IBM

IBM IT Education Services

Session E51a

VSE/ESA 2.6 and 2.7

Performance Considerations

Ingo Franzki
ifranzki@de.ibm.com

VSE Technical Conference

RETURN TO INDEX

VSE

e-business

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and / or other counties.

CICS*
DB2*
DB2 Connect
DB2 Universal Database
e-business logo*
Enterprise Storage Server
HiperSockets
IBM*
IBM logo*
IMS
Intelligent Miner
Multiprise*
MQSeries*
OS/390*
S/390*
SNAP/SHOT*
Virtual Image Facility
VM/ESA*
VSE/ESA
VisualAge*
VTAM*
WebSphere*
xSeries
z/Architecture
z/VM
zSeries

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

LINUX is a registered trademark of Linus Torvalds

Tivoli is a trademark of Tivoli Systems Inc.

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries

UNIX is a registered trademark of The Open Group in the United States and other countries.

Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.

SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

Intel is a registered trademark of Intel Corporation.

VSE

e-business

Contents

- VSE/ESA 2.6
  - Performance Items (Reminder)
- VSE/ESA 2.7
  - Performance Items (Overview)
  - Hardware support
  - HiperSockets
  - Hardware Crypto support
- Dependencies for VSE/ESA Growth
- Hints and Tips

VSE

e-business

VSE/ESA 2.6 Performance Items

- VSE/ESA 2.6 Base enhancements
  - Delete Label Function
    - saves >90% SVCs for sequential processing
  - LTA Offload for some AR commands
    - Less I/O by less FETCHes for LTA load
  - SVA-24 Phases moved above the line
    - $IJBPRTY (6KB)
  - Increased max number of SDL entries
    - Maximum value now 32765
  - SDL update from non-BG partitions
  - POWER Data file extension without reformat

VSE

e-business

VSE/ESA 2.6 Performance Items - continued

- VSE/ESA 2.6 Hardware Support
  - FICON Support (VSE/ESA 2.3 or higher)
  - New 2074 System Management Console
    - ESCON channel attached
    - Eliminates requirement for a non-SNA 3174 controller
  - OSA Express Adapter (e.g. Gigabit Ethernet)
    - Available for G5 and above
  - VSAM Support for large 3390-9 Disks (Shark)
  - Fastcopy Exploitation of ESS FlashCopy and RVA SnapShot

VSE

e-business

Hardware Support

- Queued Direct I/O
  - Designed for very efficient exchange of data
  - Uses the QDIO Hardware Facility, without traditional S/390 I/O instructions
  - Without interrupts (in general)
  - Use of internal queues
  - With pre-defined buffers in memory for asynchronous use
- Exploitation by TCP/IP for VSE/ESA
  - see TCP/IP Performance Considerations

VSE

e-business

VSAM SHROPT(4) Avoidance

- Connectors in VSE/ESA 2.5 require SHROPT(4) when updating VSAM files owned by CICS
- New VSAM-via-CICS Service avoids SHROPT(4) by routing the VSAM requests to CICS
- Communication between batch and CICS is XPCC
- Naming convention for "VSAM-via-CICS files"
  - Each CICS is treated as "virtual" catalog
  - Files defined in CICS (via CEDA DEFINE FILE) are visible within this catalog

#VSAM.#CICS.<applid>

indicates "virtual" CICS catalog

APPLID of CICS region owning the files within this catalog

VSE

e-business

VSAM Redirector

- New connector with VSE/ESA 2.6
  - VSE is client
  - PC / workstation is server
- Exploits VSAM exit IKQVEX01
- Allows to redirect one of more VSAM files to a PC or workstation
- All VSAM requests of a particular file are redirected
  - Open / close
  - Get / put / point / delete / insert
- Transparent for applications
  - Usable from batch and CICS

IBM

VSE

e-business

VSAM Redirector - Performance Implications

- Is the file redirected ?
  - No: only at OPEN time (very small overhead)
  - Yes: at each request
- Network overhead ?
  - Yes, if file is redirected
  - Depends on
    - Number of VSAM requests
    - Size of records
- Data ownership
  - OWNER=REDIR
    - no VSAM I/O

VSE

e-business

VSE/ESA 2.7 Performance Items

- VSE/ESA 2.7 hardware support
  - z800/z900, Multiprice 3000, G5/G6
  - HiperSockets
  - Hardware Crypto Support
  - 32760 cylinder 3390 support
  - 3590 buffered tape mark
- VSE/ESA 2.7 enhancements
  - new TCP/IP for VSE/ESA release 1.5
  - $IJBLBR above the line
  - II User Status Record above the line
  - VTAPE: removed DVCDN/DVCUP
  - POWER: reallocate queue file during warm start

VSE

e-business

VSE/ESA 2.7 Hardware support

- VSE/ESA 2.7 runns on the following machines
  - zSeries: z800, z900, z990
  - 9672 Parallel Enterprise Server (G5/G6)
  - Multiprise 3000 (7060)
  - equivalent emulators (Flex-ES)
- VSE/ESA 2.7 is based on the hardware instruction set described in the manual 'ESA/390 Principles of Operation' (SA22-7201).
- With VSE/ESA 2.7 it is assumed that all the ESA/390 instructions and facilities described in that manual can be used.

VSE

e-business

Supported VSE Releases

- VSE/ESA 2.4/2.3: already out of service
  - runs also on zSeries (z800, z900)
  - Does not run on z990 (Hardwait during IPL)
- VSE/ESA 2.5: end of service 31. December 2003
- VSE/ESA 2.5 and 2.6
  - runs also on zSeries (z800, z900)
  - runs also on z990 with additional PTF
- VSE/ESA 2.7
  - runs on zSeries (z800, z900, z990, G5/G6, MP3000)
- OSA Express
  - Supported with VSE/ESA 2.6 and 2.7
- HiperSockets and PCICA (Crypto)
  - Supported with VSE/ESA 2.7

VSE

e-business

zSeries Remarks

- Prior to zSeries (z800/z900/z990) there is one cache for data and instructions
- zSeries has splited data and instruction cache
- Performance implications:
  - If program variables and code thjat updates these program variables are in the same cache line (256 byte)
    - Update of program variable invalidates instruction cache
    - Performance decrease if update is done in a loop
  - See APAR PQ66981 for FORTRAN compiler

VSE

e-business

32760 cylinder 3390 support

- With announcement 101-341 at 11/13/2001 IBM announced the new 32760 cylinder 3390 volumes of the IBM TotalStorage Enterprise Storage Server (ESS).
  - This enhancement of the ESS F models was made available 11/30/2001.
- VSE/ESA 2.7 now supports these volumes.
  - helps relieve address constraints
  - improves the disk resource utilization
  - can be used to consolidate multiple disk volumes into a single address.

VSE

e-business

3590 Buffered Tape Mark support

- The 3590 control unit provides support for writing tape marks (TM) in buffered mode
- Writing TM's in "buffered" mode should enhance the performance
  - of all programs, which write many TM's as part of their file creation process (e.g. POFFLOAD)
- All the TM's written during OPEN/CLOSE (label processing) will remain to be written "UNbuffered"
  - all the programs which write TM's mainly or only during OPEN/CLOSE, will NOT benefit from this enhancement.

VSE

e-business

$IJBLBR phase moved above the line

- The $IJBLBR.PHASE has been splited into two phases
  - $IJBLBR.PHASE
  - $IJBLB31.PHASE
- $IJBLBR.PHASE will continue to reside in SVA-24
- $IJBLB31.PHASE will reside in SVA-ANY (high SVA).
  - This will free about 180k in SVA-24.

VSE

e-business

II User status record above the line

- During Logon each II user gets besides others two storage areas allocated
  - User_Status_Record USR (904 bytes)
  - Panel_Hierarchy_List PHL (1352 bytes)
  - originally located in the CICS DSA (below)
- With VSE/ESA 2.7 the USR and PHL has been moved to ESDSA (shared above)
  - frees 2.3 KB in DSA below per user.
- ICCF TCTUALOC=ANY now supported
  - ICCF transaction programs has been changed to support a TCTUA (28 bytes) above the line

VSE

e-business

HiperSockets hardware elements ('Network in a box')

- Synchronous data movement between LPARs and virtual servers within a zSeries server
  - Provides up to 4 "internal LANs" HiperSockets accessible by all LPARs and virtual servers
  - Up to 1024 devices across all 4 HiperSockets
  - Up to 4000 IP addresses
  - Similar to cross-address-space memory move using memory bus
- Extends OSA-Express QDIO support
  - LAN media and IP layer functionality (internal QDIO = iQDIO)
  - Enhanced Signal Adapter (SIGA) instruction
    - No use of System Assist Processor (SAP)

VSE

e-business

HiperSockets hardware elements ('Network in a box') - continued

- HiperSockets hardware I/O configuration with new CHPID type = IQD
  - Controlled like regular CHPID
  - Each CHPID has configurable Maximum Frame Size
- Works with both standard and IFL CPs
- No physical media constraint, no physical cabling, no priority queuing
- Secure connections

VSE

e-business

Measurement Environment

- z800 (2066-004)
  - 4 processors
- VSE/ESA 2.7 GA Driver in a LPAR (native)
  - 1 CPU active (~2066-001)
  - TCPIP00 (F7): OSA Express Fast Ethernet
  - TCPIP01 (F8): HiperSockets
- Linux for zSeries in a LPAR (native)
  - 3 CPUs active (shared)
  - eth0: OSA Express Fast Ethernet
  - hsi10: HiperSockets

VSE

e-business

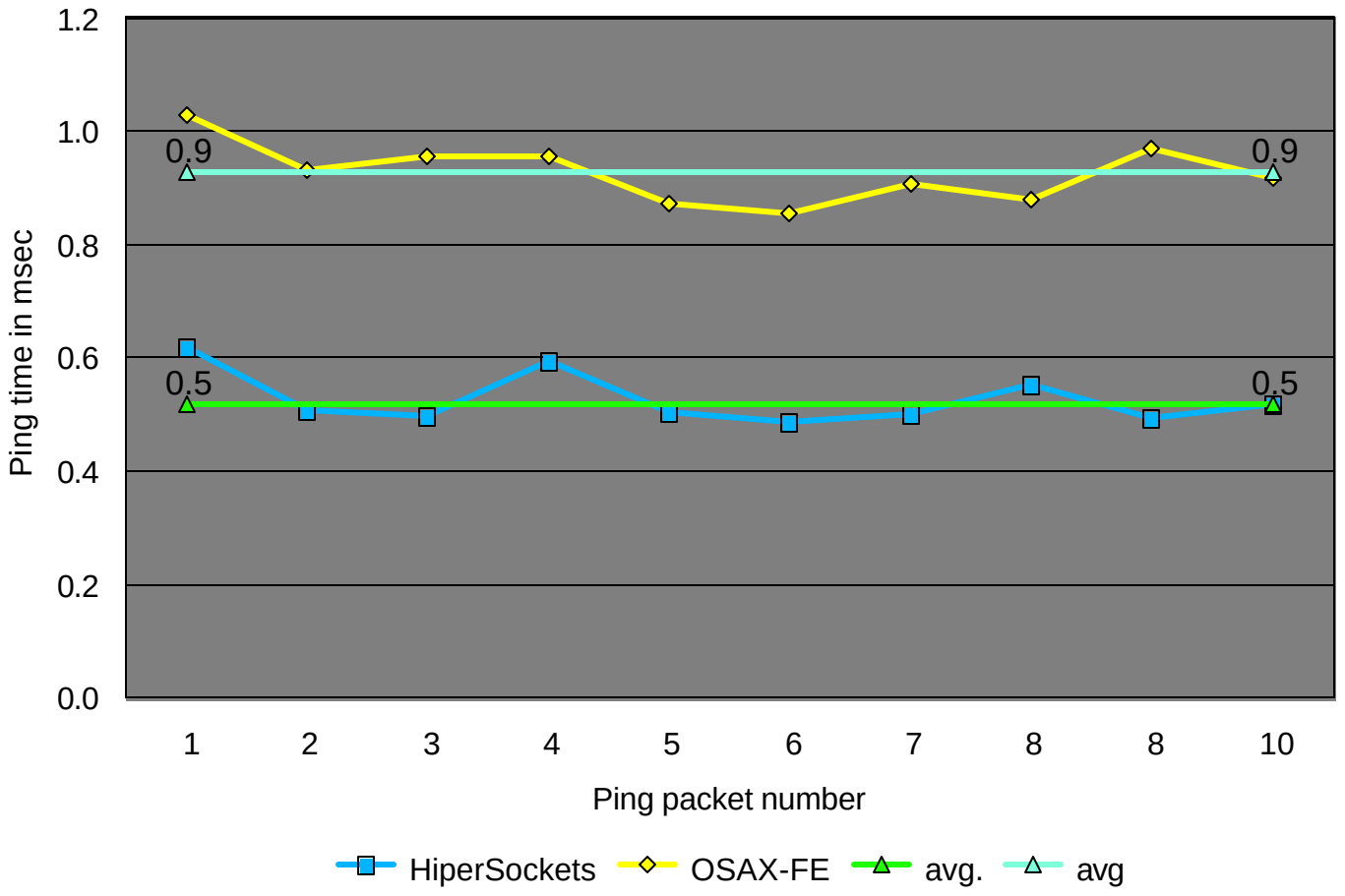
Latency (Round trip time) - results

- Measurements has been done with PING command
  - Issued at Linux side
  - 10 Pings
  - PING sends a datagram to VSE
  - VSE sends a answer back to Linux
  - Time until answer arrives is measured
    - Round trip time

