V55

The Latest and Greatest on z/VM's Control Program

John Franciscovich
francisj@us.ibm.com

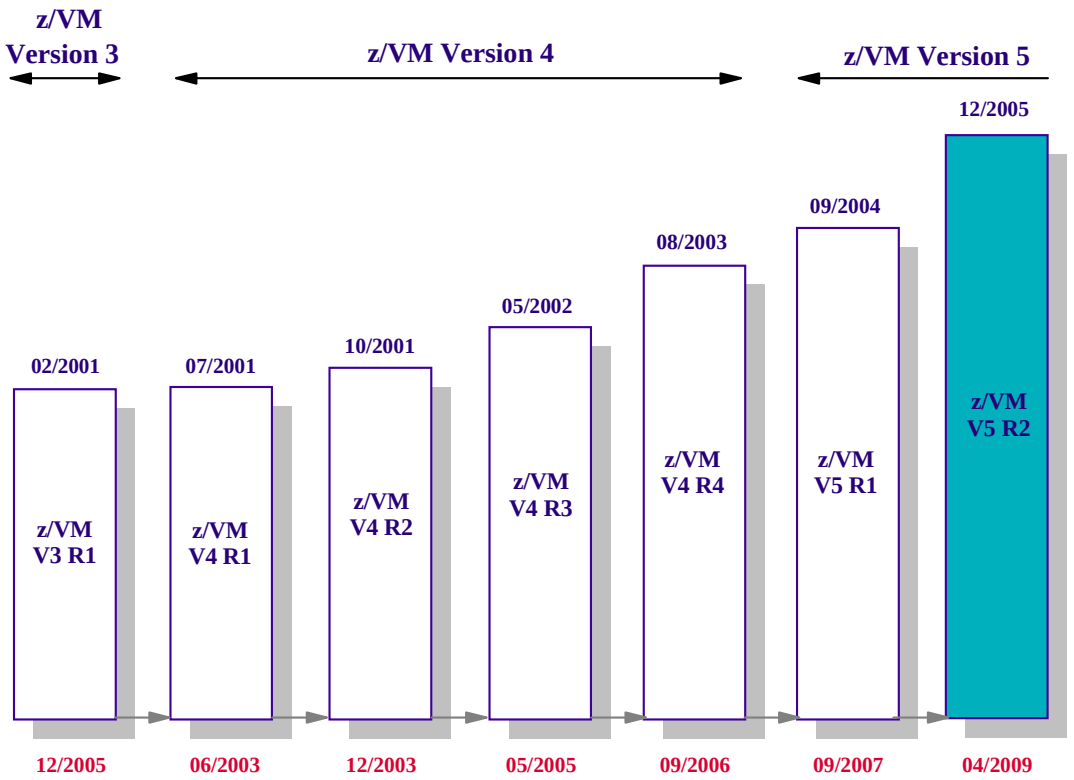
IBM
SYSTEM z9 AND zSERIES EXPO
October 9 - 13, 2006

Orlando, FL

Agenda

- **Product Evolution**
- **Release Roadmap**
- **z/VM 5.2 CP Enhancements Overview**
- **z/VM 5.2 CP Enhancements**
  - Processor and I/O Support
  - Virtualization
  - Virtual Networking
  - Other
- **Removed Functions**

