

ibm.com

Linux for IBM *e*server zSeries Technical Update



Redbooks

International Technical Support Organization

© Copyright IBM Corp. 2003. All rights reserved.

ibm.com

This information was developed for products and services offered in the U.S.A.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of 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.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

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

developerWorks®	ECKD™	Perform™
ibm.com®	ESCON®	PR/SM™
iSeries™	FlashCopy®	Redbooks™
pSeries™	FICON™	RACF®
xSeries®	GDPS®	RAMAC®
z/Architecture™	HiperSockets™	RMF™
z/OS®	Illustra™	S/370™
z/VM®	IBM®	S/390 Parallel Enterprise Server™
zSeries®	Language Environment®	S/390®
AIX®	Lotus®	SLC™
BookManager®	Multiprise®	Tivoli®
DirMaint™	OpenEdition®	TotalStorage®
DB2®	OS/390®	VM/ESA®
Enterprise Storage Server®	Parallel Sysplex®	WebSphere®

The following terms are trademarks of other companies:

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel Corporation 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.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

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



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Agenda

Introduction

z/VM 4 Release 4

Performance topics for Linux on zSeries

Security topics for Linux on zSeries



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Introduction



© Copyright IBM Corp. 2003. All rights reserved.

Objectives

International Technical Support Organization (ITSO)

- Who we are
- What we do

Linux on zSeries

- Linux background
- Linux on mainframe hardware

Where to Find This Presentation

<http://w3.itso.ibm.com/itsoapps/material.nsf/WebByDate>

Select:

- 2003 ITSO Linux for zSeries: Technical Update

Zip file password:

- itso2003



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Types of ITSO Material

Redbooks

- Broad in-depth technical manuals

Redpapers

- Specific topic HOWTOs

Hints and Tips

- Excerpts from redbooks/redpapers

Workshops

- Interactive forums based on redbooks/redpapers



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

How to Locate ITSO Information

<http://www.ibm.com/redbooks>

The screenshot shows the IBM Redbooks website interface. A search bar is located at the top right, and a 'Portals' menu is on the left side. Both are highlighted with green circles and labeled with text. The website content includes sections for 'What's New', 'Recent Redpieces (draft books)', and 'Recently Published'.

Search for: **linux AND zseries**



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ITSO Linux and zSeries Domains

Linux domain

- Linux specific topics
<http://www.ibm.com/redbooks/portals/linux>

zSeries domain

- zSeries/S390 specific topics
<http://www.ibm.com/redbooks/portals/S390>

ITSO Networking domain

- Networking specific topics
<http://www.ibm.com/redbooks/portals/Networking>



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ITSO Residencies

Usually 4-6 weeks projects

- Install and document new hardware/software

Residents can be:

- IBM employees
- Business partners
- Customers

ITSO covers resident expenses for:

- Travel to/from residency
- Living during residency

Projects advertised on Redbook web site

- Select **Residencies** on left navigation



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ITSO Mailing List

Weekly newsletter

- Customized to specific area of interest

Keep informed about newest:

- Redbooks/redpapers
- Residencies
- Workshops

Subscribe at ITSO web site

- Select **Mailing List** on left navigation



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Related Publications

- *The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this workshop.*

International Technical Support Organization Publications

- For information on ordering ITSO publications, visit us at redbooks.ibm.com (Internet Web site) or
- w3.itso.ibm.com (intranet Web site)

For Technical Support see ibm.com/support and w3.ibm.com/support

Redbooks on CD-ROMs

CD-ROM Title	Collection Kit Number
IBM Redbooks S/390 Collection	SK2T-2177-24
IBM iSeries 400 Redbooks Collection	SK2T-2849-13
IBM Redbooks Networking and Systems Management Collection	SK2T-6022-15
IBM Redbooks DB2 Information Management Collection	SK2T-8038-11
IBM Redbooks Lotus Collection	SK2T-8039-08
IBM Redbooks AIX, UNIX, and IBM ^ pSeries Collection	SK2T-8043-08
Tivoli Redbooks Collection	SK2T-8044-08
IBM ^ xSeries Redbooks Collection	SK2T-8046-07
IBM TotalStorage Redbooks Collection	SK3T-3694-07
IBM ^ Server zSeries Redbooks Collection	SK3T-7876-03
IBM Redbooks Linux Collection	SK3T-7890-01
IBM Redbooks WebSphere Collection	SK3T-8282-01



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Related Publications - Continued

Other Publications

- *These publications are also relevant as further information sources:*

Linux for zSeries Redbooks

Title	Publication Number
Linux on IBM Sserver zSeries and S/390: Performance Measurement and Tuning	SG24-6926
Linux on IBM Sserver zSeries and S/390: Best Security Practices (coming soon)	SG24-7023
Linux on IBM Sserver zSeries and S/390: System Management	SG24-6820
Linux on IBM Sserver zSeries and S/390: Large Scale Deployment	SG24-6824
Linux for IBM Sserver zSeries and S/390: ISP/ASP Solutions	SG24-6299
Linux for IBM Sserver zSeries and S/390: Distributions	SG24-6264
Linux for S/390	SG24-4987



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Related Publications - Continued

Other Publications

- *These publications are also relevant as further information sources:*

Linux for zSeries Redpapers

Title	Publication Number
Linux on IBM Sserver zSeries and S/390: VSWITCH and VLAN Features of z/VM 4.4	REDP3719
Linux on IBM Sserver zSeries and S/390: Building SuSE8 Systems Under z/VM	REDP3687
Linux on IBM Sserver zSeries and S/390: z/VM Configuration for WebSphere Deployments	REDP3661
Linux on IBM Sserver zSeries and S/390: TCP/IP Broadcast on z/VM Guest LAN	REDP3596
Linux on IBM Sserver zSeries and S/390: Server Consolidation With Linux for zSeries	REDP0222
Linux on IBM Sserver zSeries and S/390: Securing Linux for zSeries With a Central z/OS LDAP Server (RACF)	REDP0221


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Related Publications - Continued

Other Publications

- *These publications are also relevant as further information sources:*

z/VM Documentation available from <http://www.vm.ibm.com/pubs>

Title	Publication Number
Virtual Machine Operation	SC24-6036
System Operation	SC24-6000
CP Planning and Administration	SC24-6043
CP Command and Utility Reference	SC24-6008
Performance	SC24-5999
Running Guest Operating Systems	SC24-5997


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Related Publications - Continued

Useful Web sites

- *These Web sites are also relevant as further information sources:*

IBM developerWorks Linux for zSeries and S/390

<http://www.ibm.com/developerworks/oss/linux390/index.shtml>

IBM z/VM and ESA/VM Home page

<http://www.vm.ibm.com/>

Linux for Big Iron

<http://www.linuxvm.org/>

Marist Linux on S/390 mailing list

<http://www.marist.edu/htbin/wlindex?linux-390>

Linux Documentation Project

<http://www.tldp.org>

Linux Security

<http://www.linuxsecurity.org>



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ibm.com

IBM®

z/VM 4 Release 4



© Copyright IBM Corp. 2003. All rights reserved.

Objectives

z/VM evolution and value

Technology exploitation

Virtualization technology and Linux enhancements

Networking enhancements

Systems management improvements

Network performance and security

Application enablement

Packaging changes and optional features



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Version 4 Product Information

Runs on:

- IBM G5 processor technology or better
 - IBM [^]™ zSeries™ 800, 900, and 990 (including z800 Model 0LF)
 - IBM 9672 G5/G6 and Multiprise 3000
- IFL processors as well as standard processors

IPLA software product

- One-time charge license fee, priced on a per-engine basis
- Ordered via the System Delivery Option (SDO)

Optional Software Subscription & Support (S&S) product

- Required to receive telephone defect support
- Entitles customers to future z/VM releases and versions
- Annual, renewable license charge



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

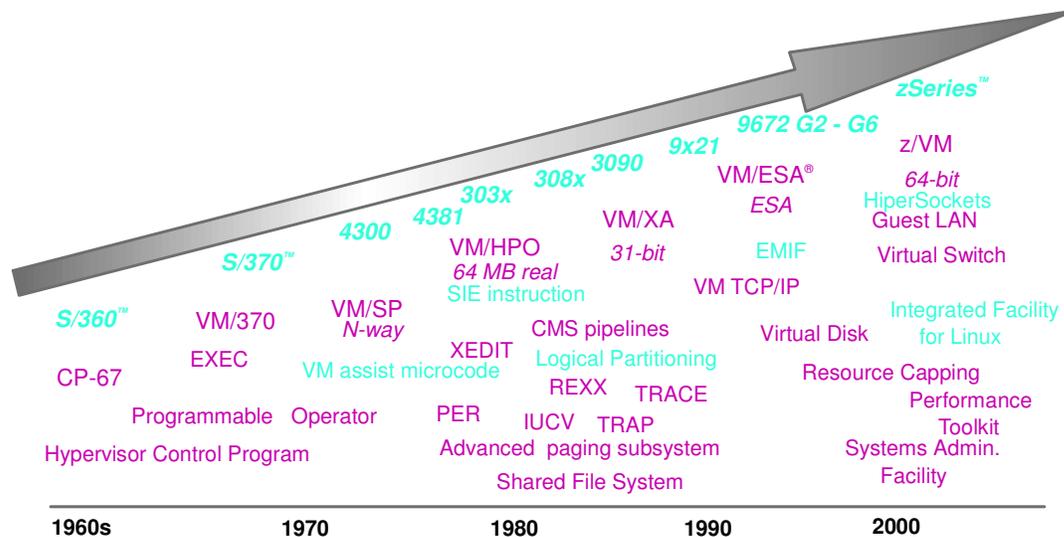
z/VM Evolution and Value



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

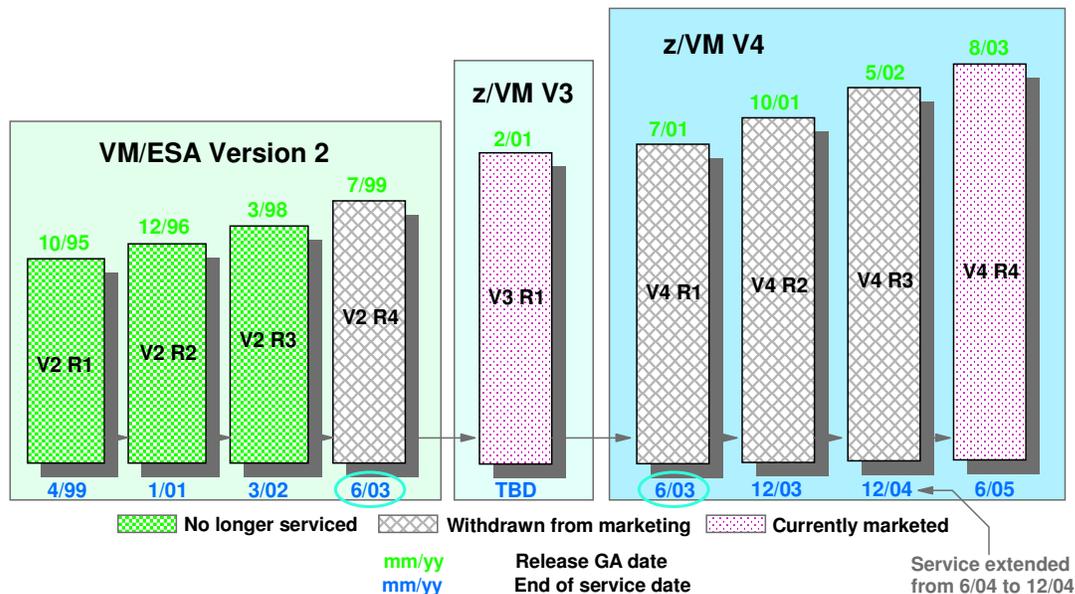
Evolution of IBM Mainframe Virtualization



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Recent z/VM History



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Why Run Linux Under z/VM?

Save Money

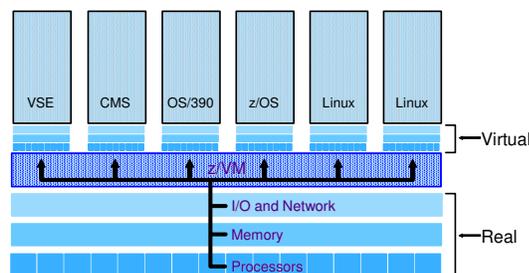
- Consolidate servers with z/VM
- Enhance profitability

Save Time

- z/VM enables:
 - Faster deployment
 - Innovative solutions

Make Linux Even Better

- Linux with z/VM is better than Linux alone
- Linux can exploit unique z/VM technology features



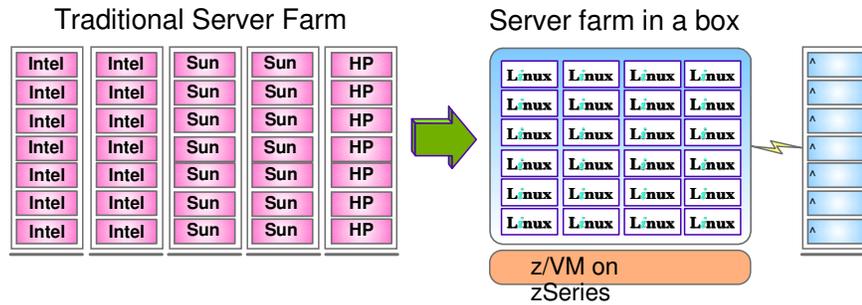
ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Realizing TCO Savings With z/VM

Total Cost of Ownership cost savings through:

- Virtual servers reduce hardware requirements
- Fewer hardware servers occupy less space
- Virtual servers can be created in minutes
- Shared code saves software, system mgmt, staffing costs
- System management tools
- Virtual networking saves hardware expense



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

Technology Exploitation



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Servers Supported by z/VM Version 4

IBM ^ zSeries (990, 900 and 800) in z/Architecture (64-bit) mode

- Also runs in ESA/390 (31-bit) mode

Servers also supported by z/VM in ESA/390 mode:

- S/390 Parallel Enterprise Server™ — Generation 5 and 6
- S/390® Multiprise 3000
- Equivalent processors



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM V4 Supports IFL Processors

IFLs are cost-effective for Linux on mainframe

- Less expensive than standard engines
- Does not affect software fees on standard processors

IFLs available on Multiprise 3000, G5/G6, z800, z900, and z990

- Allocated from the set of spare processors on MCM
- At least one standard processor must be configured before IFL(s) can be added to a mainframe
 - Exception: z800-0LF is an IFL-only server



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

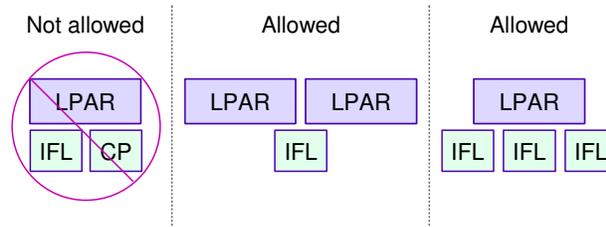
IFL Processor Support

IFL usage considerations:

- IFLs can only be used in LPAR mode
- Standard engines and IFLs cannot be mixed within an LPAR

z/VM Version 4 runs on IFL processors

- z/VM V4 features can be licensed for IFL processors
 - CMS runs on IFLs
 - Can run Linux guests
- Traditional mainframe OSs will not IPL in IFL virtual machine



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IBM ^ zSeries 990 - June 2003

Highlights:

- 2 Models (A08 and B16)
- Improved performance over the z900
- 1 - 16 way
- Up to:
 - 128 GB of central processor storage
 - 2 Logical Channel Subsystems (LCSS)
 - Spanned channel support
 - 15 LPARs
 - LPAR Mode only - No basic mode
 - 120 FICON Express cards and 512 ESCON® channels
 - Support for cascaded FICON directors
 - No parallel channels
 - 16 IFLs
 - 16 HiperSockets and 48 OSA-Express ports



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IBM ^ zSeries 990 - October 2003

New for z990:

- 2 additional models (C24 and D32)
- 1 - 32 way
- Up to 256 GB of central processor storage
- Up to 30 LPARs



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IBM ^ zSeries 900

Highlights:

- 41 general purpose models:
- FCP channel for Linux
- Integrated Facility for Linux (IFL) processors
- z/Architecture (64-bit) supported
- HiperSockets for high-speed internal TCP/IP network
- Up to:
 - 16-way (20 PUs)
 - 64 GB memory
 - 15 LPARs
 - Maximum 64 GB of storage per LPAR
 - 256 ESCON / 88 parallel channels
 - 96 FICON channels



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IBM ^ zSeries 800

Highlights:

- 10 General Purpose Models (1-4 way)
- zSeries Entry License Charge™ (zELC) Software pricing
- FCP channel for Linux
- Integrated Facility for Linux (IFL) processors
- z/Architecture (64-bit) supported
- PCI Cryptographic Accelerator and Coprocessor
- HiperSockets for high-speed internal TCP/IP network
- Up to:
 - 32 GB of central processor storage
 - 15 LPARs
 - Maximum 32 GB of storage per LPAR
 - 240 ESCON/No parallel channels
 - 32 FICON channels



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

zSeries Pricing Initiatives

Effective August 2003

- Designed to lower cost for zSeries customers

Highlights:

- Lower zSeries memory price
- Consistent IFL pricing
- Improvements to Workload License Charges (WLC)
- z/OS New Application License Charge (NALC) price reduction
- z990 pricing initiatives



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z990 Support

Multiple Logical Channel Subsystems (LCSS) support

- Dynamic I/O configuration support extended
 - Channel paths, control units, devices dynamically added, changed, deleted
- Support using HCD/HCM or CP dynamic I/O config commands
- Single LCSS support available 15 August 2003
- Multiple LCSS support planned for 31 October 2003

Extended Channel Measurement Data Support (ECMDS)

Spanned channel support planned for October 2003

- Inter-process communication among Linux images running on z/VM images in different LPARs

Support for more than 15 LPARs



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Newest z800 Family Support

z800 Model 0E1

- Deploy Linux on mainframe while hosting small traditional workloads
- Configuration:
 - One standard processor
 - One IFL processor
- Upgradeable
 - Within z800 family and into z900 family
 - Up to two additional IFL processors can be added

z800 Model 0X2

- Configuration:
 - Sub-dyadic server estimated performance up to 60% of z800 Model 0A2
 - Optional: up to two IFL processors can be configured
- Upgradeable within the z800 family and into the z900 family



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

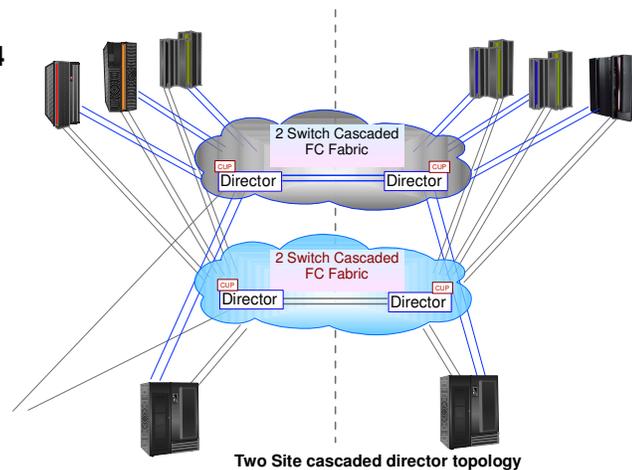
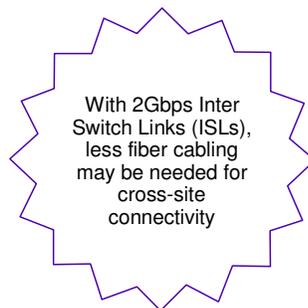
Cascaded FICON Directors

Reduce implementation cost for disaster recovery applications; GDPS™ and Remote Copy

Fewer cross site connections - Repeaters, DWDM, Fibers, Channels, Director Ports

Integrity features detect mis-cabling events and prevent data streams from being delivered to the wrong end point

