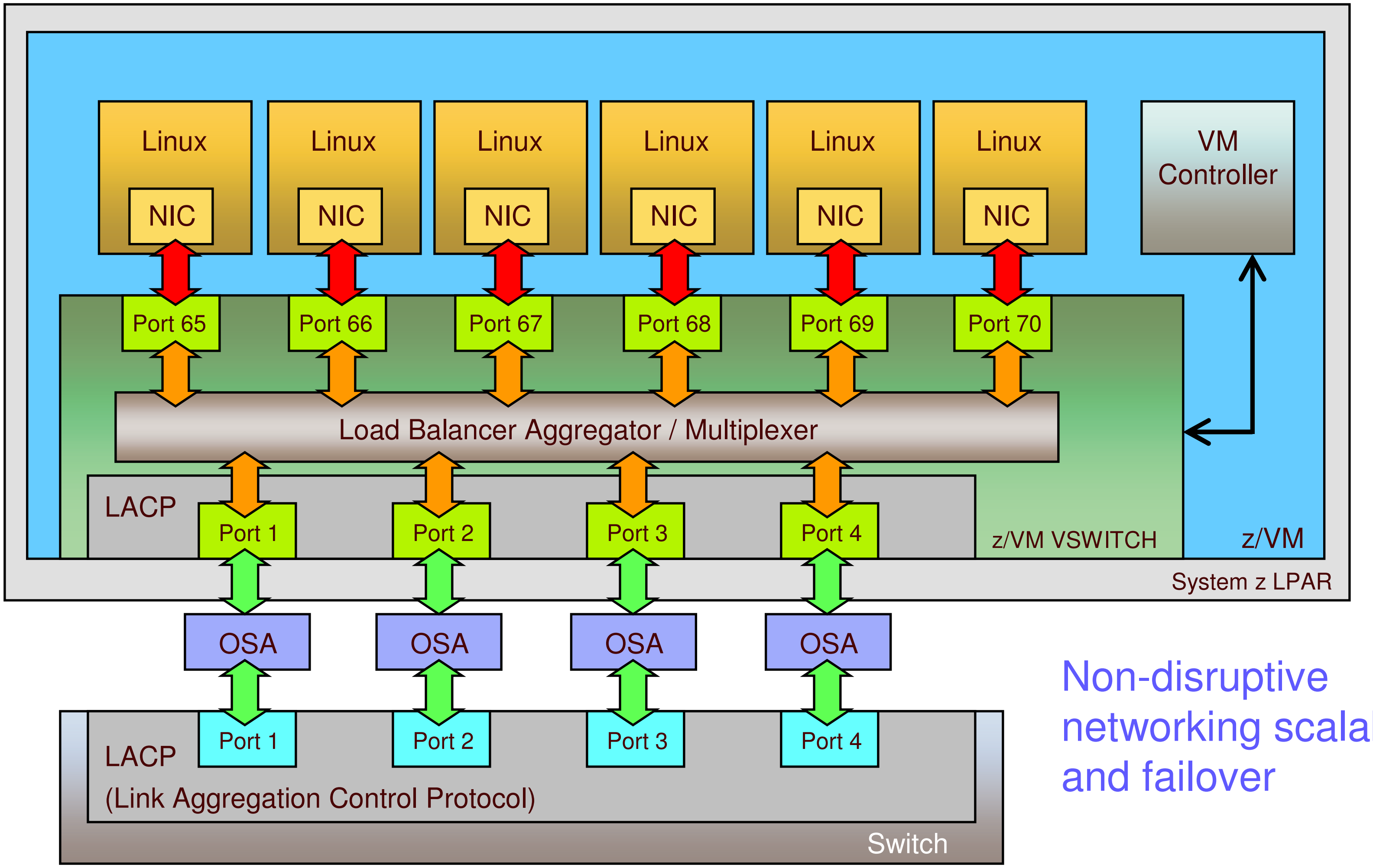
```
MODIFY VSWITCH name GRANT  userid
SET                        [VLAN vid1 vid2 vid3 vid4]
                           [PORTTYPE ACCESS | TRUNK]
                           [PROmiscuous | NOPROmiscuous]

SET    VSWITCH name REVOKE userid

Examples:
MODIFY VSWITCH SWITCH12 GRANT LNX01 VLAN 3
CP SET VSWITCH SWITCH12 GRANT LNX02 PORTTYPE TRUNK
                                    VLAN 4 20-22 29 302


CP SET VSWITCH SWITCH12 GRANT LNX02 PROMISCUOUS
```