z/VM Evolution



Withdrawn from marketing and no longer serviced

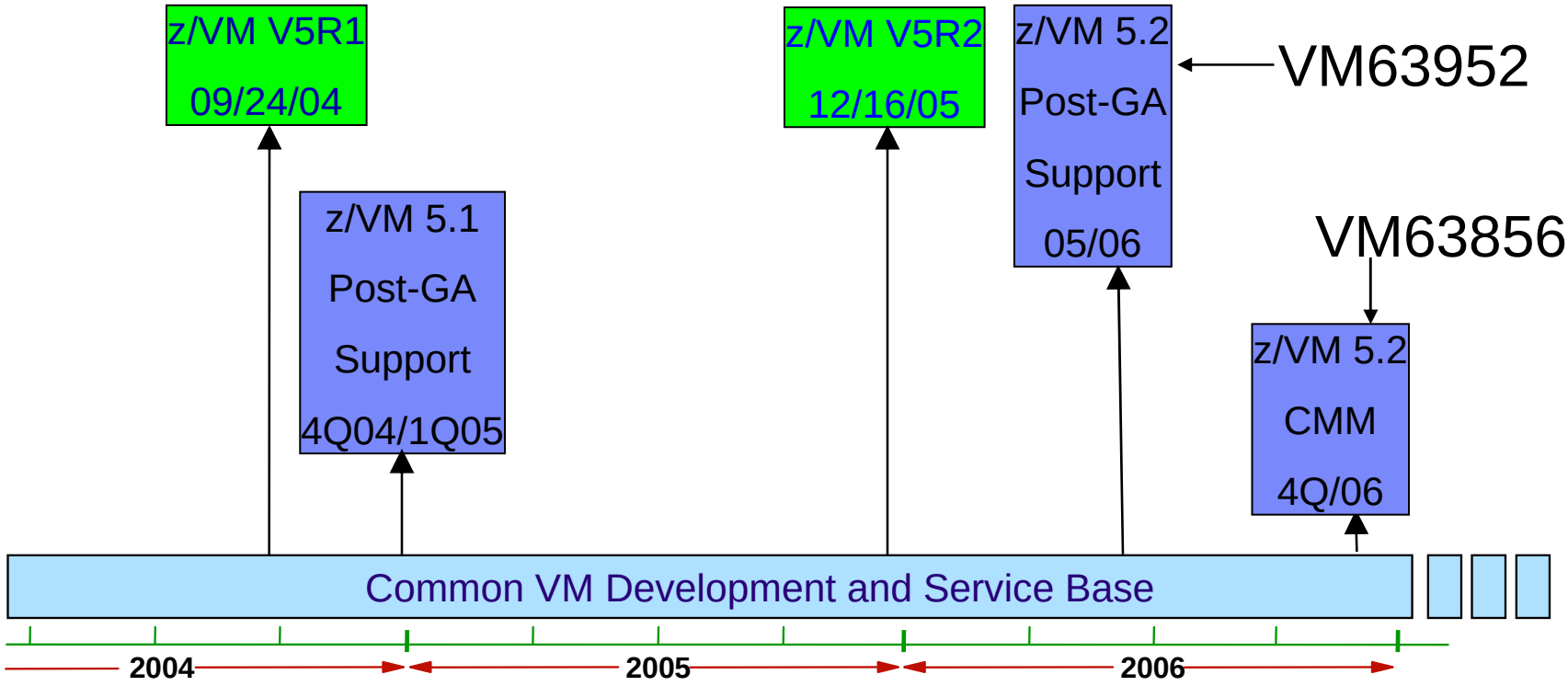
Withdrawn from marketing

Currently being marketed

XX/XX - General availability (top of box)

XX/XX - Service Discontinued (bottom of box)

z/VM Release Roadmap



z/VM 5.2 CP Enhancements Overview

- **Processor and I/O Support**
  - z9 processor support
  - Dynamic LPAR naming
  - Native SCSI disk support enhancements
  - Crypto Express2 support
  - OSA Express2 CDLC support
  - FCP N_Port ID virtualization
  - FICON Express 4
- **Virtual Networking**
  - Guest LAN and Virtual Switch sniffing
  - Virtual Switch GVRP Support
  - IPv6 Hipersockets
- **Virtualization Enhancements**
  - Large real memory exploitation
  - QDIO pass-through
  - 64-bit Diagnose 250
  - 64-bit Diagnose 98
  - Diagnose 290
  - Diagnose 9C
  - PAV support
  - Collaborative memory management
- **Other**
  - Remove S/A Dump HLASM requirement

Processor and I/O Support

- **Support for System z9 processors**
  - Hardware GA (09/24/05): compatibility support for z/VM V4.4 and V5.1
  - z/VM V5.2 GA (12/16/05): compatibility support
  - Post GA (05/26/06): exploitation support for z/VM 5.2
  - Collaborative Memory Management (4Q/2006): additional exploitation support for z/VM 5.2
- **New instruction support**
- **PER 3 Breaking-Event-Address Register (BEAR) support**

IBM System z9 EC

- **5 Models - S08, S18, S28, S38 and S54**
- **Processor Units (PUs)**
  - 12 (16 for Model S54) PUs per book
  - 2 SAPs per book, standard
  - 2 spares per server
  - 8, 18, 28, 38 or 54 PUs available
    - CPs, IFLs, ICFs, zAAPs, optional SAPs
- **Improved performance over the z990**
- **Up to 512 GB of central processor storage**
- **Up to 4 Logical Channel SubSystems (LCSSs)**
  - Up to **1024** channel paths/Up to 15 LPARs per LCSS
- **Up to 60 LPARs**
  - LPAR Mode only - No basic mode
- **FICON Express2/FICON Express**
  - Up to 84 features / 336 channels (FICON Express2)
- **OSA-Express2/OSA-Express**
  - Up to 24 features
  - Fast Ethernet, 1000BASE-T Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet
- **Up to 1024 ESCON channels**
- **Up to 8 configurable Crypto Express2 features**
- **IPL from FCP-attached SCSI disks**
- **Up to 16 HiperSockets**

IBM System z9 BC

- **Two models – R07 and S07**
  - 1 to 4-way high-performance server standard engines
  - Entry model with 1 to 3 standard engines
  - Up to a 7-way with specialty engines
- **73 capacity settings for a 2.6 times increase in flexibility over IBM eServer™ zSeries® 890 (z890)**
- **Delivers over 37% more capacity with the same low entry point as z890**
- **Up to 37% hardware performance improvement for Linux® (IFLs), Java™ (zAAPs) and coupling (ICFs)**
- **New zIIP for data serving workloads**
- **Double the memory – up to 64 GB per server**
- **Improved I/O Performance**
  - 40% more FICON® channels – up to 112
  - Up to 170% more bandwidth than z890
  - Improved FICON performance with Modified Indirect-Data-Address Word (MIDAW) facility
  - Double the FICON channel concurrent I/O operations from 32 to 64
  - Multiple Subchannel Sets (MSS) for an increased number of logical volumes