Supported by z/OS V1.4 and z/VM V4.4



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IBM TotalStorage Enterprise Storage Server

FlashCopy V2 functions:

- Data Set FlashCopy
 - Allows data to be copied onto a volume at a different cylinder location
 - The data can even be copied onto the same volume
 - Especially useful for copying a VM minidisk to another minidisk
- Multiple Relationship FlashCopy
 - Allows a source volume to be copied to many target volumes
 - Multiple copy operations can run concurrently on source/target volumes
- Elimination of Logical Storage Subsystem (LSS) constraint
 - Source / target volumes do not have to reside in the same logical control unit



PPRC V2 functions:

- Asynchronous Cascading PPRC
 - Allows secondary PPRC volume to serve as primary volume in a PPRC-XD
 - Enables a three-site, long distance disaster recovery solution



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IBM ESS Long Distance Copy

Extended Distance Peer-to-Peer Remote Copy (PPRC-XD)

- Function of IBM Enterprise Storage Server
 - Allows full volumes data to be copied *asynchronously*
- Greater distance between primary and secondary volumes
 - Well beyond 103 kilometer limit of synchronous PPRC
- Minimal effect on performance

z/VM guests can perform PPRC-XD operations

- Need datamover authority
- Guest operating system must support PPRC-XD
- Full volume copies only
- Suitable for data migration, backup, disaster recovery procedures



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

Virtualization Technology and Linux Enhancements



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Enhanced QDIO Performance

QDIO high-performance I/O interrupt

- Know as adapter interruption
 - Originally available for HiperSockets
- Extended on the z990 to include OSA-Express and FCP channels
- z/VM performance assist for the virtualization of adapter interruptions
 - Available to pageable (V=V) guests that support QDIO

Can benefit all guests that process adapter interrupts

- HiperSockets
- OSA-Express
- FCP channels
- TCP/IP for VM supports adapter interruptions
 - Beneficial when using HiperSockets and OSA-Express adapters



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Guest Capacity Improvements

New lock provided for timer request block management

- Timer request block management longer requires scheduler lock
- Avoids scheduler lock serialization
- Improve throughput of guests that issue large number of timer request interrupts (e.g., Linux)

Can increase number of Linux guests

- Reduces Control Program overhead
- Improvements are most noticeable on large multiprocessor systems



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Systems Management APIs

Intended to:

- Simplify effort of developing solutions for Linux images management

APIs:

- Enable:
 - Allocation and management of virtual machine resources
 - Virtual machine configuration changes
 - Activation and deactivation of individual or lists of virtual images
 - Connectivity management between virtual machines
- Are invoked using Remote Procedure calls (RPC)
 - Standard, platform-independent interface
 - May be called remotely (network) or from within the z/VM system
- Security and directory management functions also provided
- Require a directory manager
 - z/VM 4.4.0 DirMaint feature supports the APIs

Solution providers planning to exploit the APIs:

- Linuxcare (Levanta)
- Sine Nomine Associates (VMGUIOP)



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Virtual Machine Resource Manager

VMRM introduced in z/VM 4.3.0

- Manages performance of selected virtual machines
 - Based on customer-defined goals for CPU and I/O performance
- VMRM service virtual machine accepts:
 - Workload definitions (which can include multiple virtual machines)
 - Goal specifications and importance of achieving defined goals
- VMRM adjusts user CPU shares or I/O performance based on:
 - Velocity goals set for the user's workload class
 - Virtual machine CPU and/or I/O achievement levels

z/VM 4.4.0 enhancements:

- Monitor data is now provided to show actual workload achievements
- Wildcard characters accepted for userids in configuration file
- Improved performance of VMRM service virtual machine
- Improved messages and logging



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Guest IPL support for SCSI Disks

Boot Linux guests from FCP-attached SCSI disk

- Requires z990, z900, or z800 server
 - With corresponding IPL-from-FCP hardware function
 - IBM intends to deliver this function in the future
- Linux guests can use SCSI-only disk configuration

z/VM still requires ESCON / FICON attached disk or tape

- IPL
- Data storage



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Virtual FICON CTCA Support

Virtualizes FICON channel-to-channel adapter architecture

Enables virtual machines to use the FICON CTCA protocol

Define with new device type FCTC via:

- DEFINE command
- SPECIAL directory statement



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Integrated 3270 Console Support

Hardware Management Console (HMC)

- Provides one 3270 session per LPAR

z/VM supports the Integrated 3270 console as system console

- Device is known to z/VM as SYSG
- Eliminates need for 3174 or 2074 controller

Requires HMC level 1.8.0 or higher

- G5/G6 (CPC EC F99918)
- z800 or z900 (CPC EC J11219)
- z990



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Support for Parallel Sysplex Technology

z/VM provides a virtual Parallel Sysplex environment

- Coupling Facility Control Code (CFCC) is loaded into virtual machine(s)
 - 31-bit or 64-bit versions of CFCC
- Virtual coupling links used instead of real links
- Sysplex timer not required; virtual TOD clocks are synchronized

Real coupling links and coupling facility not required

- Or supported ...

New z/VM 4.4.0 support:

- VM/ESA and z/VM systems can run as guests of z/VM 4.4.0
 - Simulating Parallel Sysplex environments
- z/VM V4 support available on z990, z900, z800, G5/G6, Multiprise 3000



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

CP Command Response Suppression

CP commands issued without generating a response

- Using the SILENTLY option

Useful for service machines or privileged users

- Avoids the confusion a command response might create for the user

Must be authorized via:

- SYSTEM CONFIG file, or
- DEFINE / MODIFY COMMAND command

Intended for use with:

- ATTACH
- DETACH
- GIVE



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

Networking Enhancements



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

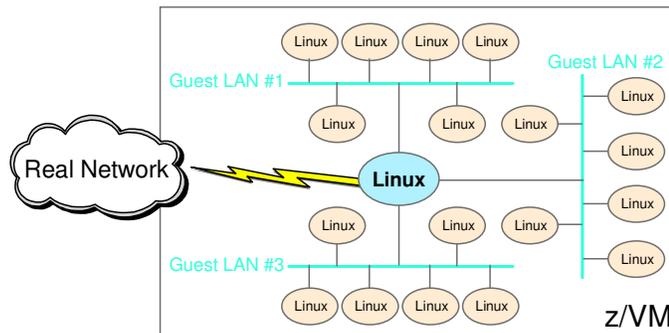
z/VM Virtual Networking - z/VM Guest LAN

Guest LAN "virtual" is LAN created by z/VM Control Program

- OSA-Express (QDIO) and HiperSockets Guest LANs can be created
 - Point-to-point, Multicast, and Broadcast (QDIO) connections are supported

Linux images can connect to one or more Guest LANs

- And connect to real network adapters at the same time



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

VLAN Support

An IEEE VLAN enables systems connected to different switches in different physical locations to be logically connected into a single Local Area Network

- Simplifies network management
- Routers connected to multiple VLANs can reduce the expense of connecting real LAN segments

z/VM TCP/IP can participate in an IEEE VLAN

- OSA-Express in QDIO mode (Ethernet) only

Virtual network adapters attached to a z/VM Guest LAN are able to participate in IEEE VLANs

- Virtual QDIO (Ethernet) or HiperSockets adapters are supported



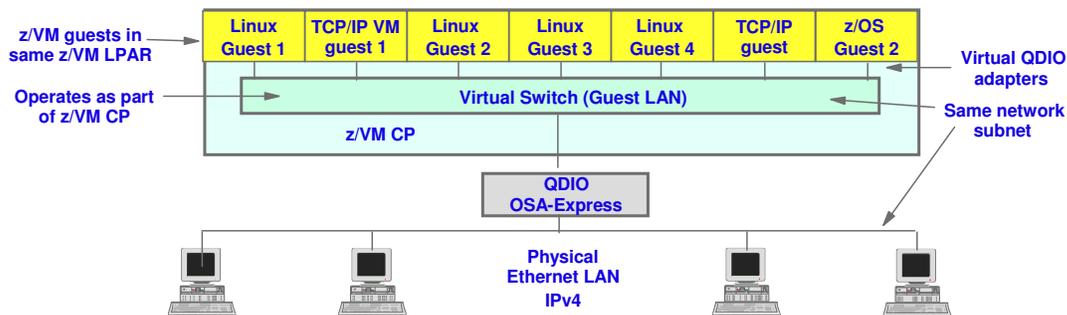
ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Virtual IP Switching

Virtual-QDIO connections to physical LAN without routing

- Allows virtual machines on the Guest LAN to be in the same subnet with the physical LAN segment
- Reduces copying of data being transported
- Provides centralized network configuration and control
- May reduce overhead associated with router virtual machines



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

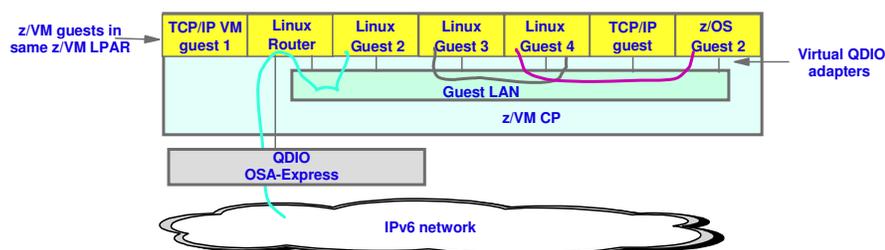
z/VM Guest LAN Support for IPv6

IPv6 uses 128-bit address space

- Significantly increases IP addressability over IPv4
- Guest LAN with virtual OSA-Express in QDIO mode
 - Linux or z/OS guest required to connect to external IPv6 network

No IPv6 support in:

- VM TCP/IP stack, or
- Virtual IP Switch



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Extended HiperSockets Support

TCP/IP broadcast support is provided in z/VM V4.4 for the HiperSockets environment when using IPv4

Applications that use the broadcast function can now propagate broadcast frames to all TCP/IP applications when using:

- HiperSockets
- OSA-Express (QDIO) adapter
- z/VM Guest LANs



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

**System Management
Improvements**



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

HCD and HCM for z/VM

HCD (Hardware Configuration Definition) ported from z/OS

- A new component in z/VM 4.4.0
- An I/O configuration definition tool
- Validates configuration definitions at data-entry time
- Creates and maintains the I/O Definition File (IODF)
- Dynamically changes I/O configuration using CP Dynamic I/O
- Maintains synchronization between dynamic I/O changes and IODF

Hardware Configuration Manager (HCM)

- Runs on a Microsoft Windows workstation
- Provides an interactive user interface to HCD
- Graphically displays I/O topology



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Toolkit for VM

A new z/VM 4.4.0 feature

- Priced on a one time charge, per processor basis (IPLA Ts&Cs)
- Based on the FCON/ESA product
- Intended to eventually replace RTM and PRF
- Can be licensed for standard and IFL processors
- No-charge upgrade to the Performance Toolkit for VM for:
 - Customers who purchased S&S for RTM or PRF features

Functional highlights

- Provides an immediate view of system performance
- Post processes its own history files or CP Monitor data
- Threshold monitoring
- User loop detection
- Can monitor remote systems
- Results can be graphically viewed by a web browser
- Processes Linux data provided by the RMF PM data collector
 - Combines and displays both VM and Linux data



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

LPAR Monitor Enhancements

z/VM V4.4 can handle more than one 4K page of LPAR data

Monitor data will:

- Indicate hardware is available for reporting LPAR data greater than 4k
- Indicate if Monitor was unable to obtain contiguous 4K pages to report on all Logical CPUs

Allow monitor processing products to avoid anomalies

- Indicates which partitions are standard and which are IFL or ICF

Performance Toolkit for VM will support this new Monitor data

- VMPRF will not be updated to support this data



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Automated SFS Shutdown

SFS supports Automated Shutdown

- First introduced in z/VM 4.3.0
- Helps maintain integrity of the Shared File System and its data

File pool virtual machines enabled for shutdown signals

- Action taken depends on new DMSPARMS
- SFS STOP command processing will occur when SHUTDOWN SIGNAL is specified (this is the system default)
- SFS will ignore shutdown signals when NOSHUTDOWN SIGNAL is specified



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Installation and Service Enhancements

Fewer decisions required during installation

- Additional automation to move a component / product into SFS
- Service minidisks for all z/VM components can reside in SFS directories

New components / features pre-installed:

- HCD and HCM for z/VM
- Language Environment
- Performance Toolkit for VM (pre-installed but disabled)
- New product levels installed:
 - OSA/SF 4.4.0
 - ICKDSF R17

VMFUPDAT enhancements:

- Manual builds can be flagged as "built"
- Service restart records can be removed



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

Network Performance and Security



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

TCP/IP Performance Improvements

Multiprocessor support for the TCP/IP for VM stack

- The CPU option is added to the Device statement to allow support for a designated device to be associated with a particular virtual processor

Optimized high use code paths

- Some Pascal rewritten in Assembler
- Improved algorithms to reduce path lengths
- Goal is to enhance environments in which VM is acting as a host (rather than as a router)



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

TCP/IP Stack Security

Logging of NETSTAT and OBEYFILE commands

New system defaults

- RestrictLowPorts
- VarSubnetting

New SECURITY trace set name has been added to control the amount of trace information collected for security related events

Port range may be specified on PORT statement

User "RESERVED" may be specified on port statement

- prevents users from listening on the port

Logon-By support included for REXECD



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

IMAP Server Authentication Enhancements

Current IMAP server implementation requires all enrolled users to have VM userids and passwords

- Limits userid to 8 characters
- CP or ESM validation does not allow for PREAUTH processing

Enhancements provide for new authentication exit

- Alternative to CP or ESM validation
- Removes restrictions on user name
- Communicates with IMAP server via CMS distributed queue
- Allows for PREAUTH processing
- Sample exit provided
- Customer defines validation and mapping algorithms for user names



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

Application Enablement



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

C/C++ for z/VM Compiler Support

Provides the necessary support in CMS to:

- Run application programs written in C/C++
- Allow greater portability of applications to z/VM from other IBM platforms

The C/C++ compiler for z/VM (5654-A22) provides:

- Support for the latest 1998 ANSI/ISO C++ standard
- Complete implementation of the 1998 ANSI/ISO C++ Standard Library, including the Standard Template Library (STL)
- Highly productive and powerful object-oriented development environment for z/VM programmers



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Language Environment for z/VM

Required for C/C++ for z/VM compiler

Equivalent to z/OS V1.4 level of Language Environment except for the following:

- No 64-bit application support
- Some z/VM OpenExtensions functions are not compatible with z/OS UNIX Systems Services (USS)
- z/OS downward compatibility is not supported
- z/OS native ASCII is not supported
- VSAM record-level sharing is not supported

Integrated into the base of z/VM V4.4



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM 4 Release 4 Topics

Packaging Changes and Optional Features



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM V4 Product Packaging Changes

TCP/IP

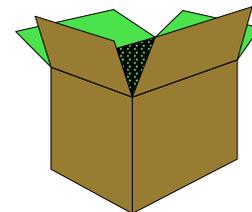
- Integrated into z/VM - no license required
- Preinstalled on base system DDRs
- NFS Server feature integrated into TCP/IP - no license required
- NFS Client integrated into TCP/IP and CMS
- Kerberos Data Encryption Standard (DES) integrated into TCP/IP
- TCP/IP source supplied with z/VM - no license required

CMS Utilities Feature

- Integrated into CMS - no license required

OpenExtensions Shell and Utilities

- Integrated into CMS
- Renamed from OpenEdition Shell and Utilities



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM V4 Product Packaging Changes

RTM, PRF, DirMaint, RACF, Performance Toolkit for VM

- Licensed under IPLA terms and conditions
- Preinstalled but disabled, license required

HCD and HCM

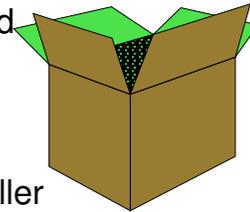
- Preinstalled on base system DDRs - no license required

Language Environment

- Integrated into base of z/VM V4.4 - no license required

Installation

- Using 3590-format tapes
- From a CD-ROM using 2074 Console Support Controller



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Other z/VM Version 4 Product Changes

Publications

- Available as PDFs from IBM Publication Center, z/VM Web site, or VM Collection CD-ROM (supplied with order)
- Some available from the IBM Publication Center (charge)
- BookManager® format from VM Collection CD-ROM or z/VM Web site
- Publications are NOT orderable from IBM distribution centers
- All optional feature publications are included in the z/VM library

Functions Removed

- CMS Vector support
- Distributed Computing Environment (DCE)
- BookManager Library
- Pre-configured CD
- ESAMIGR
- Installation to 9345, FBA, and 3380 DASD
- Installation from 4-mm DAT

LANRES/VM

- Withdrawn from marketing and is not available with z/VM Version 4



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

RACF for z/VM Feature

Licensed as an IPLA optional feature of z/VM V4

- OTC base charge
- S&S required for traditional service and no-charge upgrades
- Operates on standard engines and IFL processor features
- Will only run on z/VM V4.3 or later
- Preinstalled but disabled, license required

RACF helps meet the need for security by providing:

- Flexible control of access to protected resources
- Protection of installation-defined resources
- Ability to store information for other products
- Choice of centralized or decentralized control of profiles
- Transparency to end users
- Exits for installation-written routines



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

RealTime Monitor (RTM)

Licensed as an IPLA optional feature of z/VM V4

- OTC base charge
- S&S required for traditional service and no-charge upgrades
- Operates on standard engines and IFL processor features
- Will only run on z/VM V4
- Preinstalled but disabled, license required

RTM is a real-time monitor and diagnostic tool for:

- System monitoring, analysis, and problem-solving
 - Monitors system performance and the use of system resources
- Assists in validating the system components and establishing requirements for additional hardware or software installation

Performance Toolkit for VM is planned to replace the RTM feature

- z/VM V4.4 is planned to be the last release in which the RTM feature will be available and is planned to be withdrawn from marketing in a future z/VM release.



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Reporting Facility (PRF)

Licensed as an IPLA optional feature of z/VM V4

- OTC base charge
- S&S required for traditional service and no-charge upgrades
- Operates on standard engines and IFL processor features
- Will only run on z/VM V4
- Preinstalled but disabled, license required
- Simplifies performance analysis and resource management on your z/VM system

Analyzes your system's monitor data and produces performance reports and history files, including:

- System resource utilization, transaction response time, and throughput
- Resource utilization by userid
- DASD activity and channel utilization

Performance Toolkit for VM is planned to replace the PRF feature

- z/VM V4.4 is planned to be the last release in which the PRF feature will be available and is planned to be withdrawn from marketing in a future z/VM release.



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Directory Maintenance Facility (DirMaint)

Licensed as an IPLA optional feature of z/VM V4

- OTC base charge
- S&S required for traditional service and no-charge upgrades
- Operates on standard engines and IFL processor features
- Will only run on z/VM V4
- Pre-installed but disabled, license required
- Provides efficient and secure interactive facilities for maintaining your z/VM system directory

The required support for the Systems Management APIs are applied to the DirMaint feature supplied with the z/VM V4.4 system DDRs



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Statements of Direction for z/VM

Future releases of z/VM will:

- Support greater than 16 processors in a single VM image
- Require z/Architecture
- Provide Guest support for PCIX Cryptographic Coprocessor (PCIXCC)

z/VM V4.4 or later will:

- Support for up to 60 LPARs
- Support for up to four logical channel subsystems (LCSS)

z/VM V4.4 is planned to be:

- Last release offering RTM and PRF features
 - Future performance management enhancements to be delivered via the Performance Toolkit for VM



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ibm.com

IBM®

Performance Topics for Linux Guests



© Copyright IBM Corp. 2003. All rights reserved.