IEEE 802.3ad Link Aggregation



Non-disruptive networking scalability and failover

IEEE 802.3ad Link Aggregation

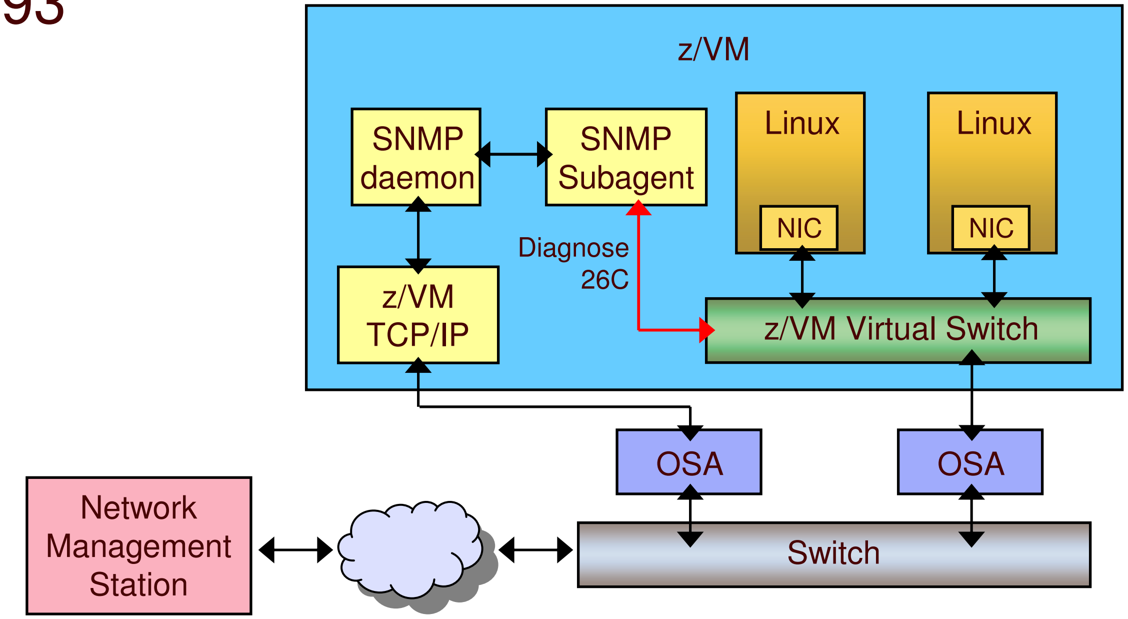
- **System z9 and later**
- **Groups available OSA-Express2/3 ports for use by the z/VM Virtual Switch**
  - Up to 8 ports per virtual switch
  - Increases Virtual Switch bandwidth and provides near seamless failover in the event of a failed controller, link or switch
  - Only supported for Layer 2 switches
- **Includes support to recover from a failed external switch**