Support for System z9 Processors

- **TRACE STFLE**
- **TRACE TABLE response includes Breaking-Event address**
- **DISPLAY BEAR**

Support for System z9 Processors …

- **z/VM 4.4 and z/VM 5.1 require PTFs for APARs**
  - VM63577 — CP (V4.4)
  - VM63646 — CP (V4.4, V5.1)
  - VM63784 — CP (V5.1)
  - PK08444 — TCP/IP (V5.1)
  - VM63721 — HCD/HCM (V4.4, V5.1)
  - VM63869 — HCD/HCM (V4.4, V5.1)
  - VM63743 — EREP (V4.4, V5.1)
  - VM63946 — EREP (V4.4, V5.1)
  - VM63744 — CP (V5.1)
  - VM63722 — CP (V5.1)
  - VM63921 — CMS IOCP (V4.4, V5.1)
  - OA15170 — OSA/SF (V4.4, V5.1)

Processor and I/O Support …

- **Dynamic LPAR naming**
  - Support for dynamic creation or deletion of Logical Partitions
  - Reserved LPAR slots for dynamic configuration can be created without real resources
    - Using the **RESOURCE** statement in IOCP
    - By deleting an existing LPAR

Dynamic LPAR Naming

```
                              |-CSS 0--|
>>-DEFine-LPAR-lparname-+-------+-MIF_id-mm-><
                              |-CSS nn-|

                              |-CSS 0--|
>>-DELete-LPAR-lparname-+-------+-MIF_id-mm-><
                              |-CSS nn-|

>>-Query-LPARs-lparname-+-------+-+---------+-><
                              |-CSS nn-| |-REServed-|
```

Processor and I/O Support …

- **Native SCSI disk support enhancements**
  - Support for 1 TB LUNs (2,147,483,640 512-byte blocks)
  - Performance improvements (code path-length reductions, I/O efficiency improvements, paging/spooling optimization)
    - FBA emulator optimized to handle data more efficiently
    - Improves format, guest IPL, read operations
  - Monitor statistics for SCSI memory pool

Native SCSI Disk Enhancements

```
>>-Set-EDEVice-rdev-+- Device Operands -+-------><
                    |-CLEAR------------|

Device Operands:
|--TYpe-FBA--ATTRibutes-+-1750--+- Path Operands -|
                        |-2105--|
                        |-2107--|
                        |-SCSI--|

Path Operands:
|--+- Paths ---------------+----------------------|
   +-ADD PATH----+- Paths -|
   |-DELete PATH-|
```

Native SCSI Disk Enhancements …

```
Paths:
  +-----------------------------------------------+
  v                                               |
|-+-FCP_DEVice-rdev--WWPN-wwpn--LUN-lun-+-------------+-+-|
                                        |-PREFerred---|
                                        |-NOTPREFerred-|
```

Note:
1750 = DS6000
2105 = TotalStorage ESS
2107 = DS8000
SCSI = generic

Native SCSI Disk Enhancements …

- **`QUERY PATHS` reports preferred paths**
- **`QUERY EDEVICE DETAILS` identifies preferred and non-preferred paths**

Processor and I/O Support …

- **Crypto Express2 Accelerator**
  - Allow use of Crypto Express2 as an accelerator card
  - Support clear-key and secure-key operations for Crypto dedicated to z/OS guest
  - Support clear-key operations for Crypto sharing among Linux guests

Crypto Express 2 Accelerator

- **New `CEX2A` AP type reported by `QUERY CRYPTO` and `QUERY VIRTUAL CRYPTO`**
- **New Processor Domain monitor records**
  - Domain 5, Record 9: MRPRCAPC - Crypto performance counters
  - Domain 5, Record 10: MRPRCAPM - Crypto performance measurement data
- **New User Directory `OPTION` statement `CRYMeasure` specification to enable guest to collect hardware measurement data**

Processor and I/O Support …

- **OSA-Express2 support for Channel Data Link Control (CDLC) connectivity for IBM 374x Communication Controller connections**
  - Designed to allow the Network Control Program (NCP) to execute as if it were running on the 374x platform
  - NCP program product has been moved to the Linux OS environment, enabling NCP to run on Linux for zSeries
  - z/VM provides the ability to attach the OSA device and communicate with the Linux NCP image
  - z/VM 5.1 requires PTF for APAR VM63722

OSA Express-2 CDLC Support

```
>>-Query-OSA-+-------------------------+--><
             |-...---------------------|
             |      |-ALL-----------|  |
             |-Type-+---------------+-|
                    |-HIPersockets-|
                    |-OSA-----------|
                    |-OSN-----------|
```

```
>>-DEFine-CHPID-...-TYpe-+--------------+-...--><
                         |-...----------|
                         |-OSA_NCP------|
                         |-OSN----------|
```

Processor and I/O Support (05/2006) …

- **FCP N_Port ID virtualization**
  - Provide N_Port Identifier Virtualization (NPIV) for guests with dedicated FCP subchannels
  - Enables each z/VM guest as well as z/VM itself to be visible to a SAN fabric (switches/controllers)
  - Allows zoning and LUN-masking on a guest image basis
  - **QUERY FCP** reports **WWPN** associated with FCP subchannel
  - z/VM 5.1 requires PTF for APAR VM63744