Objectives

Performance terms and objectives

z/VM memory and storage

Linux virtual memory

Tuning memory for Linux guests

Processor resources and the z/VM scheduler

Optimizing Linux guests processor requirements

DASD performance for Linux guests



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Topics for Linux Guests

Performance Terms and
Objectives



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Terms and Their Definitions

Internal Throughput Rate (ITR)

- Measures work per CPU second

External Throughput Rate (ETR)

- Measures work per second (wall clock)

CPU Utilization

- Measures how busy processor is
 - Related to ITR

Response Time

- Measures how long work takes to complete
 - Related to ETR
- Interactive vs. batch



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Have Performance Objectives

Know characteristics of workloads to deploy

- Plan for sufficient resources
 - Are future needs accounted for?

Add performance testing to deployment plan

- Ease the pain of deployment

Use monitoring to measure performance

- Quantitative results always preferable to qualitative impression
- Look for historical trends



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Resource Sharing Using Virtualization

z/VM guests think they 'own' physical resources

- Processor
- Memory
- I/O and network

Guests are granted access to resources through time slicing

- Memory is shared by paging

CP is the z/VM resource manager

Overcommitting resources is normal

- Problems can arise if an overcommitted resource is required by too many simultaneous guests



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Implications of Shared Resources

Which workloads perform well?

- Workloads that do not require an *entire* resource *all* the time
- Workloads that perform an action then sleep

Which workloads may not?

- Large memory footprint applications
- Applications that have continuously high CPU utilization



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Leveraging Linux Under z/VM

Maximize usage of available resources

- Many workloads tend to have spikes
- Servers often run at low CPU utilization
- z/VM can increase overall system utilization

Greater flexibility

- Easily defined virtual devices for Linux guests
 - Virtual memory
 - DASD
 - Network devices
- Simplified management



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

Have clear performance objectives

- Make performance part of deployment plan
- Monitor and track performance numbers

Understand how z/VM manages virtual machines

- Paging and time slicing is integral to z/VM

Choose the right workloads for Linux guests

- Understand the workload when considering deployment



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Topics for Linux Guests

z/VM Memory and Storage



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Memory Hierarchy

Main memory

- Programs execute in main memory
- Also known as main storage

Expanded storage

- Fast paging area
- Resides in physical memory
- Size is configurable, reduces available main memory

Paging space

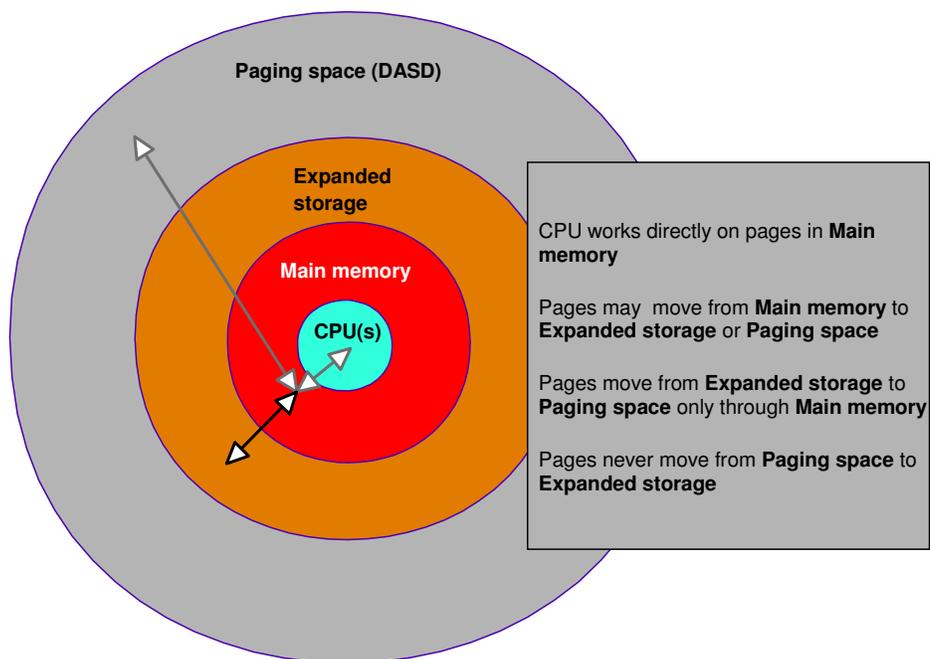
- Resides on DASD



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Storage


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Why Define Expanded Storage?

Expanded storage can improve response time

- Paging will likely occur
- z/VM paging is tuned for expanded storage
- Parts of CP must reside below 2 GB


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

How Much Expanded Storage to Define?

Rule of thumb estimate:

- Start with 25% of physical memory

Systems with low contention can reduce this ratio

When contention below 2 GB is high:

- Allocating 2-3 GB expanded storage may help

For more hints, see:

- <http://www.vm.ibm.com/perf/tips/storconf.html>



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

VDISKS

Virtual disks reside in virtual memory

VDISKS emulate real DASD

Provide fast access times



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Paging Space

Paging space resides on CP-owned DASD

For paging optimal performance:

- Use dedicated full packs for paging DASD
- Use multiple devices to permit overlapped I/O
- Define enough to allow block paging



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Name Shared Systems

What is Name Shared System (NSS)

- An OS IPL'ed by name (not device)
 - CMS is typically run from NSS
- Reentrant code and R/W data are stored in separate segments
 - Segment is a 1 MB portion of real memory
- Code is shared by multiple guests

To build a Linux NSS requires building kernel

- RedHat/SuSE do not provide NSS kernel
- Instructions at:
 - <http://www.vm.ibm.com/linux/linuxnss.html>

Warning!

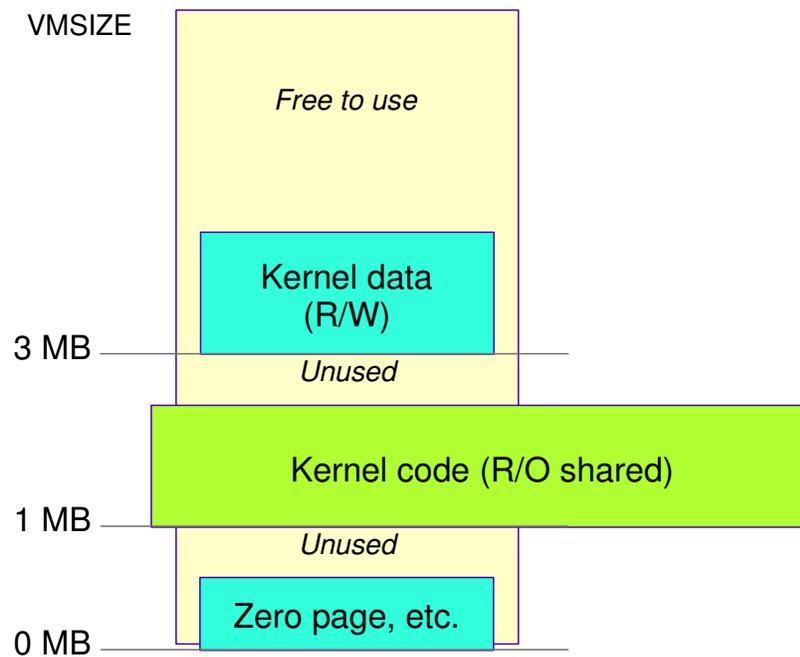
- Should be considered experimental
- May void your warranty



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Memory Map for a Shared Kernel



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Memory and Storage Guidelines

Reduce contention on paging DASD

VDISKS can offer fast DASD emulation

Use expanded storage for paging



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

z/VM virtual memory consists of

- Main memory
- Expanded storage
- Paging space

Expanded storage can improve performance

Use VDISK for fast access to emulated DASD

Reduce contention on paging DASD



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Topics for Linux Guests

Linux Virtual Memory



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Memory Usage in z/VM Linux Guests

Linux manages memory with regard to z/VM

Types of Linux memory:

- Kernel
- User memory
- Buffer and cache memory

Swap device used for paging area

Linux attempts to use all available memory

- Buffers and cache tend to fill unused memory



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Logic Behind Linux Memory Policy

On a dedicated server, unused memory serves no useful purpose

- Better to utilize free memory for buffers and cache
 - Reduce the chance of performing slow I/O operations
- Memory management (MM) algorithms are tuned to identify LRU pages

When memory is stressed, MM looks in buffers/cache for oldest accessed pages

- These may be eligible to be freed for immediate use

This has implications when real memory is shared



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Memory Management

Access counter is associated to memory pages

Memory is periodically checked for page usage:

- Pages eligible to be removed have counter decreased
- Pages that should not be removed have counter increased

Net result:

- Less recently accessed pages have lower counter value
- More recently accessed pages have higher counter value

Counter value is used during page cleaning



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Page Cleaning

Memory pages are cleaned when:

- The kswapd thread runs
- A process requests more memory than is available

When kswapd runs:

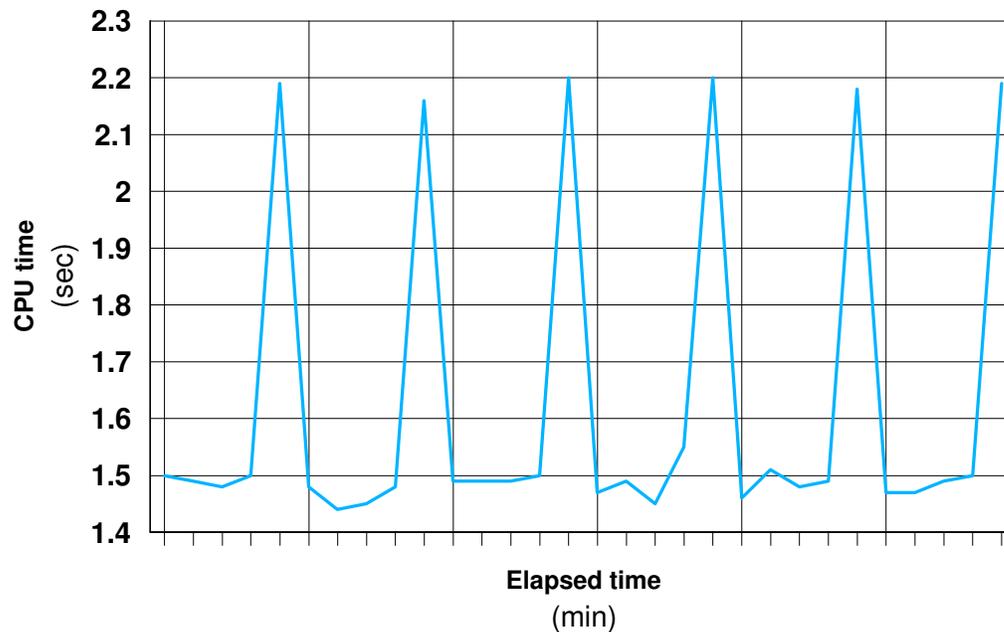
- If number of inactive pages falls below minimum, pages are moved to swap
- Inactive pages are cleaned from buffer and inode cache



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Observing Page Cleaning



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Observing Memory Usage

The `free` command:

- `$ free -k`
- Reports total, used, free, buffer, cache memory in KB

The `/proc/meminfo` kernel driver:

- `$ cat /proc/meminfo`
- Detailed memory statistics

The `vmstat` command:

- `$ vmstat 60 5`
- Continuously gathers statistics



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Be Aware!

Linux utilities assume guests owns all resources

- In reality, resources are shared by all guests

Numbers are relative to guest virtual machine

Statistics can be used to see what is happening inside the guest virtual machine



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Buffer and Cache Usage

Linux prefers to use most or all available memory

- Memory not used by applications tends to be used for buffer/cache

Rational is intended to reduce I/O operations

- Memory access is faster than I/O access

To illustrate, consider following charts:

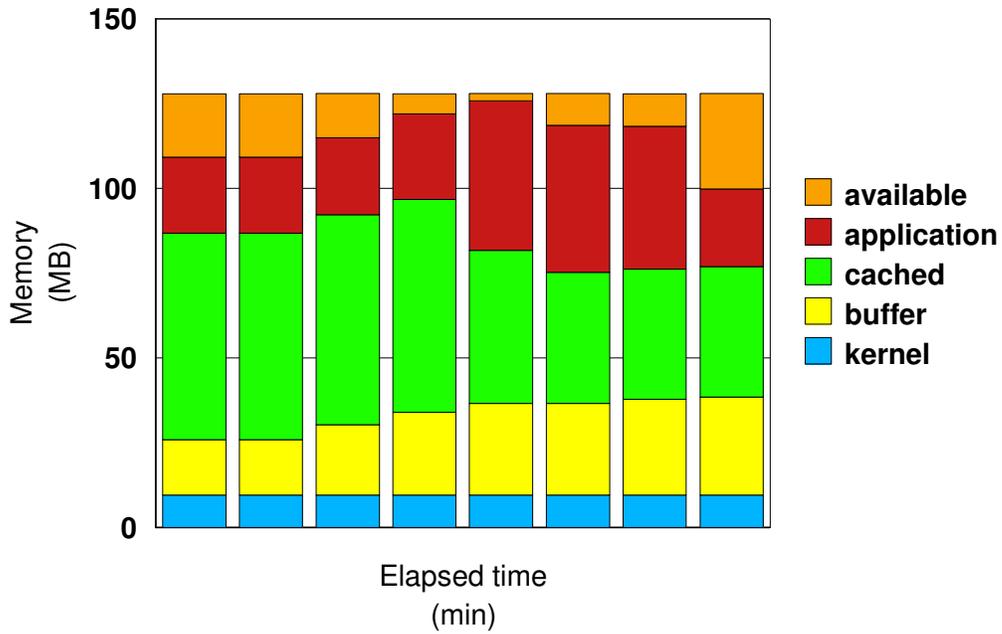
- Examine memory usage over time
- Two Linux guests
- Identical workloads
- Different virtual machine size (128 MB and 64 MB)



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

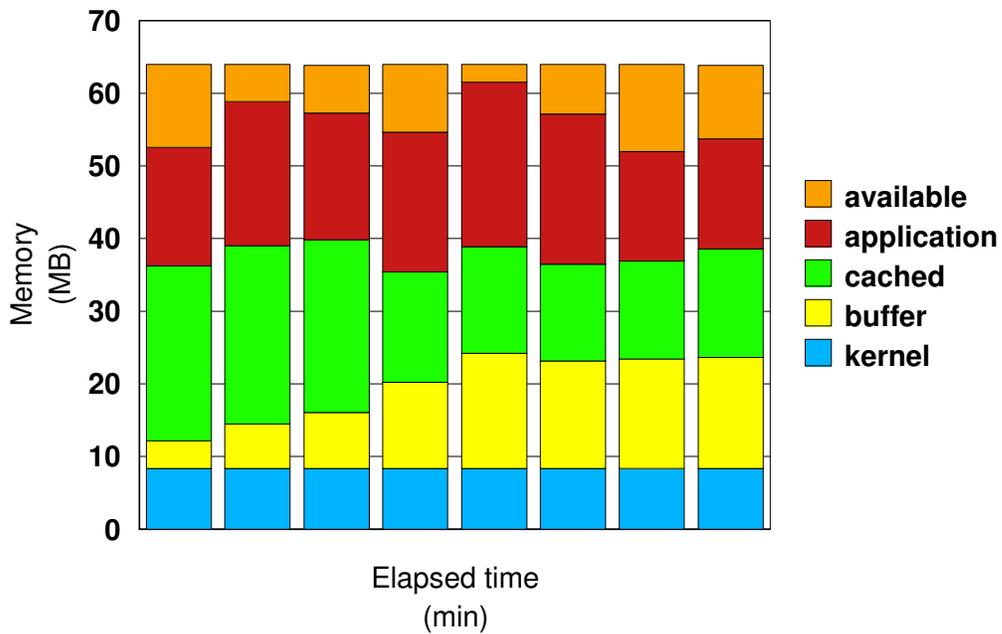
Memory Usage for 128 MB Guest



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Memory Usage for 64 MB Guest



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Conclusions

Similar memory usage pattern in both cases

- Application memory is obtained from buffer/cache

Reducing virtual machine size by 50% reduced average caching by 60%

- No appreciable effect on response time

Caching is optimal for distributed environment

- Unused memory is 'wasted' in distributed server

In shared environment, excessive caching can hurt

- Cached memory in 1 guest comes at the expense of another guest



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Swap Device

Linux pages to swap device during periods of high demand

- Use `vmstat` command to observe periods of paging

Options for Linux guests

- DASD
- VDISK

VDISK - a fast swap device option

- Virtual disk created in z/VM virtual memory
- Fastest swap device option
- If swapping does not occur, z/VM moves VDISK out of main memory

Be aware!

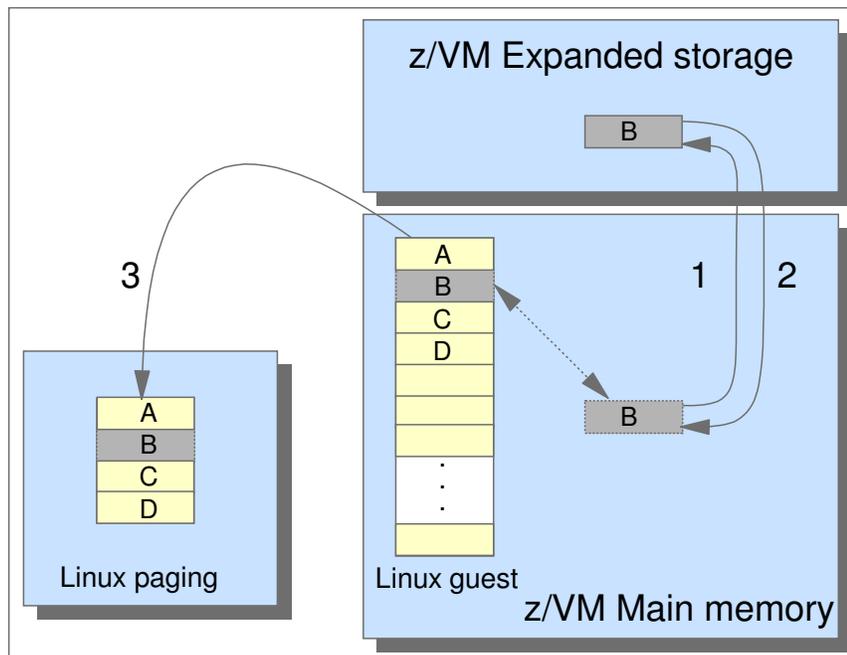
- VDISK can increase memory contention



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Double Paging Effect


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Double Paging - Cause and Solution

Double paging is not unique to Linux

- Result of 2 parties managing virtual memory

Kernel support for z/VM PAGEX included in 2.4

- Can reduce impact of doubling paging
 - Linux gives up a time slice if blocked on I/O to do double paging

To reduce the effect:

- Keep guest virtual machine size small enough to fit in main memory, or
- Make guest virtual machine size large enough so Linux will not swap


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Minidisk Cache and Linux Swap Device

Minidisk Cache (MDC) in expanded storage/main memory

- Used for:
 - Normal user minidisk, temp disks, etc.
- Not used for:
 - Shared/dedicated/attached DASD, VDISKS, etc.

