IBM Systems and Technology Group

z/VM TCP/IP Update

Miguel Diaz
IBM Corporation
Saturday, 21 May, 2005

z/VM V5.1 includes TCP/IP Function Level 510, a new level of the TCP/IP Feature that delivers significant new functions. This session gives an overview of these enhancements, as well as describing the VM TCP/IP product and the changes to it that were introduced in Function level 440 with z/VM V4.4.

Agenda

- **z/VM 4.4 TCP/IP Updates**
- **z/VM 5.1 TCP/IP Updates**
- **Futures**
- **Summary**

z/VM 4.4 TCP/IP Updates

- **Function**
- **Performance**
- **Infrastructure**

z/VM 4.4 TCP/IP Functional Enhancements

- **IMAP Authentication Extensions**
- **SSL Server Repackaging**
- **Security Enhancements**
- **SMTP Batch Mail Source Verification**

IMAP Server Authentication Enhancements

- **Eliminates requirement for enrolled IMAP users to have VM user IDs and passwords**
- **Interface to user-written authentication exit enables**
  - User names longer than 8 characters
  - Alternative to CP or External Security Manager validation
  - PREAUTH processing (pre-authenticated connections by IP address)
- **Sample exit provided with z/VM 4.4.0**
  - Customer must provide validation and mapping algorithms for user names

SSL Server Uplift

- **z/VM SSL server required out-of-service Linux**
- **Server now compatible with Linux 2.4.7 kernel**
  - Supported on SuSE 2.4.7 SLES 7.0 and SLES 8.0 distributions
  - Provided in Red Hat Package Manager (RPM) format
- **Version for 2.2.16 kernel not provided (no distributor support)**

IP Security Enhancements

- **Logging of commands issued by**
  - OBEYFILE
  - NETSTAT CP
  - NETSTAT OBEY
- **Some defaults changed for *AssortedParms* configuration statement**
  - ***RestrictLowPorts*** on by default
  - ***VarSubnetting*** on by default
- **Port range accepted on *Port* statement**
  - Ports may be reserved to prevent their use
- **REXECD supports LOGON BY**

SMTP Batch Mail Source Verification

- **New *VerifyBatchSMTPSender* configuration statement**
  - Reject batch mail with "MAIL FROM:" that does not match RSCS information
  - Designate exceptions by user identifier as operands of new statement