Processor and I/O Support (05/2006) …

- **FICON Express4**
  - Support 4Gbits/sec FICON adapter
  - Auto-negotiation of 1/2/4 GBits/sec
  - z/VM 4.4 and z/VM 5.1 require PTF for APAR VM63744

Virtualization

- **Additional large real memory exploitation**
  - Enhanced performance and scalability for memory-constrained environments
  - Significant enhancements to the z/VM Control Program
    - Decouple “logical” addresses used by z/VM CP (System eXecution Space – SXS) from real addresses
    - Dramatically reduce requirement for real storage below 2G
      - Use Access Register mode to touch guest pages
      - Assign SXS aliases for guest pages referenced by 31-bit CP code
      - Move Frame Table above 2G

Additional Large Real Memory Exploitation

```
                |-L-|
>>-Display-H-+---+-- . . . ------------------><
                |-R-|


              |-L-|
>>-DUmp-H-+---+-- . . . --------------------><
              |-R-|


>>-STore-H-+-L-+-- . . . --------------------><
            |-R-|
```

Additional Large Real Memory Exploitation …

```
>>-LOCate-SXSTE-+-addr--------+-------------><
                |-ENTRY-addr2-|

>>-Query-SXSPages-----------------------------><

>>-Query-+-SXSStorage-+----------------------><
         |-SXSStore---|
```

Additional Large Real Memory Exploitation …

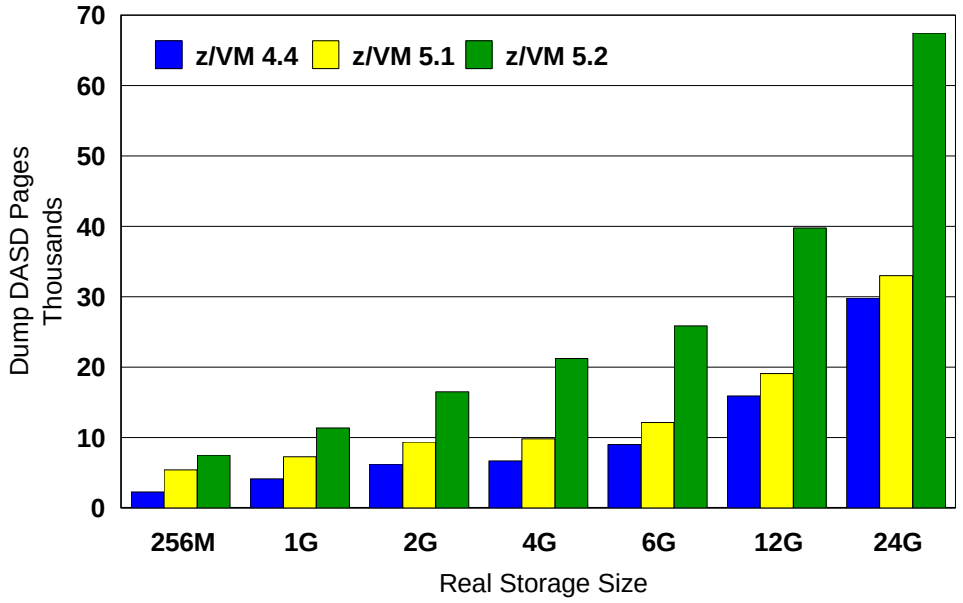
- **`QUERY FRAMES` output completely revised**
- **Many new, changed, and deleted CP termination codes**
- **Many new, changed, and deleted CP Trace Table Entries**
- **`INDICATE USER` response includes additional storage-related information**
- **No `STORAGE` information in `INDICATE LOAD` response**
- **`LOCATE FRAMETBL` reports 64-bit address**
- **Many `LOCATE` commands display host logical addresses**
- **`LOCATE VMDBK` displays host logical and real addresses**
- **Diagnose 4 handles 31-bit and 64-bit logical and real addresses**

Additional Large Real Memory Exploitation …

- **Changed Monitor records**
  - Domain 0 Record 3: MRSYTRSG - Real Storage Data (Global)
  - Domain 0 Record 4: MRSYTRSP - Real Storage Data (Per Processor)
  - Domain 0 Record 7: MRSYTSHS - Shared Storage Data
  - Domain 1 Record 7: MRMTRMEM - Memory Configuration Data
  - Domain 3 Record 1: MRSTORSG - Real Storage Management (Global)
  - Domain 3 Record 2: MRSTORSP - Real Storage Activity (Per Processor)
  - Domain 3 Record 3: MRSTOSHR - Shared Storage Management (per NSS or DCSS)
  - Domain 3 Record 14: MRSTOASI - Address Space Information
  - Domain 3 Record 16: MRSTOSHD - NSS/DCSS/SSP removed from storage
  - Domain 4 Record 2: MRSUELOF - User Logoff Data
  - Domain 4 Record 3: MRSUEACT - User Activity Data
  - Domain 4 Record 9: MRSUEATE - User Activity Data at Transaction End

Additional Large Real Memory Exploitation …

- **`SET DUMP ALL` removed**
- **Storage above 2G may be included in VMDUMPs and system dumps**
- **Additional dump DASD space may be required**
  - **One z/VM 5.2 system with 40G requires ~1G of dump space**