IEEE 802.3ad Link Aggregation

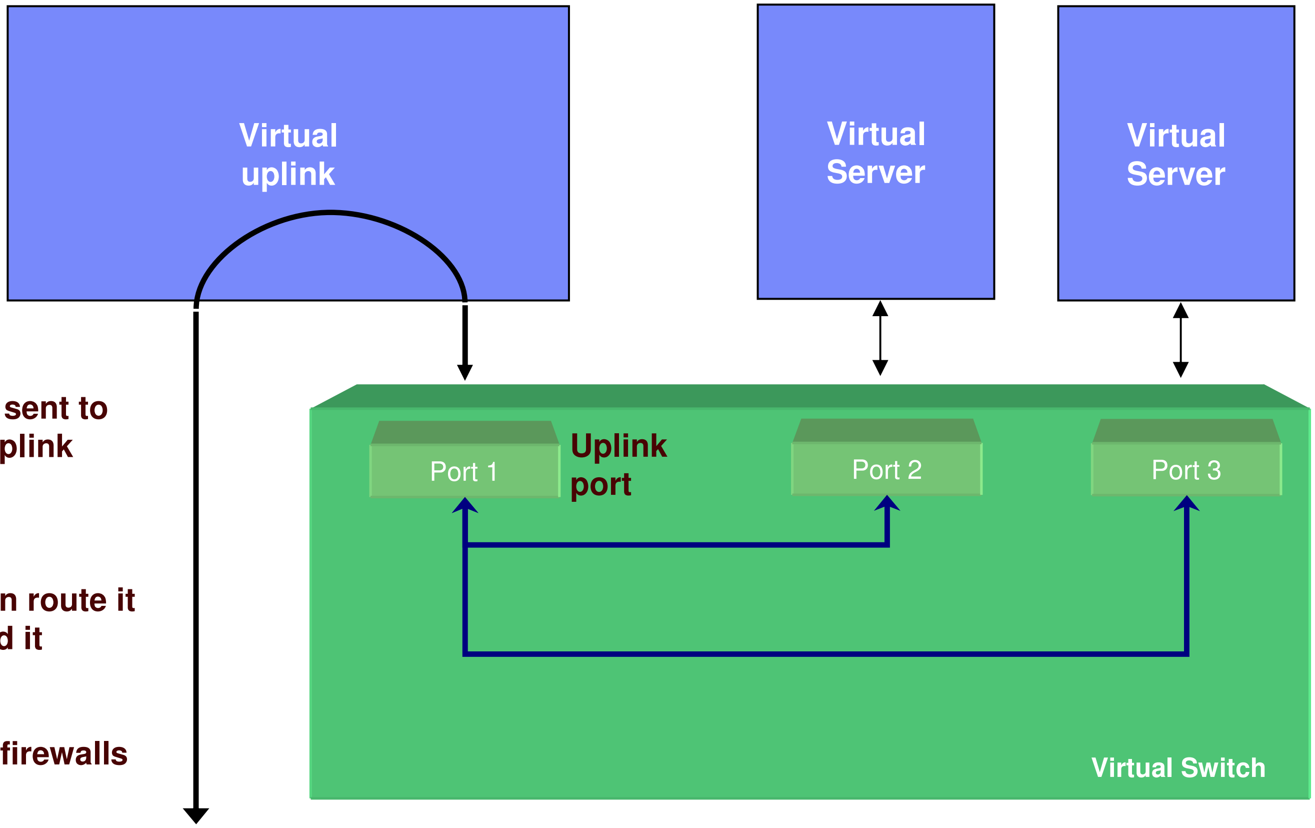
- Define an OSA port group
  - SET PORT GROUP *name* JOIN E100 E200.P1
- DEFINE VSWITCH … ETHERNET GROUP *name*
- OSAs **cannot** be shared

z/VM Virtual Switch SNMP MIB

- Integrates VSWITCH into standards-based switch management and monitoring tools
- SNMP subagent provides Bridge MIB data
  - Defined by RFC 1493



Virtual Switch Uplink Ports
“It’s not your grandfather’s VSWITCH!”



- All traffic sent to defined uplink guest
- Uplink can route it or forward it
- Great for firewalls

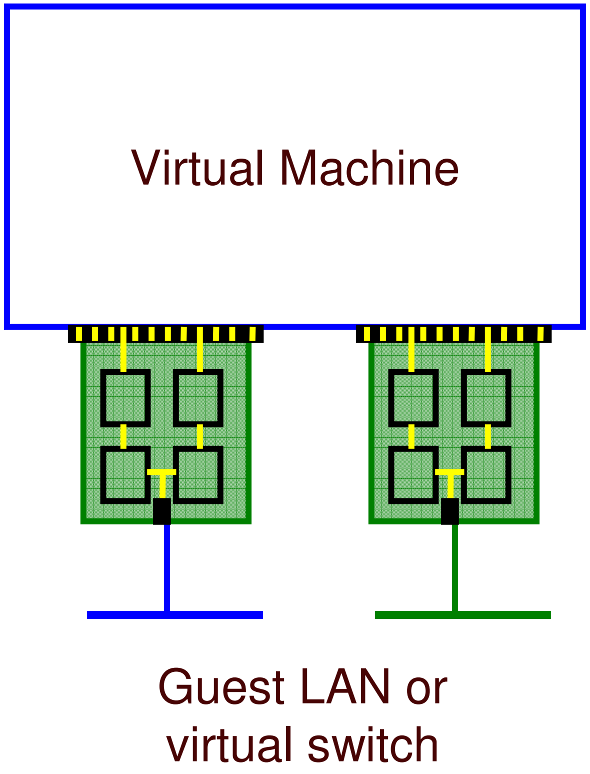
Additional security controls

- Virtual Sniffers
  - Guest must be authorized via SET VSWITCH or security server
  - Guest enables promiscuous mode using CP SET NIC or via device driver controls
    - E.g. tcpdump -P
  - Guest receives copies of all frames sent or received for authorized VLANs
- Port Isolation
  - Stop guests from talking to each other, even when in same VLAN
  - Shut off OSA “short circuit” to other users of the same OSA port

