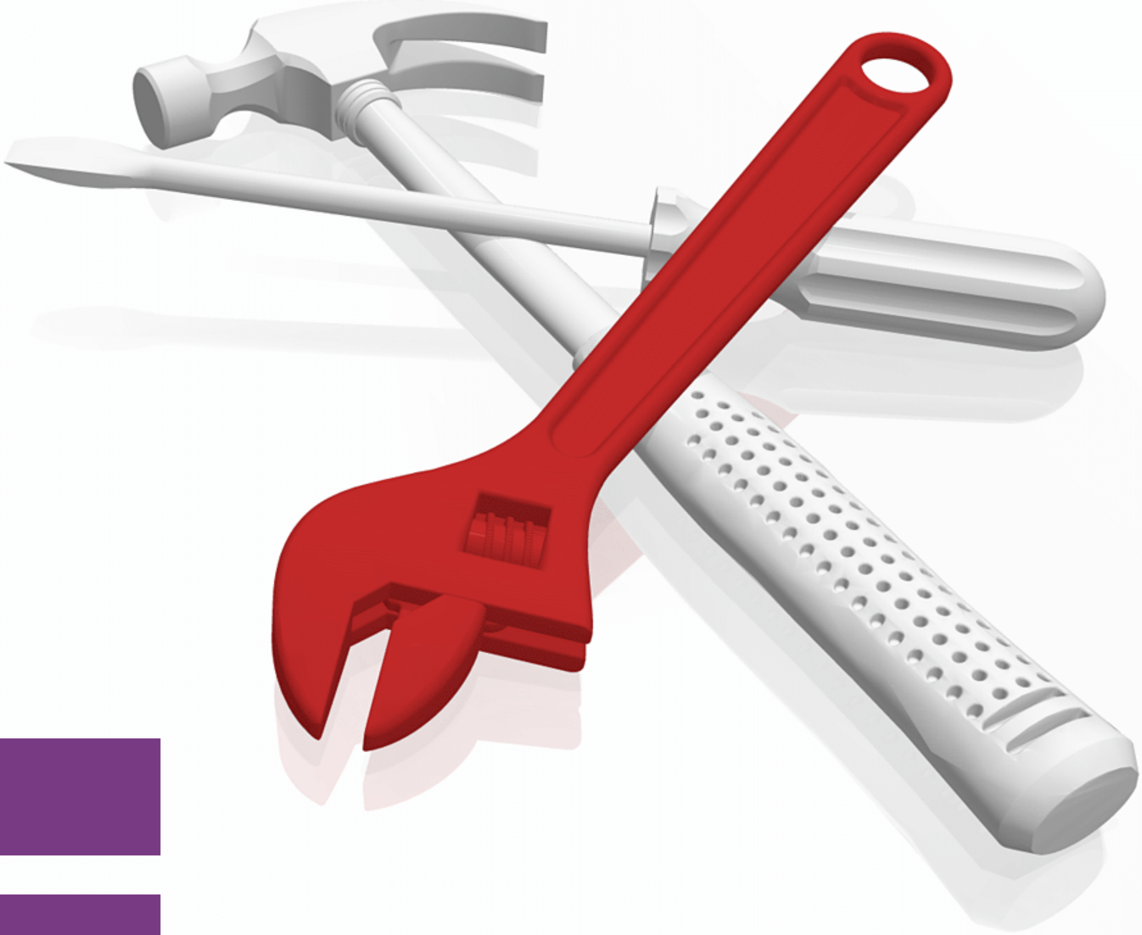
Redbooks

ibm.com/redbooks

IBM Power Security Catalog

Tim Simon
Felipe Bessa
Hugo Blanco
Carlo Castillo
Rohit Chauhan
Kevin Gee
Gayathri Gopalakrishnan
Samvedna Jha
Andrey Klyachkin
Andrea Longo
Ahmed Mashhour
Amela Peku
Prashant Sharma
Vivek Shukla
Dhanu Vasandani
Henry Vo



IBM Power

Security

IBM

Redbooks

IBM Redbooks

IBM Power Security Catalog

February 2024

SG24-8568-00

Note: Before using this information and the product it supports, read the information in “Notices” on page ix.

First Edition (February 2024)

This edition applies to:
Hardware Management Console (HMC) 10.1 or later
PowerVM Virtual I/O Server (VIOS) 3.1.3 and Virtual I/O Server 3.1.4
PowerVM Virtual I/O Server 4.1.0.10 or later
AIX 7.2 TL4 or later, and AIX 7.3 TL0 or later
IBM i 7.3 TR13, IBM i 7.4 TR7, and IBM i 7.5 or later
Red Hat Enterprise Linux (RHEL) Server 8.4 (ppc64le)
Red Hat Enterprise Linux Server 9.2 (ppc64le)
SUSE Linux Enterprise Server 15 Service Pack (SP) 5 (ppc64le)
Ubuntu 22.04.3 (ppc64le) or later
Red Hat OpenShift 4.14 or later
PowerSC 2.0 or later

Contents

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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 grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions 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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

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

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate 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. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at https://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
Db2®
DS8000®
FlashCopy®
GDPS®
Guardium®
HyperSwap®
IBM®
IBM Automation®
IBM Cloud®
IBM FlashSystem®
IBM Instana™
IBM Security®
IBM Z®
Instana®
POWER®
Power Architecture®
Power8®
Power9®
PowerHA®
PowerPC®
PowerVM®
QRadar®
Redbooks®
Redbooks (logo) ®
Satellite™
SystemMirror®
Tivoli®
WebSphere®
X-Force®
z/OS®
z/VM®

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

ITIL is a Registered Trade Mark of AXELOS Limited.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, 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 Oracle and/or its affiliates.

Red Hat, Ansible, Ceph, Fedora, JBoss, OpenShift, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

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

VMware, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

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

Preface

IT security is paramount in today's digital age. As businesses increasingly rely on technology to operate, protecting sensitive data and preventing cyberattacks is a priority. Cloud adoption introduces more security risks, which include data breaches and loss of access. A strong IT security infrastructure safeguards customer information, financial data, intellectual property, and overall business operations. By investing in robust security measures, organizations can mitigate risks, maintain trust with customers, and help ensure business continuity.

A multi-layered security architecture is essential for protection. Key areas to focus on include the following items:

- Hardware-level security: Prevent physical tampering and help ensure data integrity.
- Virtualization security: Isolate environments and control resource access.
- Management tool security: Secure hardware and cloud resources.
- Operating system security: Continuously update for robust security.
- Storage security: Protect data at rest and in transit.
- Networking security: Prevent unauthorized access and data breaches.

This IBM Redbooks® publication describes how the IBM Power ecosystem provides advanced security capabilities at each of these layers. IBM Power servers are designed with security as a core consideration.

At the hardware level, advanced technology includes tamper-resistant features that are built in to the processor to prevent unauthorized access and modifications, secure cryptographic engines to provide strong encryption of data, and Trusted Boot to help ensure that only authorized software components are loaded during system startup.

At the virtualization level, the hypervisor, which manages virtual machines (VMs), is designed to be secure and resistant to attacks. The hypervisor isolates workloads within a single physical server, which enables secure resource sharing within your infrastructure. The Hardware Management Console (HMC) provides centralized management and control of Power servers in a secure manner.

The operating systems that run on IBM Power servers (IBM AIX®, IBM i, and Linux on Power) offer robust security features, which include user authentication, access controls, and encryption support. Also, tools such as IBM PowerSC provide a comprehensive security and compliance solution that helps manage security policies, monitor threats, and enforce compliance.

Security also requires solid management and control. This book describes best practices such as conducting regular security audits, keeping operating systems and applications up to date with the latest security fixes, and implementing strong user authentication and authorization policies. Other critical elements include the implementation of data encryption for both data at rest and in transit, and strong network security processes that use firewalls, intrusion detection systems (IDS), and other security measures.

By combining these hardware, software, and management practices, IBM Power provides a robust foundation for security in your IT environment.

Authors

This book was produced by a team of specialists from around the world working at IBM Redbooks, Poughkeepsie Center.

Tim Simon is an IBM Redbooks Project Leader in Tulsa, Oklahoma, US. He has over 40 years of experience with IBM®, primarily in a technical sales role working with customers to help them create IBM solutions to solve their business problems. He holds a BS degree in Math from Towson University in Maryland. He has extensive experience creating customer solutions by using IBM Power, IBM Storage, and IBM Z® throughout his career.

Felipe Bessa is an IBM Brand Technical Specialist and Partner Technical Advocate for IBM Power. He works for IBM Technology in Brazil and has over 25 years of experience in the areas of research, planning, implementation, and administration of IT infrastructure solutions. Before joining IBM, he was recognized as a Reference Client for IBM Power Technologies for SAP and SAP HANA, IBM PowerVC, IBM PowerSC, Monitoring and Security, IBM Storage, and the Run SAP Like a Factory (SAP Solution Manager) Methodology. He was chosen as an IBM Champion for IBM Power for 2018 - 2021.

Hugo Blanco is an IBM Champion who is based in Madrid. He has been working with Power servers since 2008. He began his career as an instructor and has since taken on various roles at SIXE, which is an IBM Business Parter, where he gained extensive experience across different roles and functions. Hugo is deeply passionate about AIX, Linux on Power, and various cybersecurity solutions. He has contributed to the development of several IBM certification exams and actively participates in Common Iberia, Common Europe, and TechXchange. He enjoys delivering technical talks on emerging technologies and real-world use cases.

Carlo Castillo is a Client Services Manager for Right Computer Systems (RCS), an IBM Business Partner, and Red Hat partner in the Philippines. He has over 30 years of experience in pre-sales and post-sales support; designing full IBM infrastructure solutions; creating pre-sales configurations; performing IBM Power installation, implementation, and integration services; providing post-sales services and technical support for customers, and conducting presentations at customer engagements and corporate events. He was the first IBM-certified IBM AIX Technical Support engineer in the Philippines in 1999. As training coordinator during RCS' tenure as an IBM Authorized Training Provider 2007 - 2014, he administered the IBM Power curriculum, and conducted IBM training classes covering AIX, PureSystems, IBM PowerVM®, and IBM i. He holds a degree in Computer Data Processing Management from the Polytechnic University of the Philippines.

Rohit Chauhan is a Senior Technical Specialist with expertise in IBM i architecture. He works at Tietoevry Tech Services, Stavanger, Norway, which is an IBM Business Partner and one of the biggest IT service providers in the Nordics. He has over 12 years of experience working on the IBM Power platform with design, planning, and implementation of IBM i infrastructure, which includes high availability and disaster recovery (HADR) solutions for many customers during this tenure. Before his current role, Rohit worked for clients in Singapore and the UAE in the technical leadership and security role for the IBM Power domain. He possesses rich corporate experience in designing solutions, implementations, and system administration. He is a member of Common Europe Norway with strong focus on the IBM i platform and security. He is recognized as an IBM Advocate, Influencer, and Contributor for 2024 through the IBM Rising Champions Advocacy Badge program. He holds a bachelor’s degree in Information Technology. He is an IBM certified technical expert and also holds an ITIL CDS certificate. His areas of expertise include IBM i, IBM HMC, security enhancements, IBM PowerHA®, systems performance analysis and tuning, Backup Recovery and Media Services (BRMS), external storage, PowerVM, and solutions to customers for the IBM i platform.

Kevin Gee has over 30 years of IT experience, mostly with IBM technology solutions in support, systems engineering, consulting, and technical sales roles. He has broad experience in server and storage infrastructure and performance, HADR, enterprise backup and recovery, and product development and go-to-market strategy. Kevin has authored hundreds of white papers, training guides, and presentations for clients. He has co-authored IBM Redbooks publications, and helped develop certification exams for the AIX, PowerHA for AIX, and PowerVC products. He frequently presents at IBM technical conferences and contributes content to others' presentations. He has also published research into machine learning, graph processing, and natural language processing. Kevin holds a master's degree in Computer Science from The University of Texas at Arlington and bachelor's degrees in Computer Science and Spanish from Brigham Young University. He has been an IBM Champion since 2019.

Gayathri Gopalakrishnan works for IBM India and has over 22 years of experience as a technical solution and IT architect. She works primarily in consulting. She is a results-driven IT Architect with extensive working experience in spearheading the management, design, development, implementation, and testing of solutions.

Samvedna Jha is a Senior Technical Staff Member in the IBM Power organization, Bengaluru, India. She holds a masters degree in Computer Application and has more than 20 years of work experience. In her role as a Security Architect, IBM Power, she has a worldwide technical responsibility to handle the security and compliance requirements of Power products. Samvedna is a recognized speaker in conferences, has authored blogs, and published disclosures. She is also the security focal point for the Power products secure release process.

Andrey Klyachkin is a solution architect at eNFence, an IBM Business Partner in Germany. He has more than 25 years experience in UNIX systems, designing and supporting AIX and Linux systems for different customers worldwide. He has co-authored many IBM AIX and IBM Power certifications. He is an IBM Champion and IBM AIX Community Advocate. He is a Red Hat Certified Engineer and Red Hat Certified Instructor. He attends international and local events for IBM Power, such as IBM TechXchange, GSE, and Common Europe Congress.

Andrea Longo is a Partner Technical Specialist for IBM Power in Amsterdam, the Netherlands. He has a background in computational biology research and holds a degree in Science and Business Management from Utrecht University. He is an IBM Quantum Ambassador whose duties are to prepare academia and industry leaders to be quantum-safe and to experiment with the immense possibilities of the technology.

Ahmed Mashhour is an IBM Power Technology Services Consultant Lead at IBM Saudi Arabia. He is an IBM L2 certified Expert. He holds IBM AIX, Linux, and IBM Tivoli® certifications. He has 19 years of professional experience in IBM AIX and Linux systems. He is an IBM AIX back-end SME who supports several customers in the US, Europe, and the Middle East. His core expertise is in IBM AIX, Linux systems, clustering management, IBM AIX security, virtualization tools, and various IBM Tivoli and database products. He has authored several publications inside and outside IBM, including co-authoring other IBM Redbooks publications. He has hosted IBM AIX, Security, PowerVM, IBM PowerHA, PowerVC, Power Virtual Server, and IBM Storage Scale classes worldwide.

Amela Peku is a Partner Technical Specialist with broad experience in leading technology companies. She holds an MS in Telecommunication Engineering and is part of the IBM Power team. She works with IBM Business Partners and customers to showcase the value of IBM Power solutions. She provided technical support for next-generation firewalls, Webex, and Webex Teams, focusing on performance and networking, and handled escalations, working closely with engineering teams. She is certified in Networking, Security, and IT Management.

Prashant Sharma is the IBM Power Technical Product Leader for the Asia Pacific region. He is based in Singapore. He holds a degree in Information Technology from the University of Teesside, England, and a MBA from the University of Western Australia. With extensive experience in IT infrastructure enterprise solutions, he specializes in pre-sales activities; client and partner consultations; technical enablement; and the implementation of IBM Power servers, IBM i, and IBM Storage. He drives technical strategy and product leadership for IBM Power to help ensure the delivery of innovative solutions to diverse markets.

Vivek Shukla is a Technical Sales Specialist for IBM Power, Hybrid Cloud, artificial intelligence (AI), and Cognitive Offerings in Qatar working for GBM. He has experience in sales, application modernization, digital transformation, infrastructure sizing, cybersecurity and consulting, and SAP HANA, Oracle, and core banking. He is an IBM Certified L2 (Webexpert) Brand Technical Specialist. He has over 22 years of IT experience in technical sales, infrastructure consulting, IBM Power servers, and AIX, IBM i, and IBM Storage implementations. He has hands-on experience on IBM Power servers, AIX, PowerVM, PowerHA, PowerSC, Requests for Proposals, Statements of Work, sizing, performance tuning, root cause analysis, disaster recovery (DR), and mitigation planning. In addition to writing multiple IBM Power FAQs, he is also an IBM Redbooks author. He is a presenter, mentor, and profession champion that is accredited by IBM. He graduated with a bachelor's degree (BTech) in electronics and telecommunication engineering from IETE, New Delhi, and a master's degree (MBA) in information technology from IASE University. His areas of expertise include Red Hat OpenShift, IBM Cloud Paks, Power Enterprise Pools, and Hybrid Cloud.

Dhanu Vasandani is a Staff Software Test Engineer with over 13 years of experience, specializing in AIX Operating System Security Testing at IBM Power in Bangalore, India. She holds a Bachelor of Technology degree in Computer Science and is instrumental in testing multiple AIX releases across various Power server models. In her current role, Dhanu serves as the Component Lead for the AIX Operating System Security Guild, overseeing various subcomponents. She is responsible for conducting comprehensive system testing for pre-GA and post-GA phases of multiple AIX releases across different Power server models. Dhanu is known for her expertise in areas such as encryption, Trustchk, audit, role-based access control (RBAC), and other security aspects, contributing to IBM Lighthouse Community and IBM Docs. She is recognized for her proficiency in identifying and addressing high-impact AIX defects within the ISST System organization to help ensure the delivery of top-quality products to customers.

Henry Vo is an IBM Redbooks Project Leader with 10 years of experience at IBM. He has technical expertise in business problem solving, risk/root-cause analysis, and writing technical plans for business. He has held multiple roles at IBM that include project management, ST/FT/ETE Testing, back-end developer, and a DOL agent for NY. He is a certified IBM z/OS® Mainframe Practitioner, which includes IBM Z System programming, agile, and Telecommunication Development Jumpstart. Henry holds a Master of Management Information System degree from the University of Texas at Dallas.