Virtualization …

- **V=V Guest QDIO passthrough support**
  - Follow-on support to z/VM 4.4 Adapter Interruption Passthrough
  - Allows guest systems to perform real QDIO data transfers while in SIE
  - Enhances guest performance and system throughput in QDIO-intensive environments
  - QDIO used for Gigabit Ethernet and FCP (SCSI) I/O operations
  - Available on z890, z990, and System z9

V=V Guest QDIO Passthrough Support

- **`QUERY VIRTUAL OSA` and `QUERY VIRTUAL FCP` report subchannel `TOKEN` value and indicate when `QEBSM` is active**
- **Changed monitor records**
  - Domain 1 Record 19: MRMTRQDC - QDIO Device Configuration
  - Domain 6 Record 25: MRIODQDA - QDIO Device Activation Event
  - Domain 6 Record 26: MRIODQDS - QDIO Device Activity Sample
  - Domain 6 Record 27: MRIODQDD - QDIO Device Deactivation Event

Virtualization …

- **64-bit Diagnose 250 support**
  - Targeted for 64-bit Linux exploitation
  - Allows Diagnose x'250' to perform I/O operations using 64-bit storage addresses

64-bit Diagnose X'250' Support

Read/Write Block I/O Parameter List

00	Device Number	Flag A	00000000	00000000000000000000000000000000
08	00000000000000000000000000000000			00000000000000000000000000000000
10	00000000000000000000000000000000			00000000000000000000000000000000
18	Key	Flags	000000000000000	Block Count
20	ALET of the block–entry list			00000000000000000000000000000000
28	64 bit Interruption parameter			
30	64 bit address of the block-entry list			
38	00000000000000000000000000000000			00000000000000000000000000000000

64-bit Diagnose X'250' Support …

Block I/O Entry

00	Type	Status	0000000000000000	Data-Buffer ALET
08	Block Number (64 bit)			
10	Data-Buffer Address (64 bit)			

Virtualization …

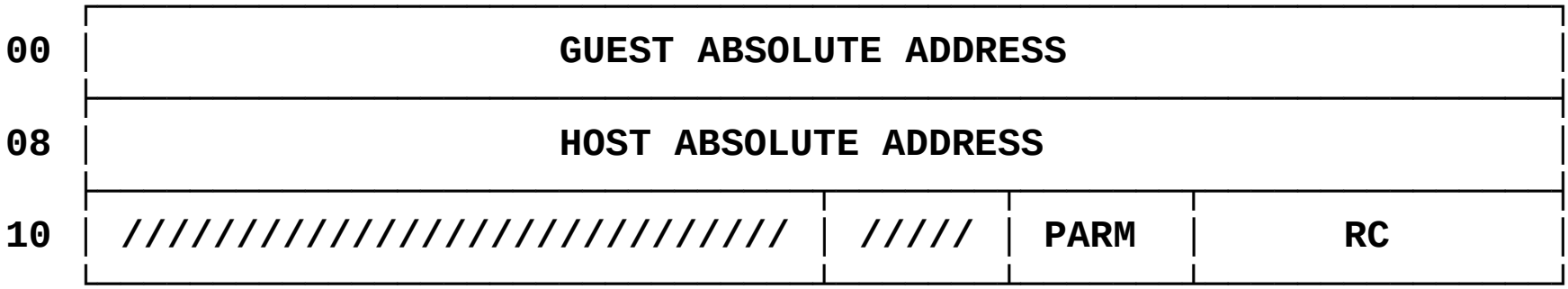
- **64-bit Diagnose 98 support**
  - Exploited by z/VM TCP/IP
  - Allows Diagnose x‘98' to lock pages using 64-bit storage addresses
  - Allows Diagnose X’98’ to lock pages above 2G in real memory

64-bit Diagnose X'98' Support

- **New "Release and Lock" sub-function of Block Diagnose 98 Request**
  - Releases any existing page contents
  - Locks page in storage
    - Parameter controls whether
      - Page locked below 2G
      - Page locked anywhere in storage, above 2G preferred

64-bit Diagnose X'98' Support …

Release and Lock Request Array Entry



Virtualization …

- **Diagnose X'290' – Perform Privileged Spool Functions**
  - Class D
  - Fetch current page of open spool file
  - Fetch virtual printer XAB data

Virtualization …

- **Diagnose X’ 9C’ – Voluntary Time Slice End With Target CPU**
  - Augments function provided by Diagnose X’44’
  - Enables guest to direct which virtual CPU should be dispatched
  - Typically used when a spin lock is held by a process dispatched on another virtual processor
  - Potential to reduce lock spin time more effectively

Virtualization … (05/2006)

- **Parallel Access Volume Support**
  - Dedicated and minidisk guest support
    - Dedicated support already exists
    - Virtual PAV aliases may be defined for a base device
    - Guest can use PAV semantics
  - Host support
    - Automatically send multiple I/O requests to a single device