VSE

e-business

Latency (Round trip time) - results



HiperSockets is about 1.8 times faster in terms of latency

VSE

e-business

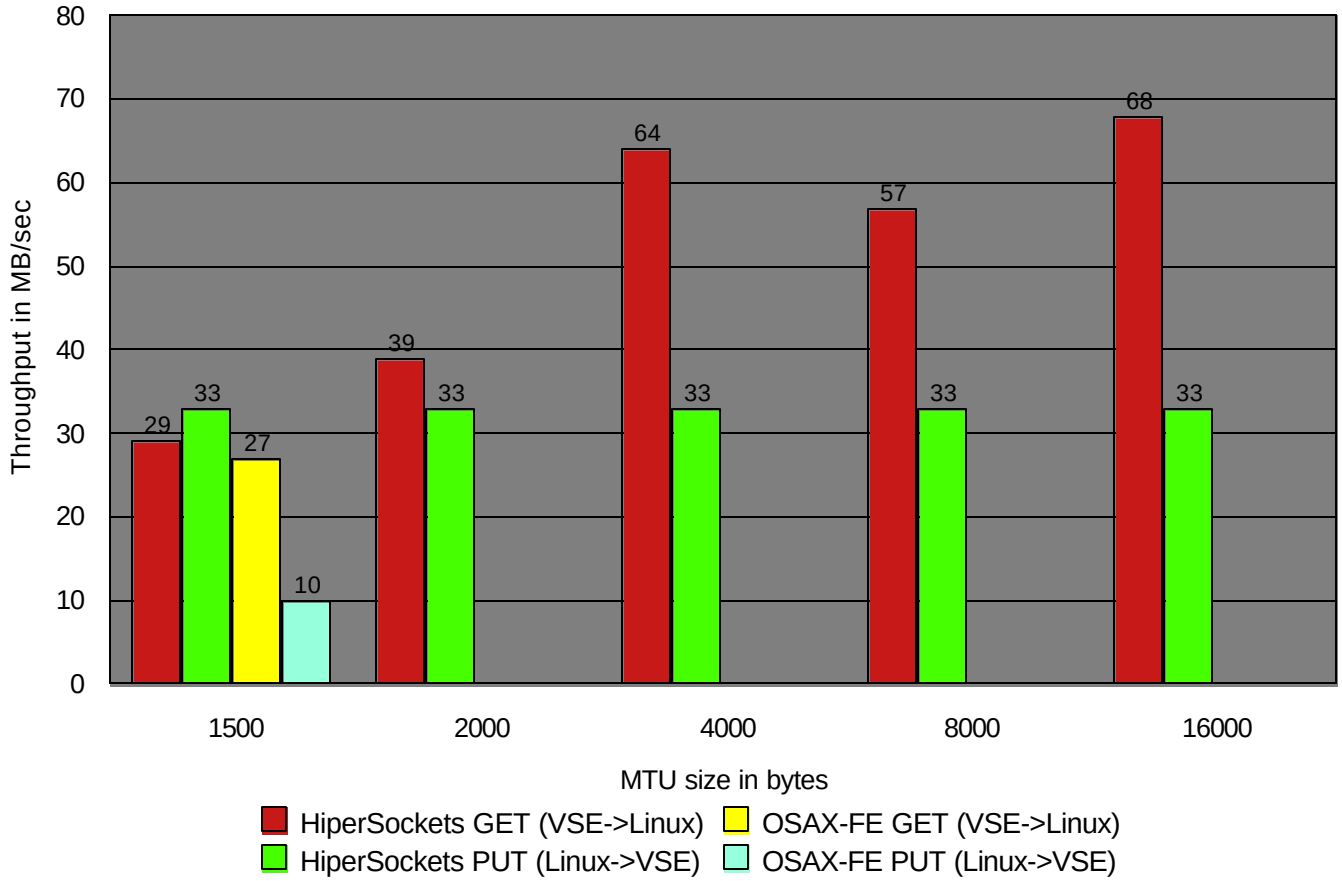
Throughput (MB/sec)

- Measurements has been done with FTP
  - Initiated at the Linux side
  - Transferring 1GB (1000MB)
    - without translation (binary)
    - 1 to 5 parallel streams
  - PUT: send data to VSE
    - VSE inbound
    - sending a 1GB file to $NULL file (in memory file)
    - No file I/O is done by VSE/Linux
  - GET: receive data from VSE
    - VSE outbound
    - receiving $NULL file (in memory file) into /dev/null
    - No file I/O is done by VSE/Linux

VSE

e-business

Throughput (MB/sec) - results

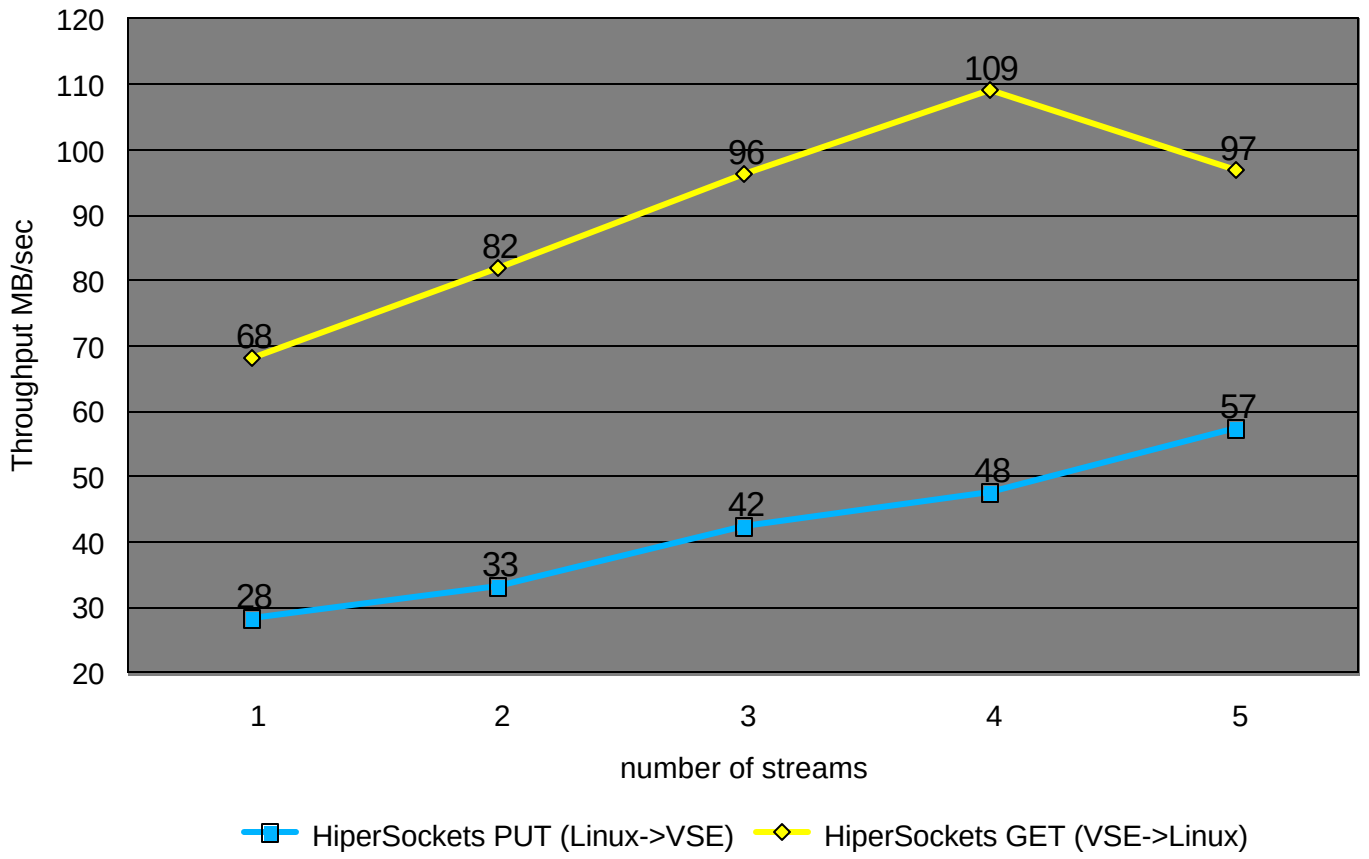


HiperSockets throughput is between 30-80 MB/sec

VSE

e-business

Throughput (MB/sec) - results (2)



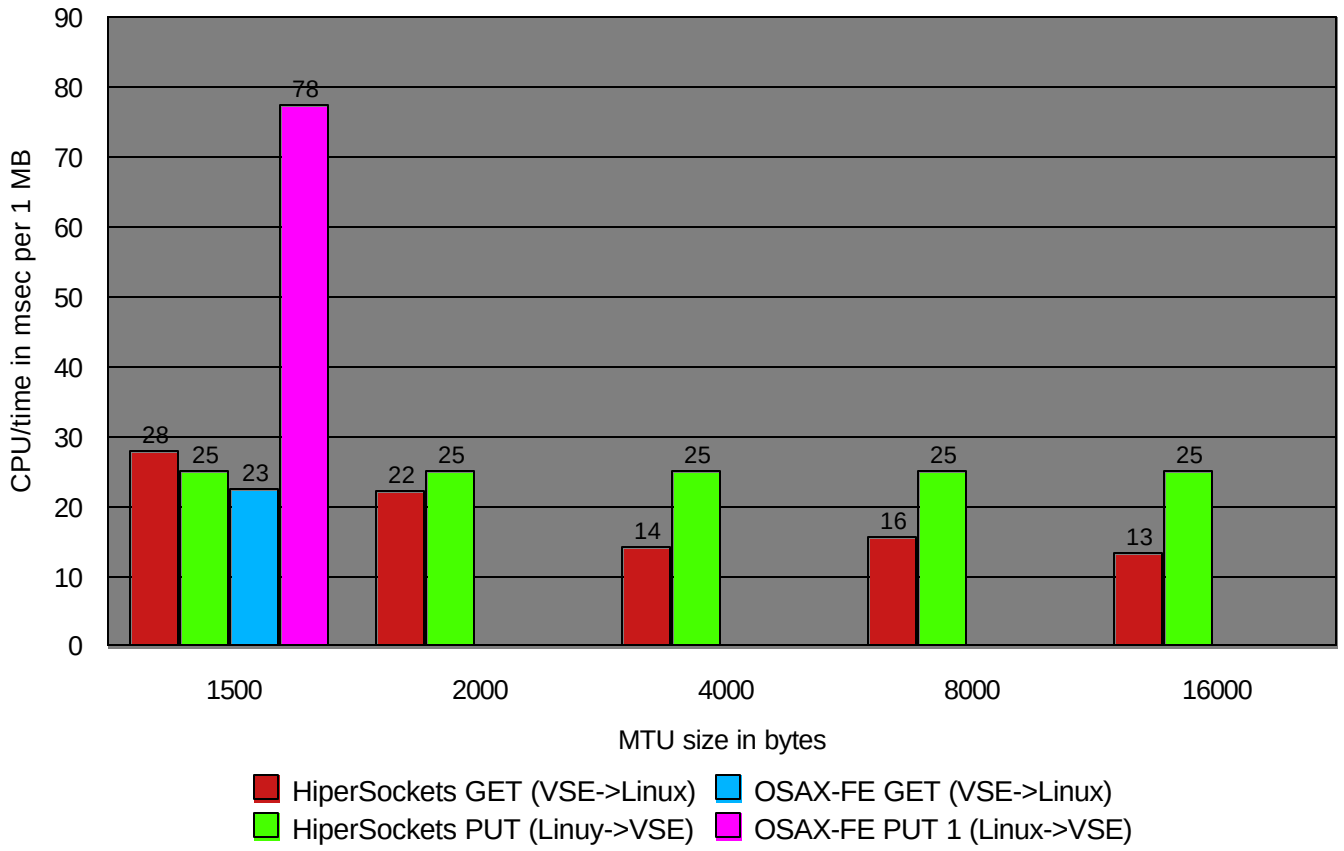
Maximum HiperSockets throughput of 109 MB/sec at 4 concurrent connections

E51a.prz

VSE

e-business

CPU time per MB - results



About 15-30 msec CPU time per MB for HiperSockets
(on a z800 2066-001)