z/VM 4.4 TCP/IP Performance Improvements

- **Virtual switch**
- **Stack performance enhancements**

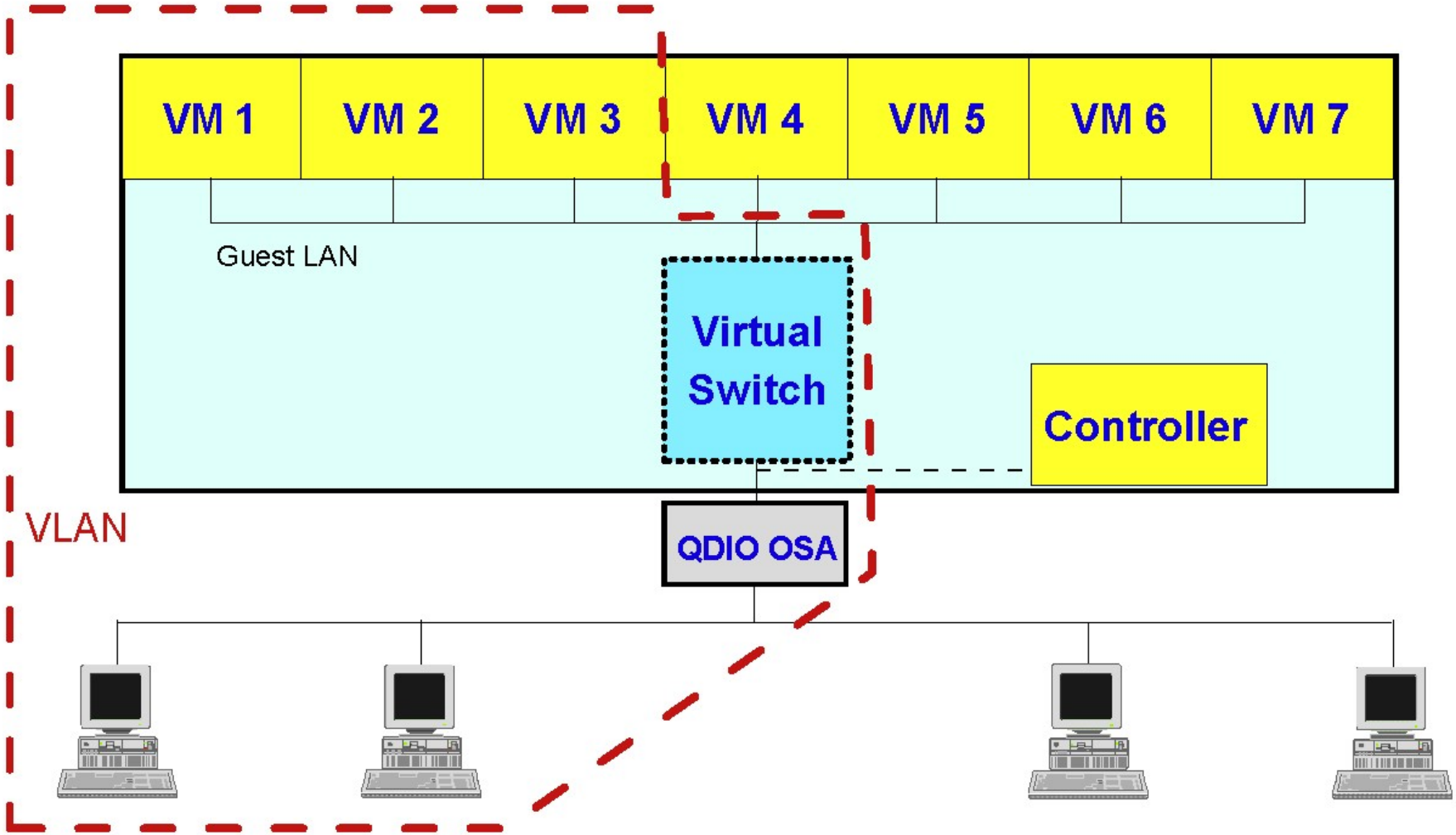
Virtual Switch

- **Layer 3 switch**
  - Switches packets between QDIO guest LAN and OSA Express physical network
  - Eliminates need for layer 3 router
  - Supports transparent VLAN specifications for guests connected to Virtual Switch
  - Packet switching performed entirely by CP
  - z/VM TCP/IP stack provides setup and control functions