Thanks to the following people for their contributions to this project:

Stephen Dominguez, WW Lead Consultant for AIX and Linux Security
IBM Technology Expert Labs, Austin, TX

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on LinkedIn:

  https://www.linkedin.com/groups/2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/subscribe

- Stay current on recent Redbooks publications with RSS Feeds:

  https://www.redbooks.ibm.com/rss.html

1

Security and cybersecurity challenges

This chapter delivers a thorough examination of the security challenges that confront organizations in today's intricate digital environment. It begins by outlining the diverse range of security issues, highlighting the growing sophistication of threats, and the imperative for strong security measures to protect sensitive data and infrastructure. As IT operations extend beyond traditional data centers into hybrid and multi-cloud settings, new vulnerabilities and challenges arise that require adjustments to security strategies.

The chapter delves into the concept of cyber resilience by focusing on the zero trust security model, which mandates continuous verification of users, devices, and network components in an environment with both internal and external threats. It also provides an in-depth look at the IBM security approach by showcasing how its advanced technologies and methodologies are designed to defend against various threats.

In summary, this chapter offers a comprehensive analysis of the security and cybersecurity challenges that are faced by organizations by presenting detailed insights into strategies and technologies to mitigate these threats and enhance overall security resilience, with a particular emphasis on the IBM response to these challenges.

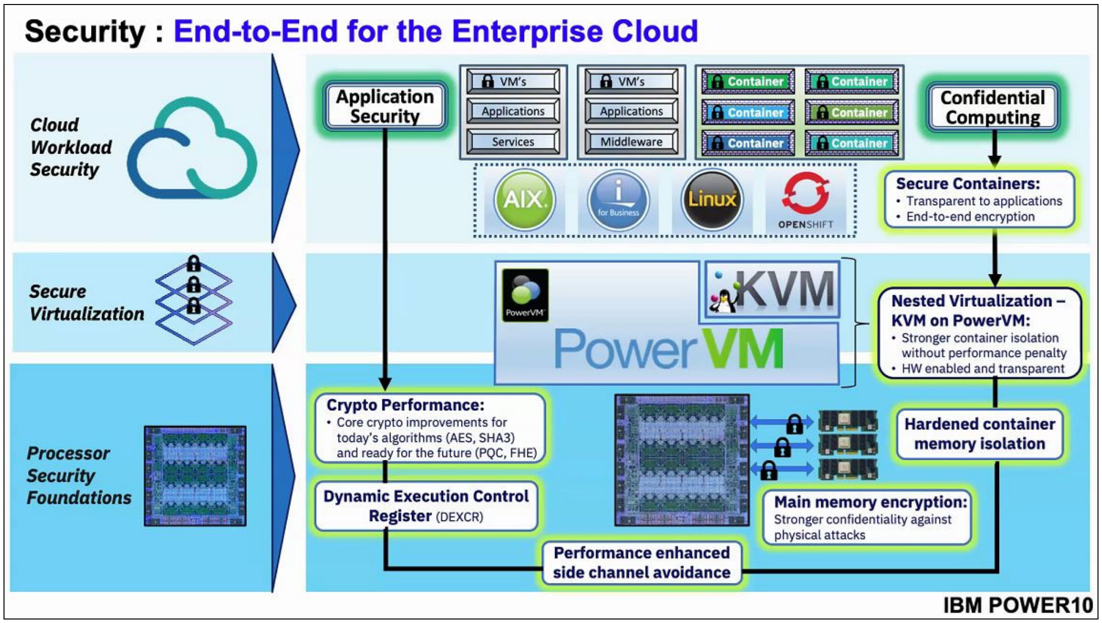
This chapter describes the following topics:

1.1 Protection at every layer

Cloud adoption, edge computing, and hybrid infrastructures offer significant benefits in terms of flexibility and scalability, but they introduce security concerns, such as data breaches, loss of control over data, and the complexities of managing security across diverse cloud platforms. Cyberattacks and ransomware attacks are becoming more prevalent and sophisticated, which can have devastating impacts on organizations. A multi-layered security architecture that encompasses several implementation layers is crucial for robust protection.

At the hardware level, built-in protections that prevent physical tampering and help ensure data integrity, are needed. Virtualization technologies must enhance security by isolating environments and controlling resource access. The security of the hypervisor, a critical component in virtualized environments, is paramount in preventing attacks that might compromise multiple virtual machines (VMs). In the IBM Power environment, logical partitioning provides strong isolation between different workloads on the same physical hardware, which enhances security.

Figure 1-1 shows how IBM Power10 works to provide protection at every layer.



Security : End-to-End for the Enterprise Cloud

Secure Containers:
- Transparent to applications
- End-to-end encryption

Nested Virtualization – KVM on PowerVM:
- Stronger container isolation without performance penalty
- HW enabled and transparent

Crypto Performance:
- Core crypto improvements for today's algorithms (AES, SHA3) and ready for the future (PQC, FHE)

Dynamic Execution Control Register (DEXCR)

Hardened container memory isolation

Main memory encryption: Stronger confidentiality against physical attacks

Performance enhanced side channel avoidance

Figure 1-1 Protection at every layer[1]

Management tools like the Hardware Management Console (HMC) and Cloud Management Console (CMC) play a vital role in securing hardware and cloud resources. Operating systems must continuously provide better security features because they are often vectors of attack, so their contribution is critical to the overall security posture of a system.

Storage security involves protecting data at rest and in transit by using techniques such as encryption and access controls. Methods for creating secure, resilient copies of data, which are known as safeguarded copies, and data resiliency are needed to protect against data corruption or loss. Finally, networking security is integral to overall security with a focus on secure network design, monitoring, and protection mechanisms to prevent unauthorized access and data breaches.

[1] Source:
https://hc32.hotchips.org/assets/program/conference/day1/HotChips2020_Server_Processors_IBM_Starke_POWER10_v33.pdf

1.2 IBM Systems: Built to protect

In today’s digital landscape, IBM Infrastructure serves as a formidable shield against increasingly sophisticated cyberthreats through its robust and integrated security solutions. IBM weaves security into the fabric of its systems and platforms so that businesses can operate confidently amid evolving risks.

At the heart of the IBM approach is the integration of security throughout its systems, which builds trust and resilience from the ground up. This approach includes safeguarding firmware integrity with Secure Boot processes and bolstering data protection through hardware-based encryption acceleration.

IBM goes beyond basic protection with a proactive cybersecurity strategy. It offers secure storage solutions and advanced threat prevention and detection mechanisms. In an incident, IBM provides rapid response and recovery options to minimize downtime and effectively manage operational risks.

Privacy and confidentiality are paramount, which are supported by IBM advanced encryption technologies. These technologies include pervasive encryption throughout the data lifecycle and quantum-safe cryptography, which is designed to guard against emerging threats such as quantum computing.

IBM simplifies regulatory compliance with continuous compliance and audit capabilities. Automated monitoring and enforcement tools help ensure adherence to industry standards, and unified security management tools facilitate consistent governance across diverse IT environments.

Collaborating closely with ecosystem partners, IBM integrates security across hybrid cloud environments, networks, software systems, architectures, and chip designs. This comprehensive approach helps ensure holistic protection and resilience across all facets of an IT infrastructure.

By consolidating security insights across various domains, IBM enables informed decision-making and proactive threat management. This integrated approach dissolves traditional security silos, turning security into a catalyst for innovation and business growth.

In summary, IBM Infrastructure sets a high standard for security excellence by embedding advanced features into its solutions and equipping businesses to address both current and future cybersecurity challenges with confidence. Through collaborative efforts with ecosystem partners and a focus on regulatory compliance, IBM delivers secure, resilient, and compliant infrastructure solutions, empowering businesses to thrive in the digital age amid evolving cyberthreats.

1.3 Overview of security challenges

Digital transformation is reshaping the landscape of modern business and driving the creation and modification of products, services, and operations through digital technology. This integration touches every area of business, which fundamentally alters operations and customer value delivery.

The necessity for digital transformation spans businesses of all sizes, from small enterprises to large corporations. This message is conveyed clearly through virtually every keynote, panel discussion, article, or study that are related to how businesses can remain competitive and relevant as the world becomes increasingly digital. However, there are many considerations, with security being one of the most important. Ensuring that the outcome of digital transformation is more secure than before, and that the transition process is handled securely, is crucial.

In the era of digital transformation, many organizations report experiencing at least one data breach due to the digital transformation process. In addition to data breaches, there are other concerns organizations must address, such as ensuring a secure expansion beyond their data centers, secure cloud adoption, and mitigating cyberattacks and ransomware.

1.3.1 Expansion beyond the data center

Expanding operations beyond the traditional data center into cloud environments, edge computing, and hybrid infrastructures introduces several security challenges. These challenges arise from increased complexity, diverse environments, and an evolving threat landscape. To navigate these complexities, organizations must address key areas such as data protection, access control, visibility, compliance, threat management, configuration, and network security.

Data protection and privacy are paramount as data flows between data centers, cloud services, and edge devices. Ensuring that data is encrypted in transit and at rest across various platforms is essential. Proper encryption and access controls safeguard stored data, and compliance with data sovereignty regulations helps ensure that data is processed and stored according to regional laws. Businesses must implement end-to-end encryption, enforce access controls, and stay updated on data protection regulations.

Access control and identity management become more complex in hybrid environments. Consistent identity and access management (IAM) across on-premises and cloud environments is crucial. Managing and monitoring privileged access helps prevent unauthorized access and insider threats. Strong authentication methods, such as multi-factor authentication (MFA), enhance security by adding extra layers of protection. Implementing robust IAM solutions, continuously monitoring access, and ensuring the use of MFA are key steps.

Visibility and monitoring across hybrid and multi-cloud environments are critical for detecting anomalies and threats. Achieving comprehensive visibility involves implementing unified monitoring solutions that provide a holistic view of the entire infrastructure. Consistent logging and auditing mechanisms are necessary to track activities and support incident response. Network monitoring helps detect and respond to threats in real time. Organizations should invest in integrated monitoring tools, establish thorough logging practices, and deploy real-time network monitoring systems.

Compliance and regulatory requirements present another set of challenges. Navigating diverse regulations across different regions and sectors requires a deep understanding of relevant laws. Ensuring that systems and processes are audit-ready demonstrates compliance, and maintaining data classification schemes helps ensure that appropriate protection measures are in place. Businesses must stay informed about regulatory changes, conduct regular audits, and implement robust data classification protocols.

Managing advanced threats is an ongoing battle. Defending against sophisticated threats such as advanced persistent threats, zero-day vulnerabilities, and ransomware requires a proactive approach. Securing endpoints, including edge devices with varying security capabilities, is crucial. Keeping systems up to date with security patches across different environments helps mitigate vulnerabilities. Organizations should adopt advanced threat detection solutions, enforce stringent endpoint security measures, and establish effective patch management processes.

Configuration management helps ensure consistent security configurations across diverse environments. Detecting and correcting misconfiguration that can introduce vulnerabilities is vital. Organizations should use automated tools to manage configurations and regularly audit systems to identify and rectify misconfiguration.

Network security involves implementing network segmentation to limit the spread of threats. Ensuring secure connections between data centers, cloud environments, and edge locations is essential. Deploying and managing firewalls and intrusion detection/prevention systems in a coordinated manner strengthens network security. Businesses should design segmented network architectures, secure connectivity channels, and maintain robust firewalls, and intrusion detection systems (IDS) and intrusion prevention systems (IPS).

Expanding operations beyond the traditional data center offers numerous benefits, but it also introduces many security challenges that organizations must proactively address. By prioritizing comprehensive data protection, robust access control, enhanced visibility, regulatory compliance, advanced threat management, consistent configuration, and strong network security, businesses can mitigate these risks and fully use the advantages of hybrid and multi-cloud environments. Security in these complex infrastructures is an ongoing process that requires vigilance, adaptability, and a commitment to staying ahead of emerging threats.

1.3.2 Cloud adoption

The migration to cloud computing has revolutionized how businesses operate, offering unparalleled scalability, cost-efficiency, and flexibility. However, alongside these advantages come security challenges that organizations must address to help ensure that their data and operations remain secure.

One of the most critical security challenges in cloud adoption is the potential for data breaches and data loss. Sensitive information that is stored in the cloud can be an attractive target for cybercriminals. Unauthorized access can lead to the exposure of confidential data, resulting in financial losses, reputational damage, and legal repercussions. To mitigate these risks, businesses should implement end-to-end encryption for data at rest and in transit, enforce strict access control policies, and conduct regular security audits and vulnerability assessments.

Compliance and regulatory issues add another layer of complexity to cloud security. Cloud environments often span multiple jurisdictions, each with its own set of regulations and compliance requirements. Ensuring that cloud operations comply with laws such as the General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and California Consumer Privacy Act (CCPA) can be complex and resource-intensive. Organizations must stay informed about relevant regulations, use compliance management tools, and engage third-party auditors to verify compliance.

Insider threats, whether from malicious intent or inadvertent actions, pose a significant risk. Employees, contractors, or third-party vendors with access to cloud systems can potentially misuse their access, leading to data leaks or disruptions. To counter these threats, businesses should implement regular security training programs, use monitoring and anomaly detection systems, and apply the principle of least privilege to limit access based on necessity.

The shared responsibility model in cloud security, where both the cloud provider and the customer share security responsibilities, can lead to confusion and security gaps. Clear definitions of security responsibilities in contracts, regular reviews of cloud provider security documentation, and ongoing collaboration between IT teams and cloud providers are essential to avoid misunderstandings and ensure comprehensive security coverage.

Application programming interfaces (APIs) are essential for cloud integration and operations, but can also introduce vulnerabilities. Poorly secured APIs can become entry points for attackers. To secure APIs, organizations should adopt secure coding practices, use API gateways to manage and secure API traffic, and implement rate limiting to prevent abuse.

Cloud accounts are vulnerable to hijacking through phishing, credential stuffing, or other attack methods. Once compromised, attackers can gain control over cloud resources and data. Enforcing MFA, implementing strong password policies, and continuously monitoring account activities for suspicious behavior are crucial steps to protect cloud accounts from hijacking.

Interfaces and APIs, as gateways to cloud services, need robust security measures. If not properly secured, they can be exploited to gain unauthorized access or disrupt services. Following best practices in API design and security, conducting regular penetration testing and vulnerability assessments, and implementing strong authentication and authorization measures are necessary to secure these critical components.

Adopting cloud technology offers numerous benefits but also introduces a range of security challenges that organizations must proactively address. By implementing robust security measures, maintaining regulatory compliance, and fostering a culture of security awareness, businesses can mitigate these risks and fully use the advantages of the cloud. Security in the cloud is an ongoing process that requires vigilance, adaptability, and a commitment to staying ahead of emerging threats.

1.3.3 Cyberattacks and ransomware

The digital transformation era has enhanced business efficiency and connectivity, but it has also introduced sophisticated cyberattacks and ransomware threats. Addressing these challenges requires a comprehensive and proactive approach to protect sensitive data and ensure business continuity.

Cyberattacks, such as phishing, malware, and advanced persistent threats, disrupt operations and compromise data. To combat these threats, organizations must implement robust firewalls, intrusion detection and prevention systems (IDPS), and use threat intelligence.

Ransomware, which encrypts data and demands a ransom for its release, has become destructive. Preventing ransomware involves regular software updates, strong email filtering, and anti-phishing solutions. Regular data backups that are stored securely and offline, along with tested recovery processes, are critical for mitigating ransomware impact. IBM Storage Defender is a storage software solution that can help protect your data and accelerate recovery in a cyberattack or other catastrophic event. It includes immutable backups, early threat detection, data copy management, and automated recovery capabilities.

Employee training and awareness are essential because many attacks exploit human vulnerabilities. Regular training can help employees recognize phishing attempts and follow best practices for data security, and act as a front-line defense against cyberthreats.

Endpoint security is crucial as employees access corporate resources from various devices. Advanced endpoint protection solutions, including anti-virus software, endpoint detection and response (EDR) tools, and mobile device management (MDM) systems safeguard endpoints from malicious activity.

Network segmentation limits the spread of ransomware and other threats. Dividing the network into smaller segments and implementing strong access controls and monitoring can contain damage and prevent lateral movement by attackers.

Incident response planning is vital for minimizing the impact of attacks. An up-to-date incident response plan (IRP) with clear communication protocols, roles, and procedures for isolating affected systems and restoring operations is essential. Regular drills help ensure the readiness of the response team.

Cyberinsurance provides extra protection, covering the costs of recovery, legal fees, data restoration, and customer notification if there is an attack.

In conclusion, addressing cyberattacks and ransomware requires a comprehensive strategy, including multi-layered defenses, regular backups, employee training, endpoint security, network segmentation, incident response planning, and cyberinsurance. By staying vigilant and continually enhancing security measures, organizations can mitigate risks and protect against these threats.

Cyber resilience: Zero trust

In today's digital landscape, organizations face increasingly sophisticated cyberthreats that require robust defense strategies. Cyber resilience and the zero trust model emerged as critical frameworks to strengthen these defenses and help ensure business continuity in the face of evolving risks.

Cyber resilience encompasses strategies to prepare for, respond to, and recover from cyberincidents effectively, which include comprehensive risk management to prioritize critical assets and threats, incident response planning with clear protocols, regular data backups, and continuous improvement through assessments and updates.

The zero trust model challenges traditional security approaches by assuming no implicit trust based on network location. Instead, it verifies and validates all devices, users, and applications attempting to connect, regardless of their location. Key principles include explicit verification, least privilege access, micro-segmentation to limit lateral movement, and continuous monitoring of network traffic and user behavior.

By integrating cyber resilience with zero trust principles, organizations enhance their ability to detect, respond to, and mitigate cyberthreats. Continuous monitoring and analysis of network activity and user behavior enable prompt threat detection and response. Dynamic, risk-based access controls based on real-time assessments improve security without hindering productivity. Robust backup and recovery measures that are combined with strict access controls help ensure data integrity and availability, even in the event of a breach.

In conclusion, cyber resilience and the zero trust model are essential for organizations striving to fortify their security posture amid a complex threat landscape. By adopting proactive strategies and integrating zero trust principles into their security framework, businesses can safeguard critical assets, maintain operational continuity, and mitigate the impact of cyberattacks. These frameworks strengthen defenses and foster a culture of security awareness and readiness across the organization, which helps ensure ongoing protection against evolving cyberthreats.