VSE

e-business

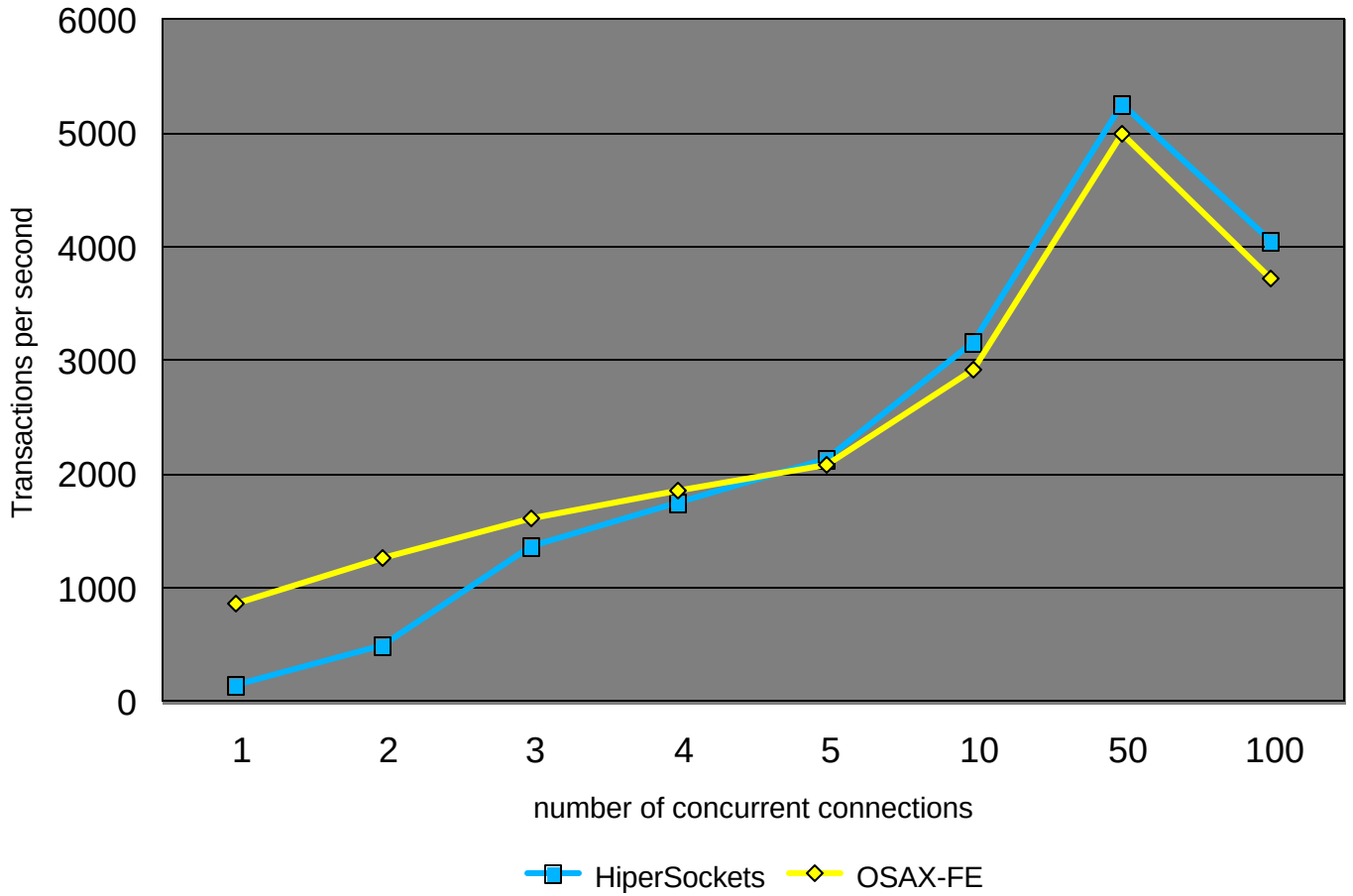
Transaction per second

- Measurements has been done with an ECHO server
  - Client on Linux sends 100 bytes to server
  - Server on VSE echoes 100 bytes
  - Per TCP connection 10000 transactions are driven
  - Variations: Number of TCP connections
    - 1,2,3,4,5
    - 10,50,100
  - Measurements
    - Transactions per second
    - CPU time per transaction

VSE

e-business

Transactions per second - results



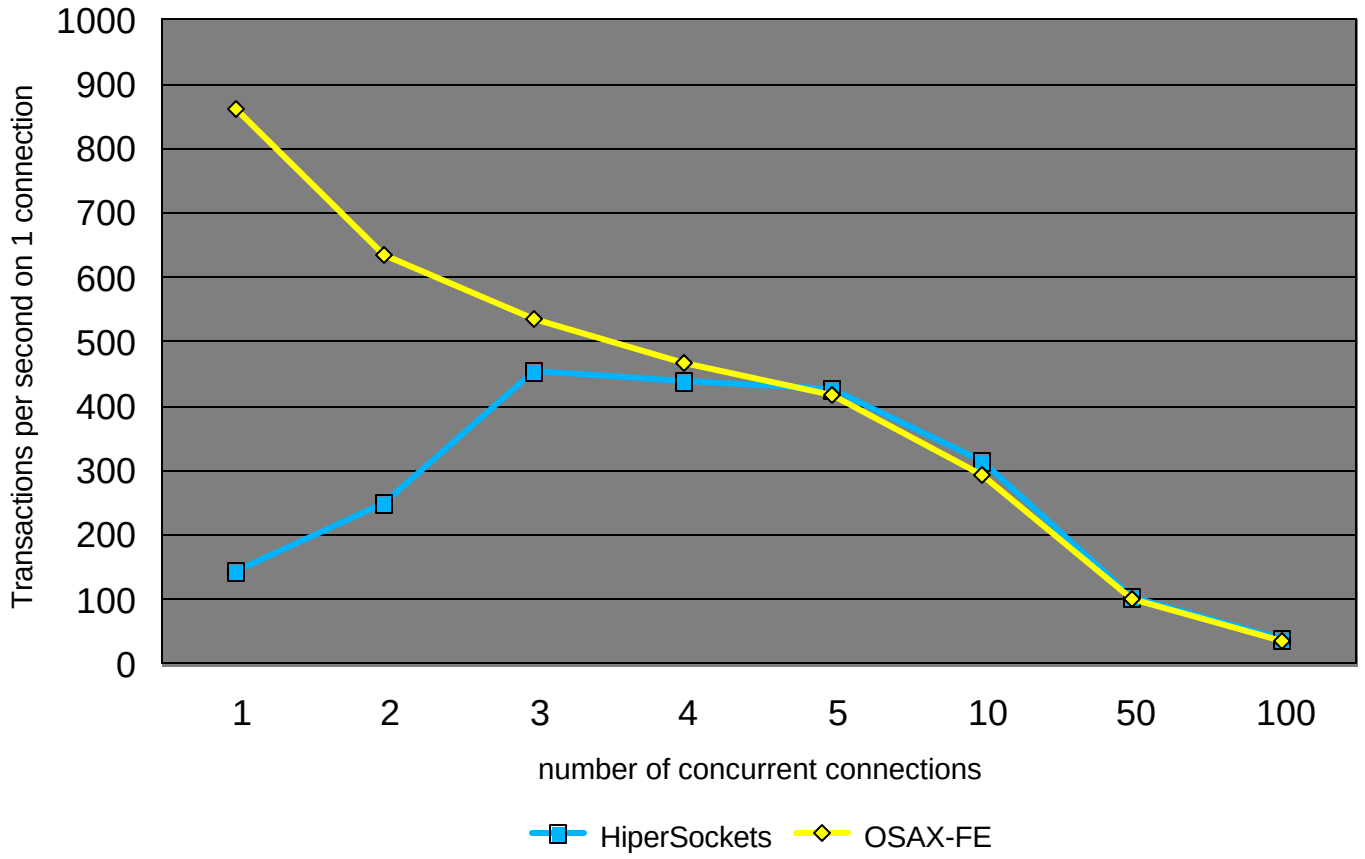
Maximum of 5200 transactions per second at 50 concurrent connections

E51a.prz

VSE

e-business

Transactions per second on 1 connection - results

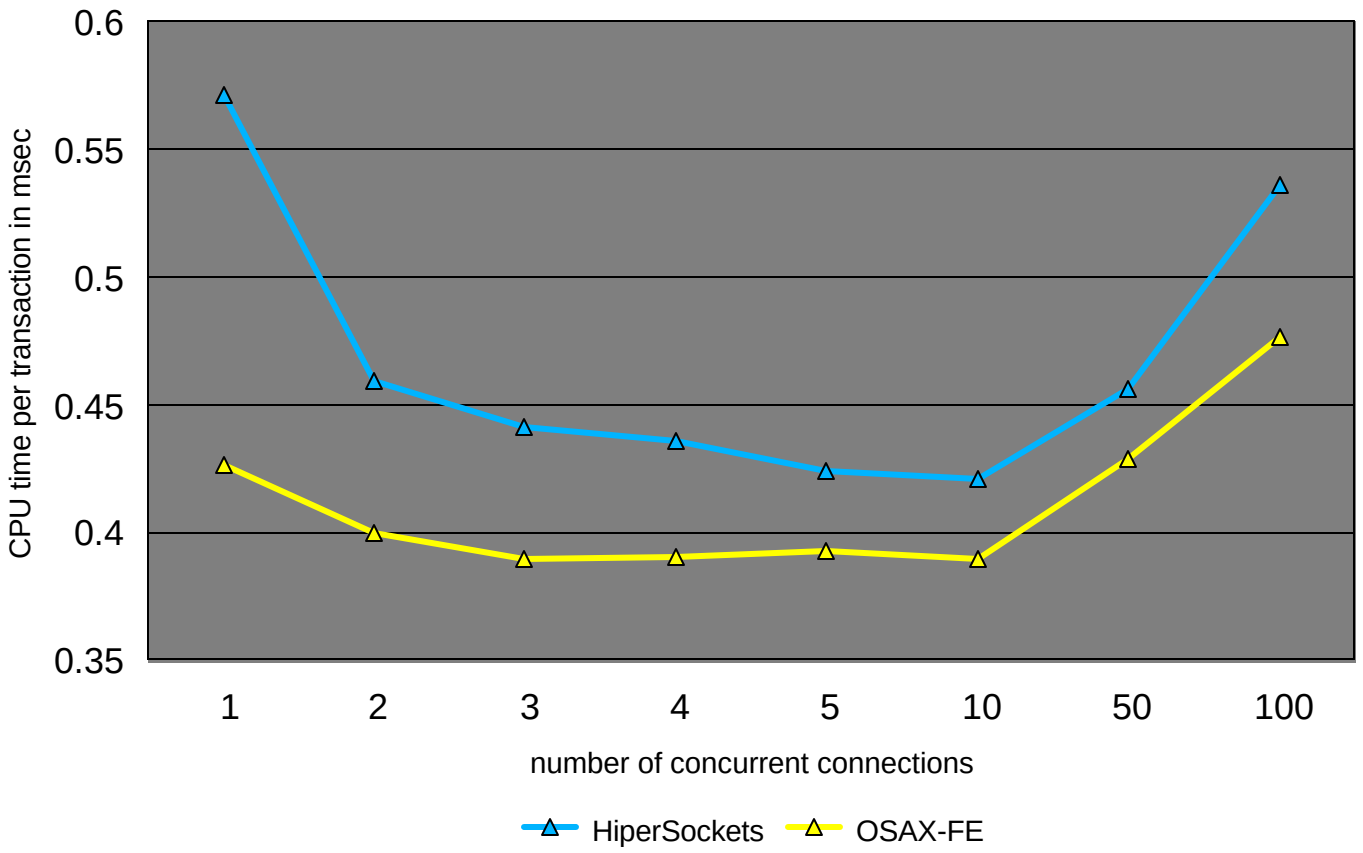


HiperSockets: Maximum of about 450 transactions per second on 1 connection (= about 2 msec response time)

VSE

e-business

CPU time per transaction



HiperSockets: About 0.45 msec CPU time per transaction for 2-50 connections

VSE

e-business

Measurement Results - conclusion

- HiperSockets
  - Throughput
    - Between 30-80 MB/sec
    - Maximum throughput of 109 MB at 4 connections
    - About 15-30 msec CPU time per MB
  - Transactions per second
    - Maximum of 5200 Transactions per second at 50 connections
    - About 0.4-0.45 msec CPU time per transaction