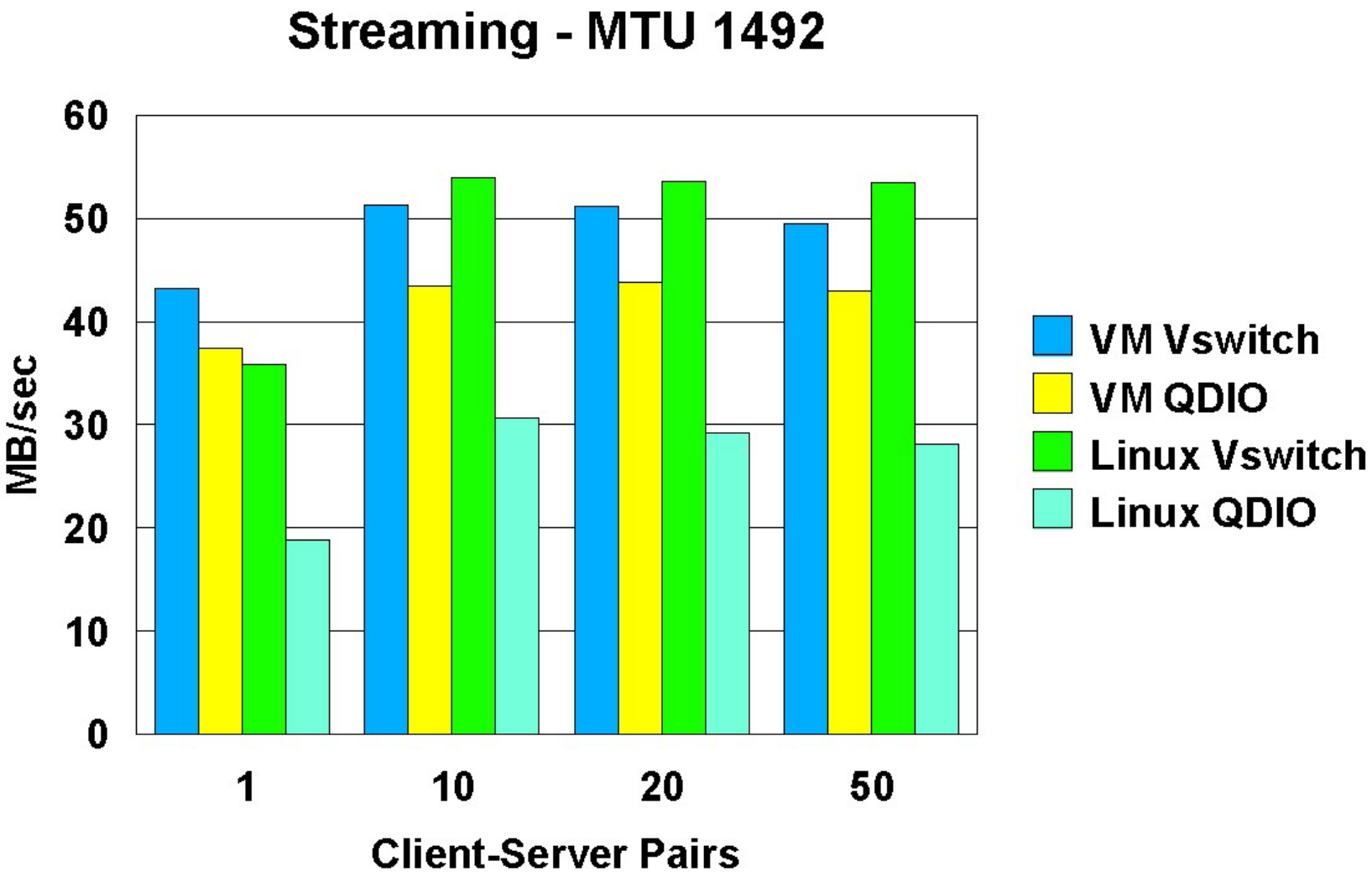
Virtual Switch (continued)

- **Provides transparent bridging**
  - Learning - automatic IP address configuration
  - Flooding - deliver packets for unknown IP addresses to all stations
  - Aging - forget learned IP addresses after some period of inactivity
- **Supports locally-administered MAC addresses**

Virtual Switch Topology



Virtual Switch Benefits



TCP/IP Stack Performance Enhancements

- **Multiprocessor support**
  - ***CPU*** option added to ***Device*** statement
  - Allows a designated device to be associated with a particular virtual processor
  - Generally allows CPU utilization (and thus throughput) to increase
- **Optimization of high-use code paths**
  - Focus on host (as opposed to router) functions
  - Some Pascal has been rewritten in Assembler
  - Path lengths for some algorithms have been reduced

TCP/IP Stack Performance Enhancements (continued)