1.3.4 Government regulations

Government regulations play a crucial role in shaping and enforcing security standards across various sectors. These regulations aim to protect sensitive information, help ensure data privacy, and maintain overall cybersecurity. This section describes some key aspects of government regulations in relation to security, and provides some examples.

Data protection and privacy laws

Different jurisdictions have specific data protection and privacy laws that regulate how data is secured and how consumers can protect their privacy. Here are a couple of examples:

- General Data Protection Regulation (GDPR)

  Enforced by the European Union (EU), the GDPR mandates stringent data protection and privacy measures for organizations handling EU residents' data. It includes requirements for data breach notifications, consent management, and data subject rights.

- California Consumer Privacy Act (CCPA)

  This law, from the United States state of California, provides residents with rights over their personal data, including access, deletion, and opt-out options for data sales.

Cybersecurity frameworks

There are groups of regulations that address cybersecurity requirements, such as the following ones:

- National Institute of Standards and Technology (NIST) Cybersecurity Framework

  Developed by the NIST, this framework provides guidelines for improving cybersecurity practices. Although it is not mandatory, many organizations adopt it to align with best practices and regulatory expectations.

- Federal Information Security Management Act (FISMA)

  In the United States, FISMA requires federal agencies and contractors to implement information security programs and comply with NIST standards.

Industry-specific regulations

There are specific industry requirements that address data management and security, such as the following ones:

- Health Insurance Portability and Accountability Act (HIPAA)

  For the healthcare industry in the US, HIPAA sets standards for protecting patient health information and requires secure handling and storage of sensitive data.

- Payment Card Industry Data Security Standard (PCI DSS)

  This set of standards applies to organizations handling credit card information and mandates secure processing, storage, and transmission of payment data.

Data breach notification laws

Many jurisdictions have laws requiring organizations to notify affected individuals and relevant authorities if there is a data breach. These laws often specify timelines for notification and the types of information that must be disclosed.

Critical infrastructure protection

In the US, the Cybersecurity and Infrastructure Security Agency (CISA) works to protect critical infrastructure from cyberthreats and provides guidance, support, and coordination for incident response.

Export control regulations

These regulations control the export of certain technologies and information, including cybersecurity tools and data to prevent unauthorized access or use by foreign entities. Some examples are the International Traffic in Arms Regulations (ITAR) and Export Administration Regulations (EAR).

Surveillance and data retention laws

Governments might impose regulations on data retention and surveillance, requiring organizations to store certain data for specified periods or provide access to law enforcement under certain conditions.

Compliance and enforcement

Regulatory bodies have the authority to enforce compliance with security regulations through audits, fines, and other penalties. Organizations must stay informed about relevant regulations and ensure that they adhere to legal requirements to avoid sanctions.

Overall, government regulations help establish a baseline for security practices, protect sensitive information, and promote trust in digital systems. Organizations must understand and comply with these regulations to safeguard their operations and avoid legal repercussions.

1.4 Architecture and implementation layers

This section delves into how security and resiliency are inherently woven into the IBM Power stack across its various layers. It also underscores key considerations for designing and implementing your IBM Power infrastructure to ensure optimal security and performance.

Figure 1-2 illustrates how the IBM Power ecosystem with IBM Power10 processors provides protection at every layer.

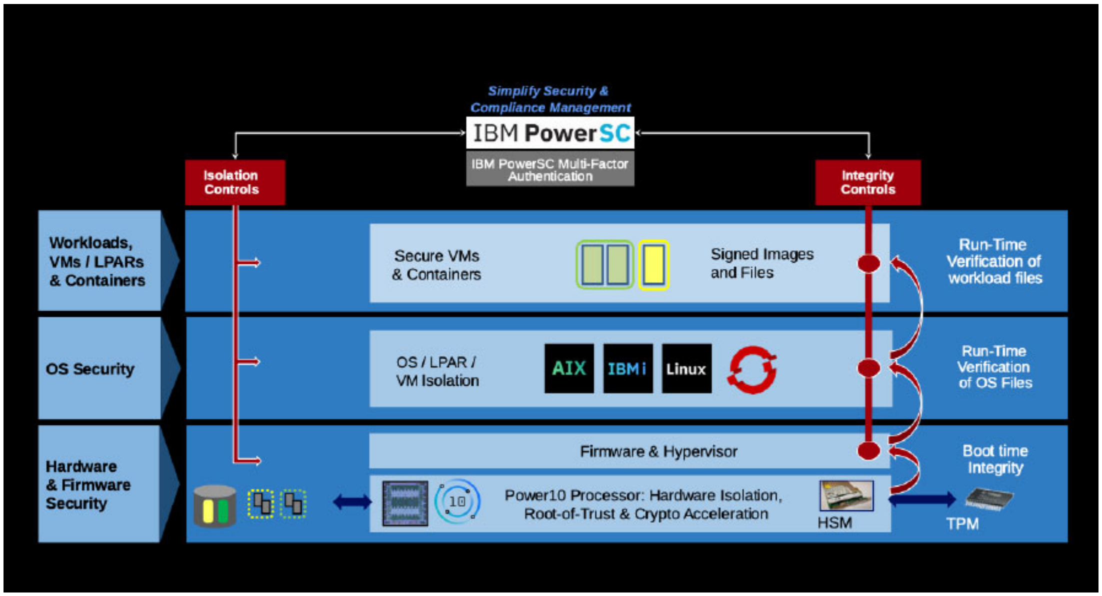

Simplify Security & Compliance Management
IBM PowerSC
IBM PowerSC Multi-Factor Authentication
Isolation Controls
Integrity Controls
Workloads, VMs / LPARs & Containers
Secure VMs & Containers
Signed Images and Files
Run-Time Verification of workload files
OS Security
OS / LPAR / VM Isolation
AIX
IBM i
Linux
Run-Time Verification of OS Files
Hardware & Firmware Security
Firmware & Hypervisor
Boot time Integrity
10
Power10 Processor: Hardware Isolation, Root-of-Trust & Crypto Acceleration
HSM
TPM

Figure 1-2 Layers of protection with IBM Power10[2]

1.4.1 Principle of least privilege

This principle involves limiting access to the minimum that is necessary to perform authorized activities, which significantly reduces the risk of an attacker gaining access to critical systems or sensitive information. To be most effective, this principle should be applied at every level of your infrastructure. Implementing this principle involves the following items:

- Access control policies

  Establishing comprehensive policies that dictate who can access specific resources based on their job requirements.

- Role-based access control (RBAC)

  Implementing RBAC systems that assign permissions to roles rather than individuals, making it simpler to manage and audit access rights across a large organization.

- Regular audits and reviews

  Periodically reviewing access rights and usage logs to ensure compliance with the principle of least privilege and detect any unauthorized access attempts or policy violations.

By adopting these practices, users of IBM Power servers can bolster their defenses against current threats and foster a more resilient posture to adapt to future security challenges.

[2] Source: https://events.ibs.bg/events/itcompass2021.nsf/IT-Compass-2021-S06-Power10.pdf

1.4.2 Hardware

The security of physical hardware is fundamental to protecting the overall integrity of IBM Power servers. This section explores the multiple facets of hardware security, from the physical measures that are used to protect equipment to the embedded technologies that are designed to safeguard data and systems from cyberthreats.

Physical security controls

Physical security controls are critical for protecting systems from unauthorized access, physical damage, or theft. This section delves into various measures and technologies that enhance the physical security of your critical infrastructure locations and components.

- Physical barriers

  Physical barriers serve as the first line of defense in protecting sensitive hardware. Barriers such as walls, gates, and secure enclosures prevent unauthorized physical access.

- Fencing and gates

  Perimeter fencing and security gates serve as the first line of defense, controlling access to your facilities. To enhance security, consider reinforced materials and integrating advanced features such as biometric locks. This comprehensive approach provides a robust barrier against unauthorized entry.

- Secure enclosures

  For shared locations, the usage of secure racks and cages might be necessary to protect critical hardware. These enclosures guard against unauthorized access and can be equipped with extra sensors and alarms for enhanced security.

Integrate physical barriers with electronic security measures, such as surveillance and access control systems to create a comprehensive security envelope around sensitive hardware. Providing access logging capabilities also helps identify personnel who have accessed the environment.

Access controls

Access controls are designed to ensure that only authorized individuals can enter specific physical areas where sensitive hardware is. There are multiple types of access control systems that work with different authentication methodologies. These systems can be broadly categorized into the following categories:

- Biometric systems

  Biometric systems use fingerprint scanners, retina scans, and facial recognition technologies to provide a high level of security by verifying the unique physical characteristics of individuals.

- Electronic access cards

  Technologies such as RFID cards, magnetic stripe cards, and smart cards grant access based on credentials stored on the card. Many of these technologies can be managed centrally to update permissions as needed.

- Personal identification number (PIN) codes and keypads

  Requiring the entry of PINs into keypads provides a method of access control that can be updated and managed remotely.

Surveillance systems

Surveillance systems are essential for monitoring physical environments to detect, deter, and document unauthorized activities. This section introduces the purpose and strategic placement of surveillance systems within power system facilities. Depending on your requirements, you might need multiple complementary surveillance systems. Here are some types of surveillance technologies that you can consider:

- CCTV cameras

  Closed-circuit television cameras provide identification and monitoring capabilities. Different types, such as dome, bullet, and pan-tilt-zoom (PTZ) cameras, should be strategically placed and monitored.

- Motion detectors

  Motion detectors that trigger alerts or camera recordings can enhance the efficiency of surveillance by focusing resources on areas where activity is detected.

- Advanced surveillance technologies

  Newer technologies like thermal imaging and night vision cameras, which capture video in low light or through obstructions, can enhance your around-the-clock surveillance capabilities.

Important: Data management and privacy considerations are involved with the collection and storage of surveillance information. Manage surveillance footage securely, including storage, access controls, and compliance with privacy laws and regulations to protect the rights of individuals.

1.4.3 Embedded security features

Embedded security features are integral components of modern hardware. They are designed to provide built-in protection against various threats. This section explores various embedded security technologies, their functions, and how they contribute to the overall security architecture of IBM Power servers.

Trusted Platform Module

The Trusted Platform Module (TPM) is a dedicated micro-controller that is designed to secure hardware through integrated cryptographic keys. TPM plays an important role in enhancing hardware security by providing hardware-based, security-related functions.

TPM enhances Secure Boot by recording measurements of the system’s firmware and configuration during the startup process. Through an attestation process, the TPM can provide a signed quote that can be used to verify the system integrity and firmware configuration at any time.

TPM also provides key storage and management. It safeguards cryptographic keys at a hardware level, preventing them from being exposed to outside threats. These keys can be used for encrypting data and securing communications (for example, during PowerVM Live Migration).

Virtual Trusted Platform Module

With PowerVM, you can configure a virtual Trusted Platform Module (vTPM) to support Trusted Boot for the OS. When enabled through the HMC, PowerVM instantiates a unique vTPM for each VM. The vTPM extends Trusted Boot support into the VM.

Secure Boot

IBM Power10 and IBM Power9® servers incorporate Secure Boot, a critical security feature that helps ensure the integrity of the system's firmware and operating system at startup. It safeguards against unauthorized modifications and potential attacks, providing a robust foundation for secure operations.

Secure Boot verifies the integrity of the firmware, boot loader, and operating system to prevent unauthorized code from running during the boot process. It helps ensure that only trusted software that is signed with a valid certificate runs, which protects against rootkits and boot-level malware that might compromise the system's security before the operating system starts.

Secure Boot uses digital signatures and certificates to validate the authenticity and integrity of firmware and software components. Each component in the boot process is signed with a cryptographic key, and the system verifies these signatures before allowing the component to run.

Organizations can manage keys and certificates that are used in Secure Boot through configuration settings, which enables them to control which software and firmware are trusted.

Secure Boot helps prevent unauthorized code execution during the boot process, which protects the system from early-stage attacks and aids in meeting compliance requirements for security standards and regulations that mandate Secure Boot processes.

Hardware encryption

Hardware encryption involves using dedicated processors that perform cryptographic operations directly within the hardware itself, which enhances security by isolating the encryption process from software vulnerabilities.

Encryption can be implemented at several layers, which provides protection of your data as it moves through the system. Power10 provides encryption acceleration that is built in to the chip. The system can encrypt memory by default within the system with no performance impact. As data leaves the processor, encryption can be used at the disk level, file system level, and network level to provide complete protection for your data.

Hardware Security Modules

Hardware Security Modules (HSMs) are physical devices that manage digital keys for strong authentication and provide secure crypto-processing. HSMs generate, store, and manage encryption keys in a tamper-resistant environment, helping ensure that keys are never exposed to potentially compromised operating systems. HSMs are integral to digital signing processes, helping ensure the integrity and authenticity of software updates and communications. For more information about HSMs, see 2.1.2, “Encryption enablement in hardware” on page 33.

1.4.4 Risk management in hardware

A significant part of security involves identifying and managing risk. Understanding your environment and its vulnerabilities is critical to creating a plan to protect your enterprise from attacks that can compromise your data or interrupt business operations.

Regular vulnerability assessments are vital for identifying weaknesses in hardware that might be exploited by attackers or fail under operational stress. These assessments should include physical inspections, cybersecurity evaluations, and testing against environmental and operational conditions. Techniques such as penetration testing and red team exercises can simulate real-world attack scenarios to test the resilience of hardware components.

Protecting your environment should include the usage of continuous monitoring technologies, including hardware sensors and network monitoring tools, which play a critical role in the early detection of potential failures or security breaches.

Regular reviews help ensure that risk management strategies and practices stay relevant as new threats emerge and business needs change. This task involves reevaluating and updating risk assessments, mitigation strategies, and response plans at defined intervals or after significant system changes.

Having detailed incident response and recovery plans is essential for minimizing downtime and restoring functions if there is a hardware failure or a security incident. These plans must include roles and responsibilities, communication strategies, and recovery steps.

Training programs for IT staff, operators, and other stakeholders that are involved in hardware management are crucial for maintaining system security. Effective documentation and reporting are also fundamental to the risk management process. Be transparent in reporting to stakeholders and regulatory bodies.

1.4.5 Virtualization

Virtualization has become a cornerstone of modern IBM Power servers by enabling enhanced flexibility and efficiency. However, the shift to virtual environments also introduces specific security challenges that must be addressed to protect these dynamic and often complex systems.

Security in a virtualized environment

A virtualized environment provides better usage of compute resources and allows more flexibility in setting up and running your environment. Virtualization also simplifies creating highly available and resilient systems so that workloads can be moved to support hardware maintenance and outages.

The function that enables virtualization in a system is called a *hypervisor*, also known as a virtual machine monitor (VMM). The hypervisor is a type of computer software that creates and runs VMs, which are also called logical partitions (LPARs). The hypervisor presents the guest operating systems with a virtual operating platform and manages the running of the guest operating systems. Hypervisors are classified into two types:

- A Type 1 hypervisor is a native hypervisor that runs on bare metal.
- A Type 2 hypervisor is hosted on an underlying operating system.

The Type 1 hypervisor is considered more secure because it can provide better isolation between the VMs and generally offers better performance to those VMs.

The IBM solution for virtualization on a Power server, PowerVM, is a combination of hardware, firmware, and software components that provide a foundation for virtualizing CPU, storage, and network resources. At the heart of PowerVM is the Power hypervisor, which is built into the Power firmware. As a Type 1 hypervisor, PowerVM enables the creation of multiple LPARs (VMs) to run on a single Power server and provides the necessary LPAR isolation. Each LPAR is assigned a set of resources, such as memory, CPU, disk space, and adapters to connect to other resources. The operating system in each LPAR is aware of only the resources that it has assigned.

Figure 1-3 illustrates how PowerVM works.

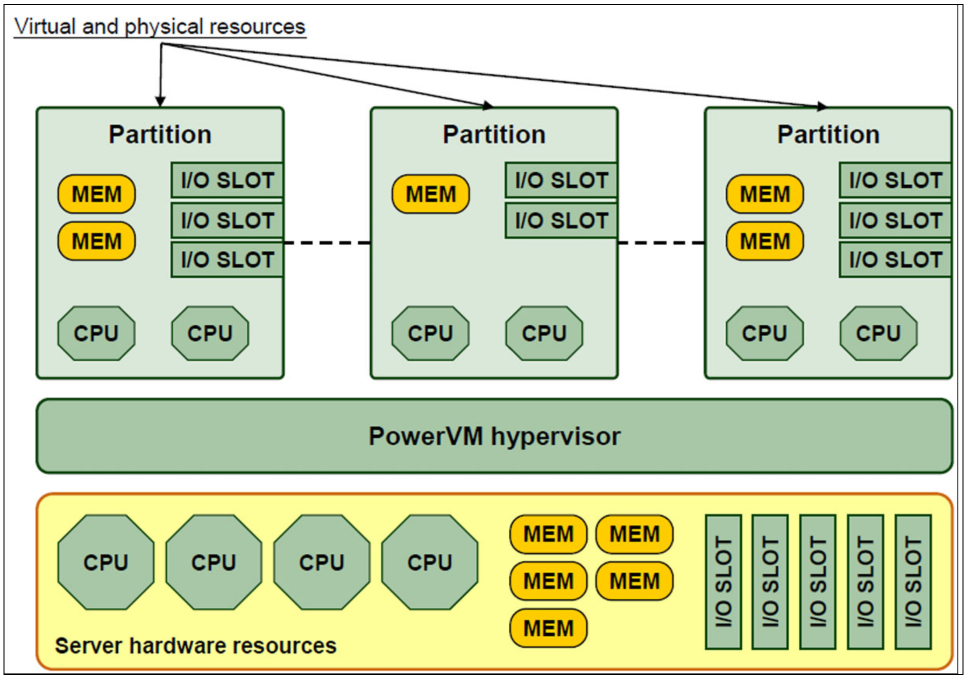


Figure 1-3 PowerVM illustration

The following list provides some security implications that must be addressed by the virtualization layer:

- Isolation failures

  Because there are multiple VMs running at any one time on the same physical hardware, it is imperative that the hypervisor maintains strict isolation between VMs to prevent a breach in one VM from compromising others.