VSE

e-business

Hardware Crypto Overview

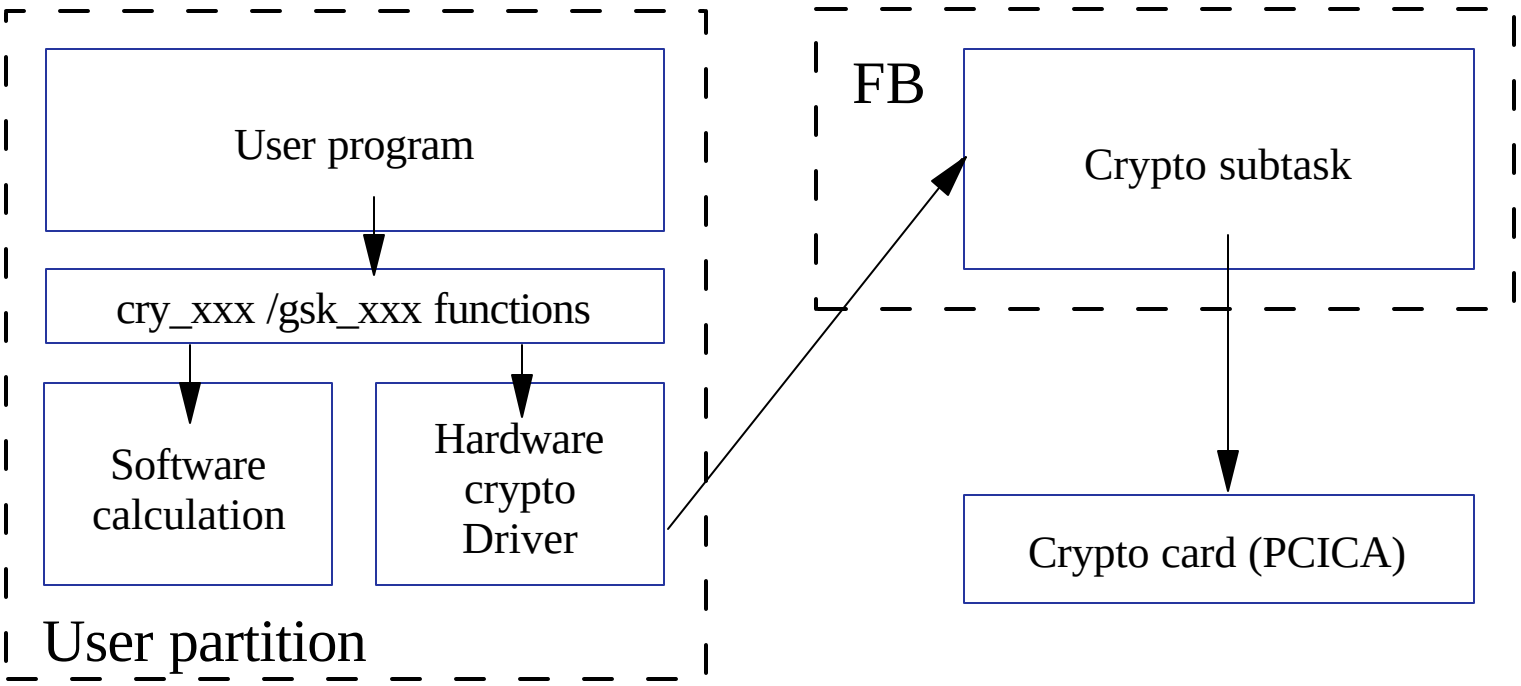
- Requires VSE/ESA 2.7 and TCP/IP for VSE/ESA 1.5
- Supported crypto cards
  - PCI Cryptographic Accelerator (PCICA)
    - Feature code 0862
    - Available for zSeries (z800, z900)
- The crypty card is plugged into the Adjunct Processor
- Currently only RSA (asymmetric) is supported
  - Of benefit for Session initiation (SSL-Handshake)
- Also supported with
  - z/VM 4.2 + APAR VM62905
  - z/VM 4.3

VSE

e-business

Hardware Crypto Overview - continued

- New crypto subtask in Security Server (SECSERV) running in FB
  - Or as separate job if no SECSERV is running
  - Crypto card is polled by crypto task



VSE

e-business

Measurement Environment

- VSE/ESA 2.7 running on a z900 (2064-109)
  - on 1 processor (~2064-101)
  - with a PCI Cryptographic Accelerator
- Testcase programs on VSE
  - Crypto operations measurements
    - calling cry_xxx functions (RSA, DES, SHA, MD5)
    - each crypto operation is performed 10000 times
  - Secured data transfer (SSL)
    - performs SSL handshake
    - performs encrypted data transfer
    - counterpart program running on Windows (SSL-client)
- All RSA operations are measured
  - with Hardware Crypto support
  - with Software Crypto
    - (support already available with TCP/IP 1.4/1.5 as shipped in VSE/ESA 2.6)

VSE

e-business

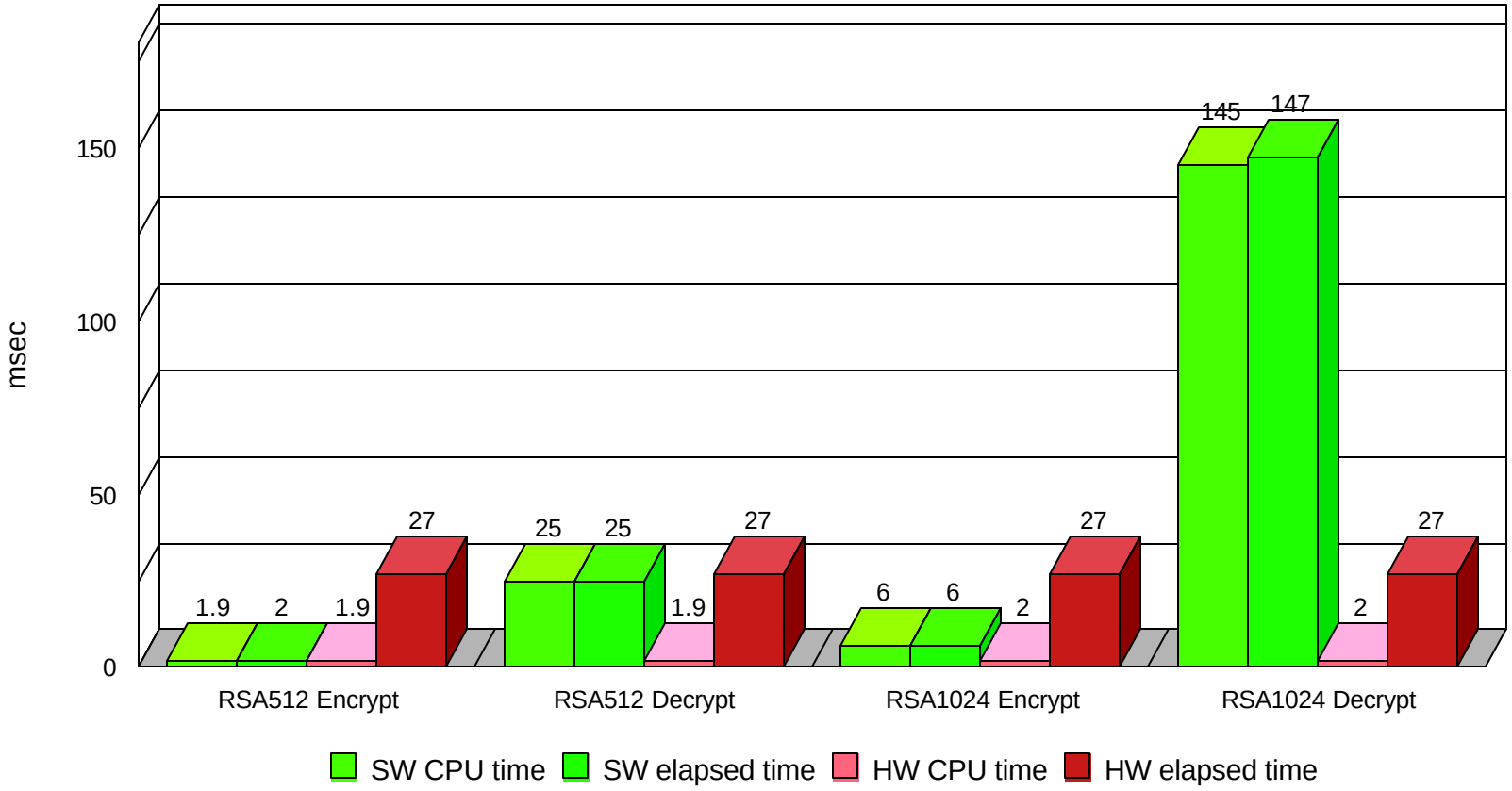
Measurement Environment - continued

- Variations
  - RSA encrypt/decrypt
    - 512 / 1024 bit key
  - DES, DES CBC, 3DES CBC encrypt/decrypt
    - software crypto only
    - message length (128, 256, 512 bytes)
  - SHA Hash, MD5 Hash, SHA HMAC, MD5 HMAC
    - software crypto only
    - message length (128, 256, 512, 1K, 2K bytes)
  - SSL handshake/data transfer
    - 01 RSA512_NULL_MD5
    - 02 RSA512_NULL_SHA
    - 08 RSA512_DES40CBC_SHA
    - 09 RSA1024_DES_CBC_SHA
    - 0A RSA1024_3DES_EDE_CBC_SHA

VSE

e-business

Measurements Results - RSA



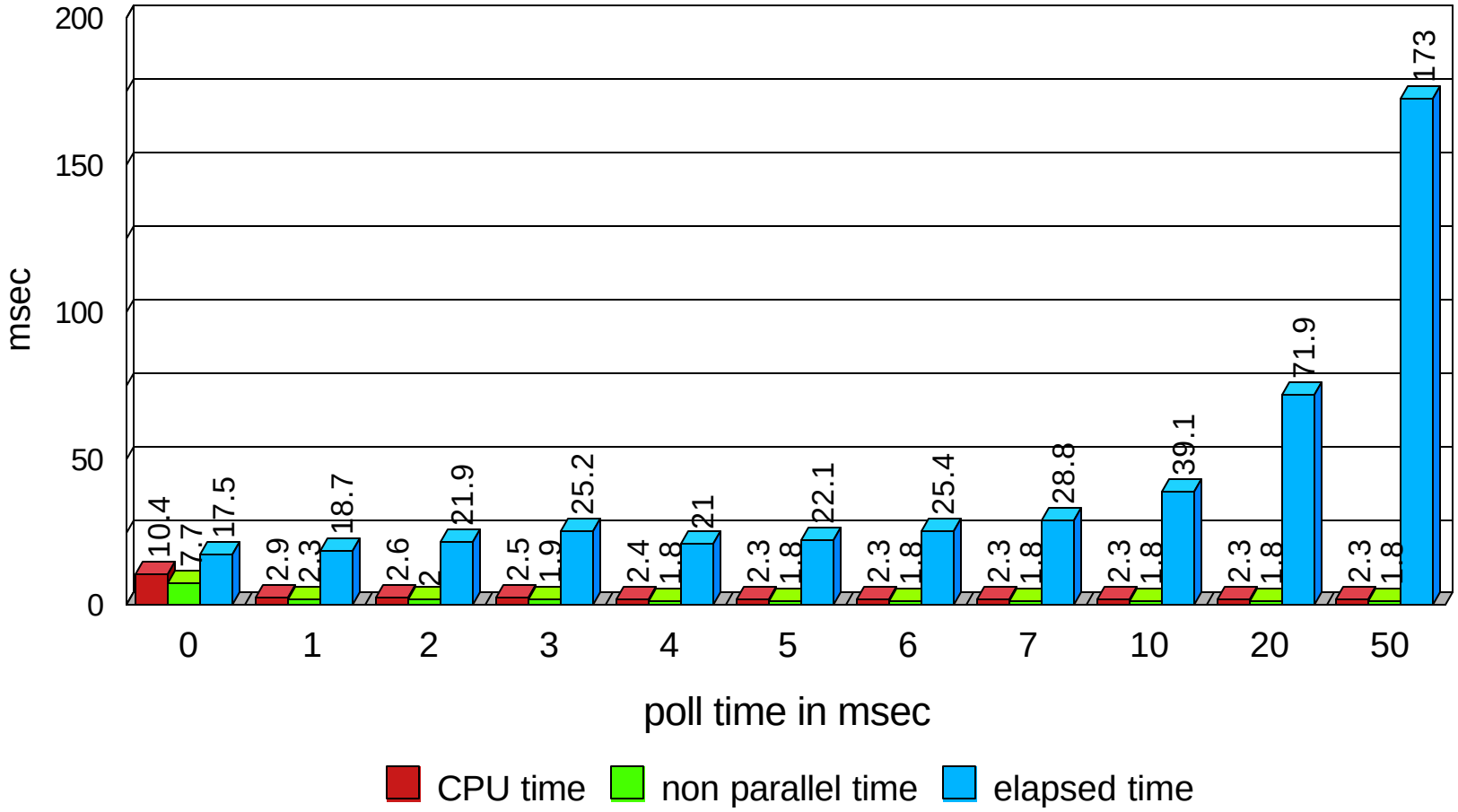
HW Crypto:

- CPU time and elapsed time is independent of operation / key length
- RSA operation takes about 2 msec CPU time and 28 msec ellapsed time
- CPU time is always less than software crypto

VSE

e-business

Measurements Results - RSA polltime



Per default a polltime of 7 msec is used.