- **Must define multiple virtual CPUs for TCPIP virtual machine (up to 7)**
- **Not effective if**
  - not enough real CPUs available
  - no CPU bottleneck
- **Results (based on preliminary measurements)**
  - Up to 65% reduction of certain pathlengths in TCP/IP
  - MP support provides up to 17% throughput improvements depending on configuration

z/VM 4.4 TCP/IP Infrastructure Improvements

- **TRACE TABLE**
- **Packet trace**
- **Stack module identification**
- **Client-stack level verification**
- **FTP Server Capacity**

TRACE TABLE

- **Internal trace tables record scheduler and interruption activity**
- **Useful for reconstructing sequence of events leading up to problems, especially in an MP environment**

Packet Trace

- **New *PacketTraceSize* statement**
  - Specifies amount of incoming and outgoing packets to record
  - Use ***TraceOnly*** statement to identify devices whose packets are to be traced
  - Uses System Trace File Interface (Diagnose X'E0')
- **Data recorded using CP (TRSOURCE TYPE GT)**

Packet Trace Example

```
NETSTAT OBEY PACKETTRACESIZE 64 64
NETSTAT OBEY TRACEONLY ETH0 ENDTRACEONLY

TRSOURCE ID TCP TYPE GT BLOCK FOR USER TCPIP
TRSOURCE ENABLE ID TCP

NETSTAT OBEY PACKETTRACESIZE 0
NETSTAT OBEY TRACEONLY ENDTRACEONLY

TRSOURCE DISABLE ID TCP
TRACERED 1234 CMS TCP TRACE A ( ALL
```

Stack Module Identification

- **Stack module file identifier reported**
  - in console log during initialization
  - in response to NETSTAT LEVEL command
- **Eases detection of configuration problems**
  - e.g., module on A-disk

```
netstat level

IBM 2064; z/VM Version 4 Release 4.0, service level 0000 (64-
bit), VM TCP/IP Level 440; RSU 0000 running TCPIP MODULE E2
dated 05/07/03 at 15:03

Ready; T=0.01/0.02 12:50:54
```

Client-Stack Level Verification

- **Clients and stack exchange level information and report mismatches**
- **Eases detection of problems caused by incompatible levels**
- **Can disable stack reporting with *AssortedParms* statement *NoLevelWarning* option**

FTP Server Capacity

- **FTP Server handles more than 256 connections**
  - Grows number of connections dynamically on demand
  - May require additional virtual storage

z/VM 5.1 TCP/IP Updates

- **Function**
- **Infrastructure**
- **Packaging**

z/VM 5.1 TCP/IP Functional Enhancements

- **IP version 6**
- **Enhanced Virtual Switch fail-over**
- **Intelligent default MTU sizes**
- **Link Forwarding Control**
- **TRACERTE Improvements**
- **PING Improvements**
- **SMTP SuppressNotification Changes**

IP version 6

- **First step towards support for IPv6 networks**
  - Address constraint relief
  - Auto-configuration
  - Other improvements
- **Support for IPv6 networks connected through OSA Express (QDIO) adapter**
  - Static routing
  - Router Advertisements
  - TRACERTE, PING, and IFCONFIG support
  - IPv6 sockets through Language Environment and OpenExtensions Callable Services

IP version 6 (continued)

- **v4 and v6 networks treated separately**
  - Separate HOME lists, filters (***BLOCK*** statement) address translation tables, static routing tables (***GATEWAY*** statement), PORT lists
  - No routing between networks
- **New *DEVICE OSD* statement options**
  - IPv6PriRouter
  - IPv6SecRouter
  - IPv6NonRouter
- **New *LINK QDIOEthernet* statement options**
  - EnableIPv6
  - DupAddrXmits

IP version 6 (continued)

- **New *RouterAdv* statement**
  - Defines characteristics of router advertisements for a link
- **New *RouterAdvPrefix* statement**
  - Defines address prefix to be used for link router advertisements and associated on-link determination, autonomous, and lifetime characteristics
- **New *AssortedParms* statement options**
  - IgnoreIPv6Redirect
  - EqualCostIPv6MultiPath