MDC is write-through cache

- Data is added to MDC when read from DASD
- Writes will update data already in MDC
 - Writes will not add new data to MDC

Do not enable MDC for swap device

- Access pattern does not justify cost



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Swap Device Recommendations

Use VDISK if feasible

- VDISK offers fastest access time

Turn off MDC for DASD swap device

Keep swap device size as small as possible

- 2x recommendation for distributed servers is probably too large
- Large swap device size may indicate more virtual memory needed



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Topics for Linux Guests

Tuning Memory for Linux Guests

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Effect of Virtual Memory Size

Shortage of virtual memory will impact performance

- Extended Linux swapping indicates virtual memory shortage

More virtual memory may not result in better performance

- Excess memory used in Linux buffers/cache

Excessive virtual memory size can impact overall system performance

- Can lead to contention when memory is shared

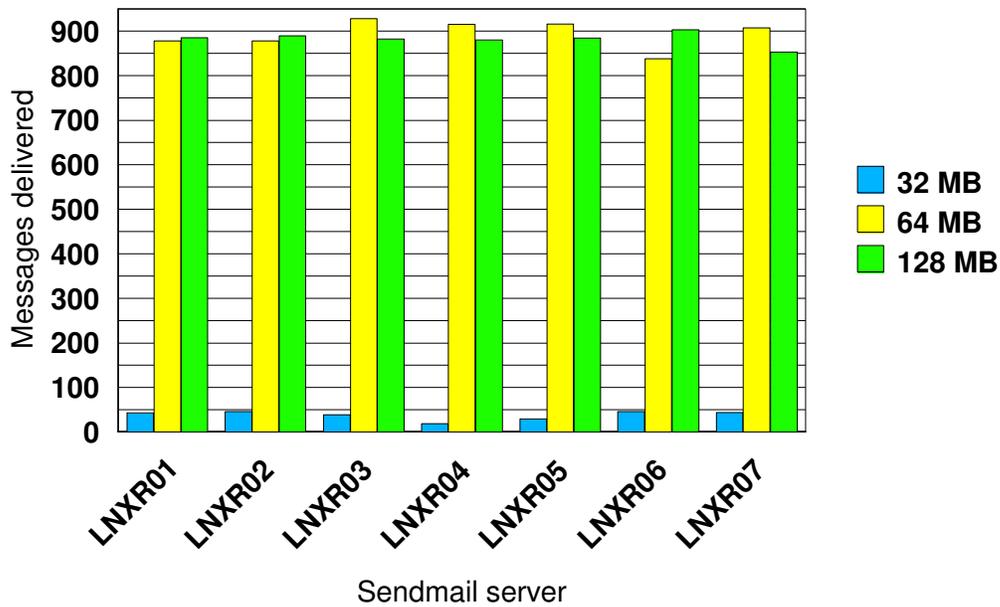
To illustrate effect of virtual machine size, consider following chart

- Identical workloads
- Varying virtual memory size

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Illustrating Effects of Memory Size


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Conclusions

Mstone benchmark used to measure performance

- More delivered messages equates to better performance

Run against Linux guests with 3 VM sizes:

- 128 MB
- 64 MB
- 32 MB

Results for 128 MB and 64 MB guests nearly identical

- Only half the memory required for 64 MB guest

Performance degrades for 32 MB guest

- Workload requires more virtual memory


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Choosing the Correct VM Size

Determine the smallest required memory footprint

- Run tests to determine this point

Find point where Linux begins to swap

- Look for swap activity using `vmstat` command

Slightly increase the guest virtual memory size

- To account for additional load on guest
- 20% may be good rule of thumb number



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

Linux attempts to use all available memory

- Unused memory is allocated to buffers and cache

Allocate sufficient virtual memory to Linux guests

- Memory shortage will degrade Linux performance

Excessive memory allocation can degrade overall performance

- Can lead to excessive contention for z/VM storage

Choose an optimal virtual memory size for Linux guests

- Look for the point where Linux begins to swap



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Topics for Linux Guests

Processor Resources and the z/VM Scheduler



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Processor Topics to be Covered

LPAR weights and options

Scheduling virtual machines in z/VM

Steps to optimize scheduling Linux guests



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

LPAR Weights

LPARs gain processor time based on weights

To calculate processor allocation to an LPAR:

- Sum weights of all LPARs
 - Total weight for all LPARs
- Divide LPAR weight by total weight
 - This is “logical share” allocated to LPAR
- Divide “logical share” by number of logical processors
 - Each processor contributes to logical share


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

LPAR Example

3 LPARs share 2 processors:

- A1 LPAR logical share is 16% of 2 shared processors (100 / 600)
- Logical share of 16% equivalent to 33% of 1 processor (16 x 2)

LPAR	Weight	Logical Share
A1	100	16%
A2	200	33%
A3	300	50%


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

LPAR Options

Capped (Yes or No)

- Limits access to processor based on weight

Wait completion (Yes or No)

- Determines if LPAR gives up processor

Dispatch slice (Dynamic or specific)

- Determined LPAR time slice



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Shared Versus Dedicated Processors

In general, use processors dedicated to an LPAR for:

- Benchmarking
- For steady workloads that justify the cost

Objectives for zSeries workloads should be high utilization to leverage zSeries:

- Reliability
- Availability
- Serviceability

Reducing system utilization reduces zSeries effectiveness



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Reduce Processor Sharing Overhead

LPAR overhead increases as ratio of logical to physical processors increases

- Logical CPs are processors shared by multiple LPARs
- Related to the cost of time slicing across LPARs

To reduce LPAR overhead:

- Use fewer LPARs and fewer logical processors



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

Processors can be shared or dedicated to LPARs

zSeries workloads are geared towards high utilization

- Shared processors leverage zSeries strengths

In general, run LPARs with:

- Capped = No
- Wait completion = No
- Dispatch slice = Dynamic

With dynamic dispatch slicing:

- Weights determine a minimum allocation for uncapped LPARs

Reduce ratio of logical to physical processors



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

The CP Scheduler

Attempts to run as many concurrent virtual machines as possible

Scheduler evaluates demand for real resources

Virtual machines classified by expected resource usage:

- **Class 1** – interactive
- **Class 2** – non-interactive
- **Class 3** – resource intensive



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Scheduler Lists

Virtual machines reside on 1 of 3 lists:

- Dormant list
 - No immediate tasks to perform
 - Move to eligible list when require servicing
- Eligible list
 - Waiting for resource availability
 - Move to dispatch list as resources become available
- Dispatch list
 - Contending for processor time



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Eligible List Queues

Virtual machines classified by expected resource usage:

- E0 - do not wait in eligible list (move immediately to dispatch list)
- E1 - short transactions (Class 1)
- E2 - medium length transactions (Class 2)
- E3 - long-running transactions (Class 3)

Classification objectives:

- Favor less resource-intensive virtual machines
- Ensure virtual machines receive designated processor share
- Control amount and type of service based on classification



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Dispatch List Queues

Virtual machines on the dispatch list inherit classification from eligible list:

- Q0 were E0 on eligible list
- Q1 were E1 on eligible list
- Q2 were E2 on eligible list
- Q3 were E3 on eligible list

Priority determines how long virtual machines control processor

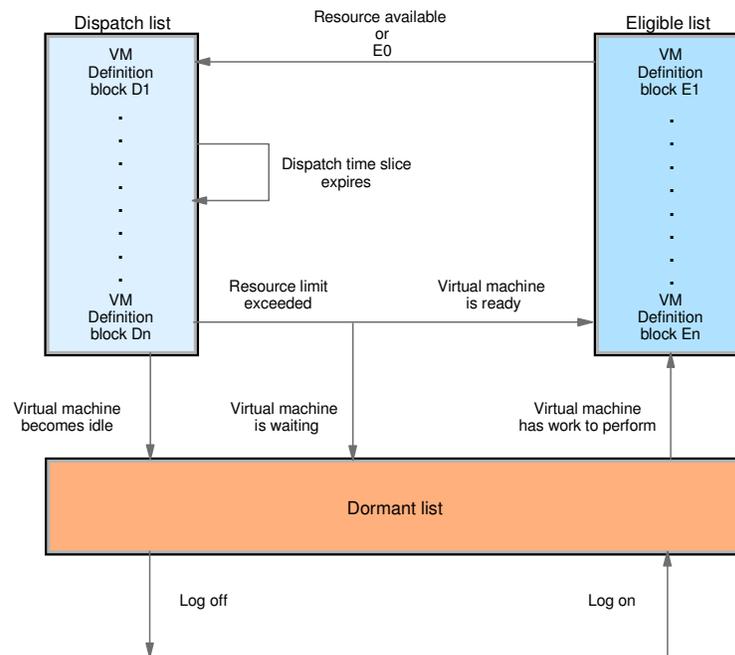
- Priority is adjusted while on dispatch list



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Scheduler Lists - State Transitions


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Leaving the Dispatch List

Virtual machines move out of dispatch list when:

- Task is complete
 - Virtual machine moves to dormant list
- Elapsed time slice has expired and work has not completed
 - Virtual machine moves to eligible list
- Long interrupt occurs
 - Virtual machine moves to dormant list

When reentering eligible list, classification is reevaluated

- Q1 drops to E2
- Q2 drops to E3
- Q3 remains at E3


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Scheduling Time Slices

Elapsed time slice

- Virtual machines remain on dispatch list for duration of elapsed time slice
- Value is dynamically adjusted

Dispatch time slice

- Virtual machine controls processor for duration of dispatch time slice
- Often referred to as *minor time slice*



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Scheduling Virtual Processors

Virtual machine can have more than 1 virtual processor

- VM definition block is assigned to each virtual processor
 - Base VM definition block
 - Additional VM definition block

Both base and additional VM definition blocks cycle through scheduler queues

- Consumed processor time is independently measured
- Base VM definition block is always at least as high in scheduler queues as any additional VM definition blocks



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Entering the Dispatch List

To move to dispatch list, virtual machine must pass 3 tests:

- Storage test
- Paging test
- Processor test

E0 virtual machines are not subject to tests

- Move immediately from eligible list to dispatch list

Tests can be adjusted using SRM controls



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Scheduling and the Linux Timer Patch

For Intel Linux, kernel timer interrupts every 10 ms

- Increments "jiffies" counter
- Results in minimal impact for dedicated servers

Timer interrupt can disrupt z/VM scheduling

- Scheduler classifies Linux guest as Q3 (long running)

To minimize impact:

- Apply timer patch from developerWorks site if not provided in your distribution
 - <http://www.ibm.com/developerworks/oss/linux390/index.shtml>



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Timer patch is provided with newer distributions

- SuSE 8.0
- SuSE 7.2
- Timer patch is not available for RedHat distributions!

To see if patch is enabled:

```
# cat /proc/sys/kernel/hz_timer
1
- 1: disabled
- 0: enabled
```



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

To Enable Timer Patch

Write to procs:

```
echo 0 > /proc/sys/kernel/hz_timer
```

Use sysctl command:

```
/sbin/sysctl -w kernel.hz_time=0
```

To make change persist across Linux IPL:

- Add to /etc/sysctl.conf file:

```
# Enable kernel timer patch
kernel.hz_timer = 0
```
- Invoke sysctl command in /etc/init.d/boot.local:

```
/sbin/sysctl -p
```



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Measuring the Effect of the Timer Patch

Timer Patch has a measurable effect on scheduling:

```

<USERID> %CPU %CP %EM ISEC PAG WSS RES UR PGES SHARE VMSIZE TYP,CHR,STAT
LNXR09 .62 .06 .56 .20 .00 28K 34K .0 0 100 128M VUX,DSC,DISP
RMHTUX01 .01 .00 .01 .00 .00 23K 24K .0 0 100 128M VUX,DSC,DISP
RMHTUX02 .01 .00 .01 .00 .00 17K 17K .0 0 100 128M VUX,IAB,DISP

```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

Scheduler attempts to run as many concurrent tasks as possible

- Virtual machines as classified for expected resource usage
- Virtual machines cycle around dormant, eligible, and dispatch lists
 - Classification is adjusted for actual resource usage

Eligible lists has tasks waiting for resources

- To move to dispatch list, 3 tests are applied

Virtual machines compete for processor on dispatch list

- Short running tasks complete work after 1 cycle on dispatch list

Virtual processors are accounted for in scheduling


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Adjusting Scheduler Parameters

It is possible to influence scheduling through CP commands

- Global commands that pertain to the scheduler
- Local commands that apply to individual virtual machines

SRM controls affect scheduler tests

- Storage test
- Paging test
- Processor test

Commands relevant to virtual machines:

- CP SET QUICKDSP
- CP SET SHARE



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

CP SRM Controls

Used to change system parameters

- Can affect scheduler behavior
- Syntax: `SET SRM option [values]`

Some options:

- STORBUF
 - Storage test
- LDUBUF
 - Paging test
- DSPBUF
 - Processor test
 - Risky control! Do not change



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Paging Test

LDUBUF stands for "Loading User Buffer"

- Loading user is high consumer of paging resources
- Expected to have high paging rate
 - 5 page faults in 1 time slice
 - Equivalent to using 1 paging device (paging exposure)
- CP `INDICATE QUEUES EXP` command displays current loading users

Scheduler restricts number of loading users on dispatch list

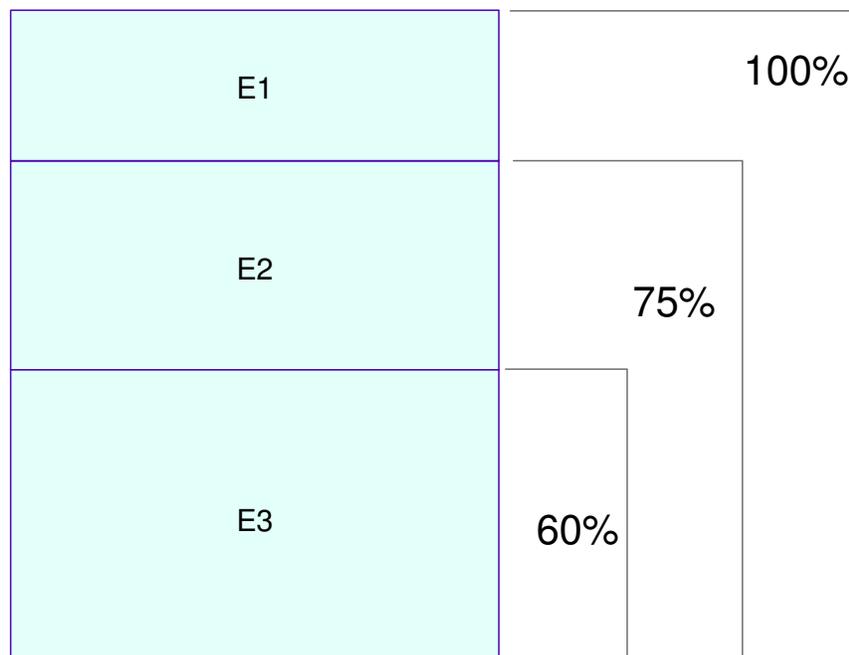
- LDUBUF specifies percentage of paging exposures allocated by transaction class

Paging resources can be overcommitted


ibm.com/redbooks

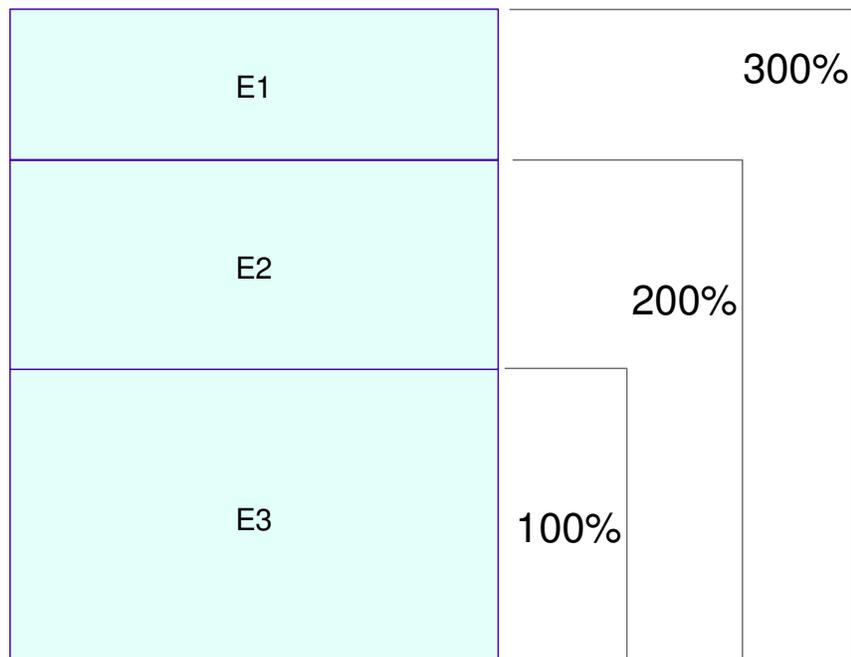
© Copyright IBM Corp. 2003. All rights reserved.

SET SRM LDUBUF 100 75 60


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SET SRM LDUBUF 300 200 100

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Storage Test

STORBUF partitions commitment of main memory

Before a virtual machine moves to dispatch list:

- Memory must be available to the virtual machine's transaction class

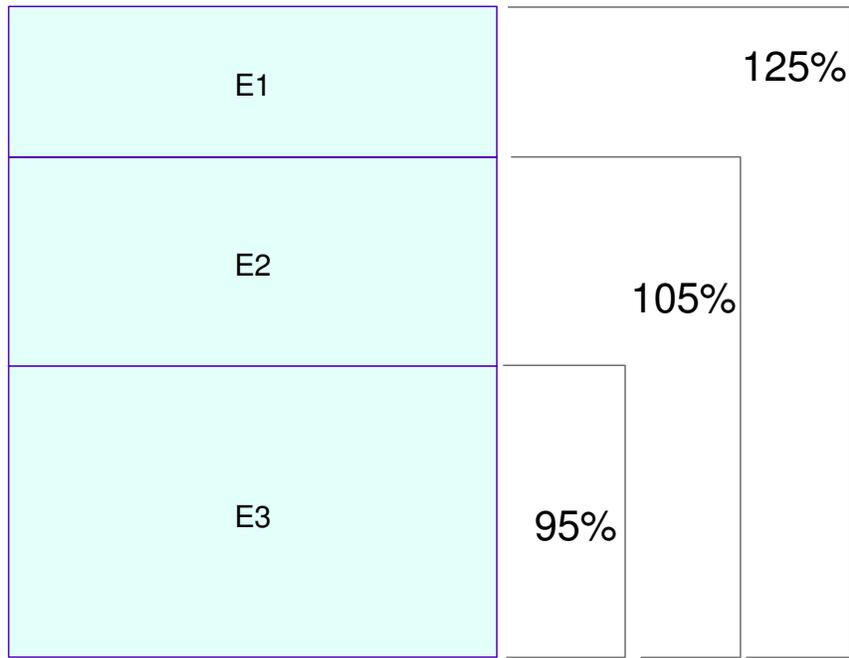
Memory can be overcommitted, however:

- Severe overcommitment can lead to thrashing
- STORBUF settings are more critical than LDUBUF

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

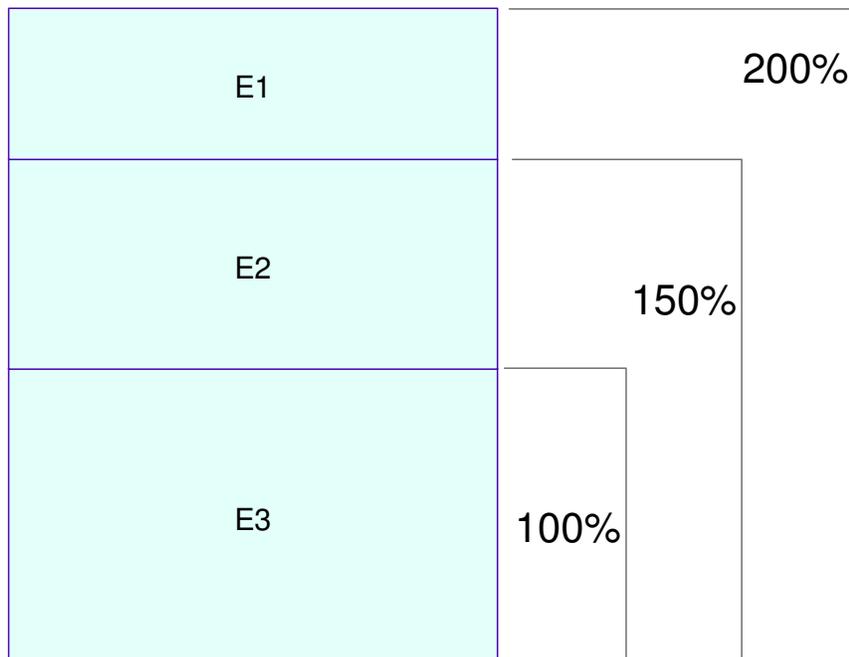
SET SRM STORBUF 125 105 95



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SET SRM STORBUF 200 150 100



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Processor Test

DSPBUF partitions controls number of virtual machines in dispatch list

Values specified by transaction class:

- **SET SRM DSPBUF 35 30 18**
 - 35 slots available for E1, E2, E3
 - 5 guaranteed for E1 (35 - 30)
 - 30 slots available for E2 and E3
 - 12 guaranteed for E1 and E2 (30 - 18)
 - 18 slots available for E3
 - These may be used by E1 and E2
- Default values infinite

Do not change!