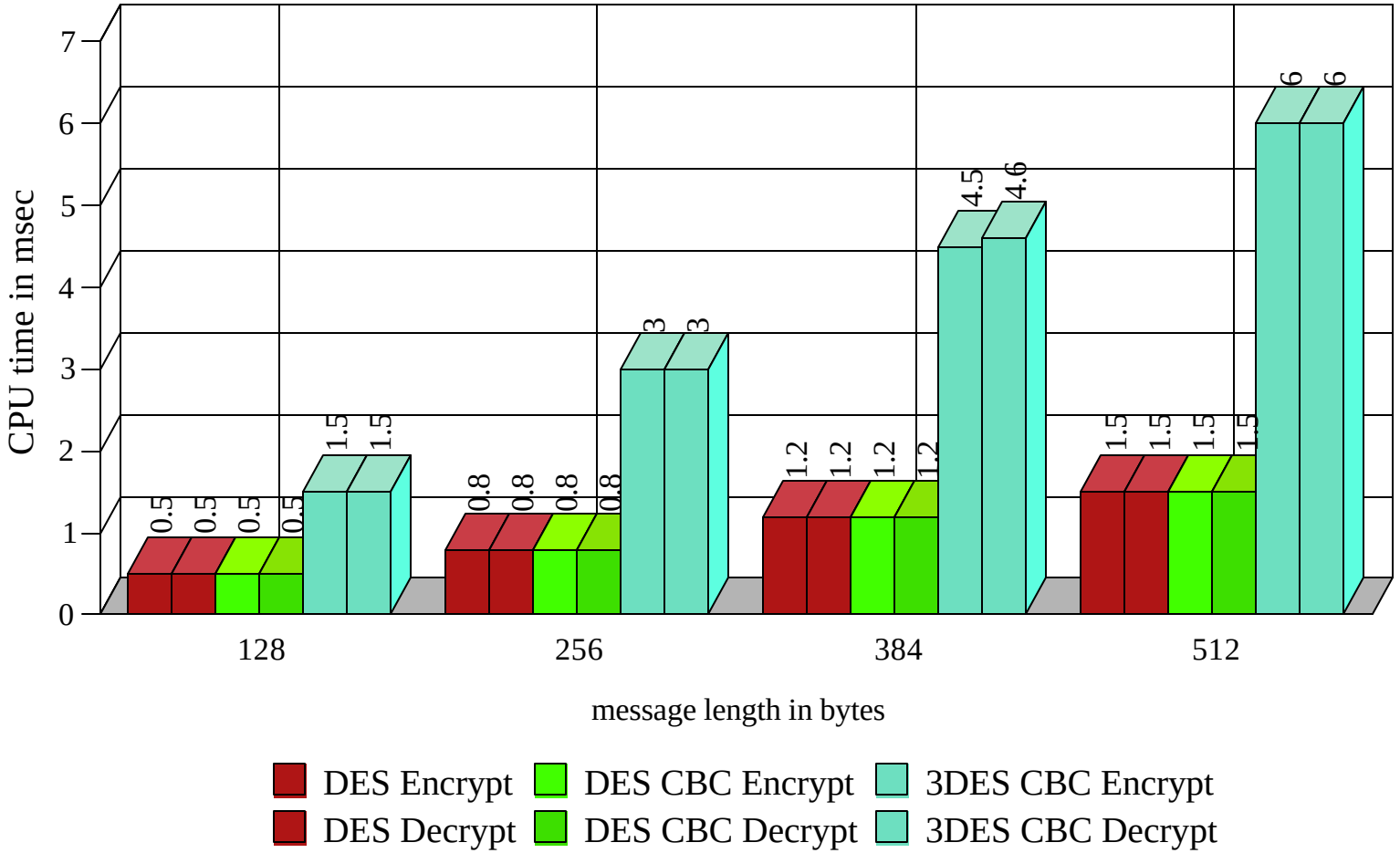
Can be changed with: MSG FB,DATA=WAITTIME=nn

Smaller values increases CPU time, higher values increases ellapsed time

VSE

e-business

Measurements Results - DES, DES CBC, 3DES CBC (symmetric)



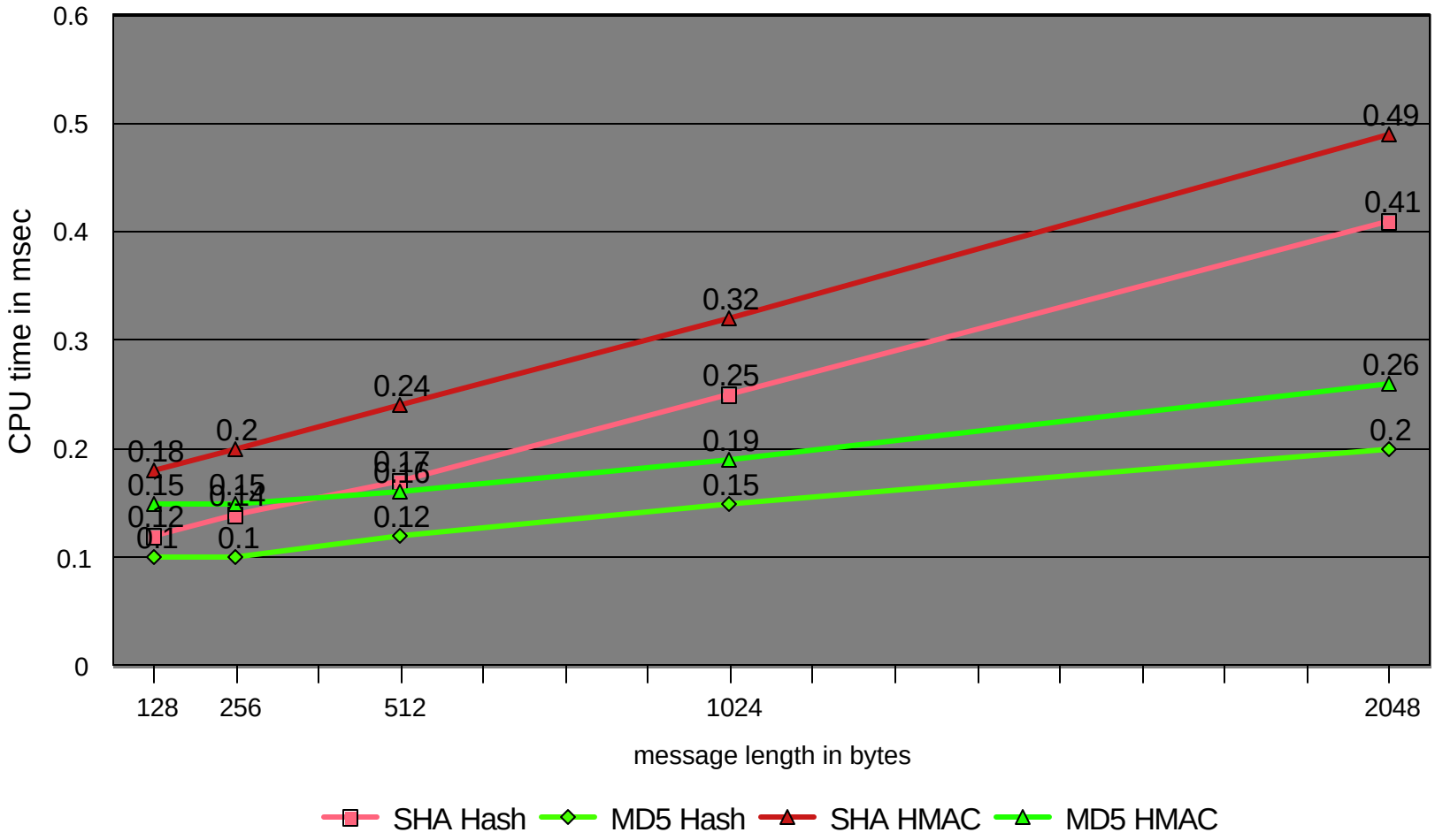
Software Crypto only!

DES and DES CBC takes similar CPU times, 3DES CBC about 3.8 times

VSE

e-business

Measurements Results - SHA, MD5

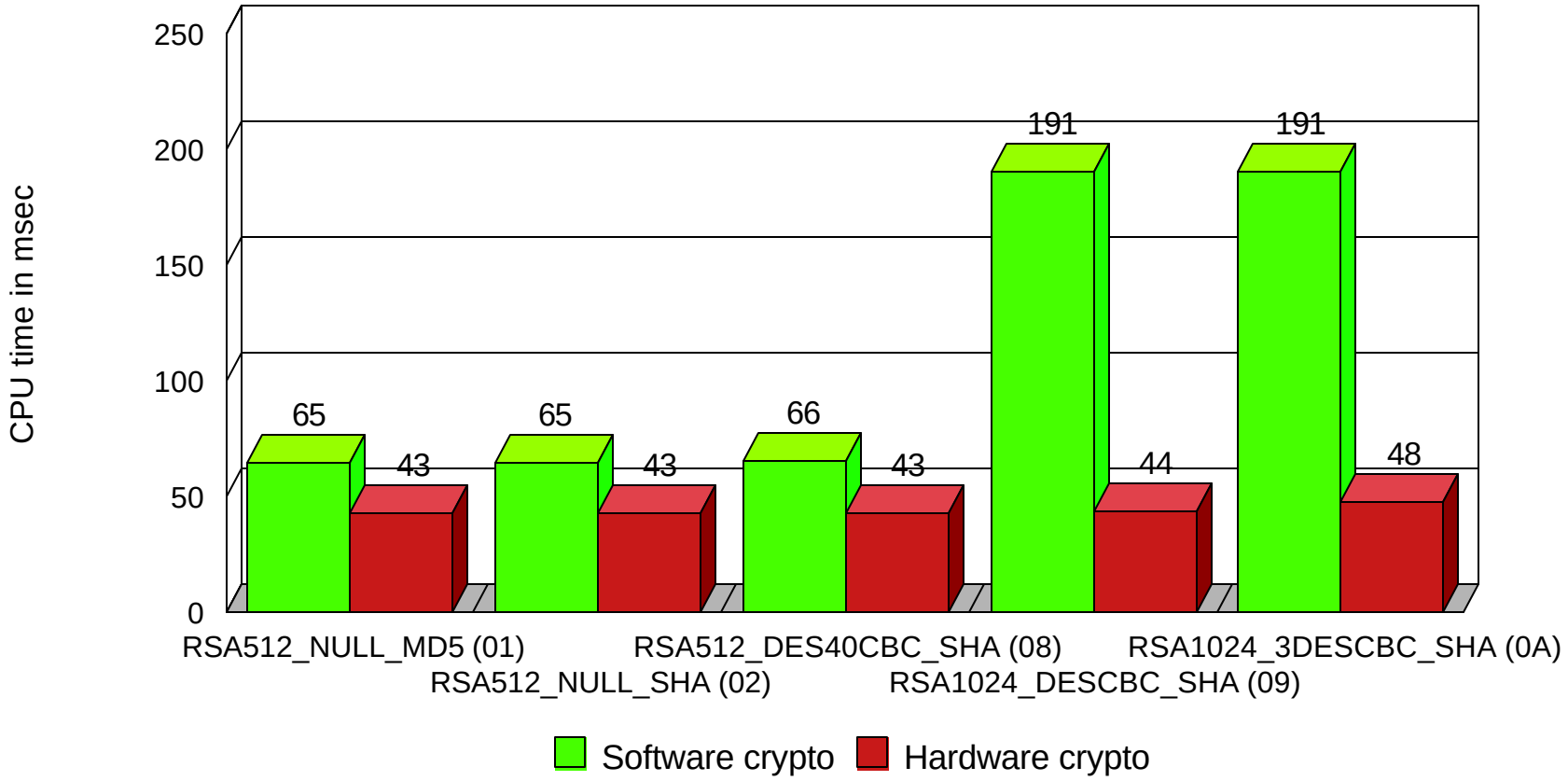


SHA takes about 1.8 times more CPU time compared to MD5
Software Crypto only!

VSE

e-business

Measurements Results - SSL Handshake



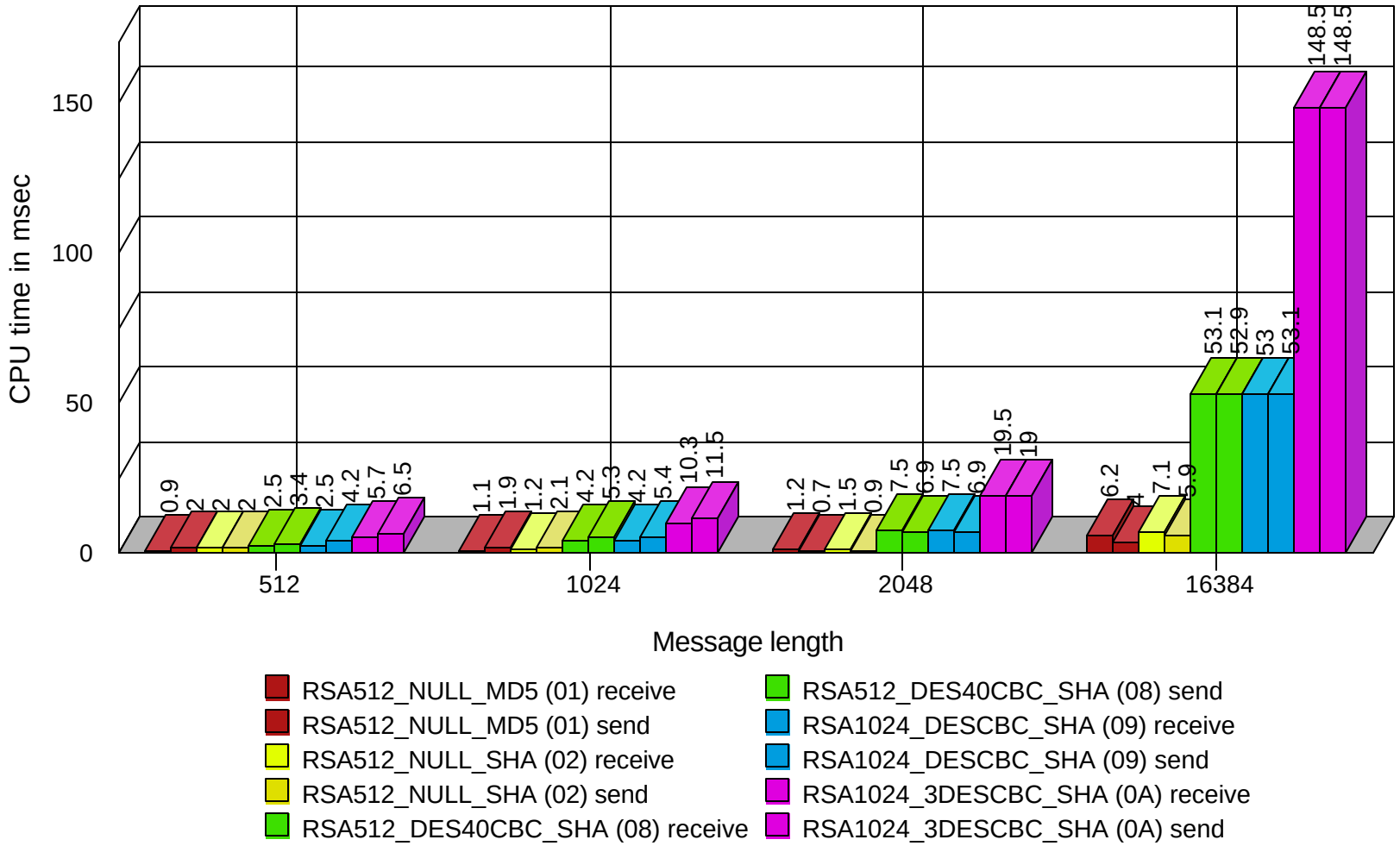
HW Crypto:

- CPU time and elapsed time is independent of cipher suite used
- SSL handshake takes about 43-48 msec CPU time (connection establishment)

VSE

e-business

Measurements Results - SSL data transfer



CPU time depends on used hashing (SHA/MD5) and encryption algorithm (DES/3DES

Software Crypto only!