- Hypervisor security

  The hypervisor is the hardware and software layer that enables virtualization, so it is a critical security focal point. Ensuring that the hypervisor is secure and kept up to date is key.

- VM sprawl

  VM sprawl refers to the rapid increase in the number of VMs as your environment grows. This rapid increase can lead to management challenges and potential security oversights. Strategies for controlling VM sprawl, such as inventory management, LPAR consolidation, and lifecycle control policies, are important.

- Secure VM configuration

  Standard processes for maintaining the security of your VMs should be followed, which includes practices such as using hardened base images, applying least privilege principles for VM access, and encrypting VM data both at rest and in transit.

Monitoring and management

Effective management and continuous monitoring are essential for maintaining the security of virtual environments. Here are some tools that you can use to accomplish this goal:

- Real-time monitoring tools

  Integrate tools that provide real-time monitoring of virtual environments, highlighting anomalies and potential security threats. This goal includes solutions that offer visibility into VM operations and network traffic patterns.

- Configuration management

  Maintain a consistent and secure configuration across all virtual assets. This goal includes using configuration management tools that can automate the application of security settings and patches.

- Log management and analysis

  Collect and analyze logs from the virtual systems. Log management solutions can help detect, investigate, and respond to security incidents within virtual environments.

- Compliance auditing

  Conduct regular audits to help ensure that virtual environments comply with relevant security standards and regulations. This goal includes using automated tools to streamline the auditing process and ensure continuous compliance.

1.4.6 HMC and CMC

The HMC and CMC are critical components in managing and monitoring the physical and virtual environments running on IBM Power servers. This section explores their roles, security challenges, and best practices to ensure their security.

HMC

The HMC is used to configure and manage IBM Power servers. Its capabilities encompass logical partitioning, centralized hardware management, Capacity on Demand (CoD) management, advanced server features, redundant and remote system supervision, and security.

The HMC provides a reliable and secure console for IBM Power servers. It is built as an appliance on a highly secured system, tied to specific hardware, and not compatible with other systems. This stringent build process includes incorporating advanced hardware and software security technologies from IBM. Furthermore, HMCs are closed and dedicated, meaning that users cannot add their own software. These features work together to create a highly secure environment.

CMC

By using the CMC, you can view information securely and gain insights about your Power infrastructure across multiple locations. Dynamic views of performance, inventory, and logging for your complete Power enterprise (on-premises or off-premises) simplify and unify information in a single location. CMC provides consolidated information and analytics, which can be key enablers for the smooth operation of infrastructure. Hosted on IBM Cloud®, the CMC is a highly secure cloud-based service that is accessible from mobile devices, tablets, and PCs.

1.4.7 Operating systems

IBM Power offers unparalleled flexibility by enabling you to consolidate diverse operating environments onto a single system. From industry-leading options like AIX and IBM i to the widely adopted Linux and Red Hat OpenShift platforms, you can harness the power of IBM Power to consolidate mission-critical applications across any number of systems. This consolidation provides enhanced reliability, availability, and security (RAS).

Supported operating systems

At the time of publication, Power10 processor-based systems support the platforms/operating system versions that are shown in Table 1-1.

Table 1-1 Power10 platform and operating system support matrix

Operating system	Supported versions
Red Hat OpenShift Container Platform	4.9 or later
PowerVM Virtual I/O Server (VIOS)	4.1.0.0 or later 3.1.4.10 or later 3.1.3.10 or later 3.1.2.30 or later 3.1.1.50 or later
AIX	7.3 TL0 or later 7.2 TL4 or later (with any I/O configuration) 7.1 TL5 or later (through VIOS only)
IBM i	7.5 or later 7.4 TR5 or later 7.3 TR11 or later
Red Hat Enterprise Linux (RHEL)	8.4 or later 9.0 or later
SUSE Linux Enterprise Server	15.3 or later 12.5 or later
Ubuntu	22.04 or later

For a full list of operating systems that run on IBM Power, see Operating systems.

Note: Table 1-1 shows the supported operating systems of the Power E1080. For more information about software maps detailing which versions are supported on which specific IBM Power server models (including previous generations of IBM Power), see System Software Maps.

1.4.8 Storage

Data is a critical asset for any organization and must be readily available. Effective data management is essential to help ensure that large volumes of operational and historical data are securely stored, accessible, and efficiently managed. This section explores storage technology, including different storage architectures, security measures, and best practices for storage management within IBM Power servers.

Storage topologies

There are multiple methods of connecting storage to your servers. The different options evolved over time to meet different requirements and each type has benefits and disadvantages. They also vary in performance, availability, and price.

Direct attached storage

Direct attached storage (DAS) is a type of storage that is directly connected to a computer or server without a network in between. DAS is in contrast to network-attached storage (NAS) or Storage Area Networks (SANs), which involve network connections. DAS as the following characteristics:

- Usage:
  - Small businesses: Small businesses often use DAS for straightforward, cost-effective storage solutions, including external hard disk drives or RAID systems that are directly connected to a server or workstation.
  - Enterprise: In enterprise environments, DAS can be used for high-performance tasks where large amounts of data need to be quickly accessible, such as for databases or VMs.
- Benefits:
  - High performance: Since DAS is directly connected, it often provides faster access speeds and lower latency compared to network-based storage solutions. This situation is crucial for applications requiring rapid data retrieval.
  - Simplicity: Setting up DAS is straightforward because it does not involve network configurations. It is simpler to install and manage, especially for non-technical users.
  - Cost-effective: Generally, DAS solutions are less expensive than NAS or SAN systems because they do not require extra network infrastructure or management tools.
  - Control and security: With DAS, data is stored directly on the device that is connected to the computer or server, which can enhance control and security because the data is not accessible over a network.
- Disadvantages:
  - Scalability: Expanding storage with DAS can be cumbersome. Adding more storage often requires physically connecting extra drives or upgrading existing hardware, which can be less flexible compared to network-based solutions.
  - Limited sharing: DAS typically does not support sharing across multiple computers or users. Each device that is connected to DAS usually has exclusive access, making it less suitable for environments that require collaborative access.

  - Management complexity: In environments with multiple DAS devices, managing and backing up data can become complex. Unlike NAS or SAN, which often include centralized management tools, DAS requires individual management of each device.
  - Redundancy and backup: Implementing redundancy and backup solutions with DAS can be more challenging. Although RAID configurations can provide redundancy, managing and monitoring these setups can be more labor-intensive compared to solutions with built-in redundancy features in networked storage systems.

In summary, DAS offers high performance and simplicity but can be limited in scalability and sharing capabilities. Its benefits make it suitable for scenarios where high speed and control are priorities, and its disadvantages suggest that it might not be ideal for environments needing extensive collaboration or large-scale storage expansion.

Network-attached storage

NAS is a dedicated file storage system that is connected to a network so that multiple users and devices can access and share data over the network. NAS devices typically contain one or more hard disk drives and have their own operating system and management interface. NAS generally has the following characteristics:

- Usage:
  - Small and medium businesses (SMBs): SMBs use NAS for file sharing, backup solutions, and as a centralized repository for documents and other business-critical data. NAS devices in this context can offer features like user authentication, access control, and remote access.
  - Enterprise environments: In enterprises, NAS systems are used for departmental file sharing, backup, and collaboration. Advanced NAS devices can support high-capacity storage, multiple RAID configurations for redundancy, and integration with enterprise applications.
- Benefits:
  - Ease of access: NAS provides a centralized location for data, making it accessible from any device on the network, which facilitates file sharing and collaboration among multiple users.
  - Scalability: NAS systems can be expanded by adding extra drives or connecting multiple NAS units, which enable scalable storage solutions as data needs grow.
  - Cost-effective: NAS is more affordable compared to SAN solutions, and offers a good balance between performance and cost, especially for SMBs and home users.
  - Centralized management: NAS devices come with management interfaces that unable setup, monitoring, and maintenance. They often include features like data encryption, access controls, and user management.
  - Data redundancy: Many NAS devices support RAID configurations, which provide redundancy and protection against data loss due to drive failure.
- Disadvantages
  - Network dependency: NAS performance depends on the network's speed and reliability. High network traffic or network issues can impact access speeds and performance.
  - Limited performance: Although NAS provides adequate performance for many applications, it might not be suitable for high-performance tasks that require fast data access, such as high-frequency trading or large-scale data processing.

  - Complexity in large deployments: In larger environments with numerous NAS devices, managing multiple units and ensuring consistent backup and security policies can become complex.
  - Security risks: Because NAS devices are network-connected, they are vulnerable to network-based threats. Proper security measures, such as encryption, firewalls, and access controls, are necessary to protect data.
  - Cost of advanced features: Although basic NAS units are affordable, those units with advanced features like high capacity, high availability (HA), or enterprise-level functions might be expensive.

In summary, NAS offers centralized, accessible, and scalable storage solutions that are suitable for a wide range of environments, from home use to enterprise settings. It excels in providing file sharing and backup capabilities but can face limitations in performance and complexity as needs grow. Proper network infrastructure and security measures are crucial for optimizing NAS performance and protecting data.

Storage Area Network

A SAN is a specialized network that provides high-speed, dedicated access to consolidated storage resources. Unlike NAS, which operates over a standard network, SAN is designed specifically for storage and often uses high-performance connections like Fibre Channel or iSCSI. SAN storage generally has these characteristics:

- Usage:
  - Enterprise environments: SAN is commonly used in large-scale enterprise environments where high performance, scalability, and reliability are critical. It is often employed for mission-critical applications, large databases, and virtualized environments.
  - Data centers: SANs are widely used in data centers to provide centralized storage for multiple servers. They support high-capacity and high-performance storage needs, which facilitate efficient data management and backup.
  - High-Performance Computing: Applications that require fast data access and large volumes of data, such as scientific simulations or financial transactions, benefit from a SAN's high throughput and low latency.
- Benefits:
  - High performance: SANs provide high-speed data access, which is essential for performance-intensive applications. Technologies like Fibre Channel offer low latency and high throughput.
  - Scalability: SANs can be easily scaled by adding more storage devices or connecting more servers, which make them suitable for growing data needs and expanding workloads.
  - Centralized storage management: SANs consolidate storage into a single, centralized system, which simplifies storage management, backup, and recovery processes.
  - Data redundancy and reliability: SANs often supports advanced redundancy features, such as multiple paths to storage devices and RAID configurations, which enhance data protection and availability.
  - Virtualization support: SANs are well suited for virtualized environments, which provide flexible and efficient storage allocation and management for VMs.

- Disadvantages:
  - Cost: SANs can be expensive to implement and maintain, especially with high-performance components like Fibre Channel switches and storage arrays. The initial investment and ongoing operational costs can be significant.
  - Complexity: SANs are complex to set up and manage. They require specialized knowledge and expertise for configuration, management, and troubleshooting. This complexity can lead to increased administrative overhead.
  - Infrastructure requirements: SANs require dedicated hardware and infrastructure, such as Fibre Channel switches or iSCSI adapters, which can add to the cost and complexity of the overall system.
  - Network congestion: Although SANs are designed to avoid network congestion, sometimes the network infrastructure supporting the SAN can become a bottleneck if not properly managed or if it is shared with other traffic.
  - Security risks: SANs are often accessed over dedicated networks, but they still require robust security measures to protect data from unauthorized access or breaches. Proper access controls and encryption are essential to safeguard data.

In summary, SAN provides high-performance, scalable, and centralized storage solutions that are ideal for enterprise and data center environments. It excels in performance and reliability, but can be costly and complex to implement and manage. Organizations that use SANs must balance their need for high-speed data access with the associated infrastructure and operational costs.

Cloud storage

Cloud storage refers to the practice of storing data on remote servers that can be accessed over the internet. Providers manage these servers and offer various services for storing, managing, and retrieving data. This model contrasts with traditional on-premises storage solutions, where data is stored locally on physical devices. Cloud storage generally has the following characteristics:

- Usage:
  - SMBs: SMBs use cloud storage for file sharing, collaboration, and remote work. It provides a cost-effective way to scale storage needs without investing in physical infrastructure.
  - Large enterprises: Enterprises use cloud storage for scalable data storage solutions, disaster recovery (DR), and global access. It supports extensive data needs, facilitates collaboration, and integrates with various enterprise applications.
  - Developers and IT professionals: Cloud storage is used for hosting applications, managing databases, and providing scalable storage solutions for big data and analytics.
- Benefits:
  - Scalability: Cloud storage offers virtually unlimited storage capacity. Users can scale their storage up or down based on their needs without needing to invest in physical hardware.
  - Accessibility: Data that is stored in the cloud can be accessed from anywhere with an internet connection. This approach supports remote work, collaboration, and access from multiple devices.
  - Cost-effectiveness: Typically, cloud storage operates on a pay-as-you-go model, so users pay only for the storage that they use. This approach reduces the upfront costs that are associated with purchasing and maintaining physical storage hardware.

  - Automatic updates and maintenance: Cloud providers handle software updates, security patches, and hardware maintenance, freeing users from these tasks and helping ensure that the storage environment is up to date.
  - DR: Many cloud storage services include built-in redundancy and backup solutions, which provide enhanced data protection and recovery options if there is data loss or system failures.
- Disadvantages:
  - Security and privacy: Storing data off site introduces concerns about data security and privacy. Users must trust cloud providers to protect their data from breaches and unauthorized access. Encryption and other security measures are essential but might not be foolproof.
  - Internet dependence: Access to cloud storage depends on a stable internet connection. Poor connectivity can hinder access to data and affect performance.
  - Ongoing costs: Although cloud storage can be cost-effective, ongoing subscription fees can add up over time, especially for large amounts of storage or high-performance requirements.
  - Data transfer speeds: Uploading and downloading large amounts of data can be slow, especially with limited internet bandwidth. This situation can impact the speed at which data is accessible or transferred.
  - Vendor lock-in: Different cloud providers use proprietary technologies and formats, which can make it challenging to migrate data between services or integrate with other systems. This situation can lead to vendor lock-in and potential difficulties if you decide to switch providers.

Cloud storage offers flexible, scalable, and accessible solutions suitable for personal, business, and enterprise needs. Its benefits include scalability, cost-effectiveness, and automatic maintenance, which makes it an attractive option for modern data management. However, concerns about security, reliance on internet connectivity, ongoing costs, and potential vendor lock-in are important considerations that users must address when opting for cloud storage solutions.

Security considerations for storage

When securing storage systems, whether on-premises or in the cloud, several key considerations are crucial to protecting data from unauthorized access, breaches, and other security threats. Here is an overview of important security considerations for storage:

- Data encryption:
  - At rest: Encrypt data that is stored on physical media to protect it from unauthorized access if the storage device is stolen or compromised.
  - In transit: Use encryption protocols like Transport Layer Security (TLS)/Secure Sockets Layer (SSL) for data that is transmitted over networks to prevent interception and unauthorized access during transmission.
- Access controls:
  - Authentication: Implement strong authentication mechanisms, such as MFA to ensure that only authorized users can access storage systems.
  - Authorization: Define and enforce access controls to limit what users can see and do based on their roles and permissions. Implement principles of least privilege to minimize access rights.

- Data integrity:
  - Checksums and hashes: Use checksums or cryptographic hashes to help ensure data integrity and detect any unauthorized alterations or corruption of data.
  - Version control: Maintain version histories of important data to recover from accidental deletions or modifications.
- Backup and recovery:
  - Regular backups: Perform regular backups of critical data and store backups securely, preferably in a different location from the primary storage to protect against site-specific disasters.
  - Test recovery: Periodically test backup and recovery processes to help ensure that data can be restored quickly and accurately if there is data loss or corruption.
- Physical security:
  - Data center security: For on-premises or colocated storage, ensure that data centers have robust physical security measures, such as restricted access, surveillance, and environmental controls.
  - Device security: Secure physical storage devices to prevent unauthorized access and theft. Consider measures like locked server rooms or safes for sensitive equipment.
- Network security:
  - Firewalls and intrusion detection: Use firewalls and intrusion detection/prevention systems to protect storage networks from unauthorized access and cyberthreats.
  - Segmentation: Segment storage networks from other networks to limit exposure and potential attack vectors.
- Monitoring and auditing:
  - Activity monitoring: Implement logging and monitoring to track access and changes to storage systems, which help identify suspicious activities and respond to potential security incidents.
  - Regular audits: Conduct regular security audits and assessments to help ensure compliance with security policies and identify vulnerabilities.
- Compliance:
  - Regulatory requirements: Adhere to relevant regulations and standards, such as GDPR, HIPAA, and PCI DSS, which mandate specific security practices for protecting data.
  - Data sovereignty: Understand and comply with data residency requirements and local laws that are related to data storage and protection.
- Vendor management:
  - Cloud providers: When using cloud storage, carefully evaluate the security measures and practices of the cloud service provider. Ensure that they meet your security requirements and comply with relevant regulations.
  - Third-party risk: Assess and manage risks that are associated with third-party vendors who have access to your storage systems or handle your data.
- Incident response:
  - Plan and preparation: Develop and maintain an IRP to address security breaches or data loss. This plan should include procedures for containment, investigation, and communication.
  - Training: Regularly train staff on security best practices and how to respond to security incidents to ensure preparedness.

Securing storage systems involves a comprehensive approach that includes data encryption, robust access controls, regular backups, physical security, network protections, monitoring, compliance, and vendor management. By addressing these considerations, organizations can significantly reduce the risk of data breaches, loss, and other security incidents.