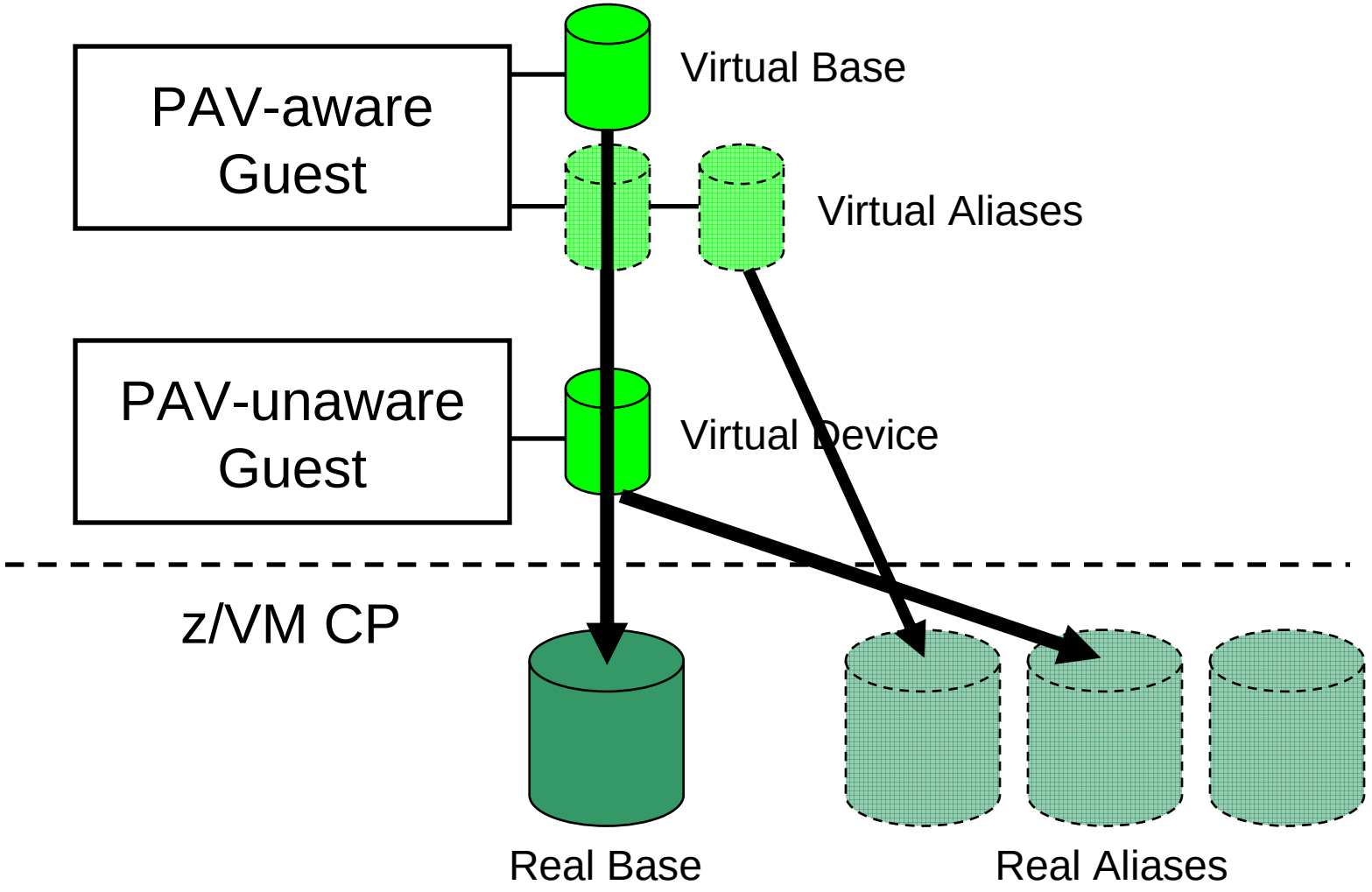
Parallel Access Volume Support

```
>>-DEFine-PAValias---vdev--------+-----+-BASE basevdev--><
                                 +-FOR-+

                          |-ALL----------|
>>-Query-+---------+-PAV-+--------------+--------------><
         +-Virtual-+     |-vdev---------|
                         |-vdev-vdev----|
```

Parallel Access Volume Support …

```
>>--DASDOPT---------+- other options ------------+------><
                    |          +-----------------+ |
                    |          v                   | |
                    |-PAValias-+-+-vdev---------+-+-|
                               |-vdev-vdev----|
                               |-vdev.numDevs-|

>>--MINIOPT---------+- other options ------------+------><
                    |          +-----------------+ |
                    |          v                   | |
                    |-PAValias-+-+-vdev---------+-+-|
                               |-vdev-vdev----+
                               |-vdev.numDevs-|
```