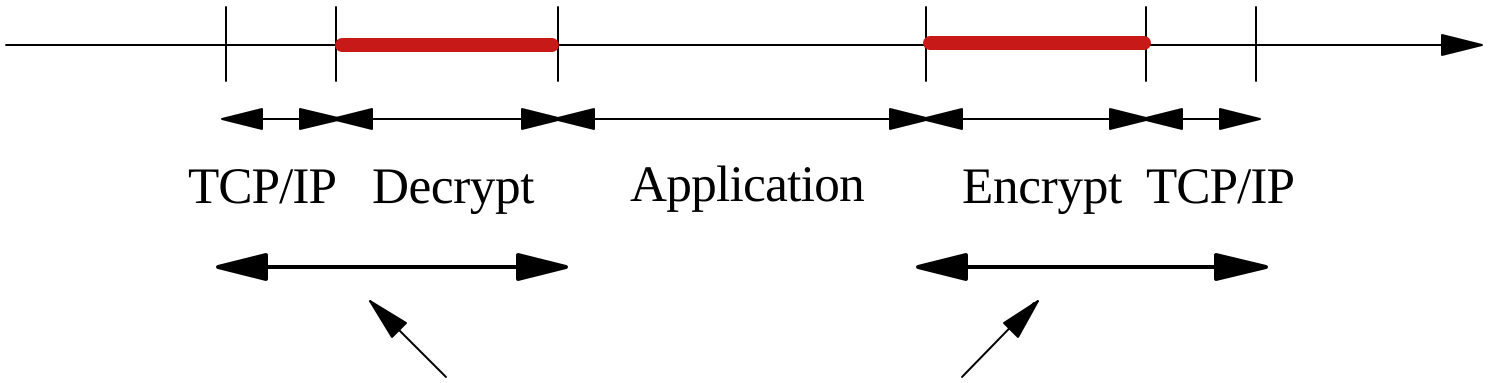
VSE

e-business

SSL data transfer overhead

Non SSL

SSL



this has been measured

VSE

e-business

Measurements Results - conclusion

- HW Crypto
  - Supports RSA operations only (e.g. used by SSL handshake)
  - CPU time/elapsed time is independent of operation and key length
  - Software RSA encryption is faster in terms of elapsed time (on large processors)
    - but hardware crypto saves CPU time
- SW Crypto
  - CPUtime /elapsed time is very dependent on CPU speed and utilization

VSE

e-business

SSL Performance Recommendations

- Use SSL only if there is a need for
  - If at least one of the follwoing is required
    - Keeping secrets
    - Proving identity
    - Verifying information
- Cipher Suites 01 and 02 has less CPU-time consumption, but NO data encryption
  - RSA512_NULL_MD5, RSA512_NULL_SHA
- If data encryption is required
  - Use cipher suites 08, 09 or 0A
  - 08 uses 512 bit keys, others 1024
  - 1024 bit RSA keylength is recomended (from a security point of view)

VSE

e-business

Dependencies for VSE/ESA Growth

- System dependecies
  - Many control-blocks etc.. still below the line
  - VTAM IOBUF areas in System GETVIS-24
  - Non-Parallel-Share limits n-way support
  - Number of tasks
    - Up to 255, 32 per partition, 208 subtasks in total
- Application dependencies
  - Integrated system concepts/functions
  - Functions/Applications dependencies
  - Number of users per TCP/IP partition

VSE

e-business

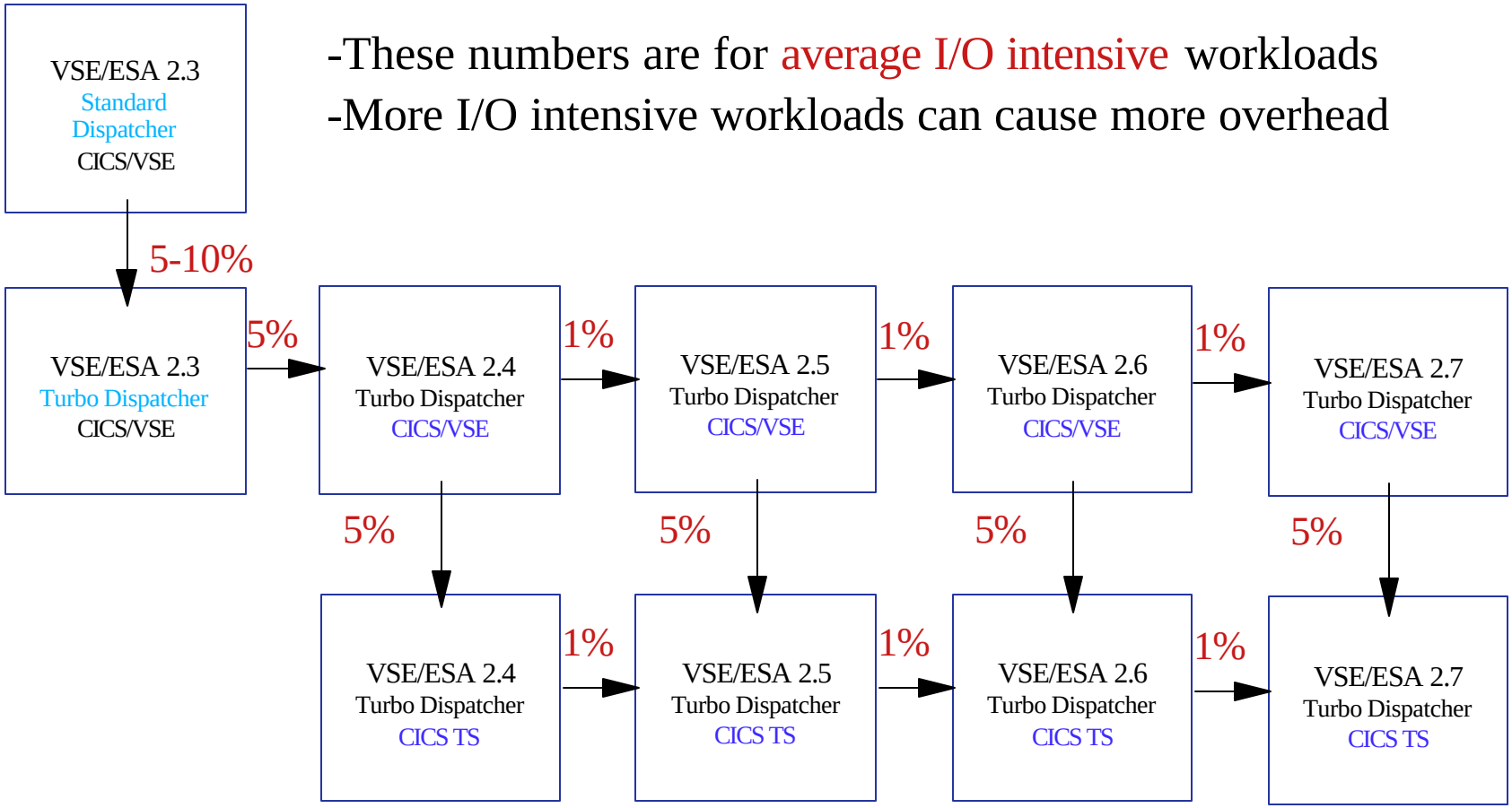
Dependencies for VSE/ESA Growth - continued

- Not being considered to be a limit
  - Number of partitions
    - 12 static + 150-200 dyn. partitions
  - Real storage (max. 2 GB)
  - Total virtual storage (max. 90 GB)
  - Total number of devices (3 digit CUU)
    - Max. 1024 devices (and 16 channels)
  - Total number of logical units
    - 255 per partition and 12x255=3060 in total
  - Label area
    - Max. about 9000 in total, and 712 in sub areas

VSE

e-business

Overhead Deltas for VSE Releases



-These numbers are for average I/O intensive workloads

-More I/O intensive workloads can cause more overhead

New functions may cause more system overhead

BUT: Exploitation of new functions helps to improve performance

VSE

e-business

VSE Health Check

- Goals
  - Recognize actual/upcoming problems
  - Optimize the system for new/current workload
- A-B-C analysis
  - A - concentrate on the essentials
    - 20 % work for 80 % results
  - B - more detailed analysis
    - 30 % work for 15 % results
  - C - analyze all details
    - 50 % work for 5 % results
- A-B analysis takes about 2 days
- C analysis takes about 1 week
- Should be done about once a year

VSE

e-business

VSE Health Check - continued

- What should be checked?
  - Processor (utilization, dispatching, z/VM, ...)
  - DASD, Tapes (I/O rate, cache, ...)
  - Network (network load, missrouted packets, ...)
  - System software
    - Turbo Dispatcher (PRTY, PRTY SHARE, ...)
    - VSAM (CA/CI sizes, shareoptions, buffers, ...)
    - CICS (MXT, DSA/EDSA sizes, SOS, ...)
    - Storage Layout (GETVIS 24, SVA, partitions, DSPACE, ...)
    - VTAM (bufferpool)
    - POWER (DBLK, DBLKGP, ...)
    - LE runtime options (Heap size, ...)
  - Application software

VSE

e-business

Hints and Tips for Performance

- Try to exploit Turbo Dispatcher functions
  - Priority settings
  - Partition balancing
  - Partition balancing groups
- Use as much data in memory (DIM) as possible
  - CICS Shared Data Tables
  - Large/many VSAM Buffers (with buffer hashing)
  - Virtual Disks
- Switch tracing/DEBUG off for production

VSE

e-business

Hints and Tips for Connector- and TCP/IP-Performance

- Reduce amount of data transferred
  - Transfer only data that is needed
  - Issue only requests that are needed
- Use connection pooling
  - Reduce overhead of connection establishment
- Performance of connectors depends on
  - Network performance
  - Performance of "server"
  - Performance of "client" or middle tier
- Reduce misrouted packets
- Use a packet filter
  - Unwanted packets increases TCP/IP and CPU load

VSE

e-business

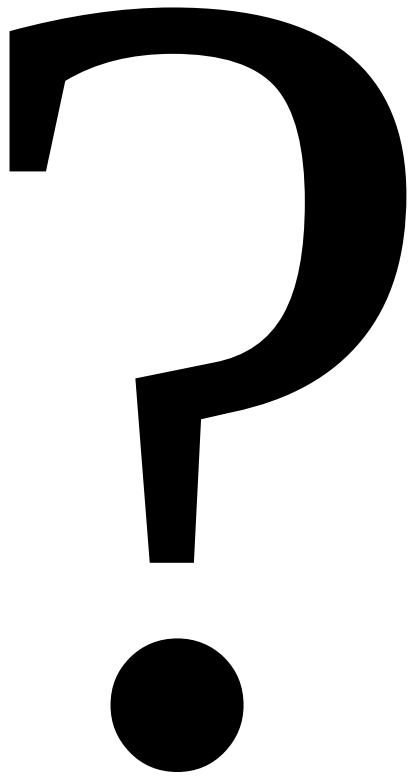
Further Information

- VSE Homepage:
  http://www.ibm.com/servers/eserver/zseries/os/vse/
- VSE Performance Homepage:
  http://www.ibm.com/servers/eserver/zseries/os/vse/library/vseperf.htm
- Performance Documents from W. Kraemer
  - available on the Performance Homepage

VSE

e-business

Questions



IBM IT Education Services

VSE/ESA Turbo Dispatcher

Session E51b

Ingolf Salm
VSE/ESA Design

e-mail: salm@de.ibm.com

VSE Technical Conference

Trademarks

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

CICS*
DB2*
DB2 Connect
DB2 Universal Database
e-business logo*
Enterprise Storage Server
HiperSockets
IBM*
IBM logo*
Multiprise*
MQSeries*
OS/390*
S/390*
VSE/ESA
VTAM*
WebSphere*
z/Architecture
z/OS
z/VM
zSeries

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

LINUX is a registered trademark of Linus Torvalds

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries

UNIX is a registered trademark of The Open Group in the United States and other countries.