Best practices for storage management

Beyond the security aspects of your data, there are some other considerations that must be addressed in your storage environment, including compliance requirements and planning for recovery if there is a cyberattack or ransomware. Effective storage management helps ensure that data is stored securely, efficiently, and cost-effectively.

Here are some best practices for managing storage across various environments, including on-premises and cloud-based solutions:

- Capacity planning:
  - Assess needs: Regularly evaluate current and future storage needs based on data growth projections, application requirements, and business goals.
  - Optimize usage: Use tools to monitor storage usage and optimize space. Consider implementing data deduplication and compression to reduce the amount of storage required.
- Data classification and organization:
  - Classify data: Categorize data based on its importance, sensitivity, and usage patterns. This best practice helps with applying appropriate storage and security policies.
  - Organize efficiently: Structure storage systems to facilitate access and retrieval. Use logical grouping and hierarchical storage management to keep data organized.
- Data backup and recovery:
  - Regular backups: Implement a consistent backup schedule to help ensure that data is regularly backed up. Consider full, incremental, and differential backups based on the needs.
  - Test recovery: Periodically test backup and recovery procedures to help ensure that data can be restored quickly and accurately if there is a loss or corruption.
- Data retention policies:
  - Define policies: Establish clear data retention policies based on legal requirements, business needs, and data usage patterns. Determine how long different types of data should be kept before deletion or archiving.
  - Automate management: Use automated tools to enforce retention policies, manage the data lifecycle, and handle the archiving or deletion of obsolete data.
- Performance optimization:
  - Monitor performance: Regularly monitor storage performance metrics, such as I/O operations and response times to identify and address bottlenecks.
  - Optimize storage: Use techniques such as tiered storage to allocate high-performance storage to critical applications while using lower-cost storage for less critical data.
- Cost management:
  - Budgeting: Develop and adhere to a storage budget that aligns with business needs and growth projections.
  - Cost optimization: Regularly review storage costs and explore cost-saving options, such as moving infrequently accessed data to lower-cost storage tiers or cloud storage solutions.

- Disaster recovery planning:
  - Plan development: Create a DR plan that outlines procedures for data backup, restoration, and continuity of operations if there is catastrophic events.
  - Regular updates: Review and update the DR plan regularly to adapt to changes in the business environment or technology.
- Compliance and auditing:
  - Ensure compliance: Stay compliant with relevant regulations and standards (for example, GDPR, HIPAA, and PCI DSS) regarding data storage and protection.
  - Conduct audits: Perform regular security and compliance audits to identify and address any gaps or issues in storage management practices.
- Vendor management:
  - Evaluate providers: When using third-party storage solutions, thoroughly evaluate vendors based on their security, reliability, and performance.
  - Manage contracts: Clearly define service-level agreements (SLAs) and terms in contracts with storage vendors to help ensure that they meet your performance and security requirements.
- Documentation and training:
  - Document procedures: Maintain comprehensive documentation of storage management policies, procedures, and configurations.
  - Train staff: Provide training for the staff that are involved in storage management to ensure that they are knowledgeable about best practices, tools, and security measures.
- Automation and tools: Implement automation: Use storage management tools and automation to streamline tasks such as provisioning, monitoring, and maintenance.

Effective storage management involves strategic planning, organization, and implementation of best practices in data classification, backup, security, and performance optimization. By adhering to these practices, organizations can help ensure that their storage systems are reliable, secure, and aligned with business objectives while managing costs and compliance requirements efficiently.

Safeguarded Copy and data resiliency

Safeguarded Copy is a term that is used in data management and backup solutions to refer to backup copies of data that are protected to ensure their reliability and integrity. The concept focuses on creating and maintaining backup copies that are reliable, and secure from various threats, including corruption, tampering, and unauthorized access. It is an essential part of a robust data protection strategy, especially in environments where data integrity and availability are critical.

Here are the key aspects of Safeguarded Copy:

- Data integrity:
  - Consistency: Safeguarded copies are created to reflect a stable state of the data at a specific point in time. This approach can involve techniques like snapshots or consistent backups to help ensure that all parts of the data are accurately captured.
  - Validation: Integrity checks, such as checksums or cryptographic hashes, are used to verify that backup data has not been altered or corrupted. This approach helps ensure that the data in the backup matches the original data.

- Security:
  - Encryption: Backup copies are encrypted both at rest and during transmission. Encryption protects data from unauthorized access and helps ensure that even if a backup is compromised, the data remains secure.
  - Access controls: Strict access controls are enforced to limit who can access or manage the backup copies. This best practice helps prevent unauthorized access or tampering with the backup data.
- Protection from ransomware:
  - Immutable backups: Implement features such as write-once, read-many (WORM) to make backup copies immutable. Once a backup is created, it cannot be altered or deleted by ransomware or malicious actors.
  - Isolated storage: Store backup copies in a separate location or isolated environment that is not directly accessible from the primary network. This best practice reduces the risk of backups being affected by ransomware or other attacks.
- Automated management:
  - Scheduling: Automated scheduling helps ensure that backups are performed regularly and consistently without relying on manual intervention. This best practice helps maintain up-to-date backup copies.
  - Verification: Automated verification processes check the integrity of backup copies to ensure that they are usable and not corrupted.
- Disaster recovery readiness:
  - Testing: Regularly test backup and recovery processes to ensure that safeguarded copies can be restored effectively. This best practice includes performing periodic restore tests to verify the accuracy and completeness of backups.
  - Documentation: Maintain detailed documentation of backup procedures, configurations, and recovery plans. This best practice helps ensure that backup and restoration processes are clear and can be run quickly in an emergency.
- Compliance with regulatory requirements: Ensure that safeguarded copies meet regulatory and industry standards for data protection and privacy. This best practice includes adhering to requirements for data retention, security, and access control.

In summary, Safeguarded Copy refers to backup copies of data that are protected to ensure their integrity, security, and reliability. This feature involves creating consistent and reliable backups, encrypting and securing backup data, protecting against threats like ransomware, and automating management processes. By implementing safeguarded copies, organizations can ensure that their data backups are robust, secure, and capable of supporting effective DR and data protection strategies.

Important: A Safeguarded Copy is not just a physical copy of the data. It involves automation and management to take regular copies, validate that they are valid and stored so that they cannot be modified. Equally important is the ability to quickly recognize when your data is compromised and recover to a last good state. It also involves business processes to recover applications and databases to minimize data loss.

IBM Storage provides a Safeguarded Copy capability in both the IBM DS8000® and the IBM FlashSystem® systems. For more information about the IBM solutions, see the following resources:

- *IBM Storage DS8000 Safeguarded Copy: Updated for DS8000 Release 9.3.2*, REDP-5506
- *Data Resiliency Designs: A Deep Dive into IBM Storage Safeguarded Snapshots*, REDP-5737
- *Cyber Resiliency with IBM Storage Sentinel and IBM Storage Safeguarded Copy*, SG24-8541

1.4.9 Networking

Security considerations for networking involve several key aspects to protect data integrity, confidentiality, and availability across networked systems. Whether using physical networking connections or virtualizing the network functions, the considerations are generally the same. Here are some essential considerations:

- Network segmentation

  Dividing a network into segments can limit the spread of attacks and contain potential breaches. Segmentation helps isolate sensitive data and systems from less critical areas.

- Firewalls

  Firewalls act as barriers between internal networks and external threats. They filter incoming and outgoing traffic based on predefined security rules.

- Intrusion detection and prevention systems

  These systems monitor network traffic for suspicious activity and can either alert administrators or block potential threats.

- Encryption

  Encrypting data that is transmitted over the network helps ensure that even if data is intercepted, it remains unreadable without the proper decryption keys.

- Access controls

  Implementing strong access controls, including MFA and least privilege principles, helps ensure that only authorized users and devices can access network resources.

- Regular updates and patching

  Keeping network devices and software up to date with the latest security patches helps protect against known vulnerabilities and exploits.

- Network monitoring

  Continuous monitoring of network traffic and device behavior helps detect and respond to anomalies and potential security incidents in real time.

- Secure configuration

  Ensuring that network devices (for example, routers and switches) are securely configured according to best practices reduces the risk of exploitation through bad actor connections and potential threats, and it helps reduce the risk of human error and social engineering attacks.

- Incident response planning

  Having a plan in place for responding to network security incidents helps minimize damage and recover quickly from breaches.

- Security policies and training

  Establishing clear security policies and training employees on best practices and potential threats helps reduce the risk of human error and social engineering attacks.

Addressing these considerations helps build a robust network security posture and protect against various cyberthreats.

2

Protection across every layer

An infrastructure that is built by using IBM Power servers benefits from the robust security technologies that are integrated into both the hardware and software stacks. IBM Power servers offer advanced security features at every level of the system, which help ensure comprehensive protection for sensitive data and applications. These features include advanced encryption technologies, Secure Boot capabilities, and integrated firmware updates. Also, IBM Power servers use IBM's extensive expertise in securing mission-critical workloads, making them a popular choice for organizations seeking a secure environment for their digital assets.

Workloads on the IBM Power10 server see significant benefits from improved cryptographic accelerator performance compared to previous generations. Specifically, the Power10 chip supports accelerated cryptographic algorithms such as AES, SHA2, and SHA3, resulting in considerably higher per-core performance for these algorithms. This enhancement enables features like AIX Logical Volume Encryption to operate with minimal impact on system performance.

The processor core technology of the Power10 incorporates integrated security protections:

- Improved cryptographic performance

  Integrated cryptographic support reduces the performance impact of encrypting and decrypting your data so that you can make encryption pervasive to protect all your critical data.

- Increased application security

  Hardened defenses against return-oriented programming (ROP) attacks.

- Simplified hybrid cloud security

  Setup-free hybrid cloud security administration with a single interface.

- Enhanced virtual machine (VM) isolation

  Providing the industry's most secure VM isolation technology, this technology defends against attacks that exploit operating system or application vulnerabilities in the VM to access other VMs or the host system.

This chapter describes the following topics:

- 2.1, “Encryption technologies and their applications” on page 30
- 2.2, “Compliance automation and real-time monitoring” on page 39
- 2.3, “Endpoint detection and response” on page 40
- 2.4, “Malware prevention technologies” on page 40
- 2.5, “Secure Boot and Trusted Boot” on page 41
- 2.6, “Hypervisor security” on page 48

2.1 Encryption technologies and their applications

Power10 emphasizes comprehensive security throughout its design by offering multiple encryption options. Key among them are Transparent Memory Encryption (TME), Fully Homomorphic Encryption (FHE), and Quantum-Safe Encryption (QSE).

- Transparent Memory Encryption encrypts data in memory to protect it from unauthorized access and tampering during run time. Operating at the hardware level, TME uses the Power10 processor's cryptographic engines to perform encryption and decryption tasks efficiently. TME helps ensure pervasive protection of data in memory with minimal impact on system performance because encryption and decryption are managed at the chip level. Its integration into normal operations is seamless and automatic.
- Fully Homomorphic Encryption enables computations to be performed directly on encrypted data without decrypting it first to help ensure that sensitive data remains confidential even during processing. FHE operates at the software level and involves sophisticated mathematical algorithms to enable computations on ciphertexts. Implementing FHE requires specialized libraries and frameworks. However, FHE is computationally intensive and can introduce performance overhead compared to conventional hardware-only encryption methods due to the complexity of the algorithms.
- Quantum-Safe Encryption is resistant to quantum attacks, securing data against the computational capabilities of future quantum computers that might potentially break current cryptographic algorithms. QSE employs cryptographic algorithms that are believed to be resistant to quantum attacks, such as lattice-based, hash-based, and multivariate-quadratic-equations-based cryptography. Many quantum-safe algorithms are still undergoing testing and standardization to ensure that they provide robust security in the face of future quantum advancements. QSE is typically used for securing long-term data, sensitive communications, and critical infrastructure.

The relevant features and differences in these technologies are shown in Table 2-1.

Table 2-1 Key differences

Feature	TME	FHE	QSE
Encryption scope	Secures data in memory.	Allows computations on encrypted data.	Prevents against future quantum computing threats.
Implementation level	Implemented in hardware.	Implemented by using a combination of hardware and software.	Implemented by using a combination of hardware and software.

Feature	TME	FHE	QSE
Performance impact	Hardware that is accelerated through the Power10 cryptographic engines and designed to have minimal performance impact.	Involves substantial computational overhead.	The impact of QSE varies. Some quantum-safe algorithms might introduce performance overhead. This topic is a subject of ongoing research.
Use cases	Used to protect data in memory.	Used for performing secure computations on sensitive data without decrypting it.	Provide long-term data protection and secure communications in the future.

2.1.1 Quantum-Safe Encryption

QSE, also known as Post-Quantum Cryptography (PQC), refers to encryption methods that are secure against both classical and quantum computers. As quantum computers advance, they might pose a threat to existing cryptographic systems, potentially compromising their security. QSE is essential for protecting sensitive data, communication channels, and user identities in the age of quantum computing.

The urgency of adopting QSE stems from two primary concerns:

- Advanced quantum computers might enable adversaries to intercept and decrypt protected digital communications through Harvest Now, Decrypt Later (HNDL) strategies, even before reaching Q-Day. (Q-Day is the anticipated point in time when quantum supremacy becomes widespread and many of the current encryption algorithms are no longer effective.)
- Migrating to QSE might require over a decade due to the complexities of organizational structures and IT infrastructure.

Therefore, organizations should start evaluating and implementing QSE solutions immediately to ensure continued protection and maintain trust among their stakeholders.

Delaying QSE adoption might have severe consequences. Legacy cryptographic systems left unaltered might be compromised if there is a successful quantum attack, which can expose sensitive data and risk confidential business transactions and individual privacy. Financial institutions, critical infrastructure providers, and government agencies might face significant challenges in maintaining operational integrity and confidentiality. Therefore, prioritizing QSE implementation is crucial for long-term cybersecurity resilience.

Power10 supports these quantum-safe algorithms to help ensure robust security even as quantum computing advances.

Here are some of the quantum-safe features that Power10 supports:

- Data encryption breakage protection:
  - Risk: Quantum computers might break widely used cryptographic algorithms, such as RSA, Elliptic Curve Cryptography, and traditional Diffie-Hellman key exchange protocols. Shor's algorithm, for example, could efficiently factor large integers and solve discrete logarithms, which can compromise the security of these algorithms.
  - Protection: Power10 supports quantum-safe algorithms like lattice-based, hash-based, code-based, and multivariate quadratic equations-based cryptography, which are believed to be resistant to quantum attacks. The crypto engines in Power10 enhance the performance of these algorithms to help ensure secure encryption and key exchange processes with minimal performance degradation.
- Secure communications:
  - Risk: Quantum computers might intercept and decrypt secure communications, which can undermine protocols that currently rely on classical encryption methods.
  - Protection: Power10 secures communication channels with quantum-resistant protocols to help ensure data confidentiality even in the presence of quantum adversaries. End-to-end encryption is maintained throughout the data lifecycle, from storage to transmission, by using algorithms that are resistant to quantum attacks.
- Data integrity and authenticity:
  - Risk: Quantum computers might forge digital signatures or tamper with data.
  - Protection: Power10 can support quantum-safe digital signature algorithms such as eXtended Merkle Signature Scheme (XMSS) and Unbalanced Oil and Vinegar (UOV), which provide strong security against quantum attacks.
- Long-term data protection:
  - Risk: Sensitive data that is stored today might be harvested and decrypted in the future as quantum computers become more powerful, which threatens long-term confidentiality.
  - Protection: Implementing QSE methods helps ensure that data remains secure over time, even as quantum computing capabilities evolve. The Power10 architecture supports updates to cryptographic libraries and protocols to enable the adoption of new quantum-safe algorithms as they are developed and standardized.
- Physical and memory attack protection:
  - Risk: Physical attacks on memory, such as cold start attacks, might expose sensitive data if it is not adequately protected.
  - Protection: The Power10 TME helps ensure that data in memory is encrypted to protect it from physical attacks during run time.

Quantum-safe algorithms that are supported by Power10

The following quantum-safe algorithms are supported by Power10:

- Lattice-based cryptography:
  - Algorithms: ML-DSA (Dilithium), ML-KEM (Kyber), and NTRUEncrypt.
  - Characteristics: Secure against quantum attacks. Based on lattice problems Learning With Errors (LWE), and Ring-Learning With Errors (Ring-LWE). Relatively efficient for hardware and software implementations.
- Hash-based cryptography:
  - Algorithms: Merkle Signature Scheme (MSS), XMSS, and SPHINCS+.
  - Characteristics: Secure based on hash functions, though generally produces larger signatures and keys.
- Code-based cryptography:
  - Algorithms: McEliece Cryptosystem, Bit Flipping Key Encapsulation (BIKE), and Hamming Quasi-Cyclic (HQC).
  - Characteristics: Quantum-resistant based on decoding random linear codes, although public keys can be large.
- Multivariate Quadratic Equations:
  - Algorithms: UOV, and Rainbow.
  - Characteristics: Secure against solving systems of multivariate quadratic equations and efficient in signature generation, but it might involve larger key sizes.

Power10 implementation

Power10 processors support these quantum-safe algorithms by using the following features:

- Crypto engines: Multiple engines per core enable efficient execution of cryptographic operations.
- Software updates: The architecture enables updates to cryptographic libraries, which help ensure the integration of new quantum-safe algorithms as they become standardized.

The Power10 design and capabilities help ensure robust security against future quantum threats by using hardware acceleration and flexible software updates to maintain high-security standards as the cryptographic landscape evolves.

2.1.2 Encryption enablement in hardware

There are two options for accelerating encryption in an IBM Power10 server:

- Use the built-in encryption acceleration that is built in to the Power10 chip.
- Use a PCI Express (PCIe)-based encryption accelerator.