- LDUBUF and STORBUF are preferable



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Dropping from Queue

Default scheduler controls give more access to resources to Class 1 virtual machines

- In general, short transactions use less resources

As virtual machines complete transactions:

- They are moved from dispatch list to dormant list
 - This is referred to as 'dropping from queue'
- Reclassified as Class 1 when new work needs to be performed

Dormant virtual machines are not using real resources

- Timer patch is needed for Linux guest to drop from queue
- Other considerations for Linux guests to drop from queue



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

STORBUF and Linux Guests

When Linux guest do not drop from queue:

- Scheduler may perceive storage requirements to be larger reality
- Result:
 - Linux guests help in eligible queue longer than needed

In this case:

- Overcommitting memory may help
- Things to try:
 - Use STORBUF to increase overcommittment
- But remember!
 - Avoid thrashing



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

When to Use SRM Controls

SRM controls attempt to manage resource over-commitment

- Access to resources by virtual machine classification

Use SRM to:

- Overcommit paging resources
 - Look for full utilization of paging devices
- Overcommit memory resources
 - When Linux virtual machines do not drop from queue, this may help
 - Avoid thrashing!



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Other SRM Controls

DSPSLICE

- Changes the dispatch (minor) time slice (default is determined by model at IPL)

XSTORE

- Sets percentage of extended storage scheduler considers for dispatching purposes (default is 0)

MAXWSS

- Sets maximum working set a normal user is allowed to have (turned off by default)

IABIAS

- Sets interactive bias for virtual machines on dispatch list



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Options for Adjusting Local Controls

Scheduling can be adjusted for specific z/VM guests

- These provide finer granularity

Options to consider:

- SET QUICKDSP
- SET SHARE



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

The QUICKDSP option

Use CP SET QUICKDSP command to bypass scheduler tests

- Syntax: `SET QUICKDSP userid ON | OFF`
- Virtual machine is classified as E0 on eligible list
- E0 virtual machines move directly to dispatch list

Intended to give priority to critical virtual machines

- May be useful for Linux guests

Be aware!

- Overuse may not produce intended results



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

CP SET SHARE Command

SHARE factors in both eligible and dispatch priority

Two types of settings:

- ABSOLUTE
 - Can guarantee a percentage of system resources
- RELATIVE
 - Gives priority after ABSOLUTE share are assigned

Values are assigned as numbers

- ABSOLUTE ranges from 0% - 100%
- RELATIVE ranges from 1 - 10000

Syntax: `SET SHARE userid type value`



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SHARE Setting Recommendations

With **RELATIVE** share:

- As more users log on, share of resources drops

With **ABSOLUTE** share:

- Share remain fixed up to sum of all ABSOLUTE shares = 100%

Use **ABSOLUTE** for critical servers (i.e.. TCPIP, DNS)

If needed, use **RELATIVE** for all others. But remember....

- A server assigned high share will consume resources at the expense of other users



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

SRM settings can globally influence scheduler settings

- LDUBUF and STORBUF are first to consider

Local controls can influence behavior of specific virtual machines

- QUICKDSP
- SHARE

Goodness is when idle virtual machines drop from queue!

- Timer patch is a prerequisite for Linux guests



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Optimizing Linux Guests Processor Requirements



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ibm.com

Effect of Unneeded Linux Daemons

Default Linux installation starts many daemons which may not be used in the guest

- Installer offers a prepackaged selection of applications
- Users typically do not select individual services

Implications for both:

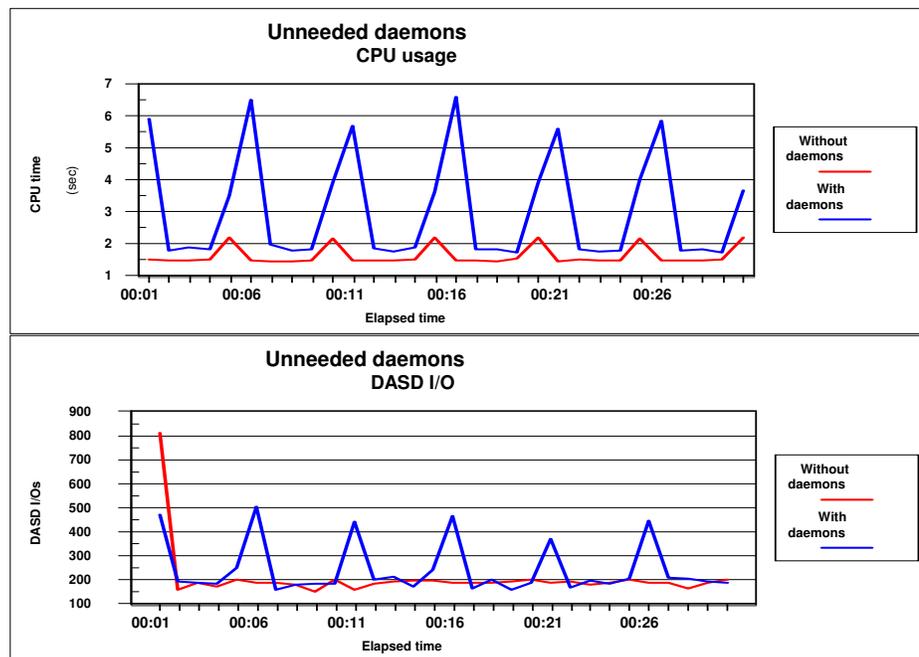
- Virtual memory used by the guest
- Processor resources used when the daemon runs



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Effect of Daemons - CPU and DASD



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

What Constitutes an Unneeded Service?

In a dedicated environment, resources used by unused daemons have minimal impact

- Convenience of starting the service outweighs the cost of running it

For a shared environment, this is not the case

- Every active memory page subtracts from amount available to other virtual machines
- To get CPU cycles requires scheduling

Unneeded services can be anything that is not actively used

- The `chkconfig` command is useful to see what is enabled
 - Can be used to turn off/on services as well



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Services to Consider

sendmail

- If sending/receiving mail not required

anacron, atd, cron

- Stop these if no regularly scheduled jobs run

autofs, nfs, nfslock, portmap

- Provide file system services, RPC support

lpd, xfs

- Printing and X-Windows

inetd/xinetd

- Manages Internet services



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Corollary:

Avoid 'Are you there?' pings

- May cause unnecessary scheduling

Do not monitor idle Linux guests

- The act of monitoring requires CPU cycles

Note:

- The `top` command is a notorious consumer of CPU time!



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

QDIO Patch

Even with the timer patch applied:

- Linux guests using QDIO network driver may not drop from queue
- Effect seen when creating Performance redbook

Linux guests using IUCV network driver drop from queue

- Indication something different occurring in QDIO driver

After discussion with z/VM development:

- Identified issue with QDIO
 - Read I/O event awaiting completion
 - No real pending read outstanding
- Fix to APAR VM63282 for z/VM 4.3
- Incorporated into z/VM 4.4


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Guests with QDIO Fix

With PTF for APAR VM63282, QDIO guests drop from queue:

CP INDICATE

```
AVGPROC-003% 02
XSTORE-000000/SEC MIGRATE-0000/SEC
MDC READS-000001/SEC WRITES-000001/SEC HIT RATIO-090%
STORAGE-052% PAGING-0001/SEC STEAL-000%
Q0-00000 (00000) DORMANT-00094
Q1-00020 (00000) E1-00000 (00000)
Q2-00003 (00000) EXPAN-002 E2-00000 (00000)
Q3-00014 (00000) EXPAN-002 E3-00000 (00000)
```

```
PROC 0000-004% PROC 0001-002%
```

```
LIMITED-00000
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Defining Virtual Processors

Real processors defined to z/VM LPAR are shared by all virtual machines

By default, a virtual machine defines one virtual processor

- Base processor

Additional virtual processors can be added

- Adjunct processors
- True multi-processing requires at least two virtual processors

As discussed in scheduler section:

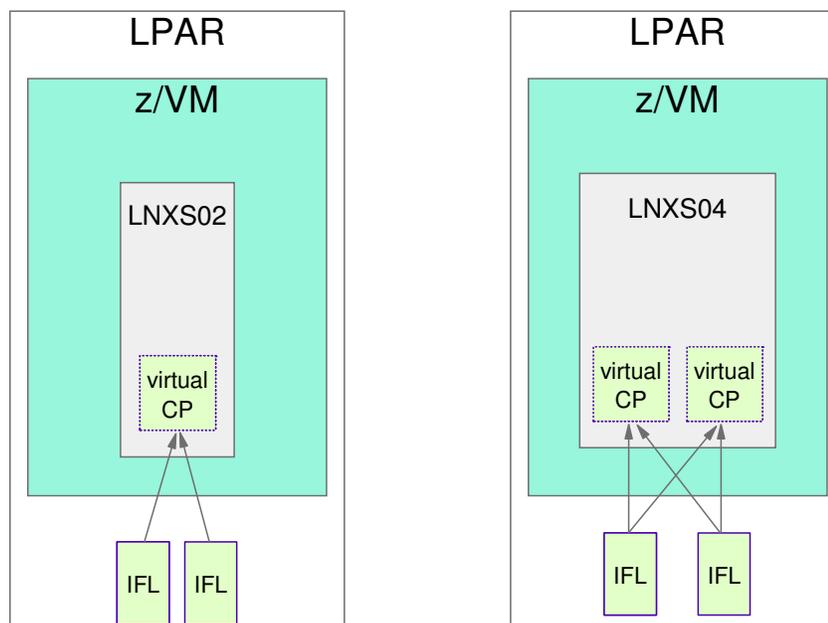
- Each virtual processor cycles through scheduler lists
- Affinity exists between base and adjunct processor(s)



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Virtual Processors in a z/VM Guest



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

When and How Many Virtual Processors?

Virtual processor can help CPU-bound guests

- When CPU utilization is high, virtual processors may help
 - Multiprocessing is then possible in the Linux guest

Define as many virtual processors as real processors

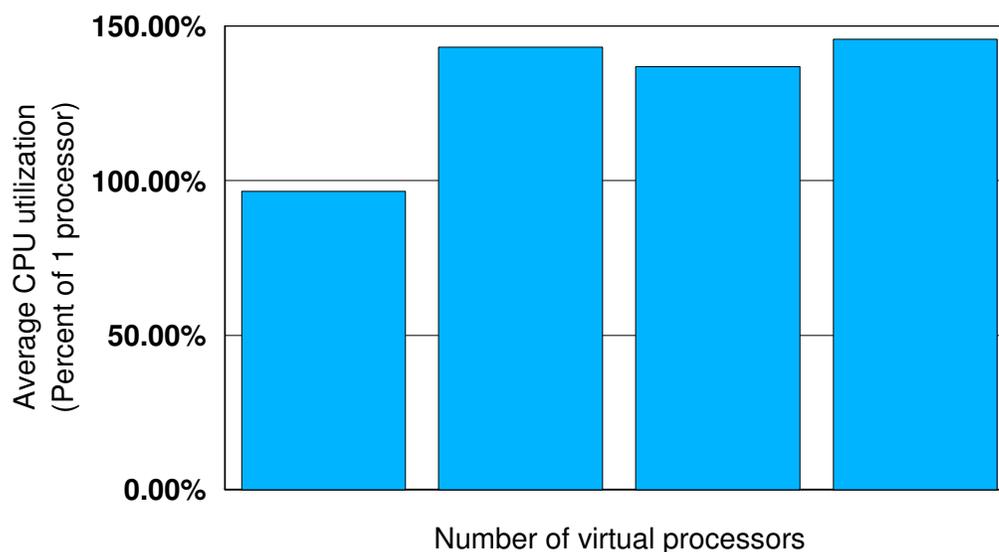
- If two processors available, use two virtual processors
- Do not define more virtual than real processors

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Utilization of Virtual Processors

Processor Utilization

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Processor Performance Recommendations

Stop unneeded Linux services

- Default Linux installation will start more services than required
- Remove scheduled tasks
 - Check for unneeded tasks started by cron

Reduce processor usage by idle Linux guests

- Install timer patch
- Eliminate "are you there" pings
- Do not measure performance on idle Linux guests

Virtual processors can help CPU-bound virtual machines

- Define only as many virtual processors as real processors available to z/VM LPAR



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Performance Topics for Linux Guests

DASD Performance for Linux Guests



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Factors That Influence DASD Response

Speed of physical devices

- DASD and control units
 - IBM Enterprise Storage Server (ESS)
 - Older RAMAC Virtual Array (RVA)
- Channels
 - Enterprise Systems Connection (ESCON)
 - Fibre Connection (FICON)
 - SCSI over Fibre Channel Protocol (FCP)

Contention

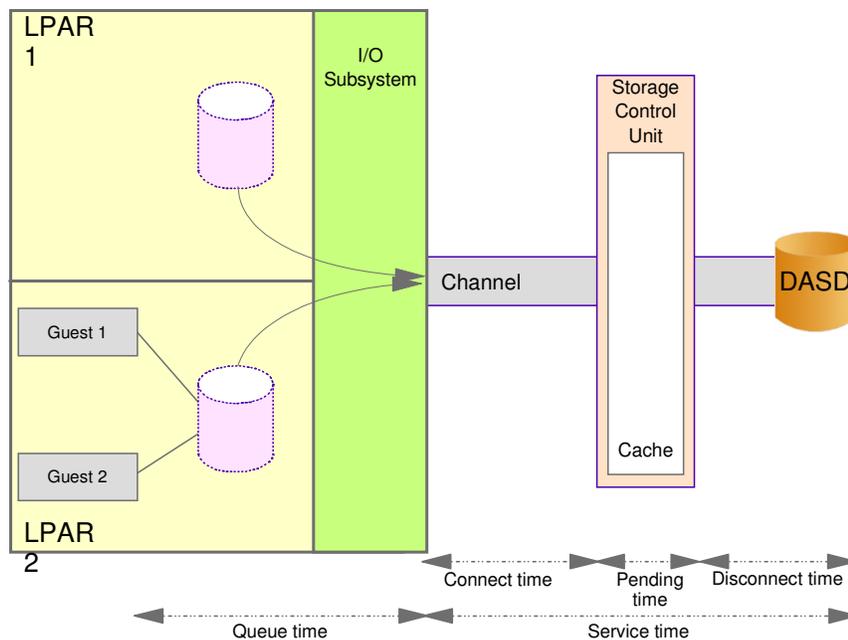
- On the physical devices
- Between z/VM guests



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Components of Overall Response Time



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

DASD Unit Options

IBM Enterprise Storage Server

- The premier storage solution for all IBM Sserver brands
 - Including IBM zSeries
- Based on SCSI technology
 - Presents some performance opportunities for Linux
- Configuration is critical for maximum performance
 - Objective is to maximize parallel access

RVA

- Older technology
- Still commonly used



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Comparing Channel Types

ESCON

- Older ESCON has a maximum data transfer rate of 17 MB/s

FICON

- Transfer rate up to 100 MB/s

Both are based on fiber optic technology

- All else being equal, FICON is faster



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Logical Volume Manager

Disk management system for Linux

- Operates below file system level

Relevant point for this discussion:

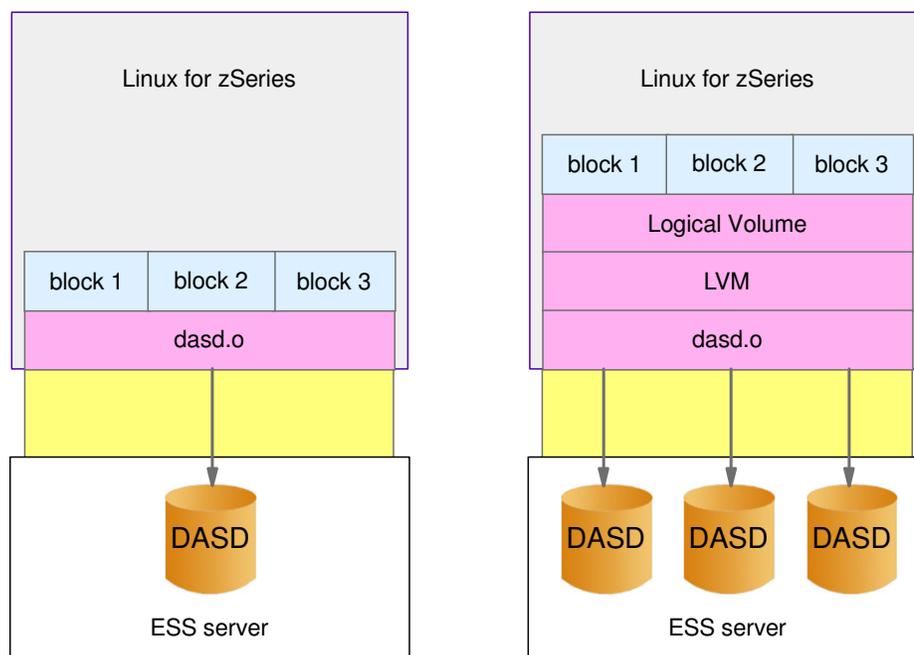
- Enables assembling a file system from multiple physical disks
- Can be used to maximize parallel access to data
 - Can be used to reduce contention



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Increasing Parallel Access With LVM



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Realizing the Benefits of Parallel Access

To illustrate benefits of multiple DASD devices:

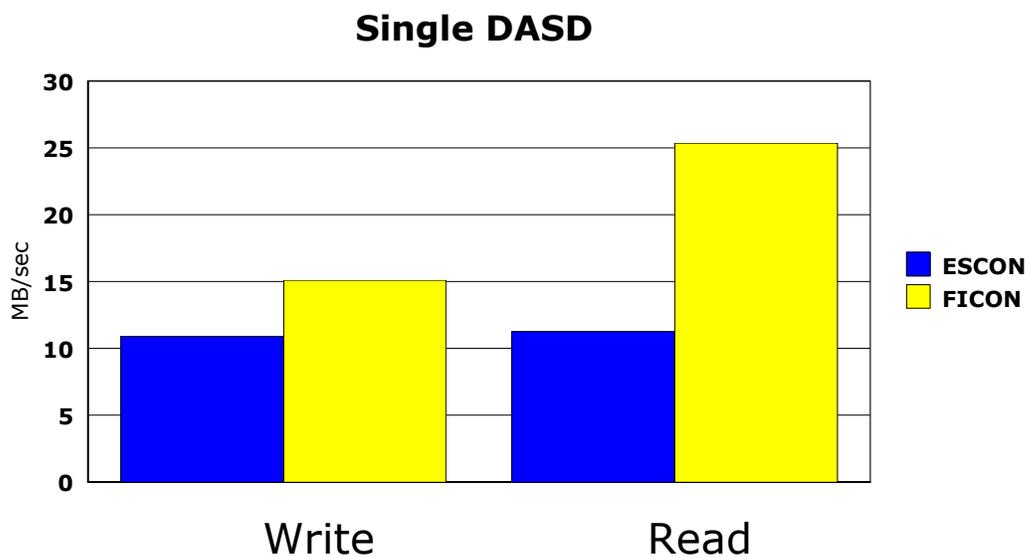
- Compare ESCON and FICON throughput
 - Using a single disk
 - Using up to 8 disks assembled by LVM into a single file system

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ESCON vs. FICON Measurements

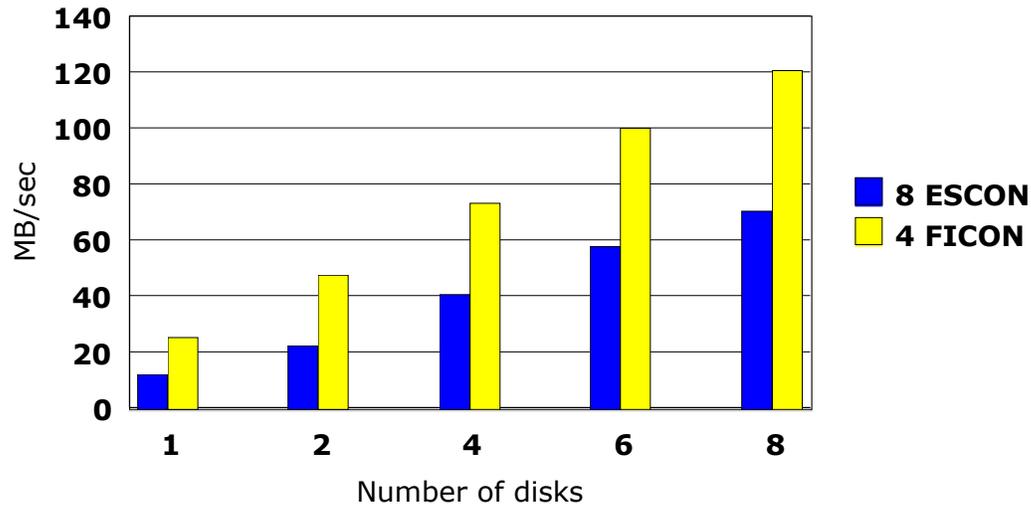
ibm.com

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

ESCON vs. FICON Measurements

Multiple DASD



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SCSI over FCP

IBM Enterprise Storage Server stores data in block format

- Data transfers use SCSI protocol

Linux reads buffer/cache data in block format

- SCSI is block format protocol

ESCON and FICON use ECKD protocol

- Translation occurs in the dasd.o driver and at ESS

With SCSI over FCP translation can be avoided

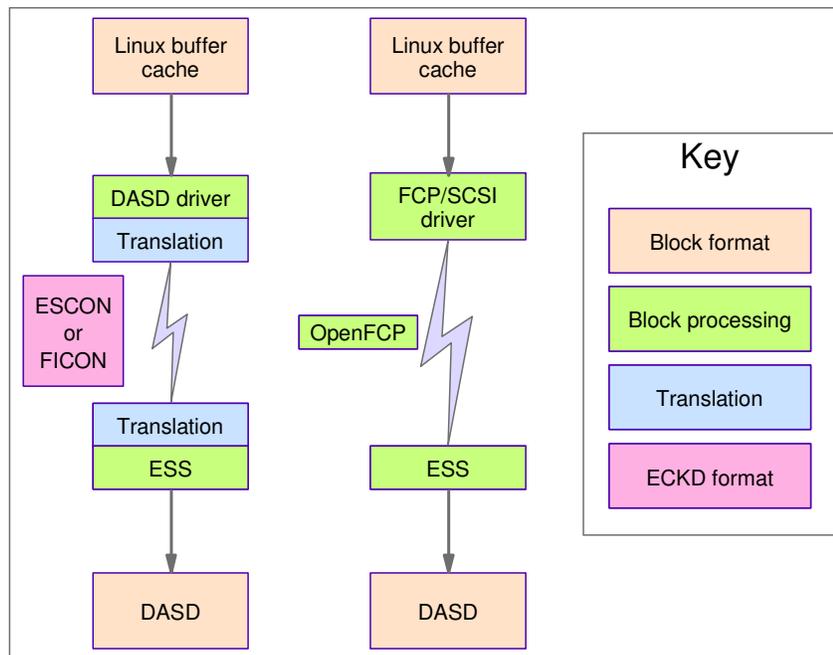
- Presents highest performance opportunity



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Advantages of SCSI Over FCP


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Recommendations for DASD Performance

Use more, smaller disks

- I/O to a single physical disk is serialized

Use LVM for parallel data access

- With LVM, disks can be accessed in parallel

SCSI over FCP offers the best performance for Linux

- No translation overhead
- Highest channel throughput

Reduce contention points

- DASD
- Control unit
- Channel


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux Security on zSeries



© Copyright IBM Corp. 2003. All rights reserved.

Objectives

- Security objectives and policies
- zSeries and z/VM system integrity
- Securing z/VM
- Linux user access and authentication
- Monitoring and hardening Linux
- Virtual Private Networks and Firewalls

Security Objectives and Policies



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Factors to Consider For Security

ibm.com

Integrity

- Can the computing environment reliably protect data?

Confidentiality

- Is sensitive data involved?

Risk

- What is the potential cost of unauthorized access?

Threat

- Who is most likely to gain unauthorized access?

Vulnerability

- Where is an attack likely to succeed?



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Know Your Security Objectives

What constitutes a 'secure' installation?

- Answer often depends on who is asked
- The most secure machine is:
 - Locked in separate room of bombproof bunker
 - Disconnected from any network
 - Powered off

Some level of paranoia is required

- Choose the correct level of security based on business objectives
- But, DO NOT take security for granted!

Choose the correct security level for your system



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Security Policy

For risk analysis, identify:

- Vulnerabilities
 - Where are the weak point?
- Threats
 - Where are the bad guys most like to attack?

Adopt a general policy

- That which is not expressly permitted is forbidden
- That which is not expressly forbidden is permitted



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Creating a Security Policy

Security policy:

- States operating procedures for secure computing environment
- Should be in writing!
- Includes guidelines for System administrators and users

For reference, see RFC2196: *Site Security Handbook*

<http://www.ietf.org/rfc/rfc2196.txt?number=219>

<http://www.sans.org/resources/policies/>



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Some Types of Vulnerabilities and Attacks

Physical compromise

- Shoulder surfing / password guessing

Executable weaknesses

- Trojan horse
- Back door
- SUID executables
- Buffer overflows

Network attacks

- Denial of Service (DoS) attacks
- Scanning
- IP address spoofing
- Session hijacking



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

zSeries and z/VM System Integrity



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Topics to be Covered

Facilities to ensure system integrity across LPARs

HiperSockets and OSA-Express network integrity

z/VM system integrity

Virtual machine isolation



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

LPAR Integrity

Processor Resource / System Manager (PR/SM)

- Hardware facility
 - Enables partitioning of central processor complex (CPC)
- Real storage is always dedicated to LPAR
 - Dynamic Address Translation always in effect (hardware function)

Start Interpretive Execution (SIE)

- Enables context switch between LPARs
- Resources are always cleared / reset before used in other context

Memory is:

- Directly associated with LPAR
- Reconfiguration possible
 - Memory is cleared before reassignment


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

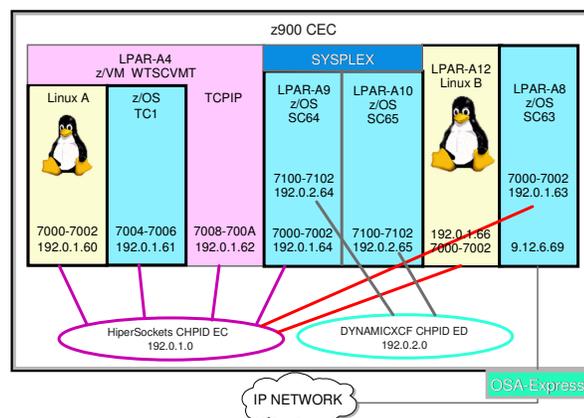
zSeries Network Security Benefits

HiperSockets provide secure zSeries networking

- Network traffic contained within Central Electronics Complex (CPC)
- Data transferred through direct memory-to-memory copy
 - Cannot be 'sniffed' by third party

OSA-Express

- Secure 'within the box'


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM System Integrity

Virtual machines are isolated from each other

- Control Program (CP) manages machine isolation

Linux creates separate address space for each process

- z/VM provides an additional level of address translation

Virtual machine execution utilizes SIE instruction

- CP dispatches virtual machine using SIE
- Control returns to CP when:
 - Virtual machine time slice expires, or
 - Execution cannot be virtualized



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Virtual Memory and Devices

Previously unreferenced memory pages are always zero

- Virtual machine cannot see another virtual machine's memory

CP mediates virtual machine access to real devices

- Virtual machine I/O requests are intercepted by CP
- CP examines and validates I/O requests

For access to minidisks:

- DEFINE EXTENT channel command inserted into real channel program
 - Limits I/O to minidisk cylinder range

For R/O device access:

- CP inserts command to disable write operations



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Virtual Machine Isolation

Without special privilege, a virtual machine may not:

- Obtain higher privileges than assigned by system administrator
- Disable CP system resource access control
- Obtain control of CPU in real supervisor state
- Circumvent CP isolation of real memory
- Access disk extent limitations imposed by CP
- Disrupt another virtual machine



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Summary

SIE instruction permits partitioning of CPC

- z/VM relies on SIE to dispatch virtual machines

PR/SM ensures resources are cleared on context switch

- Memory is assigned to an LPAR, cleared when reassigned

zSeries networks are memory-to-memory copy

- Cannot be sniffed

CP maintains virtual machine isolation

- Access to real devices are checked
- General users may not compromise CP or other virtual machines



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Securing z/VM



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Topics to be Covered

ibm.com

Virtual machine privileges and CP commands

Authentication and authorization

z/VM user directory

Directory Maintenance Facility (DirMaint)

External Security Manager

z/VM system configuration file

Securing virtual networks



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Virtual Machine Privileges Classes

IBM defined user classes:

- System operator (A)
 - Controls z/VM system
- System resource operator (B)
 - Controls real resources
- System programmers (C)
 - Updates / changes system-wide parameters
- Spooling operator (D)
 - Control spool files
- System analyst (E)
 - Examines / saves system operation in z/VM storage
- Service representative (F)
 - Examines system details (reserved for IBM use)
- General user (G)


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Privilege Class and CP Commands

Access to CP commands is based on user privilege class

- Attempts to execute unauthorized commands are ignored

Privilege class maintains system integrity

- No class G command can affect CP or other virtual machines
- System operator is usually class A, B, and D

Important!

- Class A, B, or C users can bypass security controls
- Define Linux virtual machines no higher than class G

Privilege class may be altered

- Classes I-Z and 1-6 are available for customer use
- Use CP **MODIFY COMMAND**


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Authentication and Authorization

Userid and password must be provided for access

- login, ftp, nfs, rexec
- Anonymous access is possible but must be explicitly granted

Authorization:

- Is based on VM user (class and user directory entry)
- Can be supplemented by External Security Manager (ESM)
 - RACF for instance

Be aware!

- Linux login is distinct from VM login
- Console login does not hide Linux password!
 - Limit console login for Linux guests
 - Watch out for shoulder surfing



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM User Directory

Defines virtual machine to CP

- User privileges
- Virtual devices

Statements to consider:

- USER
- LOGONBY
- MDISK
- LINK
- IUCV
- OPTION



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM User Directory Entry