Microsoft, Windows and Windows XP are registered trademarks of Microsoft Corporation.

SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.

Intel is a registered trademark of Intel Corporation.

Contents

- Design
- Operation
- Migration
- Performance Considerations
- Performance Results
- Documentation

Supported Environments

- Turbo Dispatcher (TD) available since 1995
  - Latest TD change in
    - VSE/ESA 2.6.2 (APAR DY45869)
    - VSE/ESA 2.7.0 (APAR DY45926)

- VSE/ESA supports multiprocessors
  - Basic (native), in LPARs and z/VM guests

- TD runs on all ESA/390 and zSeries processors
  - On uni-processors
  - On n-way processors

Turbo Dispatcher Design

- TD allows to exploit uni-, 2-, 3-way systems
  - CPUs with shared real/virtual memory
  - CPUs have "equal" rights
- VSE/ESA 2.1 - 2.3: standard and Turbo Dispatcher
  TD can be selected at IPL time
- Since VSE/ESA 2.4.0: Turbo Dispatcher only
- Job accounting always active (SYS JA=YES)
- Additional CPUs started after IPL complete
  - Via operator or startup procedure
  - CPUs can be stopped any time

Turbo Dispatcher Design ...

- Turbo Dispatcher improvements by VSE/ESA release
  - VSE/ESA 2.1: Partition balancing improvements
    - Equal time slices for static/dynamic partitions
  - VSE/ESA 2.2 and 2.3: Relative CPU share
    - Limits CPU usage of static/dynamic partitions
    - To combine different workload types
  - VSE/ESA 2.4 - Adaptations to CICS TS for VSE/ESA 1.1
  - VSE/ESA 2.5 – 2.7: Minor adaptations

Turbo Dispatcher Design ...

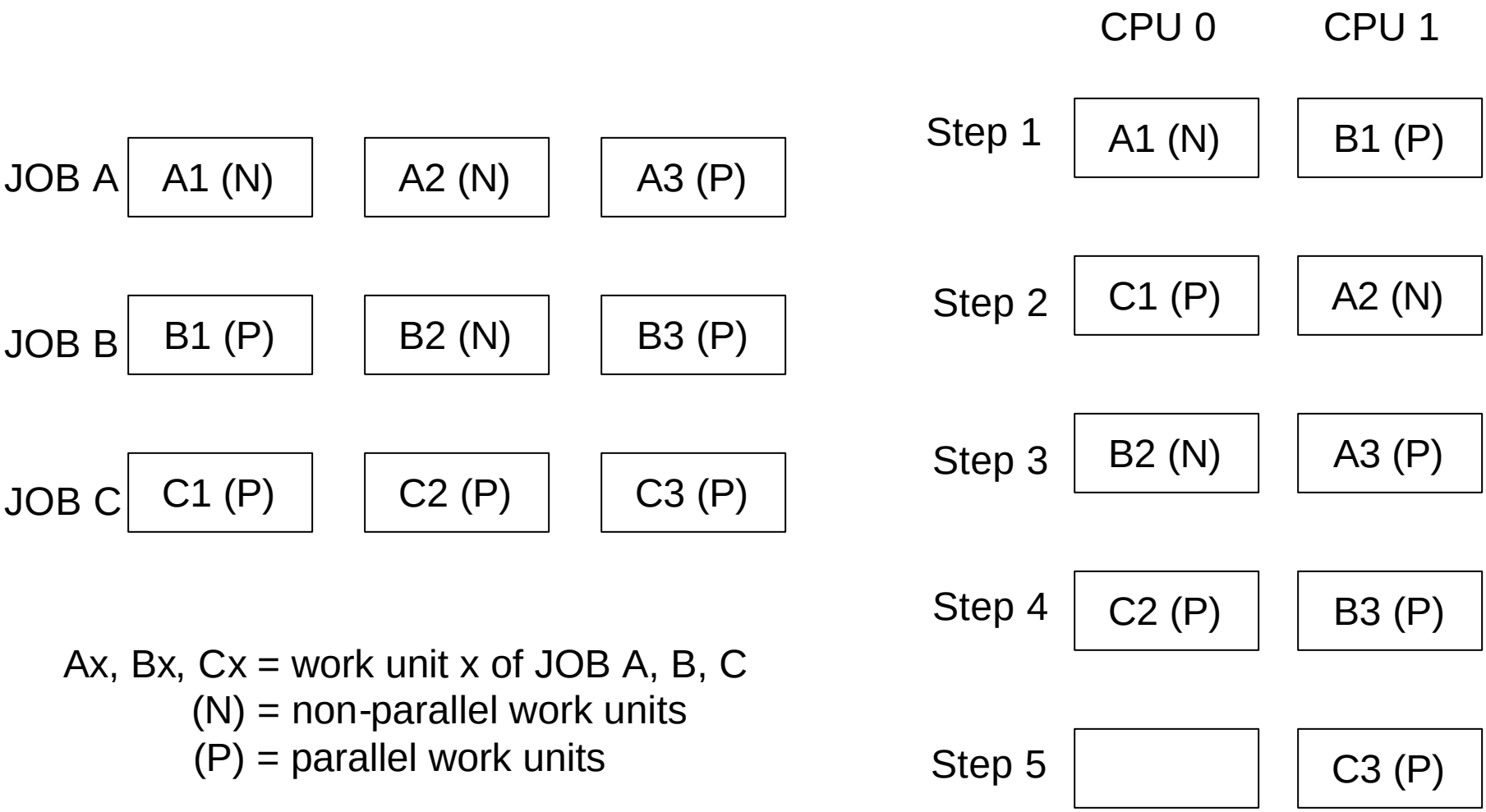
- TD dynamically assigns partitions to CPUs
  - Assignment to one CPU lasts
    from dispatcher selection to next interrupt = work unit
  - If one task of a partition is active,
    no other task of the same partition can be selected

- A partition (VSE/POWER job)
  processes many work units

Turbo Dispatcher Design ...

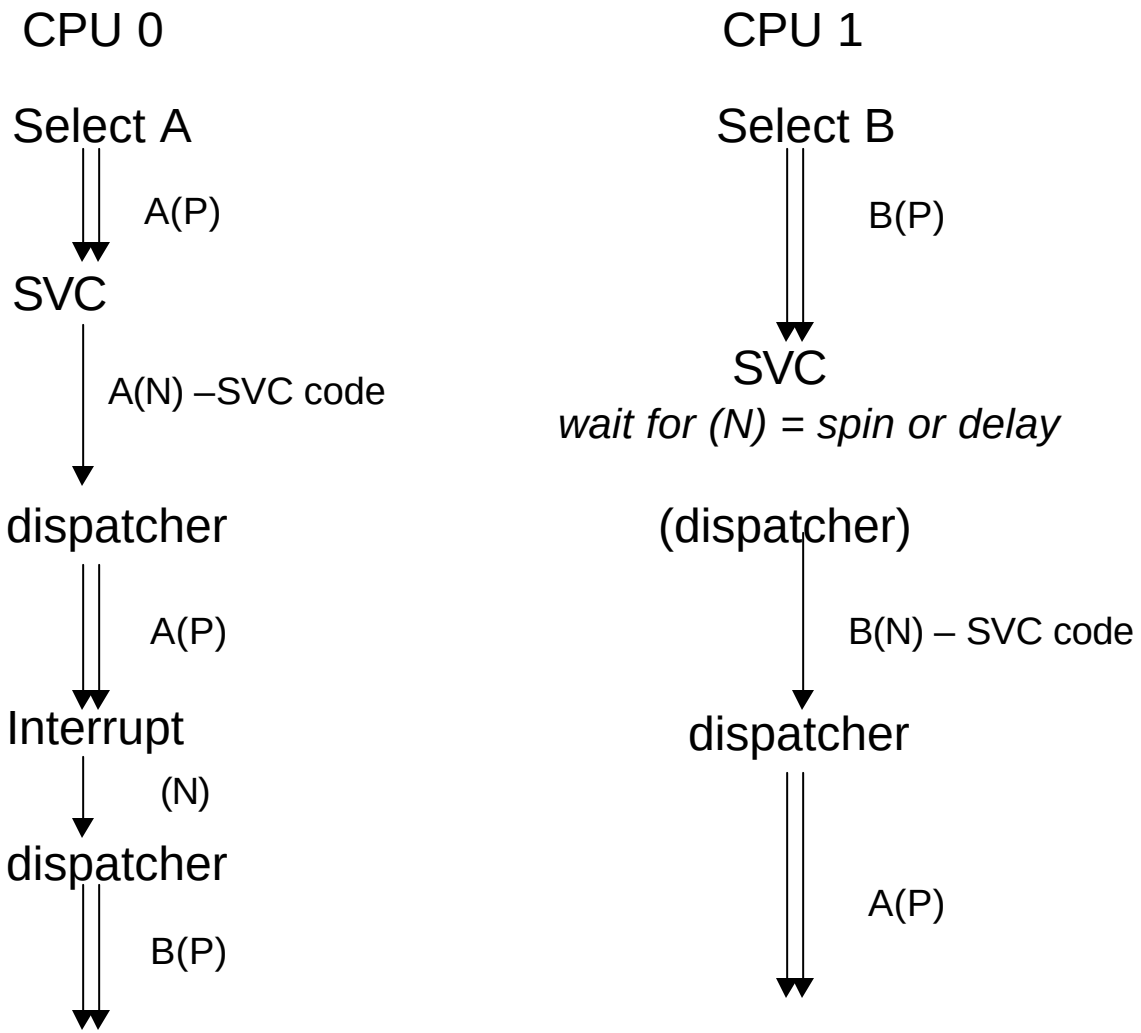
- Work units types:
  - parallel work unit
    - Application code (CICS/VSE, batch)
    - A parallel work unit may run on any CPU concurrently with other parallel or non-parallel work units.
  - non-parallel work unit
    - System code (services, ACF/VTAM)
      - As long as one non-parallel work unit is active on one CPU, no other non-parallel work unit can execute on any other CPU.
- VSE/POWER maintask has parallel or non-parallel work units

Turbo Dispatcher Design …



Ax, Bx, Cx = work unit x of JOB A, B, C
(N) = non-parallel work units
(P) = parallel work units

Turbo Dispatcher Design …



Turbo Dispatcher Design ...

Transparency

- Transparent to most user applications
- Low impact on operating environment
  - No impact on system structure
  - No additional system administration
  - Tuning for multiprocessor exploitation required
- Vendor exits for transparency

Turbo Dispatcher Design ...
Exploitation

- Exploiting uni-processors
  - New partition balancing concept
  - Determination of non-parallel work units

- Exploiting multiprocessors
  - System tuning required for exploitation
  - Full exploitation for VSE/ESA workload expected up to 3-way CECs
  - Increased capacity
  - Exploitation increases by reduction of non-parallel work units (e.g. by data in memory)

Turbo Dispatcher Operation

- Performing IPL
- Start additional CPUs: SYSDEF TD,START=n|ALL
  - After IPL complete
  - One CPU by address or all available CPUs
- Stop started CPUs: SYSDEF TD, STOP=n|ALL
  - one CPU by address or all available CPUs with the exception of the IPLed CPU
- Reset system counters: SYSDEF TD,RESETCNT

Turbo Dispatcher Operation ...

- Quiesce CPU support (SYSDEF TD,STOPQ=...)
  to quiesce specified CPU
  - Implemented for z/VM guest systems:
    - Not started guest CPUs stop IOASSIST. STOPQ leaves IOASSIST active and avoids TD overhead, if CPU(s) can not be exploited.
  - Quiesced CPU
    - will no longer participate in work unit selection
    - can be started via SYSDEF TD,START=...
    - disabled for I/O interrupts
  - STOPQ implies SYSDEF TD,RESETCNT
  - Additional status information on SYSDEF TD command

SYSDEF Examples

Start all available/defined CPUs:

```
sysdef td,start=all
AR 0015 1YH7I  NUMBER OF CPU(S)  - ACTIVE: 4  - QUIESCED: 0  - INACTIVE: 0
AR 0015 1I40I  READY
```

Stop one active CPU:

```
sysdef td,stop=3
AR 0015 1YH7I  NUMBER OF CPU(S)  - ACTIVE: 3  - QUIESCED: 0  - INACTIVE: 1
AR 0015 1I40I  READY
```

Stop one active CPU:

```
sysdef td,stopq=2
AR 0015 1YH7I  NUMBER OF CPU(S)  - ACTIVE: 2  - QUIESCED: 1  - INACTIVE: 1
AR 0015 1I40I  READY
```

Turbo Dispatcher Operation ...