On-chip encryption support in Power10

The IBM Power10 chip is designed to effectively support future encryption (including FHE and QSC) to be ready for the quantum age. The Power10 processor-chip instruction set architecture (ISA) is tailored for these solutions' software libraries, which are available at the time of writing or will soon be made available in the corresponding open source communities.

Workloads on the Power10 benefit from cryptographic algorithm acceleration, which enables higher per-core performance than Power9 processor-based servers for algorithms like Advanced Encryption Standard (AES), SHA2, and SHA3. Features like AIX Logical Volume Encryption can be activated with minimal performance overhead because of this performance enhancement.

With four times as many AES encryption engines, Power10 processor technology is designed to offer quicker encryption performance. Power10 is more advanced than IBM Power9 processor-based servers, with updates for the most stringent standards of today and future cryptographic standards, which include post-quantum and FHE. It also introduces extra improvements to container security. By using hardware features for a seamless user experience, TME aims to simplify encryption and support end-to-end security without compromising performance.

IBM PCIe Cryptographic Coprocessor cards

IBM PCIe Cryptographic Coprocessors are a family of high-performance Hardware Security Modules (HSMs). These programmable PCIe cards work with certain IBM Z®, x86-64, and IBM Power servers to offload computationally intensive cryptographic processes, such as secure payments or transactions from the host server.

With these coprocessors, you can accelerate cryptographic processes that safeguard and secure your data while protecting against various attacks. The IBM 4769, IBM 4768, and IBM 4767 HSMs deliver security-rich, high-speed cryptographic operations for sensitive business and customer information with the highest level of certification for commercial cryptographic devices.

Cryptographic Coprocessor cards relieve the main processor from cryptographic tasks. The IBM HSMs have a PCIe local-bus-compatible interface with tamper responding, programmable, cryptographic coprocessors. Each coprocessor contains a CPU, encryption hardware, RAM, persistent memory, hardware random number generator, time-of-day clock, infrastructure firmware, and software. Their specialized hardware performs AES, DES, DES, RSA, Elliptic Curve Cryptography, AESKW, HMAC, DES/3DES/AES mandatory access control (MAC), SHA-1, SHA-224 to SHA-512, SHA-3, and other cryptographic processes. This hardware relieves the main processor from these tasks. The coprocessor design protects your cryptographic keys and any sensitive customer applications.

Customizable to meet special requirements

The firmware that is running in the coprocessor together with the software that is running on your host can be customized to meet any special requirements that your enterprise has. For the IBM 4769 and IBM 4767, the Cryptographic Coprocessor Toolkit (CCTK) is available for purchase from IBM, subject to the export regulations of the United States government. With the CCTK, developers can build applications for the HSM, authenticate programs, and load programs into the HSM. The custom programming toolkit includes a custom software interface reference that describes the function calls that the applications that are running in the HSM use to obtain services from the HSM operating system, and from the HSM host system device driver. Another included reference provides a method for extending the Common Cryptographic Architecture (CCA) host application programming interface (API) and the API reference for the user-defined extensions programming environment. Finally, the Interactive Code Analysis Tool (ICAT) is provided so that developers can debug applications that are running on the HSM. Frequently, a custom contract provides consultation to hasten application development, and sometimes provides for initial development by IBM. Whenever needed, IBM can bid on developing your custom solution or extension.

Secure administration of HSMs

For the IBM 4769 and IBM 4767, IBM offers GUI-based utilities to administer the HSM cards, which include the loading of initial keys and the setup of the access control system. Each of these utilities can use smart cards as part of the administrative process to carry key parts securely, and to identify administrators and enable them to perform sensitive functions. On IBM Power servers running AIX (and on Intel x86-64 systems), the Smart Card Utility Program (SCUP), Cryptographic Hardware Initialization and Maintenance (CHIM), and (Cryptographic Node Management (CNM) (4767 only) utilities are provided with the HSM software.

The CHIM workstation connects through secure sessions to the cryptographic coprocessors to enable authorized personnel to perform the following tasks:

- View the coprocessor status.
- View and manage the coprocessor configuration.
- Manage coprocessor access control (user roles and profiles).
- Generate and load coprocessor master keys.
- Create and load operational key parts.

Figure 2-1illustrates the secure management of the IBM HSM cards.

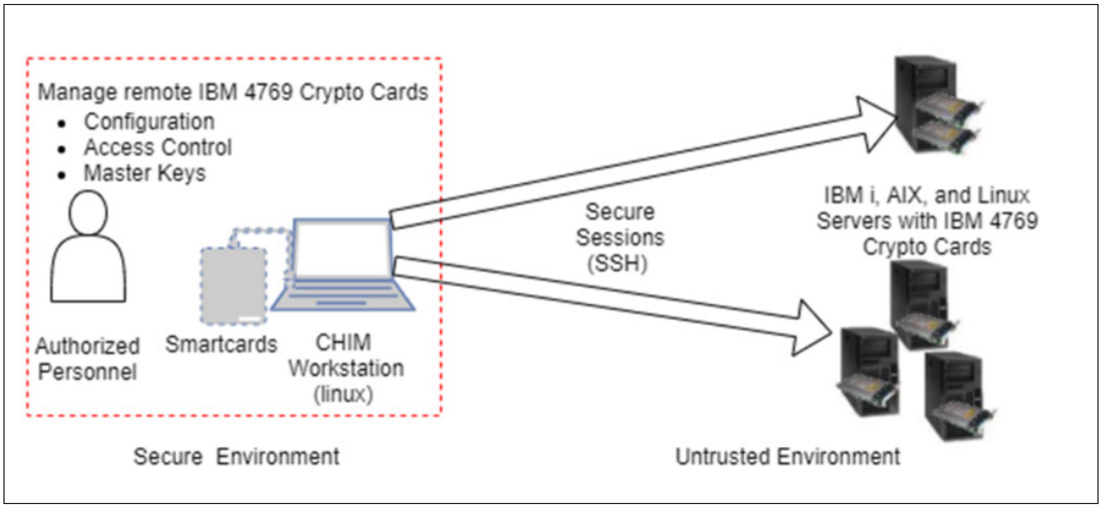


Figure 2-1 Secure management of IBM 4769 crypto cards

Cryptographic Hardware Initialization and Maintenance on IBM i

CHIM is a PCI-compliant interface to configure the IBM 4769 Cryptographic Coprocessor. With CHIM, you can securely manage remote IBM PCIe Cryptographic Coprocessors with a secure connection even in an environment where not all devices are trusted. Management tasks are done by using a specialized workstation (the CHIM workstation). CHIM uses smart cards for profile authentication and the storage of coprocessor master key parts.

IBM i CCA Cryptographic Service Provider (CSP), which is delivered as IBM i Option 35, supports IBM CHIM Catcher. This support is provided by IBM i Product Temporary Fixes (PTFs).

Here are the requirements to use CHIM to manage IBM 4769 Cryptographic Coprocessors in IBM i systems:

- The following products are installed (with the appropriate PTF levels):
  - 5770SS1 option 35 - CCA Cryptographic Service Provider
  - 5733CY3 - Cryptographic Device Manager
  - 5733SC1 option 1 - OpenSSH, OpenSSL, and zlib
- The Secure Shell (SSH) server daemon must be active (use `STRTCPSVR *SSHD`) and configured to allow local port forwarding from the CHIM workstation to the CHIM catcher port (which defaults to 50003) on localhost. The SSH daemon must have logging that is configured at the INFO level (the default) at a minimum.
- The CHIM catcher must be active (use `STRTCPSVR *CHIM`). The CHIM catcher will not start successfully if the previous requirements are not met.
- Cryptographic device descriptions must be created for each IBM 4769 Cryptographic Coprocessor that is managed (use `CRTDEVCRP`) and have the `*ACTIVE` status (use `VRYCFG` or `WRKCFGSTS`).
- The IBM i user profile that is used when you authenticate from the CHIM workstation must have **IOSYSCFG* special authority and have `*USE` authority for the cryptographic device descriptions for each IBM 4769 Cryptographic Coprocessor that is managed.

The CHIM catcher is controlled like all other TCP servers on IBM i. Use the `STRTCPSVR`, `ENDTCPSVR`, and `CHGTCPSVR` commands to manage the CHIM catcher. The server application value for CHIM is `*CHIM`. The CHIM catcher port is configured with service name “chim”, which is set to port 50003. The CHIM catcher listens only for incoming connections on localhost. The CHIM catcher ends itself if no server activity occurs for 1 hour.

Smart cards on Linux

For the 4769 and 4767, IBM provides the SCUP and CHIM applications to manage smart cards by using an IBM HSM. SCUP and CHIM run on x86-64 systems with Linux and can target IBM HSMs that are installed in x86-64 and Power servers that run AIX. Customers can use SCUP to initialize smart cards that can be used with CHIM to generate and store CCA master key parts on supported smart cards; load CCA master key parts that are stored on supported smart cards; and log on to CCA by using smart card CCA profiles that are tied to an RSA key pair that is associated with a particular smart card and user profile. Smart cards are available for purchase from IBM.

Benefits

IBM PCIe Cryptographic Coprocessors have the following benefits:

- Keep data safe and secure.

  Safeguard data with a tamper-responding design and sensors that protect against module penetration and power or temperature manipulation attacks.
- Choose your platform.

  Available on select IBM Z servers with z/OS or Linux; IBM LinuxONE Emperor with Rockhopper; IBM Power servers; and x86-64 servers with certain Red Hat Enterprise Linux (RHEL) releases.

> **Note:** At the time of writing, IBM Power supports both the 4769 and 4767 HSMs. The 4769 is available, but the 4767 was withdrawn from marketing.

Features

IBM Cryptographic Coprocessor cards provide the following features:

- High-end secure coprocessors

  Deliver high-speed cryptographic functions for data encryption and digital signing, and secure storage of signing keys or custom cryptographic applications.

- Highest level of certification: Federal Information Processing Standards (FIPS) PUB 140-2, Level 4

  Validated to FIPS PUB 140-2, Security Requirements for Cryptographic Modules, Overall Security Level 4, which is the highest level of certification that is achievable.

- Performance and architectural improvements

  IBM 4769 can exceed 23,000 personal identification number (PIN) conversion operations per second, contains custom symmetric key and hashing engines, and supports asymmetric algorithms.

- Tamper-responding design

  Sensors protect against various attacks on the system and immediately destroy all keys and sensitive data if tampering is detected.

- CCA Enterprise Public Key Cryptography Standards (PKCS) #11 APIs

  Performs cryptographic functions that are common in the finance industry and business applications, with custom functions that are available through a programming toolkit.

- Embedded certificate for external verification

  Generates a unique public or private key pair with a certificate that is stored in the device, with safeguards to ensure that the HSM is genuine and untampered.

The remainder of this section covers the 4769 Cryptographic Coprocessor, which is the available HSM option for IBM Power servers at the time of writing.

IBM 4769 Cryptographic HSM highlights

Each IBM HSM device offers the highest cryptographic security that is available commercially. FIPS publication 140-2 defines security requirements for cryptographic modules. It is issued by the US National Institute of Standards and Technology (NIST) and is widely used as a measure of the security of HSMs. The cryptographic processes of each IBM HSM are performed within an enclosure on the HSM that is designed to provide complete physical security.

The 4769 Cryptographic Coprocessor is a PCIe Generation 3 (Gen3) x4 adapter. The secure-key adapter provides both cryptographic coprocessor and cryptographic accelerator functions in a single PCIe card. The 4769 Cryptographic Coprocessor is suited to applications that require high-speed, security-sensitive, and RSA accelerated cryptographic operations for data encryption and digital signing. Also, the adapter is useful for secure management, cryptographic keys, or custom cryptographic applications. It provides secure storage of cryptographic keys in a tamper-resistant HSM that is designed to meet FIPS 140-2 level 4 security requirements. The adapter runs in dedicated mode only.

The IBM 4769 is available as Customer Card Identification Number (CCIN) C0AF (without a blind-swap cassette custom carrier) (Feature Code EJ35) and as CCIN C0AF (with blind-swap cassette custom carrier) (Feature Code EJ37) on IBM Power10 servers, either on IBM AIX, IBM i, or Linux (RHEL or SUSE Linux Enterprise Server) operating systems. It is also available as Feature Codes EJ35 and EJ37 on IBM Power9 servers, either on IBM AIX or IBM i.

Figure 2-2 shows an image of the 4769 Cryptographic Coprocessor.

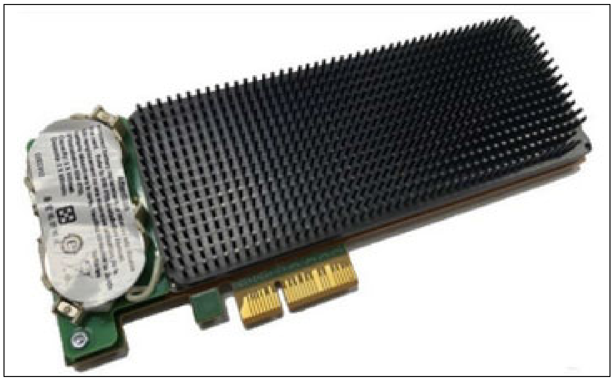


Figure 2-2 4769 Cryptographic Coprocessor

The IBM 4769 hardware provides significant performance improvements over its predecessors while enabling future growth. The secure module contains redundant IBM PowerPC® 476 processors and custom symmetric key and hashing engines to perform AES, DES, TDES, SHA-1 and SHA- 2, MD5 and HMAC, and public key cryptographic algorithm support for RSA and Elliptic Curve Cryptography. Other hardware support includes a secure real-time clock, a hardware random number generator, and a prime number generator. The secure module is protected by a tamper-responding design that protects against various attacks against the server.

IBM CEX7S / 4769

The IBM 4769 is validated by NIST to FIPS 140-2 Level 4, which is the highest level of certification that is achievable for commercial cryptographic devices.[1] FIPS 140 defines security requirements for cryptographic modules. It is issued by NIST and widely used as a measure of the security of HSMs.

The “Payment Card Industry HSM” standard, PCI HSM, is issued by the PCI Security Standards Council. It defines physical and logical security requirements for HSMs that are used in the finance industry. The IBM CEX7S with CCA 7.x has PCI HSM certification.[2]

The following attributes are provided by the 4769:

- Supported cryptographic mode: CCA.
- PPC 476 processors run in lockstep, and the outputs of each core are compared cycle by cycle.
- Error Checking and Correction protection on DDR3 memory.
- Cryptographic key generation and random number generation.
- Over 300 cryptographic algorithms and modes.
- Byte-wide parity protection on all internal registers and data paths wider than two bits.
- RSA and Error Correcting Code (ECC) engines are protected by a duplicate engine that predicts the CRC of the result.
- SHA, MD5, AES, and DES engines are protected by running the same operation on two independent engines and the outputs are compared cycle by cycle.

[1] Source: https://csrc.nist.gov/projects/cryptographic-module-validation-program/certificate/4079

[2] Source: https://listings.pcisecuritystandards.org/popups/pts_device.php?appnum=4-20358

The IBM 4769 is designed for improved performance and security-rich services for sensitive workloads, and to deliver high throughput for cryptographic functions. For more information about the capabilities and specifications of the IBM 4769, see the IBM 4769 Data Sheet.

- Reliability, availability, and serviceability (RAS)

  The hardware supports the highest level of RAS requirements, which enables the secure module to self-test at any time. This self-test is achieved by running a pair of PowerPC processors in lockstep and comparing the result from each cycle by cycle. Also all interfaces, registers, memory, cryptographic engines, and buses are protected always by using parity, ECC, or CRC. Power-on self-tests that are securely stored inside the secure module verify that the hardware and firmware that is loaded on the module is secure and reliable at every power-on. Then, the built-in RAS features check it continuously in real time.

- Embedded certificate

  During the final manufacturing step, the coprocessor generates a unique public/private key pair that is stored in the device. The tamper detection circuitry is activated now and remains active throughout the useful life of the coprocessor, protecting this private key and other keys and sensitive data. The public key of the coprocessor is certified at the factory by an IBM private key, and the certificate is retained in the coprocessor. Then, the private key of the coprocessor is used to sign the coprocessor status responses that, with a series of public key certificates, demonstrate that the coprocessor remains intact and is genuine.

- Tamper responding design

  NIST validates that the IBM 4769 HSM meets the FIPS 140-2 Level 4 requirements to protect against attacks that include probe penetration or other intrusions into the secure module, such as side-channel attacks, power manipulation, and temperature manipulation. From the time of manufacture, the hardware self-protects by using tamper sensors to detect probing or drilling attempts. If the tamper sensors are triggered, the HSM destroys critical keys and certificates, and is rendered permanently inoperable. Therefore, the HSM must always be within the specified temperature, humidity, and barometric pressure ranges.

2.2 Compliance automation and real-time monitoring

IBM Power10 servers offer tools to automate compliance tasks and monitor your IT environment in real time, which helps you meet regulations and stay secure. You can use IBM PowerSC for automated compliance management. For more information, see 9.1, “Compliance automation” on page 238.

PowerVM, the virtualization management tool for IBM Power, provides real-time monitoring of virtualized environments. It helps track performance, resource utilization, and security status across VMs and physical servers. Also, the Hardware Management Console (HMC) offers real-time monitoring and management of IBM Power servers. It provides insights into system health, performance metrics, and potential security issues.

Power10 servers can be configured to generate real-time alerts for various events, including security incidents, system performance issues, and hardware faults. These alerts can be integrated with enterprise monitoring solutions for centralized management. Integration with Security Information and Event Management (SIEM) systems enables real-time analysis of security events and incidents, which helps in detecting and responding to potential threats as they occur.

IBM Power10 servers provide a robust framework for compliance automation and real-time monitoring through integrated tools and features. By using solutions like IBM PowerSC, advanced monitoring tools, and real-time alert systems, organizations can ensure continuous compliance with security standards, automate policy enforcement, and monitor system performance and security in real time. This combination of capabilities helps maintain a secure and compliant IT environment, which reduces risks and enhances operational efficiency.