IP version 6 (last one...I promise)

- **New *NCBPoolSize* statement**
  - Defines size of IPv6 Neighbor Control Block pool
- **New *ICMPErrorLimit* statement**
  - Define maximum rate per second of IPv6 ICMP error packets transmitted on a link
- **New *Neighbor* and *DelNeighbor* functions of NETSTAT**
  - Display/delete neighbor cache entries
- **NETSTAT DEVLINKS reports**
  - Maximum frame size (Hipersockets links)
  - MTU size
  - IPv6 status

Enhanced Virtual Switch Failover

- **New *VSWITCH CONTROLLER* statement options**
  - Failover_Enabled
  - Failover_Disabled
- **If enabled, CP ensures that TCP/IP controller is responding to requests**
  - Switches to backup controller and OSA if unresponsive

Intelligent Default MTU Sizes

- **Specify MTU size as zero**
  - Default selected based on link type, envelope size, and frame size (CLAW)
  - More likely to get it right
- ***MTU* option added to *LINK* statement for all interface types**

Link Forwarding Control

- **New *NoFwd* or *NoForward* option on *LINK* statement**
  - Controls whether datagrams are forwarded from that link to others

TRACERTE Improvements

- **New *ADDRTYPE* option**
  - Select IPv4 or IPv6 format
- **New *SOURCEIP* option**
  - Sets source IP address in datagrams
  - Must be valid home address
- **New *LINK* option**
  - Determines link used to send datagrams
- **New *NONamelookup* option**
  - Suppresses address-to-name resolution

PING Improvements

- **New *ADDRTYPE* option**
  - Select IPv4 or IPv6 format
- **New *SOURCEIP* option**
  - Sets source IP address in datagrams
  - Must be valid home address
- **New *LINK* option**
  - Determines link used to send datagrams

SMTP SuppressNotification Changes

- **Can suppress "Received from ..." as well as "Mail delivered" messages**
- **New *RECEIVED*, *DELIVERED*, and *ALL* options on *SuppressNotification* statement**

z/VM 5.1.0 TCP/IP Infrastructure Enhancements

- **DTCPARMS improvements**
- **User-level stack profiles**
- **Improved OSD device restart**

DTCPARMS Improvements

- **New *:vnic* tag**
  - Defines virtual NIC
  - Couples to designated Guest LAN or VSWITCH

:vnic.*vdev* [TO] *owner LANname*, ...

User Level Stack Profiles

- **Search for `userid TCPIP` before `nodename TCPIP` and `PROFILE TCPIP`**

Improved OSD Device Restart

- **OSD device automatic restart procedure**
  - If restart-eligible failure, try to restart immediately
  - If unsuccessful, try again every 30 seconds
  - Stop restart either when successful or if device is stopped (e.g., with NETSTAT OBEY STOP)

z/VM 5.1 TCP/IP Packaging Enhancements

- **Configuration files not supplied**
  - Can be created using ***TCP2PROD TCPCONFIG***

Future Candidates

- **Path MTU Discovery**
- **Upgraded MPRoute Level**
- **VLAN Support Enhancements**
- **...**

Summary

- **TCP/IP for VM is alive and well**
- **Levels 440 and 510 delivered major advances**
- **Future levels will continue the trend**
- **We still have more to do**
  - Anticipate where most VM TCP/IP customers are going next
  - Your requirements are important to us

Contact Information

- **By e-mail:** migueld@us.ibm.com
- **In person:** USA 607.429.3504
- **Mailing lists:**

  A number of listservs relevent to z/VM are availible. Information on how to subscribe can be found at the following website:

  http://www.vm.ibm.com/techinfo/listserv.html

  Of particular interest:
  VMESA-L@listserv.uark.edu
  IBMTCP-L@vm.marist.edu
  LINUX-390@vm.marist.edu