- Retrieve CPU time values: QUERY TD

```
 CPU   STATUS     SPIN_TIME     NP_TIME TOTAL_TIME NP/TOT
  00   ACTIVE             0      237100     416698  0.568
  01   ACTIVE             0      157556     415229  0.379
  02   QUIESCED           0           0          0  *.***
  03   INACTIVE
                 ----------------------------------------
TOTAL                     0      394656     831927  0.474

               NP/TOT: 0.474       SPIN/(SPIN+TOT): 0.000
 OVERALL UTILIZATION: 179%         NP UTILIZATION:  85%

 ELAPSED TIME SINCE LAST RESET:        463433
```

TOTAL_TIME = CPU time used by workload
NP_TIME = non-parallel CPU time, contained in TOTAL_TIME
SPIN_TIME = CPU time needed to wait for a non-parallel work unit
All above values given in milliseconds.

NP/TOT = ratio NP_TIME / TOTAL_TIME = non-parallel share
SPIN/(SPIN+TOT) = spin time ratio

Turbo Dispatcher Operation ...

Relative CPU Shares

- PRTY SHARE command allows
  - To set and retrieve the SHAREs for the balanced group: which may hold static partitions/dynamic classes. Balanced group defined via e.g. PRTY BG,C=F5=F8,F2,F3,F1
    - Each member of the balanced group has a default SHARE.
    - Dynamic partitions have the SHARE of the corresp. dynamic class.

- Set a share: PRTY SHARE,<mem>=n,

  where <mem> = static partition or dynamic class

  n = any value out of 1.. 9999

  n = 0 = <mem> receives lowest priority within balanced group

  e.g. PRTY SHARE,C=50

- Display the shares: PRTY

Turbo Dispatcher Operation ...

How to Monitor TD

- System Activity Dialog
  - IUI dialog (host based): shows numbers of active CPUs, CPU utilization, non-parallel share, SHARE values, etc.
- VSE/ESA Console display
  - shows that TD is active and number of active CPUs
- Performance monitor from vendor

Turbo Dispatcher Operation ...

- SIR Attention Routine Command

```
 VSE/ESA          VSE/ESA 2.7     TURBO (03)      USER:  SYS
ID:VSETB                                          TIME:  14:20:10


CPUID  VM = 0021981420640000      VSE = FF00000B20640000
VM-SYSTEM = VM   (LPAR)  4.3.0       0202
PROCESSOR = 2064-00                  USERID    = VSETB
PROC-MODE = ESA (64-BIT)    IPL(200)    10:47:31      04/04/2003
SYSTEM    = VSE/ESA           2.7.0 GA                02/25/2003
              VSE/AF            6.7.0       DY45926       02/26/2003
              VSE/POWER         6.7.0       DY46048       02/25/2003
IPL-PROC  = $IPLESA           JCL-PROC  = $$JCL
SUPVR     = $$A$SUPX          TURBO-DISPATCHER (40) ACTIVE
                              HARDWARE COMPRESSION ENABLED
SEC. MGR.= BASIC              SECURITY  = ONLINE and BATCH
VIRTCPU     = 0000:05:34.817        CP = 0000:00:27.684
CPU-ADDR. = 0000(IPL)   ACTIVE
  ACTIVE      = 0000:00:00.002   WAIT = 0000:00:05.079
  PARALLEL= 0019:05:19.476    SPIN = 0000:00:00.000
CPU-ADDR. = 0001          ACTIVE
  ACTIVE      = 0000:00:00.000   WAIT = 0000:00:05.080
  PARALLEL= 0000:00:00.000    SPIN = 0000:00:00.000
CPU-ADDR. = 0002          ACTIVE
  ACTIVE      = 0000:00:00.000   WAIT = 0000:00:04.750
  PARALLEL= 0000:00:00.000    SPIN = 0000:00:00.000
```

Turbo Dispatcher Operation ...

- SIR MON Attention Routine Command

```
                    MONITORING REPORT
            (BASED ON A 0000:01:07.713 INTERVAL)
                    SVC SUMMARY REPORT
EXCP      =      3623  FCH-$$B  =        48  SVC-03   =        42
LOAD      =       204  WAIT     =      4717  SETIME   =        34
SVC-0B    =        24  SVC-0C   =        15  SVC-0D   =        38
EOJ       =        42  SYSIO    =       703  EXIT IT  =        81
STXIT OC  =         2  SETIME   =        37  SVC-1A   =        12
WAITM     =       223  COMREG   =       255  GETIME   =       208

                SVC-X'6B' DETAIL REPORT
FC-02 =      3289      FC-03 =        17      FC-06 =        90
FC-07 =        24      FC-08 =        23      FC-09 =        17
FC-0D =        17      FC-0E =        63      FC-0F =        60
FC-15 =         6      FC-1B =        26      FC-31 =        50
```

Turbo Dispatcher Operation ...

- How to gather monitored information:

  1) SIR MON=ON - starts monitoring
  2) SYSDEF TD,RESETCNT - resets TD counters
  3) <monitor interval - e.g. 1 hour at peak>
  4) SIR MON=OFF - stops monitoring
  5) QUERY TD - displays CPU counters
  6) SIR MON - displays SVC counters
  7) To start next interval begin with 1)

- Monitored data can be retrieved from VSE Console

Turbo Dispatcher Operation ...

VSE/ESA Command Restrictions

- DSPLY command

  If at least one additional CPU is started, displays for address range 0 to X'FFF' are no more unique.

- Following command not allowed,

  if more than one CPU active:

  – ALTER command for first page

- DLF command (DASD sharing)

  always from the same CPU (CPU id)

Migration Aspects

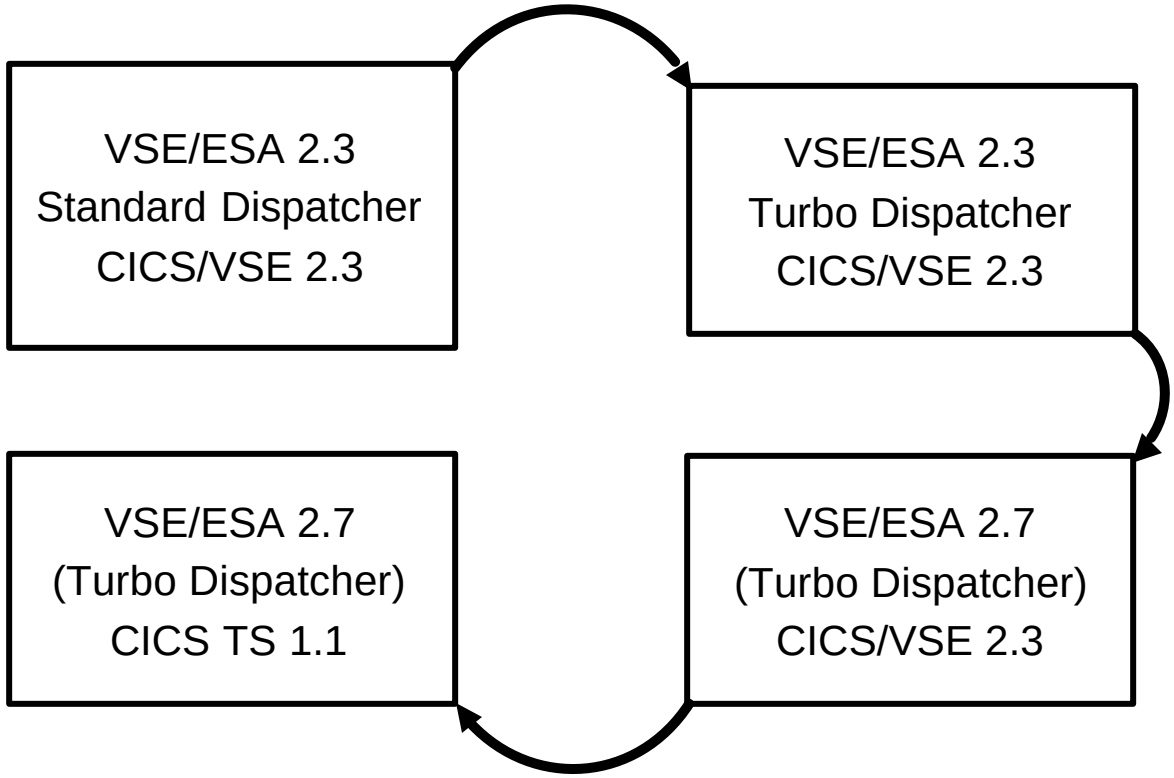
- Consider your hard-/software requirements:
  - Does my largest partition still fit into a single CPU of the target processor ?
  - Is the processor capacity and speed still sufficient to run the workload ?
  - Does multiprocessing help to run the workload ?
    - Is there a need to remove an I/O bottleneck or to add devices ?
    - What is my expectation level ?
  - Do my vendor products run on or exploit TD ?
  - Do I have system applications that interface with system routines or areas ?

Migration Overhead

- Uni-processor:
  - CPU time increase on uni-processor caused by
    - Release migration: e.g. VSE/ESA 2.6 to VSE/ESA V2.7
    - Turbo Dispatcher overhead on uni-processor
    - CICS/VSE to CICS TS migration

- Multiprocessor:
  - CPU time increases,
    when moving from uni-processor and TD to multiprocessor
    - TD overhead for multiprocessor exploitation
      (including n-way hardware overhead)
    - z/VM overhead for switch from uni- to multiprocessor.
      This overhead will also apply for standard dispatcher.
    - z/VM overhead, if guest uses multiple CPUs

Migration Steps



N-way Processor Environments

- VSE/ESA native
- VSE/ESA guest(s) under z/VM
  - VSE/ESA should run as V=R or V=F guest
  - VSE/ESA may run as V=V guest
  - CPUs may be dedicated or shared
  - IOASSISTs active only if all CPUs started/quiesced
- VSE/ESA system(s) in LPAR(s)
  - CPUs may be dedicated/shared between LPARs
- VSE/ESA guest(s) under z/VM in LPAR
  - NOT recommended because of performance reasons
  - CPUs may be dedicated/shared between LPARs
  - VSE/ESA may run as V=R guest
  - CPUs may be dedicated to VSE/ESA guest or shared
  - No IOASSIST

Performance Hints

- One partition can only exploit the power of a single CPU
- Use as many partitions as required for selected n-way
- Use/define only as many CPUs as really needed
- Partition setup
  - Set up more batch and/or (independent) CICS partitions
  - Split CICS production partitions into multiple partitions (MRO)
  - Use a database (DB2)

Non-Parallel Components

- A single CPU must be able to handle the non-parallel part of the total workload.
- Non-parallel code limits the maximum MP exploitation.
- QUERY TD command shows non-parallel share (NPS).
- System code (Key 0) code increases NPS.
  - Vendor code can have significant impact.
- TD searches for parallel work, when non-parallel resource is occupied.
- Overhead increases when NP code limits throughput.

Limited Multiprocessor Benefits

- 'Largest' VSE partition requires more CPU power as available on a single CPU of the n-way
- VSE system limited by system resources other than CPU utilization, e.g. I/O, LTA, System GETVIS (24 bit), ...
- New bottleneck because of more capacity,

would also appear on faster uni-processor

- Overall workload's non-parallel share too high
- Not enough partitions concurrently active

CICS Implications

- Single CICS
  - Can consume processing power of one CPU only
- Multiple CICS partitions (MP exploitation)
  - E.g. non-parallel share of 30 %
    - max. exploitable CPUs = 3
  - Multiple CICS workload alternatives
    - Independent CICS partitions
    - MRO transaction routing
    - MRO function shipping to file owning region
    - Mixtures of transaction routing and function shipping

Performance Measurements

- Additional CPU time cost for exploitation of
  TD on uni versus standard dispatcher: 5 to 10 %

- Quiesced CPU costs up to 5 % overhead

- 2 or 3 CPUs can be fully exploited
  - Where non-parallel share ranges from 0.5 to 0.25

- Measurements with our workloads
  - Batch workload (16 partitions):
    - TD overhead: 15 %, NPS: 0.48, MP factor (2-way): 1.4
  - Online workload, TD overhead 4%, NPS: 0.27
    - 2-way, 3xCICS: MP factor: 1.75, utilization: 93%
    - 3-way, 4xCICS: MP factor: 2.35, utilization: 84%

z/VM-VSE Considerations

- On-line workload (DSW) measurement results for z/VM
  - z/VM guest/native ratio for TD
    - On 2-way: about 4% lower versus uni
- z/VM can provide
  - Real multiprocessing by dedication of CPUs
  - Virtual multiprocessing
    - Definition of more virtual CPUs than real CPUs
      results in poor guest performance
    - No performance reasons to define > 1 virtual CPU,
      if z/VM runs on a uni-processor
- All z/VM guest defined CPUs must be started or quiesced, otherwise IOASSISTs are not available.

Documentation

- VSE/ESA Turbo Dispatcher Guide and Reference, SC33-6797
- VSE/ESA Turbo Dispatcher Performance, SC33-6749
- ITSO VSE/ESA 2.1 Turbo Dispatcher, SG24-4674
- VSE/ESA 2.7 Release Guide, SC33-6718
- Hints and Tips for VSE/ESA, SC33-6757
- VSE/ESA home page
  http://www-1.ibm.com/servers/eserver/zseries/os/vse/