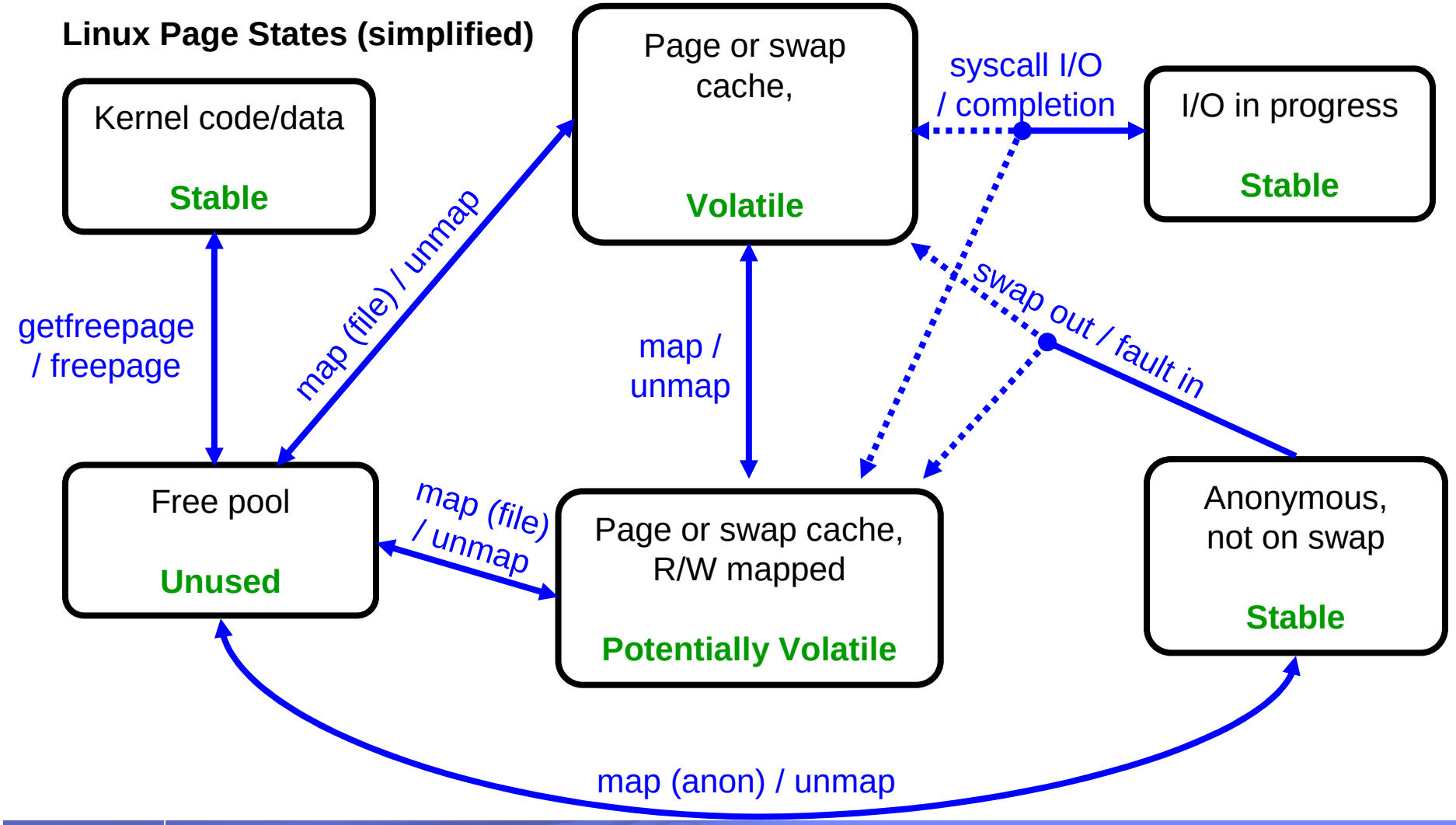
Parallel Access Volume Support …



Virtualization … (4Q/2006)

- **Collaborative memory management (CMM)**
  - Coordinates memory state and page management between Linux and z/VM at the level of individual pages
  - New ESSA instruction
  - Linux exploitation under discussion with Open Source community

Collaborative Memory Management (CMM)