```
USER USER1 USER1PW 128M 1G G
  INCLUDE IBMDFLT
  IPL 190 PARM AUTOOCR
  LOGONBY USER2
  MACHINE ESA
  MDISK 0191 3390 21 5 440U1R
  MDISK 0201 3390 26 1087 440U1R WR READ WRITE MULTIPLE
  MDISK 0202 FB-512 V-DISK 5000 WV
  LINK TCPMAINT 592 592 RR
```

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Directory Maintenance Facility (DirMaint)

Provides secure facility to maintain user directory

- Command interface provides error checking

Restricts directory maintenance to authorized users

- Authorized users can be tailored for specific commands

Directory maintenance logging

- Commands to service machines are auditable
- Some commands execute in caller's virtual machine

Supports External Security Manager

- Password control
- Minidisk access

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

External Security Managers

Enhances auditing, authentication, access control

Encrypts user passwords

Use Access Control List for minidisk access

- Instead of minidisk password

RACF/VM is packaged with z/VM 4.4

- License required to enable



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

The SYSTEM CONFIG File

Some statements to consider:

- DEFINE LAN / DEFINE VSWITCH / VMLAN
 - Defines virtual networks and sets global attributes for VM Guest LANs
- PRIV_CLASSES
 - Changes privilege class authorization to CP commands
- SYSTEM_USERIDS
 - Specifies special userids (operator, accounting, etc.)
- FEATURES CLEAR_TDISK
 - Automatically clears TDISK on initialization and when detached
- FEATURES PASSWORDS_ON_COMMANDS
 - Controls password prompting for AUTOLOG, LINK, and LOGON



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

z/VM Virtual Networks

Access to virtual networks can be restricted to specific users

- VM Guest LAN
 - CP DEFINE / MODIFY LAN command
- VSWITCH
 - CP DEFINE VSWITCH command
- Virtual CTC
 - CP DEFINE CTC command / SPECIAL user directory entry

Be Aware!

- Any user can create and connect to TRANSIENT Guest LAN
- To prevent TRANSIENT Guest LAN creation,
 - Use VMLAN statement in SYSTEM CONFIG:

VMLAN LIMIT TRANSIENT 0



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Recommendations

Assign z/VM users to lowest possible privilege class

- Define Linux virtual machines no higher than class G
- Avoid root login to Linux guest from console if possible

Use DirMaint to maintain user directory

Use an External Security Manager such as RACF

Restrict access to system-owned virtual networks

- Remove unneeded IUCV ANY or IUCV ALL directory statements
- Use RESTRICTED Guest LAN - grant authorization as needed
- Prohibit TRANSIENT Guest LAN using VMLAN statement in SYSTEM CONFIG
- Restrict virtual CTC to specific users
 - DEFINE CTC command, or SPECIAL user directory entry



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux User Access and Authentication



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Topics to be Covered

Validating RPM packages

User and group management

Pluggable Authentication Modules (PAM)

Discretionary access control

Access control lists and extended attributes

Delegating authority root authority with sudo



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Validating RPM Packages

RedHat Package Manager (RPM)

- Manages software packaging for RedHat and SuSE
- Many open source applications use RPM
 - Be cautious when installing packages from Internet!

Check RPM GPG signatures:

```
# rpm -K tripwire-1.2-385.s390.rpm
tripwire-1.2-385.s390.rpm: md5 gpg OK
```

To verify one executable:

```
# rpm -vf /usr/sbin/tripwire
```

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux User and Group Management

Users defined in /etc/passwd

- Passwords are hashed, not encrypted
 - Hashed values stored in /etc/shadow
 - 13 characters in length

Default user attributes set in /etc/login.defs

- Password aging, \$PATH, login failures, etc.

To disable login by a specific user:

- Change user password entry in /etc/shadow to 1 character

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Superuser Account

Superuser (root) is identified by UID of 0 in /etc/passwd

- Multiple superusers are permitted
 - Simply specify UID = 0

Recommendations for superuser:

- Never login as superuser
 - Use `su` or `sudo` when authority is required
 - Use SSH for remote access
- Exercise caution when executing as superuser
 - Do what is needed, then logout
- Never include `.` in PATH
 - Know what is being executed
 - Use absolute paths


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Restricting Access With /etc/securetty

Limit root login to the console

- /etc/securetty identifies secure devices for root access

Recommendations:

- Include nothing more than system console (ttyS0)
 - If file exists but is empty, no root login is permitted anywhere!
- Exclude pseudo-terminals (pts*)
 - Use `who` command to see login terminals:

```
# who
root      ttyS0      Sep 12 08:39
geiselha pts/1      Sep 12 11:46 (greg01.itso.ibm.com)
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Discretionary Access Control (DAC)

Linux uses Discretionary Access Control (DAC)

- Access to files is granted based on user and file ownership
 - Permissions based on file owner (u), group (g), and others (o)
- Owner has complete control over file access
- Only two types of users: owner / superuser

Groups offer some level of granularity, but:

- Adding users to groups can introduce vulnerabilities
- Managing groups is cumbersome


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

The SUID, SGID, and Sticky Bits

SUID (set UID)

- Executes file with authority of file owner
 - Set using `chmod u+s filename` command
- Dangerous but sometimes necessary:


```
-rwsr-xr-x 1 root shadow 67072 Nov  5 2002 /usr/bin/passwd
```

SGID (set GID)

- Executes file with authority of file group
 - Set using `chmod g+s filename` command
- If set on a directory, new files / subdirectories inherit group ownership

Sticky bit (Save Text)

- Only meaningful for directories: `chmod +t /tmp`
- Designed for world-writable directories
 - Files can only be deleted by owner (and superuser)


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Some Useful find Commands

Find all SETUID / SETGID files:

```
find / -type f -perm +6000 -ls
```

Find all world-writable files:

```
find / -perm +2 ! -type l -ls
```

Find all files not belonging to a group or user:

```
find / -nouser -og -nogroup
```

Find all files changed in last 24 hours on ext3 filesystems:

```
find / -fstype ext3 -ctime -1 -ls
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Restricting SUID / SGID

/dev/dasda1	/	ext3	defaults	1 1
/dev/dasdb1	/home	ext3	defaults, nosuid	1 1
/dev/dasdc1	swap	swap	pri=42	0 0
devpts	/dev/pts	devpts	mode=0620,gid=5	0 0
proc	/proc	proc	defaults	0 0

Other useful flags:

noexec

– No executable files allowed on filesystem

nodev

– No device files allowed on filesystem


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Access Control Lists

ACLs provide finer granularity for granting file permissions

- Arbitrary users / groups can be granted / denied file access
- User, group, and other are part of the ACL

Enabled ACLs on a filesystem basis

- ACLs support is included in SuSE 8

Use `getfacl` command to view ACLs

Use `setfacl` command to change ACLs


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Enabling Access Control Lists

Configured in `/etc/fstab` file

```

/dev/dasda1 /          ext3      defaults,acl      1 1
/dev/dasdb1 swap          swap      pri=42             0 0
devpts      /dev/pts     devpts    mode=0620,gid=5    0 0
proc        /proc        proc      defaults            0 0

```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Setting Access Control Lists

```
# ls -l
drwxr-s---  2 root    wheel           4096 2003-09-16 03:47 acls
# setfacl -m mask:rw acls
# ls -l
drwxr-x---+  2 root    wheel           4096 2003-09-16 03:47 acls
# getfacl acls
# file: acls
# owner: root
# group: wheel
user::rwx
group::r-x
mask::r-x
other:---
# setfacl -n -m user:user1:rw acls
# getfacl acls
# file: acls
# owner: root
# group: wheel
user::rwx
user:user1:rwx           #effective:r-x
group::r-x
mask::r-x
other:---
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Extended Attributes

EAs provide additional layer on file permission

- To set EAs, use `chattr` command
- To view EAs, use `lsattr` command

Attributes:

- Immutable (*i*)
 - File may not be modified, removed, or linked
- Append only (*a*)
 - File may only appended to
- Secure delete (*s*)
 - On delete, zeroed blocks are written back to disk