Virtual Network Interface Card

Virtual Network Interface Card (NIC)

- A simulated network adapter
- 3 or more devices per NIC
  - More than 3 to simulate port sharing on 2nd-level system or for multiple data channels
- Provides access to Guest LAN or Virtual Switch
- Created by NICDEF or CP DEFINE NIC command



Virtual NIC - User Directory

- One per interface in USER DIRECT file:

```
NICDEF vdev [TYPE HIPERS | QDIO]
            [LAN owner name]
            [DEVICES nn]
            [CHPID xx]
            [MACID xxyyzz]

Example:

NICDEF 1100 LAN SYSTEM SWITCH1 CHPID B1 MACID B10006
```

Combined with VMLAN MACPREFIX to create virtual MAC

- This is the only way to pre-assign the MAC address!

Virtual NIC - CP Command

- May be interactive with CP DEFINE NIC and COUPLE commands:

```
CP DEFINE NIC vdev
        [[TYPE] HIPERsockets|QDIO]
        [DEVices devs]
        [CHPID xx]

CP COUPLE vdev [TO] owner name

Example:

CP DEFINE NIC 1200 TYPE QDIO
CP COUPLE 1200 TO SYSTEM SWITCH12
```

NIC CHPID parameter

CHPID xx

- Specifies the Channel Path ID number (in hex) to use for this NIC
  - Default is any available unused real CHPID number

- Needed for z/OS guests only when connecting to HiperSockets Guest LAN

- **This is a virtual CHPID number**

Some Final Thoughts...

Network Configuration

- Guest LANs require a new subnet and the use of a virtual router
    - Can use a Disconnected VSWITCH instead

- A Virtual SWITCH extends the subnets you already have

- By having virtual and real configurations be the same, you can easily test network configuration before deployment with real hardware

Built-in Diagnostics

- **CP QUERY VMLAN**
  - to get global VM LAN information (e.g. limits)
  - to find out what service has been applied
- **CP QUERY LAN ACTIVE**
  - to find out which users are coupled
  - to find out which IP addresses are active
- **CP QUERY NIC DETAILS**
  - to find out if your adapter is coupled
  - to find out if your adapter is initialized
  - to find out if your IP addresses have been registered
  - to find out how many bytes/packets sent/received

Support Summary

z/VM 6.1	▪Uplink port can be OSA or guest
z/VM 5.4	▪Port isolation ▪Native VLAN id defaults to 1 ▪z/VM TCP/IP support for Layer 2
z/VM V5.3	▪Link aggregation ▪Separation of default VLAN id from native VLAN id ▪SNMP monitor
z/VM V5.2	▪Virtual SPAN ports for sniffers
z/VM V5.1	▪Virtual trunk and access port controls ▪Removal of VLAN ANY ▪Layer 2 (MAC) frame transport ▪Improved virtual switch error detection & recovery ▪External security manager access control
z/VM V4	▪IPv4 Virtual Switch with IEEE VLANs ▪IPv4 HiperSocket Guest LAN ▪IPv4 and IPv6 QDIO Guest LAN

References

- Publications:
  - z/VM CP Planning and Administration
  - z/VM CP Command and Utility Reference
  - z/VM TCP/IP Planning and Customization
  - z/VM Connectivity
- Links:
  - http://www.ibm.com/servers/eserver/zseries/os/linux/
  - http://www.linuxvm.org/

Contact Information

▪ By e-mail:	Alan_Altmark@us.ibm.com
▪ In person:	USA 607.429.3323
▪ On the Web:	http://ibm.com/vm/devpages/altmarka
▪ Mailing lists:	IBMTCP-L@vm.marist.edu IBMVM@listserv.uark.edu LINUX-390@vm.marist.edu http://ibm.com/vm/techinfo/listserv.html