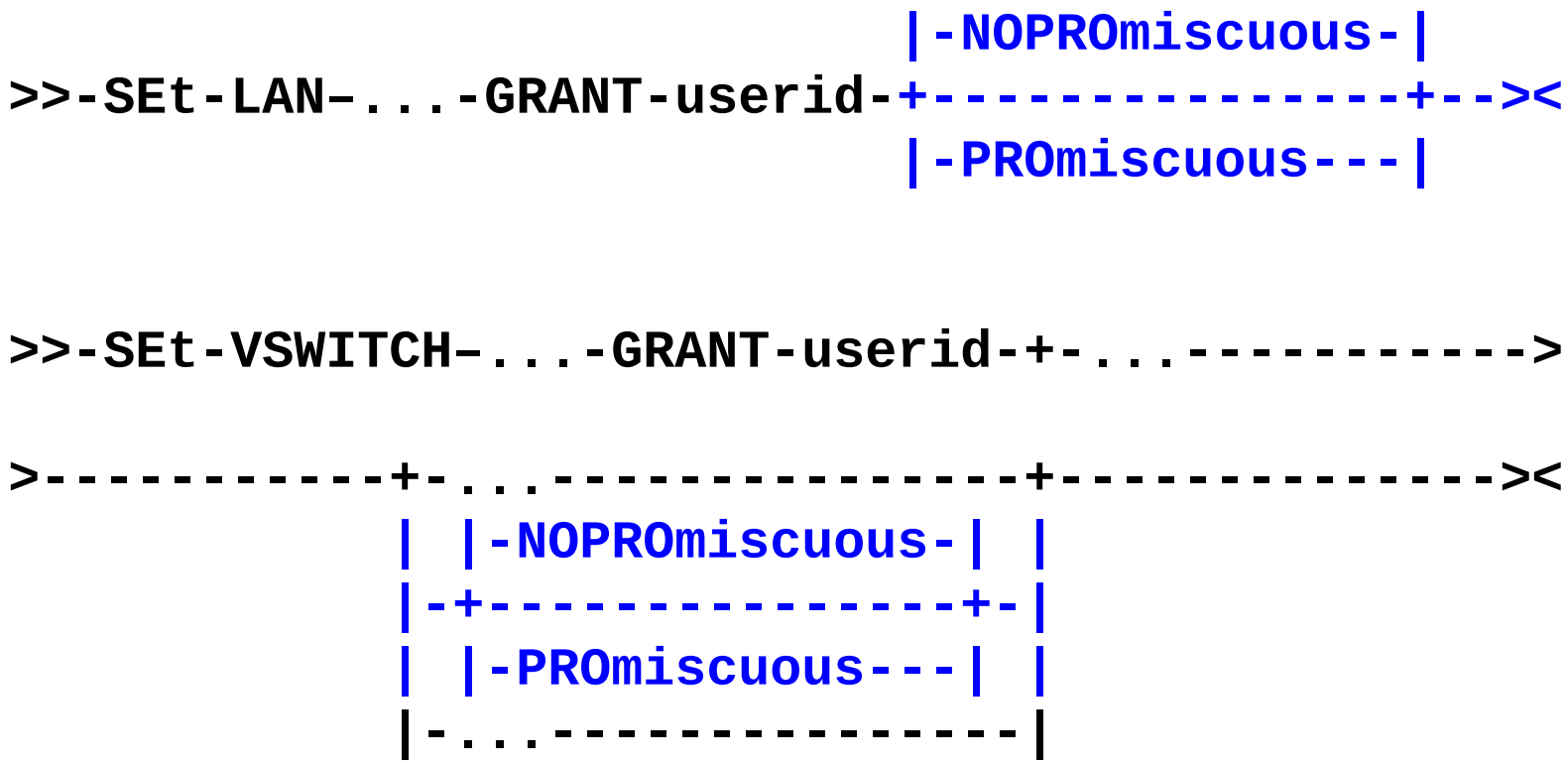
Virtual Networking

- **Guest LAN and Virtual Switch sniffing**
  - Authorized guests
    - Controlled by External Security Manager or CP
  - Aids in diagnosing networking problems
  - Supported by Linux **`tcpdump`** tool, z/VM **`TRSOURCE`** command and **`TRACERED`** utility, and z/VM TCP/IP **`IPFORMAT`** command

Guest LAN and Virtual Switch Sniffing

```
>>-TRSOurce-. . .-TYPE-LAN-OWNER-+-SYSTEM--+--->
                                 |-ownerid-|

                        |-VLAN-ALL------|
>--+-LANNAME-lanname-+--------------+--------->
                     |      +------+ |
                     |      v       | |
                     |-VLAN-+-vlan-+-|

   |-LENGTH 512--|
>--+-------------+-+-----------------+--------><
   |-LENGTH FULL-| |-NIC userid-vdev-|
   |-LENGTH size-| |-TRUNK-----------|
                 |-DROPPED---------|
```

Guest LAN and Virtual Switch Sniffing …



Guest LAN and Virtual Switch Sniffing …

```
>>-SEt-NIC–vdev-+-NOPROmiscuous-+--------><
                |-PROmiscuous---|

>>-Query-LAN-...-+-------------+----------><
                 |-PROmiscuous-|

>>-Query-TRSOurce-...-+--------------+----><
                      |-TYPE-+-LAN-+-|
                      |      |-...-| |

>>-Query-VSWitch-...-+-------------+-...--><
                     |-PROmiscuous-|
```

Virtual Networking … (05/2006)

- **Virtual Switch VLAN GVRP support**
  - GARP (Generic Attribute Registration Protocol) VLAN Registration Protocol
  - Provides more of standard switch semantics
  - **`QUERY VSWITCH`** and **`QUERY CONTROLLER`** report **`GVRP`** status or attribute

Virtual Switch VLAN GVRP Support

```
                                          |-GVRP---|
>>-DEFine-VSWITCH-. . .-VLAN nnn-+--------+-><
                                          |-NOGVRP-|

>>-SEt-VSWITCH-. . .-+-GVRP---+-><
                     |-NOGVRP-|
```

Virtual Networking … (05/2006)

- **IPv6 HiperSockets**
  - IPv6 datagrams can flow over virtual and real HiperSockets
  - **`QUERY LAN`** and **`QUERY NIC`** can report IPv6 information

Other

- **Remove standalone dump HLASM requirement**
  - No need to require HLASM to assemble HCPSDC
  - Eliminated use of HLASM-dependent features
  - Also available for z/VM 4.4 and 5.1
    - PTF for APAR VM63713

Removed Functions

- **Start-stop virtual consoles (2741/TTY)**
- **Device support: 3705/3720/3725/8232 LCS**