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Using Extended Attributes

```
# chattr +i test1
# chattr +a test2
# lsattr
----i----- ./test1
-----a----- ./test2
# su - user1
$ echo "Try to overwrite">test2
-bash: test2: Operation not permitted
$ echo "Try to append">>test2
$ rm test1
rm: remove write-protected regular file `test1'? y
rm: cannot remove `test1': Operation not permitted
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

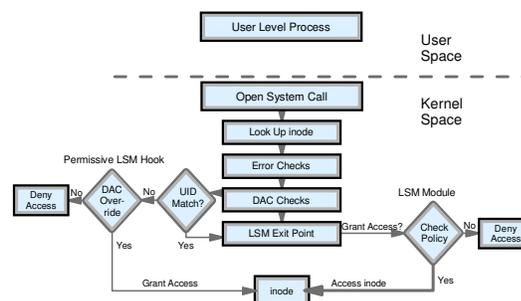
Linux Security Module (LSM)

Mandatory Access Control (MAC)

- Access control is removed from users
- Security administrator assigns access control

LSM is a MAC-based security framework

- Operates at the kernel level
- Suggested by Secure-Enhanced Linux (SELinux)
- Possible inclusion in 2.6 kernel


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

User Authentication Using PAM

Pluggable Authentication Modules (PAM)

- Standard method for Linux authentication

With PAM, you get:

- Easily authentication configuration
- Standard authentication methods
- Ability to individually configure authentication for PAM-aware applications

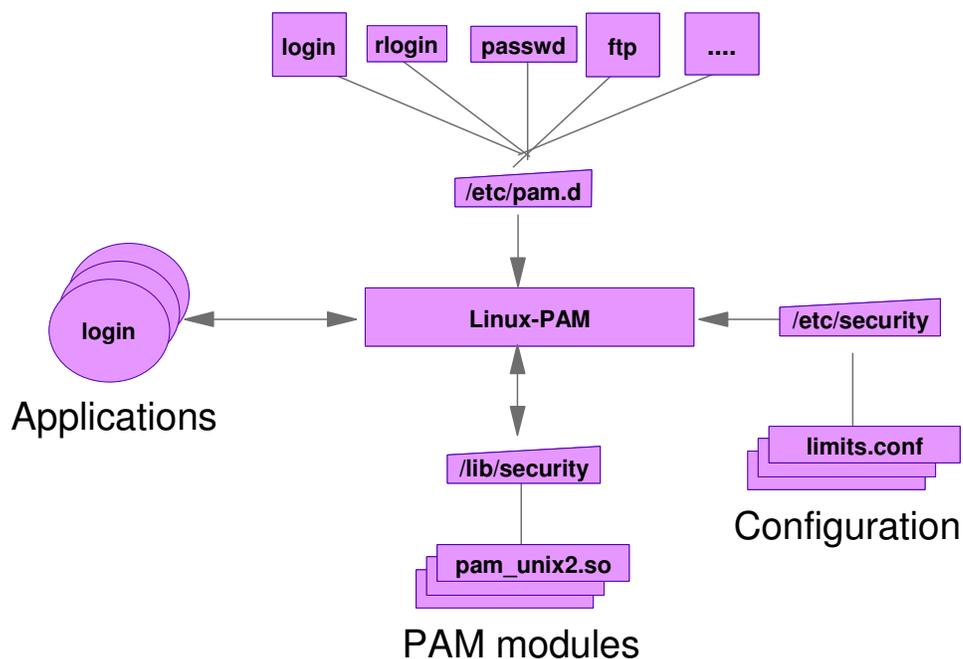


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Illustrating PAM Operation

ibm.com



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

PAM Configuration Files

Found in /etc/pam.d directory

- 1 file per PAM-aware application

Specifies:

- Type of authentication check (service type)
- How check return is interpreted (module type)
- Which module performs checking (PAM module)
- Optional arguments passed to PAM module

Multiple modules may be specified

- 'Module stacking' - modules are executed in order

Be careful when modifying configuration!

- Could be locked out


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Default /etc/pam.d/login

```

auth requisite      pam_unix2.so      nullok
auth required      pam_securetty.so
auth required      pam_nologin.so
auth required      pam_env.so
auth required      pam_mail.so
account required   pam_unix2.so
password required  pam_pwcheck.so    nullok
password required  pam_unix2.so      nullok use_first_pass use_authtok
session required   pam_unix2.so      none
session required   pam_limits.so

```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

PAM Service Types

auth

- Prompts for and authenticates passwords

account

- Checks user account specifics
 - Password aging, permitted login times, permitted login terminals etc

passwd

- Updates passwords

session

- Sets environment variables



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

PAM Module Types

required

- Module must return success
 - Next stacked module is always executed

requisite

- Module must return success
 - On failure, status is returned to application

sufficient

- If module succeeds, access is granted
 - On failure, next stacked module is executed

optional

- Success is optional, execution continues
 - Unless it is only module in stack



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

PAM Arguments

Arguments are passed to PAM modules

- Global arguments may be provide in /etc/security

Some common arguments:

- debug
 - Print debug messages to syslog
- audit
 - Lots of messages to syslog
- use_first_pass
 - Use password from previous module - no additional prompting
- try_first_pass
 - Use password from previous module but prompt for another password if attempt fails

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Some Commonly Used PAM Modules

pam_unix.so / pam_unix2.so

- Traditional password checking module

pam_cracklib.so

- Password checking module

pam_limits.so

- Limits system resources available globally or based on user

pam_listfile.so

- Can limit access based on configuration file
 - Limit by user
 - Limit by user

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Restricting root Access With pam_listfile

To prevent SSH access by root

- Add as first line in /etc/pam.d/sshd:

```
auth required pam_listfile.so onerr=fail item=user sense=deny \  
file=/etc/security/nossh.conf
```

- Create /etc/security/nossh.conf:

```
root
```

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Setting Variables Using pam_env Module

Global variables can be set in /etc/security/pam_env.conf

- To set the DISPLAY environment variable, add:

```
REMOTEHOST  DEFAULT=localhost  OVERRIDE=@{PAM_RHOST}  
DISPLAY     DEFAULT=${REMOTEHOST}:0.0  OVERRIDE=${DISPLAY}
```

- Enable pam_env in /etc/pam.d/sshd

```
auth required  pam_env.so
```

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Delegating Root Authority With sudo

Use sudo to safely delegate root authority

- Non-root user can be authorized to execute specific commands
- Configuration in /etc/sudoers
 - Configured using `visudo` command

Advantages:

- Provides granular control for specific users to specific commands
- Provides extensive logging (auditable)

Use cautiously!

- Delegated user can escape to shell as root in some cases
 - For instance, using sudo for `vi`



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Recommendations

Check signatures when installing new RPM packages

Exercise caution when executing as superuser

- Never login as root
 - Login on non-root id, then `su` to root (or use sudo)
- Never include current directory (`.`) in PATH

Restrict access to sensitive files / executables

- Avoid world-writable files,
- Restrict SUID / SGID executables
- Use Access Control Lists if possible
- Enable Extended Attributes as a second line of defense

Use PAM to customize authentication



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Monitoring and Hardening Linux



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Objectives

ibm.com

Securing system log files

Monitor changes to critical system files

Securing network services

Secure network access using SSH

Hardening Linux using Bastille

Network firewalls



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux System Logging

System logging is performed by syslogd daemon

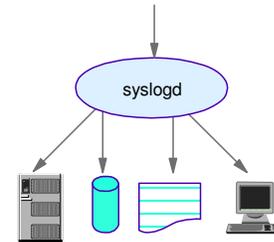
- Accepts incoming log messages from:
 - Daemons, applications, kernel
- Messages are directed to locations indicated in /etc/syslog.conf
 - File, console, remote host, logged-in users, named pipe (FIFO)

Messages identified by:

- Facility - who generated message
- Level - how important is the message

Log to centralized log server for

- Simplify log management
- Increase security
 - Harden central log server


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Sample syslog.conf

```

# write to console
kern.*;*.warn;news.emerg;mail.alert;authpriv.none /dev/console
# write to FIFO
kern.*;*.warn;news.err;mail.err;authpriv.none |/dev/xconsole
# write to all logged in users
*.emerg *
# write to remote host
*.* @greg01
# write to file
mail.* -/var/log/mail
news.crit -/var/log/news/news.crit
news.err -/var/log/news/news.err
news.notice -/var/log/news/news.notice
*.*;mail.none;news.none -/var/log/messages
local0,local1.* -/var/log/localmessages
local2,local3.* -/var/log/localmessages
local4,local5.* -/var/log/localmessages
local6,local7.* -/var/log/localmessages
  
```

To test logging, use logger command:

```
logger "This is a test"
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Monitoring Failed Login Attempts

Login failures are recorded in /var/log/faillog

- To view login failures, use `faillog` command:

```
# faillog
Username  Failures  Maximum  Latest
root      1         0        Mon Sep 22 11:55:04 -0700 2003 on 2
geiselha  0         0        Mon Sep 22 12:03:30 -0700 2003 on 2
```

- To set maximum failed login attempts:

```
# faillog -u geiselha -m 3
```

- To reset a userid:

```
# faillog -r -u geiselha
```

Never limit failed logins for root


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Monitoring Changes to Files and Directories

File / directory changes may indicate an intrusion

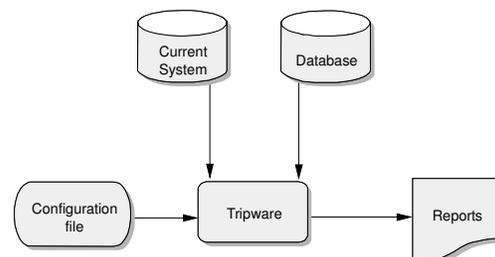
- Subversion of critical files, attempts to cover tracks

Tripwire - policy-based filesystem monitoring utility

- Policies are configured for directories and files
- Last state of monitored objects stored in encrypted database

Files can be fingerprinted based on:

- Permission
- Ownership
- File size
- Modification timestamp
- MD5 hash
-


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Securing Network Services

Most network services have inherent weaknesses

- Unencrypted traffic (telnet, ftp for example)
 - Passwords are exposed, sensitive data can be examined
- Default SuSE installation disables inetd services

Network services can be started from:

- Internet daemon (inetd)
 - Configured in /etc/inetd.conf
- Startup scripts
 - /etc/init.d/portmap, /etc/init.d/nfsserver

Default SuSE installation disables inetd services


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Controlling Incoming Network Connections

If inet services are required, use TCP wrappers

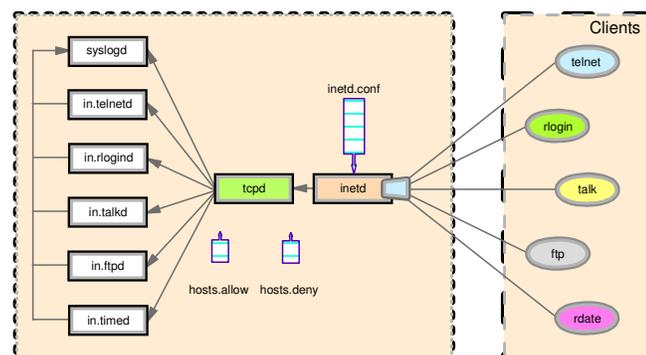
- Provides Access Control Lists for network services invoked from inetd
 - Invoked from /usr/sbin/tcpd

Access Control Lists

- /etc/hosts.allow
- /etc/hosts.deny

Disallow trusted hosts!

- /etc/hosts.equiv


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SSH - The Secure Shell

Access using telnet, ftp, r-commands is inherently insecure

- Data and passwords data are passed in cleartext
- Password challenges can be bypassed

Always use a secure SSH equivalent

- ssh for remote access, scp / sftp for file transfers

With SSH:

- Network data is always encrypted
- Data integrity is checked at the application level
 - CRC for SSH 1, MD5 for SSH 2
- Both host and users can be authenticated



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SSH Encryption Keys

SSH uses both private/public key and symmetrical encryption

- Public/private keys for host authentication and key/password exchange
- Symmetrical encryption for speed

Two public/private keys pairs:

- Host key to identify known SSH hosts
 - Generated by `ssh-keygen` command
 - Stored in `/etc/ssh/ssh_host_key` and `/etc/ssh/ssh_host_key.pub`
- Server key to prevent session playback
 - Generated hourly by `sshd`
 - Never stored on disk

Important!

- Fingerprint host key files
- Ensure `/etc/ssh/ssh_host_key` can only be read by root



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Remote Access with ssh

```

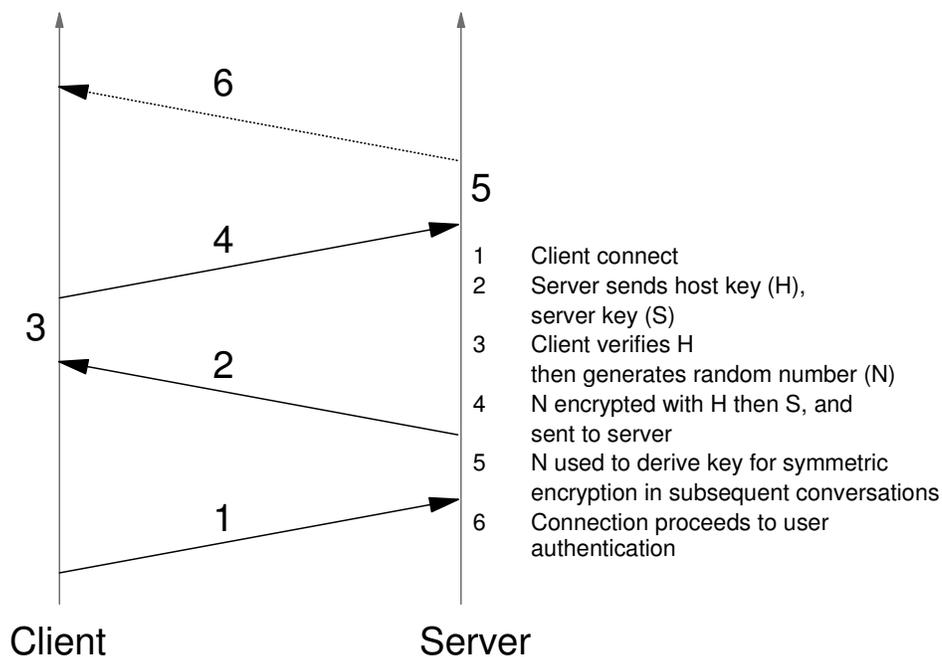
$ ssh geiselha@lnxsuse
The authenticity of host 'lnxsuse (9.12.9.34)' can't be established.
RSA key fingerprint is
c4:76:eb:82:5d:3a:5f:e1:19:fb:91:a9:5a:7d:13:bf.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'lnxsuse,9.12.9.34' (RSA) to the list of
known hosts.
geiselha@lnxsuse's password:
Last login: Wed Sep 24 12:30:40 2003 from greg01.itso.ibm.com
geiselha@lnxsuse:~> exit
logout
Connection to lnxsuse closed.
$ ls .ssh
known_hosts
$ cat .ssh/known_hosts
lnxsuse,9.12.9.34 ssh-rsa AAAAB3NzaC1yc2EAAAABIwAAAI.....
$

```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SSH Version 1 Host Authentication


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SSH User Authentication

User authentication configured in /etc/ssh/sshd_config

- 3 nonpassword user authentication mechanisms available:
 - Trusted host authentication (RhostsAuthentication)
 - Rhosts and RSA authentication (RhostsRSAAuthentication)
 - RSA authentication (RSAAuthentication)
- If enabled, each method is tried until success
 - Fallback is to challenge for user password

Only RSA authentication should be considered secure!

- Others are disabled by default in SLES 8


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Configuring RSA Authentication

```
$ ssh-keygen -b 1024 -C "geiselha@itso.ibm.com" -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/geiselha/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
identification has been saved in /home/geiselha/.ssh/id_rsa.
Your public key has been saved in /home/geiselha/.ssh/id_rsa.pub.
The key fingerprint is:
04:f5:d9:16:e1:93:ea:20:6d:81:2a:b0:03:9b:7a:9c geiselha@itso.ibm.com
$ scp .ssh/id_rsa.pub geiselha@lnxsuse:~/.ssh
authenticity of host 'lnxsuse (9.12.9.34)' can't be established.
RSA key fingerprint is c4:76:eb:82:5d:3a:5f:e1:19:fb:91:a9:5a:7d:13:bf.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'lnxsuse' (RSA) to the list of known hosts.
geiselha@lnxsuse's password:
id_rsa.pub      100% |*****|          231      00:00
$ scp .ssh/id_rsa.pub geiselha@lnxsuse:~/.ssh/authorized_keys
geiselha@lnxsuse's password:
id_rsa.pub      100% |*****|          231      00:00
$ ssh geiselha@lnxsuse
Enter passphrase for key '/home/geiselha/.ssh/id_rsa':
Last login: Wed Sep 24 12:21:35 2003 from greg01.itso.ibm.com
geiselha@lnxsuse:~>
```


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Hardening Linux Using Bastille

Bastille is a hardening tool for Linux

- Graphic user interface allows administrator to choose hardening level
 - Run `/usr/sbin/InteractiveBastille` command
- Consists of Perl modules tailored to specific security settings
 - File permissions
 - Account security
 - Boot security
 - Inetd security
 - ..

Note! Bastille distributed with SLES8 does not work

- Can be installed from source
 - Use Mandrake 1.3.0 source (from <http://rpmfind.net>)