2.3 Endpoint detection and response

Endpoint detection and response (EDR) is a system that monitors and analyzes security threats from endpoints, such as computers and mobile devices. It uses machine learning and analytics to identify patterns that indicate suspicious activity or known threats in real time. The goal of EDR is to find security breaches as they happen and respond quickly to potential or discovered threats. An EDR solution proactively and automatically blocks and isolates malware while equipping security teams with the right tools to confidently deal with these challenges. A modern EDR can ensure business continuity by effectively mitigating fast-growing, automated, and advanced threats, such as ransomware or other attacks, without increasing analyst workloads or requiring highly skilled security specialists.

Here are some options within the IBM Power ecosystem to support EDR:

- IBM PowerSC is a security and compliance solution that is optimized for virtualized environments on IBM Power servers running AIX, IBM i, or Linux. PowerSC sits on top of the IBM Power server stack, which integrates security features that are built at different layers. You can now centrally manage security and compliance on Power servers for all IBM AIX and Linux on Power endpoints.

  For more information, see 9.3, “Endpoint detection and response” on page 239.
- IBM Security® QRadar® EDR remediates known and unknown endpoint threats in near real time with intelligent automation that requires little-to-no human interaction.

  For more information, see “IBM QRadar Suite (Palo Alto Networks)” on page 268.

2.4 Malware prevention technologies

ROP is a type of software attack technique that exploits vulnerabilities in programs that do not properly validate user-supplied input. By carefully crafting input data, attackers can manipulate the program's control flow to run arbitrary code, which often leads to severe consequences like remote code execution.

The Power10 processor architecture incorporates several features to enhance control flow security and mitigate the risk of ROP attacks. These features include improved hardware-based encryption and advanced protection against side-channel attacks. Although these features can mitigate some attacks, they do not make ROP attacks impossible, but rather more challenging.

Modern compilers and operating systems for Power10 can include extra security features and mitigations. Developers should ensure that their software is built with the latest security practices and that the operating system is up to date with relevant patches.

2.5 Secure Boot and Trusted Boot

The ability to boot a system into a known and verifiable state is a requirement for protecting your systems form unauthorized code and malware. Although these terms are sometimes incorrectly used interchangeably, they are distinct concepts that can coexist and are complementary.

Secure Boot protects the initial program load (IPL) by ensuring that only authorized modules (ones that are cryptographically signed by the manufacturer) are loaded during the boot process.

Trusted Boot starts with the platform that is provided by the Secure Boot process and then builds on it by recording measurements of the system's firmware and configuration during the startup process to the Trusted Platform Module (TPM). Through an attestation process, the TPM can provide a signed quote that can be used to verify the system firmware integrity at any time.

Figure 2-3 illustrates how the different layers work together to support Secure Boot within Power10.

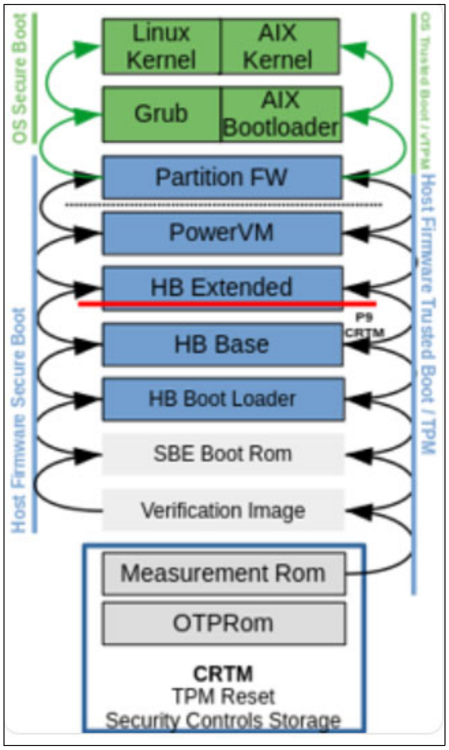


Figure 2-3 Secure Boot architecture

Power Secure Boot and Trusted Boot use the following components:

- Firmware Secure Boot

  Integrity validation of all firmware components from the hardware root of trust up through PowerVM and partition firmware (PFW).

- Firmware Trusted Boot

  Firmware measurements that are recorded to the TPM from the hardware root of trust up through PowerVM and PFW.

- OS Secure Boot

  Integrity validation of the OS boot loader starting from PFW. This feature is configurable from the HMC.

- OS Trusted Boot by using a virtual Trusted Platform Module (vTPM)

  OS measurements that are recorded to the vTPM starting from PFW. This feature is configurable from the HMC.

2.5.1 Secure Boot in PowerVM

IBM Power servers provide a highly secure server platform. IBM Power9 and IBM Power10 processor-based hardware and firmware include PowerVM features to provide a more secure platform for cloud deployment. Secure Boot seeks to prevent unauthorized access to customer data. This unauthorized access might come from unauthorized firmware running on a host processor; by using security vulnerabilities in authorized service processor firmware; or by using hardware service interfaces from the Flexible Service Processor (FSP).

Secure Boot does not provide protection against the following attacks:

- OS-software based attacks to gain unauthorized access to customer data
- Rogue system administrators
- Hardware physical attacks (for example, chip substitutions, or bus traffic recording)

Secure Boot implements a processor-based chain of trust that is based in the IBM POWER® processor hardware and enabled by the IBM Power firmware stack. Secure Boot provides for a trusted firmware base to enhance the confidentiality and integrity of customer data in a virtualized environment.

Secure Boot establishes trust through the platform boot process. With Secure Boot, the system starts in a trusted and defined state. *Trusted* means that the code that runs during the IPL process originates from the platform manufacturer, and is signed by the platform manufacturer and has not been modified since. For more information about Secure Boot processing in PowerVM, see Secure Boot in IBM Documentation or this Secure Boot PDF.

2.5.2 Secure Boot in AIX

The AIX Secure Boot feature extends the chain of trust to the AIX logical partition (LPAR) by digitally verifying the following AIX and PFW codes:

- OS boot loader
- Kernel
- Runtime environment
- Device drivers, which include boot device drivers
- Kernel extensions
- Applications
- Libraries

The AIX boot image includes digital signatures of the boot loader and kernel so that the PFW can validate them. The PFW also validates the digital signature of the boot code in the adapter microcode. If an adapter's boot code lacks a valid digital signature, it cannot be used as a boot device for the trusted LPAR.

The boot loader validates the kernel's digital signature. The AIX Secure Boot feature uses Trusted Execution (TE) technology, which relies on the Trusted Signature Database (TSD), which stores digital signatures of device drivers, application binary files, and other AIX codes. The feature checks the integrity of boot and initialization codes up to the end of the `inittab` file.

The AIX Secure Boot feature includes the following enhancements:

- The feature starts validating code integrity before the TE feature by loading the TSD earlier in the boot process before the kernel loads the first application.
- At run time, the TE feature verifies the cryptographic hashes of the boot and initialization codes.
- The feature verifies the digital signatures of codes that must run.

The AIX Secure Boot feature is configured by using the management console, with the HMC supporting it.

The AIX operating system supports the following basic Secure Boot settings:

0 Secure Boot disabled.

1 Enabled (or log only).

2 Enforce (abort the boot operation if signature verification fails).

3 Enforce policy 2 and avoid loading programs or libraries that are not found in TSD, which also disables write access to `/dev/*mem` devices.

4 Enforce policy 3 and disable the kernel debugger (KDB).

If file integrity validation fails during the boot operation in Audit mode, the LPAR continues to boot, but the system administrator logs errors in `/var/adm/ras/Secure Bootlog` for inspection after the LPAR starts. When digital signature verification of files fails during the boot in Enforce mode, the boot process aborts, and the LPAR status is displayed in the HMC with a specific LED code.

2.5.3 Secure Boot in Linux

The Linux Secure Boot feature extends the chain of trust to the Linux LPAR by digitally verifying the following Linux and PFW codes:

- GRUB bootloader
- Linux kernel
- Device drivers, including boot device drivers

Linux boot images are signed by distributions like Red Hat and SUSE so that PFW can validate them by using PKCS7 (Cryptographic Message Syntax (CMS)). PKCS7 is one of the PKCS family of standards that was created by RSA Laboratories. and it is a standard syntax for storing signed or encrypted data. PowerVM includes the public keys that are used by PFW to validate the GRUB boot loader.

The PFW verifies the appended signature on the GRUB image before handing control to GRUB. Similarly, GRUB verifies the appended signature on the kernel image before starting the OS to ensure that every image that runs at boot time is verified and trusted.

Figure 2-4 shows how Secure Boot works with Linux.

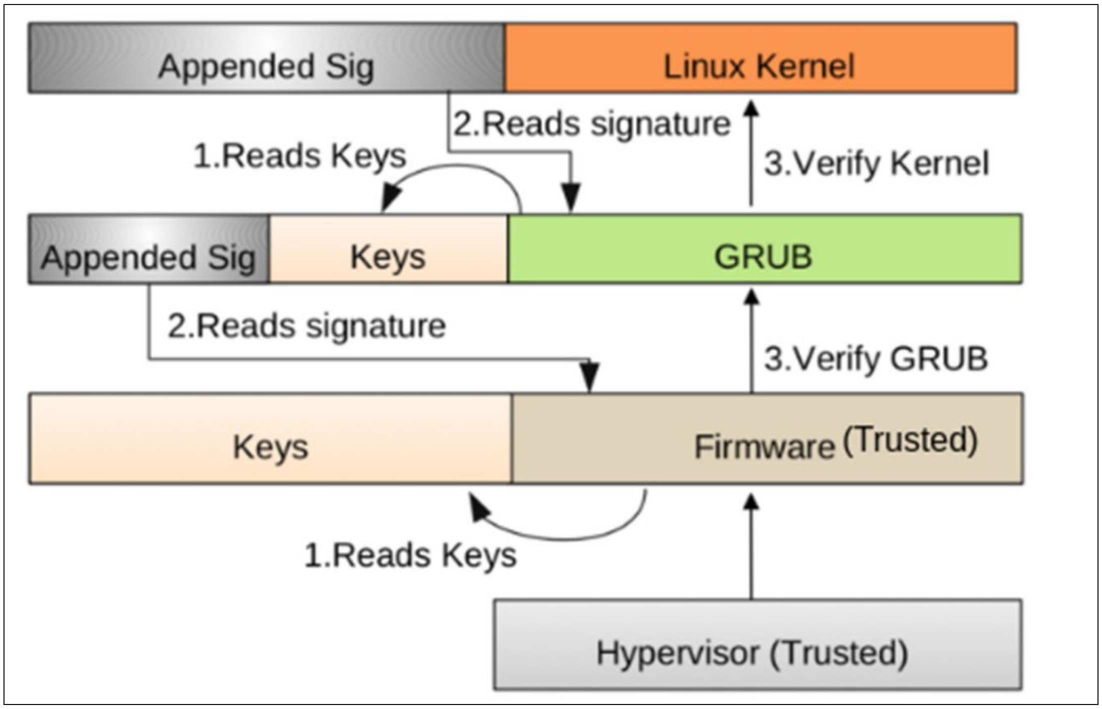


Figure 2-4 Secure Boot process in Linux

Limitations

Here are the limitations for Secure Boot with Linux at the time of writing:

- Key rotations for the GRUB or kernel require a complete firmware update.
- Administrators cannot take control of the LPAR and manage their keys.
- User-signed custom builds for kernel or GRUB do not start by using static key management.
- Secure Boot enables lockdown in the kernel to restrict direct or indirect access to the running kernel, which protects against unauthorized modifications to the kernel or access to sensitive kernel data.
- Lock-own impacts some of the IBM Power platform functions that are accessible by using the user space RTAS interface.

Turning on Secure Boot for Linux

Administrators can configure Secure Boot from the HMC for each LPAR. The default setting is Disabled. This setting is available under Advanced Settings.

The HMC provides three Secure Boot modes:

- Disabled
- Enabled and log only
- Enabled and enforced

Linux supports two out of these three modes:

- Disabled
- Enabled and enforced

When the mode is set to Disabled, Secure Boot is not activated. When digital signature verification fails during the boot in Enforce mode, the boot process aborts, and the LPAR status appears in the HMC as a specific LED code.

Table 2-2 shows the supported combinations of firmware and Linux distribution.

Table 2-2 Supported firmware and Linux distributions

Firmware release version	Distribution release version	Key management mode
FW 1010	► Red Hat Enterprise Linux (RHEL) 9.2, 9.3, and 9.4 ► SUSE Linux Enterprise Server 15 Service Pack (SP) 4, 5, and 6	Static
FW 1020	► RHEL 9.2, 9.3, and 9.4 ► SUSE Linux Enterprise Server 15 SP 4, 5, and 6	Static
FW 1030	► RHEL 9.2, 9.3, and 9.4 ► SUSE Linux Enterprise Server 15 SP 4, 5, and 6	Static
FW 1040	► RHEL 9.2, 9.3, and 9.4 ► SUSE Linux Enterprise Server 15 SP 4, 5, and 6	Static
FW 1050	► RHEL 9.2, 9.3, and 9.4 ► SUSE Linux Enterprise Server 15 SP 4, 5, and 6	Static[a]
FW 1060	► RHEL 9.2, 9.3, and 9.4 ► SUSE Linux Enterprise Server 15 SP 4, 5, and 6	Static[a]

[a] HMC provides an option to enable dynamic key management, but it is not supported at the time of writing. Dynamic key management is expected in future Linux distributions.

Table 2-3 shows the supported levels of GRUB.

Table 2-3 Supported levels of GRUB

Firmware version	Supported GRUB verification keys
FW1010.00 FW1010.10	Red Hat Secure Boot 602 Serial Number=00d39c4133dd6b5f45 KeyID=e86a1cab2c48f96036a2f07b8ed29db42a2898c8
FW1010.20	Red Hat Secure Boot 602 Serial Number=00d39c4133dd6b5f45 KeyID=e86a1cab2c48f96036a2f07b8e SUSE Secure Boot signing key 20210225 Serial Number=00ed8785b78ffc127e KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1010.30 FW1010.40 FW1010.50	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20220525 Serial Number=00ed8785b78ffc1280 KeyID=f33fa22ef28fcb9dc18d43d20bc7ef65c1c565e4

Firmware version	Supported GRUB verification keys
FW1010.60 and later	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1020.00 FW1020.10 FW1020.30	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20220525 Serial Number=00ed8785b78ffc1280 KeyID=f33fa22ef28fcb9dc18d43d20bc7ef65c1c565e4
FW1020.40 and later	Red Hat Secure Boot 702 Serial Number =00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1030.00 FW1030.10	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20220525 Serial Number=00ed8785b78ffc1280 KeyID=f33fa22ef28fcb9dc18d43d20bc7ef65c1c565e4
FW1030.20 FW1030.30 FW1030.40 FW1030.50	Red Hat Secure Boot 702 Serial Number =00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1030.60 and later	Red Hat Secure Boot 702 Serial Number =00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f SUSE Secure Boot signing key 20230510 Serial Number=00cafcb5d75ec58983 KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1040.00 and later	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1050.00 FW1050.10	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f

Firmware version	Supported GRUB verification keys
FW1050.20 and later	Red Hat Secure Boot 702 Serial Number =00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f SUSE Secure Boot signing key 20230510 Serial Number=00cafcb5d75ec58983 KeyID=ecab0d42c456cf770436b973993862965e87262f
FW1060.00 and later	Red Hat Secure Boot 702 Serial Number=00e41f32362a936b1e KeyID=c442130fde4c50fa1686bbf0692e3ebc64f5db3e SUSE Secure Boot signing key 20230301 Serial Number=00cafcb5d75ec58982 KeyID=ecab0d42c456cf770436b973993862965e87262f SUSE Secure Boot signing key 20230510 Serial Number=00cafcb5d75ec58983 KeyID=ecab0d42c456cf770436b973993862965e87262f

For more information about Secure Boot with Linux, see Guest Secure Boot with static keys in IBM Documentation.

2.5.4 Trusted Boot

Trusted Boot is complementary to Secure Boot. After the Secure Boot process verifies the integrity of the firmware components of the system, the system measures and extends a hash of the firmware component to the TPM. The firmware component starts only after the integrity validation and measurement successfully complete. Trusted Boot helps organizations meet security compliance requirements by helping ensure that their systems are running trusted software and firmware.

Here are the key components of any Trusted Boot process:

- Measurement

  Trusted Boot records measurements of the system’s configuration and a hash of each component in the boot process, from the boot loader to the operating system kernel.

- Root of Trust

  Trusted Boot relies on a Root of Trust, which is a secure, hardware-based mechanism that serves as the foundation for all other trust decisions. It is typically implemented in a TPM or similar secure hardware component.

- Chain of Trust for Measurement

  The Trusted Boot process creates a Chain of Trust for Measurement, where each successive component in the boot sequence is measured against the TPM before passing control to it. This component helps ensure that the entire boot process is recorded securely.

- Attestation

  The TPM provides an attestation mechanism that provides a signed quote that can be used to verify the system integrity and firmware configuration at any time.

For AIX, the Trusted Boot function is handled by the TE functions, as described in 4.9, “Trusted Execution” on page 115.

2.6 Hypervisor security

Hypervisors are the backbone of virtualized environments. They require robust security measures to protect against various threats and vulnerabilities that can compromise the entire infrastructure. This section delves into the security aspects of hypervisors, detailing the potential attack vectors, security best practices, and the latest technologies for safeguarding these critical components.

Hypervisor vulnerabilities

The first step in securing hypervisors is understanding the unique vulnerabilities that they face. The following sections describe some of these vulnerabilities and how to avoid them in your IBM Power environment.

Hyperjacking

`Hyperjacking` is a type of advanced cyberattack where threat actors take control of the hypervisor, which handles the virtualized environment within a main computer system. The actors ultimate aim is to deceive the hypervisor into running unauthorized tasks without leaving traces elsewhere on the computer.