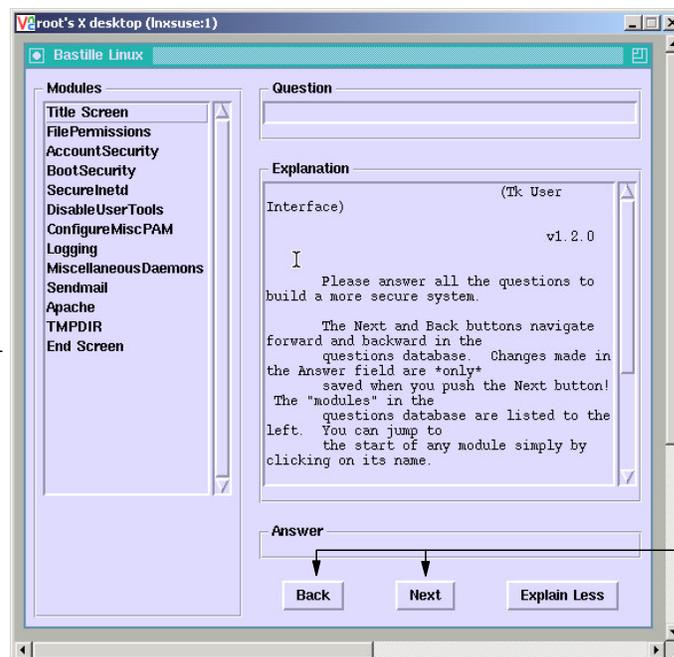

ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Bastille Main Menu

ibm.com

Modules

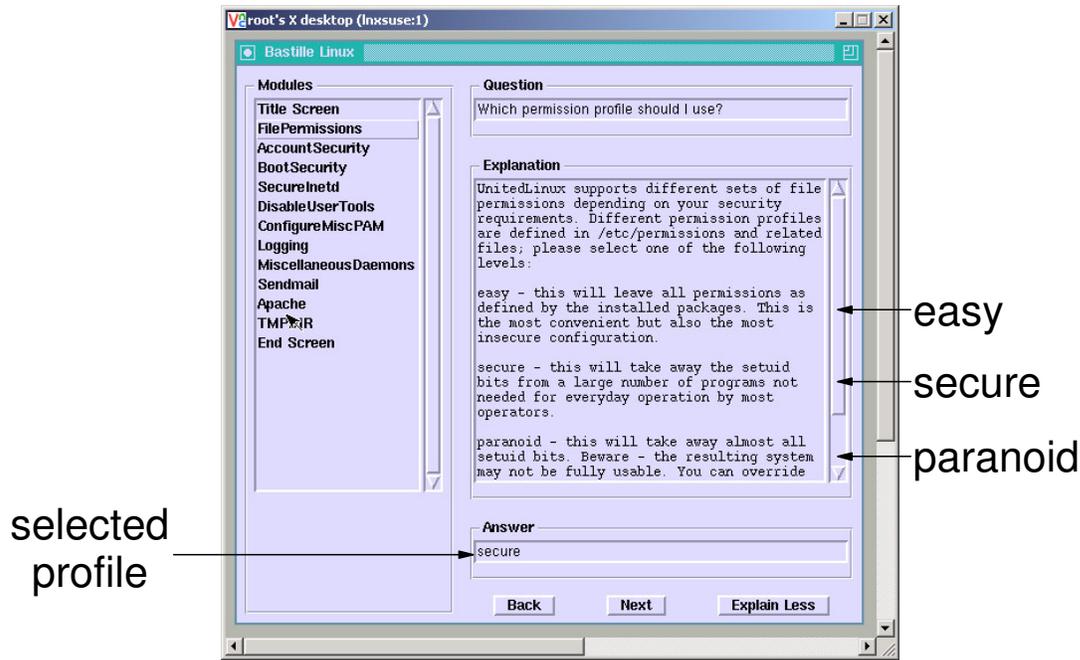


Navigation


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

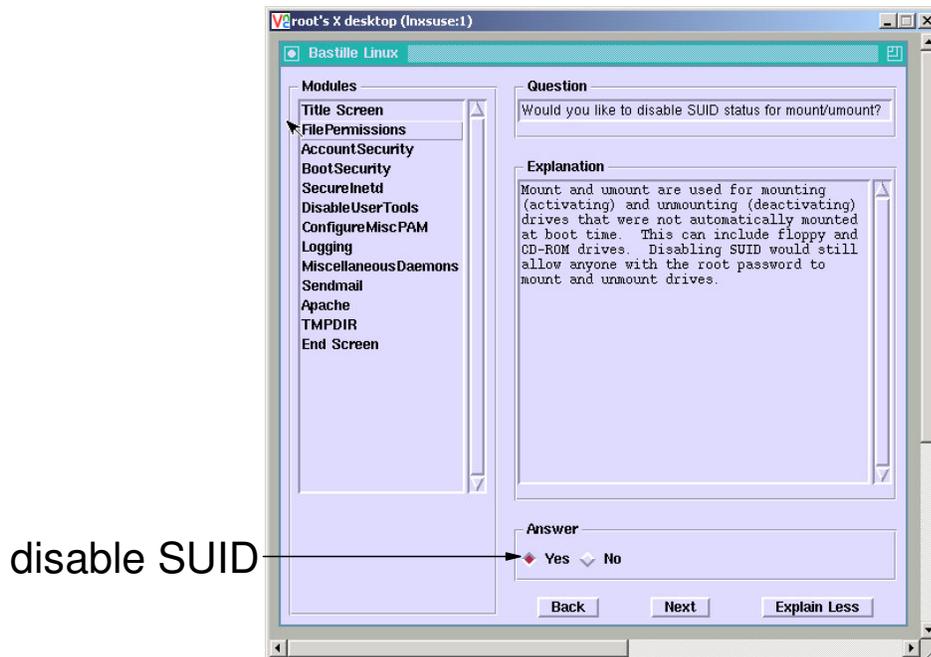
Bastille - File Permissions Profiles



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

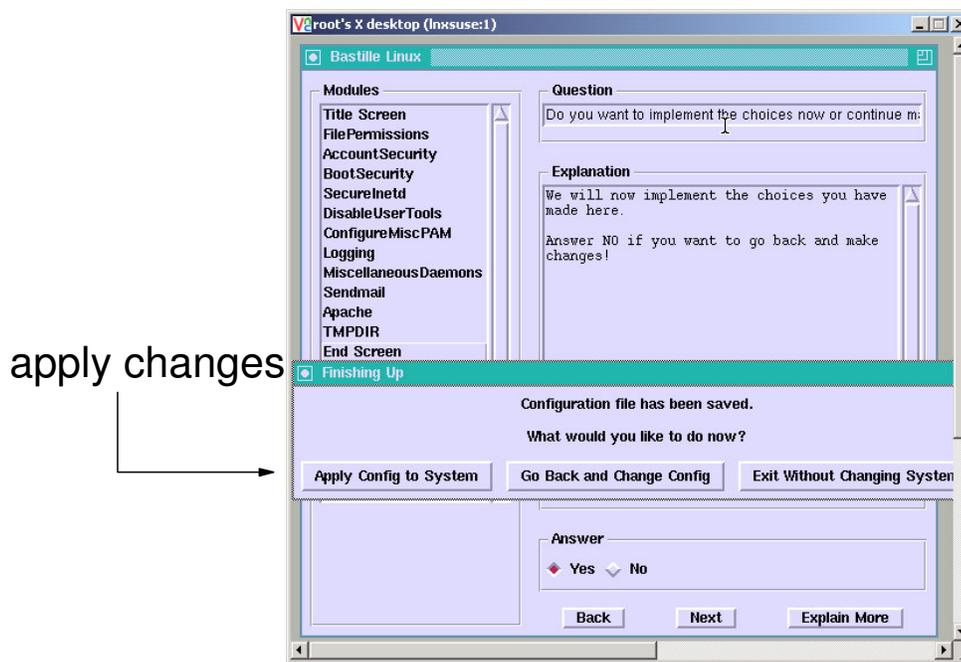
Bastille - SUID for Mount Command



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Bastille - Committing Changes


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Bastille - Recovering System Changes

Bastille remembers changes applied to system

- Actions recorded in `/var/log/Bastille/actions-log`

Bastille knows how to reverse actions:

- Execute `UndoBastille` command


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Recommendations

Use centralized system logging

Monitor critical files and directories using Tripwire

- Fingerprint critical files

Shutdown insecure networks services

- Use SSH as secure replacement for telnet, ftp, rexec
- If network server is required, use TCP wrappers to control access

Use SSH for network access and file transfers

- Use only password-based authentication

Use Bastille to tighten security settings



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Linux for zSeries Security Topics

Networks Firewalls



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Check Linux /proc Filesystem Settings

Some security settings:

- For non-routing hosts:
`echo '0'>/proc/sys/net/ipv4/ip_forward`
- To ignore ICMP echo requests:
`echo '1'>/proc/sys/net/ipv4/ip_echo_ignore_all`
- To ignore ICMP echo broadcast / multicast:
`echo '1'>/proc/sys/net/ipv4/ip_echo_ignore_broadcasts`

To set values on startup, use `/etc/sysctl.conf`



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

What is a Firewall?

Network security device

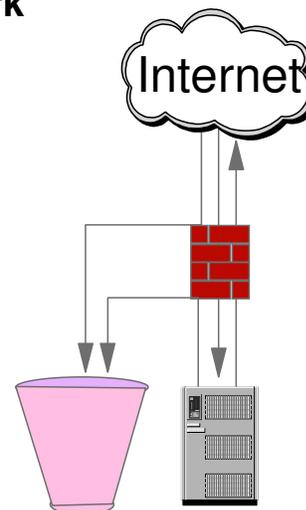
- Separates secure (internal) network from insecure (external) network

Restricts entry to / exit from internal network

- Firewall examines network traffic
 - Determines if traffic is acceptable
- IP traffic is packet-switched

Firewalls can be:

- Servers dedicated to policing traffic
- Multipurpose servers



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

How do Firewalls Work?

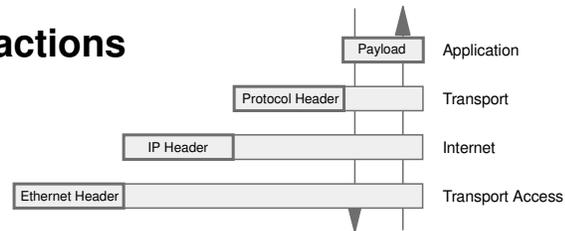
Examining packet headers

- Packets constructed / disassembled in TCP/IP stack

Firewalls examine packet headers for:

- Source IP address / port
- Destination IP address / port
- Protocol
- Flags
-

Filtering rules specify firewall actions



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Packet Filtering Firewall

Packets examined at network layer

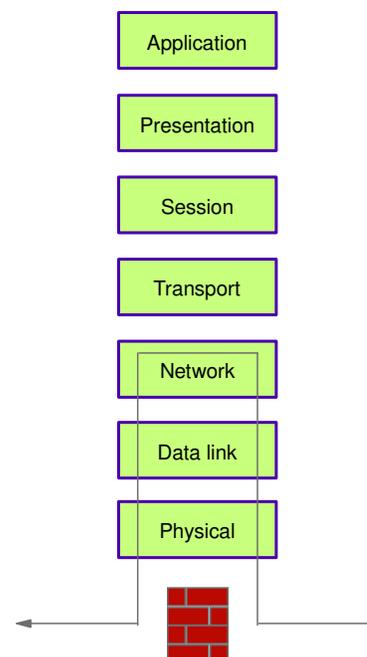
- Filtering looks at packet headers
 - Protocol, IP address, ports, flags, etc.
- Decisions to grant/deny based on access rules

Advantages:

- Fast when traffic is low

Disadvantages:

- Filtering easily fooled
 - Fragmentation, bogus header, spoofed IP addresses
- Cannot examine higher levels of protocol stack



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Proxy Firewall

Firewall running proxy server

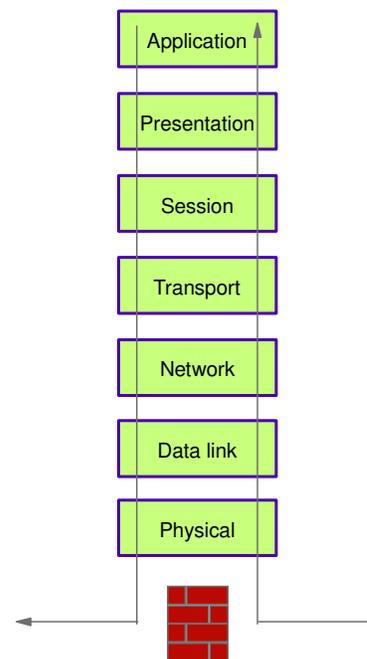
- Proxy initiates connection to server on behalf of client

Advantages:

- Offers more security
 - Packets pass through every protocol layer

Disadvantages:

- Overhead of initiating connections
- Limited number of supported protocols



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Stateful Packet Inspection

Improves on packet filtering

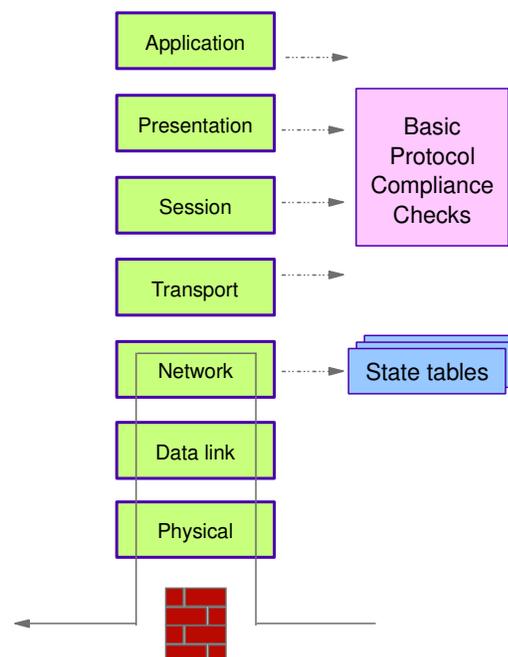
- State tables maintain connection information

Advantages

- Packets can be analyzed in context
 - Improves security

Disadvantages

- High overhead in maintaining state tables
- Limited ability to inspect higher layers



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Packet Filtering Using iptables

Native Linux packet filtering mechanism

- Packets may be examined at various routing points
 - Referred to as Chains

Rules are applied to Tables associated to Chains

- Rules specify actions to be taken when filtering criteria is met
 - Packets may be accepted, rejected, dropped, modified

Complex filtering rules can be built

- Requires expertise to design secure iptables firewall!



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Chains and Tables

Chains

- PREROUTING
- INPUT
- OUTPUT
- FORWARD
- POSTROUTING

Tables

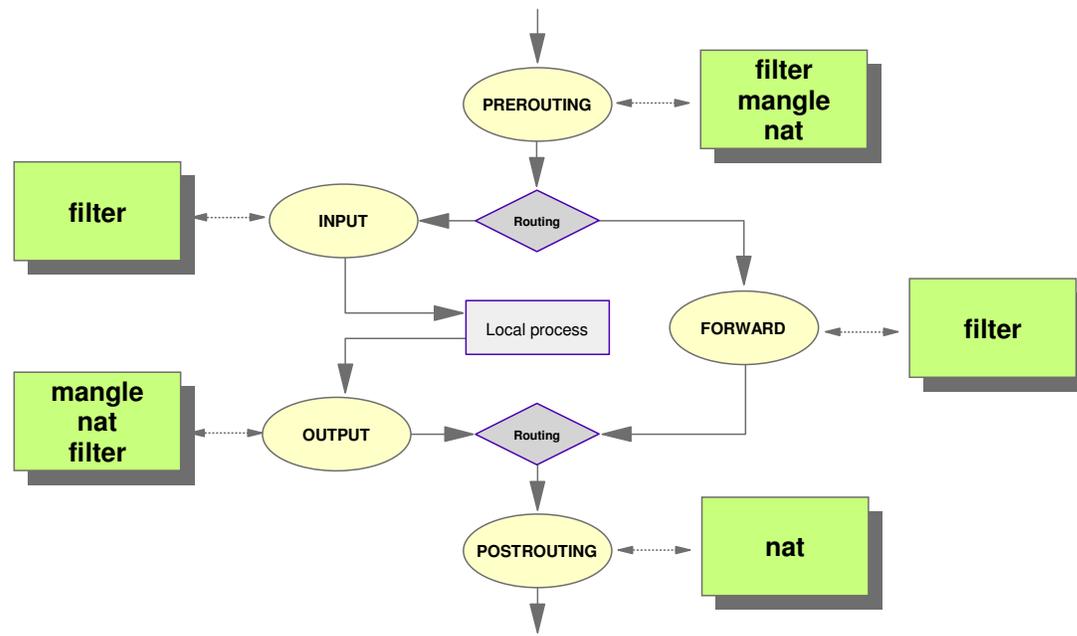
- filter
- mangle
- nat



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Traversing iptables Chains


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

SuSEfirewall2

SuSEfirewall2 is iptables rules generation tool

- Provided with SuSE 8
- Simplifies generating iptables rules
 - Rules are set from variables specified in /etc/sysconfig/SuSEfirewall2
- Custom rules can be added

Advantages:

- Relatively simple configuration
 - Commentary in config file
- Yast2 interface

Disadvantages:

- Default rules may be too simplistic for your installation
 - In-depth knowledge of iptables required to fully customize rules


ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

StoneGate Firewall

Distributed firewall on Linux for zSeries

- Commercial product from StoneSoft

StoneGate advantages:

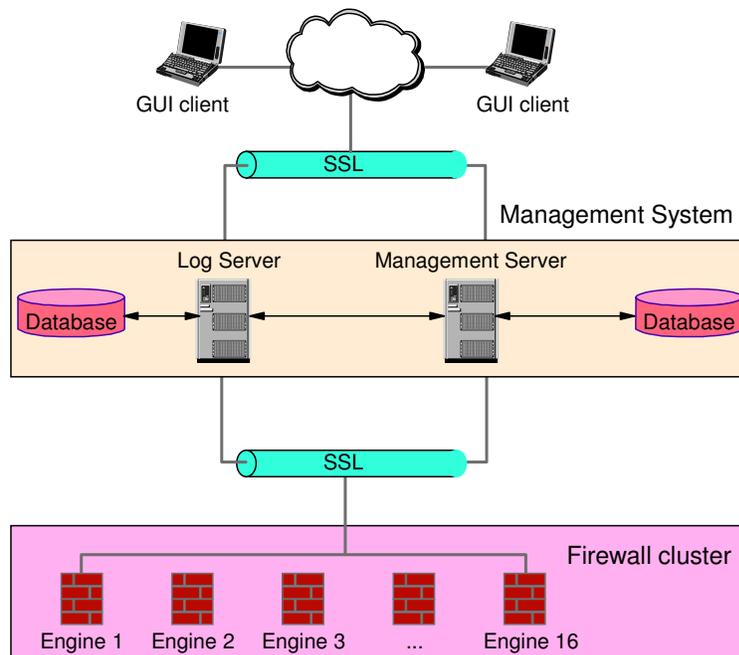
- Simple management system
 - Superuser
 - Editor
 - Operator
- High availability clustering
- Multi-Layer Inspection technology



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

StoneGate Architecture



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

StoneGate Components

Administration client

- Graphic interface to configure and administer StoneGate
 - Easily configure firewall
 - Manage users, authentication, network services
 - Monitor firewall, manage log data

Management System

- Management Server
 - Central administration point for enterprise StoneGate cluster
- Log Server
 - Manages log data from firewall nodes

Firewall engines

- Runs on hardened, Linux-based OS

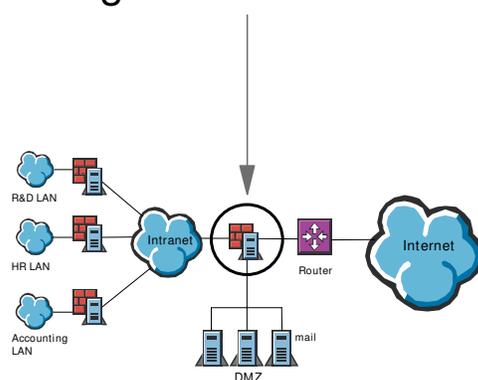


ibm.com/redbooks

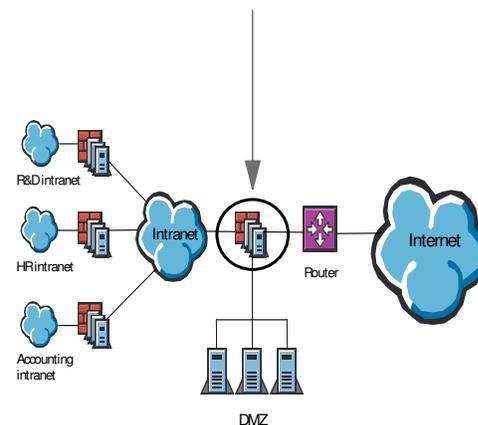
© Copyright IBM Corp. 2003. All rights reserved.

Firewall Clustering

Single Point of Failure



Clustered Firewall



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

StoneGate Multi-Layer Inspection

Uses stateful packet inspection

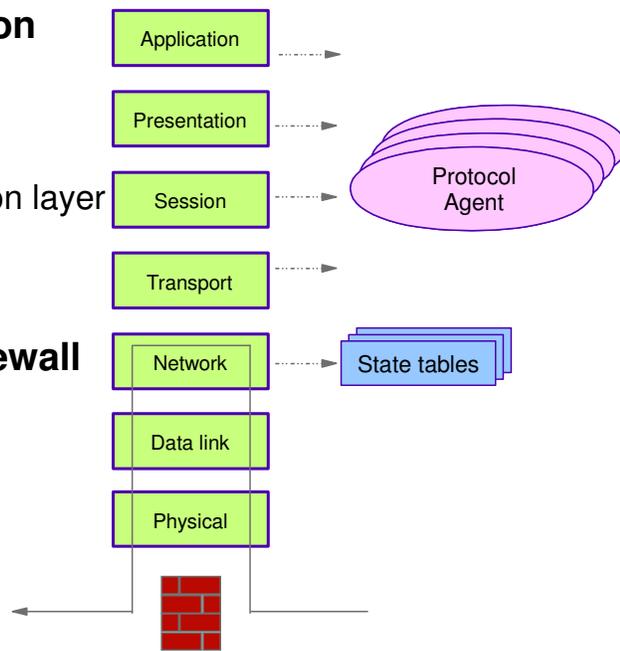
- Can act as simple packet filter

Uses protocol agents

- Inspect packets up to application layer
- Offers security of proxy firewall
 - Without overhead

More flexibility than proxy firewall

- Easily add new protocol agents



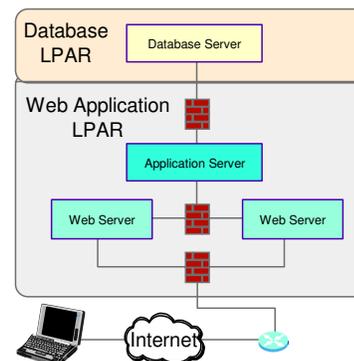
ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.

Deploying StoneGate on Linux for zSeries

Removes the need for external firewalls

- Between front-end and back-end servers
- Between zSeries and external network



ibm.com/redbooks

© Copyright IBM Corp. 2003. All rights reserved.