Hyperjacking involves exploiting a vulnerability in a user's browser extension or add-on to gain control over their computer without their knowledge. By hijacking the user session, cybercriminals can conduct surveillance, manipulate connected devices remotely, and potentially steal sensitive information.

Virtual machine escape

VM escape is a security vulnerability that lets attackers breach VMs and obtain unauthorized access to the underlying physical hardware, such as the hypervisor or host system. It circumvents the virtualization layer's isolation barriers, enabling potential exploits.

IBM Power and PowerVM provide many options to help prevent a VM escape type attack. PowerVM has excellent LPAR isolation that prevents an LPAR from seeing resources outside of its defined VM.

Resource exhaustion

Attackers can target the resource allocation features of a hypervisor, which can lead to a denial of service (DoS) by exhausting resources such as CPU and memory, which affects all VMs that are hosted on the hypervisor.

Within PowerVM, when a resource is defined to an LPAR there are limits that are enforced by the hypervisor to protect them from overallocation. Defining the minimum, maximum, and requested values for memory and CPU correctly help ensure that you avoid resource exhaustion attacks.

Ensure that the administrator credentials that used for configuring PowerVM are protected and use role-based access control (RBAC) to limit the scope of changes that can be made.

Protecting from hypervisor vulnerabilities

To mitigate the risks that are associated with hypervisor vulnerabilities, use the same strategies that are used to protect other components in your environment:

- Turn on Secure Boot.

  The IBM PowerVM hypervisor is built into the firmware on the server. Hypervisor Secure Boot, which cannot be disabled without a significant amount of effort, protects you by preventing a bad actor from inserting a false hypervisor into the system.

- Protect access to PowerVM.

  Most hypervisor intrusions are due to an actor gaining access to the hypervisor by using exposed credentials. Implement good authentication control and define users who are allowed to configure PowerVM, and use RBAC to limit their access to resources.

- Protect access to the VMs.

  Ensure that you follow best practices for protecting each VM. Manage credentials, use multi-factor authentication (MFA), and define user credentials with the minimum capabilities that they need to do their job.

- Plan for isolation of networks and storage.

  Protect your VMs by isolating network traffic between VMs to prevent eavesdropping and network attacks. Use techniques such as VLANs, firewall rules, and virtual network appliances.

  Isolate storage access among VMs to prevent data leakage or corruption. Use separate storage accounts for sensitive data and implement robust access controls. Use encryption to avoid improper access to storage.

  Isolate your VMs from the management network that is used by the HMC and service processor to mitigate risks against the management infrastructure.

- Maintain currency in your firmware.

  Check for new firmware versions and schedule any updates regularly.

- Maintain currency in your VM operating systems.

  Monitor updates that are available for the operating systems in your VMs and schedule any updates regularly.

For more information about securing your PowerVM environment, see 3.1, “Hardware Management Console security” on page 52 and 3.3, “VIOS security” on page 72.

2.6.1 LPAR isolation

Logical partitioning is a technology that is used primarily in enterprise computing environments to divide a computer's resources, such as CPU, memory, and storage, into multiple, separate VMs. Each LPAR operates as a stand-alone environment with its own operating system and applications, which makes it an invaluable tool for optimizing resource use, improving system security, and increasing availability in Power servers.

LPAR isolation is a basic tenet of IBM PowerVM, which is the IBM Power hypervisor. PowerVM can share resources from a single machine across all LPARs that are defined on that machine. Also, PowerVM has extra capabilities that allow LPARs to be non-disruptively moved from one host machine to another one to provide load balancing and high availability (HA) configurations. LPAR restart technologies support disaster recovery (DR) options to restart workloads at another site if there is a site failure.

One of the strengths of PowerVM is its flexibility in sharing processing resources. CPUs can be defined to an LPAR as dedicated or shared, capped or uncapped, and donating or not donating. You effectively allocate resources among LPARs, which helps ensure that each partition receives the necessary resources to perform optimally without affecting the performance of others. This approach uses dynamic resource allocation techniques that enable you to reallocate resources based on workload demands.

3

Security in the virtualization and management layer

Virtualization is a keystone of the IBM Power ecosystem. Clients use the performance, reliability, and security that is built in to IBM Power servers while they reduce the cost and complexity of running isolated workloads on dedicated machines. Virtualization by definition involves sharing hardware infrastructure across multiple workloads. This approach provides significant benefits but also creates challenges for keeping different workloads isolated and secure.

The task of managing the virtualization layer in the IBM Power ecosystem is divided into two distinct areas: hardware management, and I/O virtualization. The hardware management aspect is handled by the Hardware Management Console (HMC), which is an appliance that defines the logical partitions (LPARs) in each server, dividing and sharing the installed resources across the various virtual machines (VMs) that are supported. An HMC can manage multiple servers, but as your infrastructure grows across multiple locations and many servers, the Cloud Management Console (CMC) provides a single tool for consolidating information across several HMCs.

The Virtual I/O Server (VIOS) is a special partition that runs in an IBM Power server that shares physical devices across multiple LPARs. The purpose of the VIOS is to virtualize the physical adapters in the system to reduce the number of adapters. Systems with virtualized I/O can move to other servers as needed for load-balancing and high availability (HA) during planned or unplanned outages, which provides a more available and resilient environment.

This chapter describes the following topics:

- 3.1, “Hardware Management Console security” on page 52
- 3.2, “Cloud Management Console security” on page 64
- 3.3, “VIOS security” on page 72

3.1 Hardware Management Console security

The HMC is a specialized device for configuring and managing IBM Power servers. It facilitates basic virtualization management by supporting the setup of LPARs and dynamic resource allocation, which includes adjustments to processor and memory settings for IBM Power servers. Also, the HMC offers advanced service functions such as guided repair and verification, concurrent firmware updates for managed systems, and continuous error reporting through the Electronic Service Agent for expedited support. The latest model, the 7063-CR2, operates on an IBM Power9 server. This dedicated device is exclusively used for controlling and servicing IBM Power servers and cannot be used as a general-purpose computing resource.

HMC packaging

Initially, the HMC was delivered solely as a traditional hardware appliance, with the software and hardware bundled together and installed onsite. As client environments grew, there was a demand to virtualize the HMC function to minimize infrastructure needs. In response, IBM introduced the virtual Hardware Management Console (vHMC), where you can use your own hardware and server virtualization to host the IBM-provided HMC virtual appliance. The vHMC image is available for both x86 and IBM Power servers and supports the following hypervisors:

- For x86 virtualization:
  - Kernel-based Virtual Machine (KVM) on Ubuntu 18.04 LTS or Red Hat Enterprise Linux (RHEL) 8.0 or 9.0
  - Xen on SUSE Linux Enterprise Server 12
  - VMware ESXi 6.5, 7.0, or 7.0.2
- For Power virtualization: PowerVM

The distribution of HMC Service Packs (SPs) and fixes is consistent for both hardware and vHMCs. However, for vHMC on PowerVM, Power firmware updates are managed by IBM. For vHMC on x86 systems, if security vulnerabilities arise, consult with the hypervisor and x86 system vendors for any necessary updates to the hypervisor and firmware. The steps for enabling Secure Boot differ between hardware and vHMCs due to architectural differences. For more information and detailed instructions about enabling the Secure Boot function, see 3.1.9, “Secure Boot” on page 61.

For more information about the vHMC, see Virtual HMC appliance (vHMC) Overview.

HMC functions

With the HMC, you can create and manage LPARs, which include the ability to dynamically add or remove resources from active partitions. The HMC also handles advanced virtualization functions such as Capacity Upgrade on Demand and Power Enterprise Pools.

Also, the HMC provides terminal emulation for the LPARs on your managed systems. You can connect directly to these partitions from the HMC or configure it for remote access. This terminal emulation feature helps ensure a reliable connection, which is useful if other terminal devices are unavailable or not operational. It is particularly valuable during the initial system setup before you configure your preferred terminal.

By using its service applications, the HMC communicates with managed systems to detect, consolidate, and relay information to service and support teams for analysis. For a visual representation of how the HMC integrates into the management and serviceability of IBM Power servers, see Figure 3-1 on page 53.

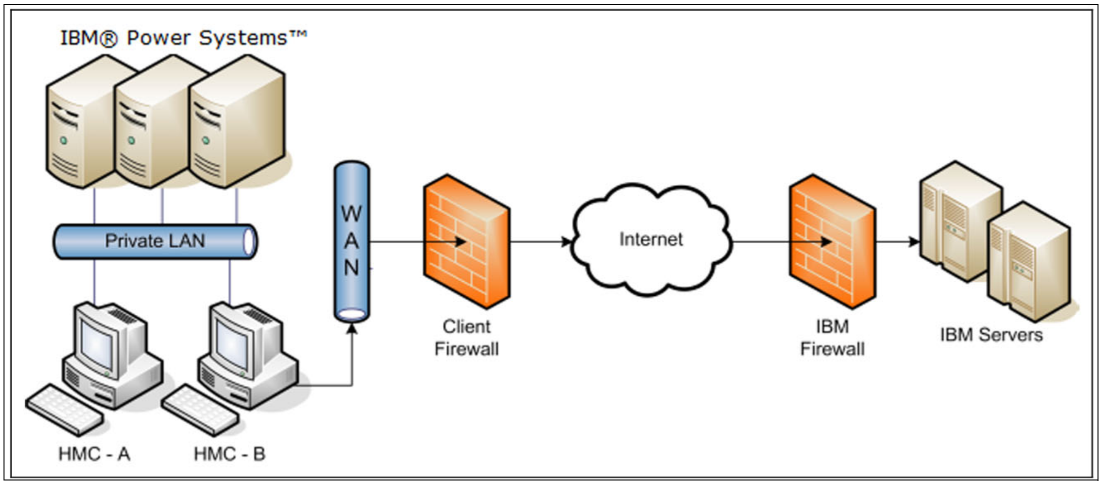


Figure 3-1 The HMC that is used for configuration and serviceability functions

One HMC can oversee multiple servers, and multiple HMCs can connect to a single server. If a single HMC fails or loses connection to the server firmware, the server continues to operate normally, but changes to the LPAR configuration is not possible. To mitigate this situation, you can connect an extra HMC as a backup to help ensure a redundant path between the server and service and support.

Each HMC comes preinstalled with the HMC Licensed Machine Code to help ensure consistent function. You have two options for configuring HMCs to provide flexibility and availability:

- Local HMC

  A local HMC is situated physically close to the system that it manages and connects by using a private or public network. In a private network setup, the HMC acts as a DHCP server for the system's service processors. Alternatively, it can manage the system over an open network, where the service processor's IP address is manually assigned through the Advanced System Management Interface (ASMI).

- Remote HMC

  A remote HMC is located away from its managed systems, which might be in a different room, building, or even a separate site. Typically, a remote HMC connects to its managed servers over a public network, although it can also be configured to connect through a private network.

IBM created a document that provides a starting point about understanding the connectivity that is used by the HMC and how to make it secure. The HMC 1060 Connectivity Security white paper is a good starting point for enabling a secure HMC environment in your enterprise.

3.1.1 Security levels for the HMC

The HMC offers REST, GUI, and command-line interface (CLI) interfaces for user interaction. Security requirements vary depending on the model and the interface that is used. The HMC provides recommendations for different levels of security, which are outlined in this section.

Level 1

Level 1 defines the security actions that you must have. Here are the minimum measures that are recommended to secure your HMCs:

- Change the default password of the `hscroot` user.
- Enable the GRUB password if your HMC is not in a physically secure environment by running the following command:

  ```
  chhmc -c GRUBpasswd -s enable --passwd <new GRUB password>
  ```

- If you configured the Integrated Management Module (IMM) on the HMC, set a strong IMM password.
- Set a strong password for the admin and general users on all servers.
- Keep HMC updated with all released security fixes. Fixes are available at IBM Fix Central.

Level 2

Level 2 defines some actions that you should consider when you have multiple HMC users that are defined in the environment. If you have multiple users that use the HMC, consider the following items:

- HMC supports fine-grained control of resources and roles. Create an account for each user on the HMC.
- Assign only the necessary roles to users.
- Assign only necessary resources (systems, partitions, and others) to users.
- Both resources and roles that are assigned to the users must follow the least privilege principle. Create custom roles if necessary.
- Enable user data replication between HMCs with different modes.
- Import a certificate that is signed by a certificate authority (CA).
- Enable Secure Boot.
- Enable multi-factor authentication (MFA).
- Enable a PowerSC profile.

Level 3

Level 3 defines extra considerations when you have multiple HMCs in the environment. If you have many HMCs and sysadmins, consider the following items:

- Use centralized authentication by using Lightweight Directory Access Protocol (LDAP) or Kerberos (HMC does not support the single sign-on (SSO) feature).
- Enable user data replication between HMCs.
- Put HMC in National Institute of Standards and Technology (NIST) SP 800-131A mode so that it uses only strong ciphers.
- Block unnecessary ports in a firewall.

For more information, see this IBM Document on HMC Security.

3.1.2 Port security

The HMC is primarily a Java-based application running on Linux that uses various open-source components. It communicates with different services through various ports, which are encrypted by using Transport Layer Security (TLS) 1.2 or later. For remote access, expose only the following ports:

- Secure Shell (SSH) (port 22)
- HTTPS (port 443)
- VTerm (port 9960)

All other ports should be kept within a private or isolated network for security purposes.

3.1.3 Securing connections to Power servers

The HMC connects to managed systems through an integrated service adapter that is built in to the system. Depending on the model of the Power server, this adapter might be a Flexible Service Processor (FSP) or an Enterprise Baseboard Management Controller (eBMC). Connectivity varies slightly based on the endpoint:

- For FSP and a hypervisor: Management uses a proprietary binary protocol that is known as NETC, with communication encrypted by using TLS 1.2 or later.
- For eBMC: The connection is established through the Redfish REST application programming interface (API), with encrypted communication that is supported by TLS 1.2 or later, and appropriate certificates.

3.1.4 NIST SP 800-131A compliance mode

The HMC can operate in two modes: legacy and nist_sp800_131a. Once you set the HMC as the nist_sp800_131a mode, only strong ciphers that are listed by NIST SP 800-131A are used.

To set the NIST SP 800-131A mode, run the following command on the HMC:

```
chhmc -c security -s modify --mode nist_sp800_131a
```

If you want to return the HMC to legacy mode, run the following command:

```
chhmc -c security -s modify --mode legacy
```

3.1.5 Encryption

All communication channels that are used by the HMC are encrypted. By default, the HMC employs TLS and HTTPS with secure cipher sets that are bundled with the HMC. The default ciphers provide strong encryption and are used for secure communication on ports 443, 17443, 2301, and 5250 proxy, and for internal HMC communication.

> **Note:** For more information about the encryption ciphers that are used by the HMC, run the `lshmcencr` command in the HMC CLI. If your organization's corporate standards require different ciphers, use the `chhmcencr` command to modify them. For more information, see Managing the HTTPS ciphers of the HMC web interface by using the HMC.

The HMC supports both self-signed and CA-signed certificates for encryption. Starting with HMC 10.2.1040.0 and later, you can select the key size for certificates when generating a certificate signing request (CSR), with options for 2048 bits, 3072 bits, or 4096 bits. When using CA-signed certificates, use 2048-bit RSA encryption at a minimum. By default, the HMC uses a self-signed certificate with the SHA256 algorithm and 2048-bit RSA encryption.

3.1.6 Certificate management

Security certificates are crucial for helping ensure that the HMC operates securely in client/server mode, where the managed machines act as servers and the managed users are clients. Communication between the server and client occurs over the Secure Sockets Layer (SSL) protocol, which provides server authentication, data encryption, and data integrity.

When a user seeks remote access to the HMC user interface through a web browser, they initiate a request for the secure page by using `https://<hmc_hostname>`. Then, the HMC presents its certificate to the remote client (web browser) during the connection process. The browser verifies the certificate by checking that it was issued by a trusted authority, that it is still valid, and that it was specifically issued for that HMC.

3.1.7 User management

On an HMC, a user can be a member of various task roles. Each task role enables the user to access different parts of the HMC and to perform different tasks on the managed system.

HMC task roles are either predefined or customized. When you create an HMC user, you must assign a task role to that user. Each task role grants the user varying levels of access to tasks that are available on the HMC interface. You can assign managed systems and LPARs to individual HMC users so that you can create a user that has access to managed system A but not to managed system B. Each grouping of managed resource access is called a *managed resource role*.

Table 3-1 lists the predefined HMC task roles, which are the defaults on the HMC.

Table 3-1 HMC predefined roles

Task role	Description
`hmcservicerep`	A service representative is an employee who is at your location to install, configure, or repair the system.
`hmcviewer`	A viewer can view HMC information, but cannot change any configuration Information.
`hmcoperator`	The operator is responsible for daily system operation.
`hmcpe`	A product engineer helps support situations but cannot access HMC user management functions. To provide support access for your system, you must create and administer user IDs (UIDs) with the product engineer role.
`hmcsuperadmin`	The super administrator acts as the root user or manager of the HMC system. The super administrator has unrestricted authority to access and modify most of the HMC system.

You can create customized HMC task roles by modifying predefined HMC task roles. Creating customized HMC task roles is useful for restricting or granting specific task privileges to a certain user.

Authentication

User authentication is the first step to protecting your HMC and helping ensure that only authorized users can access the management console. The HMC supports various authentication methods to validate users:

- Local Authentication

  If you use Local Authentication, then the password and the number of days that the password is valid must be set.

- Kerberos Authentication

  If you use Kerberos Authentication, specify a Kerberos remote UID and configure the HMC to use Kerberos. When a user logs in to the HMC, authentication is first verified against a local password file. If a local password file is not found, the HMC can contact a remote Kerberos server for authentication. Configure your HMC so that it uses Kerberos remote authentication.

  For more information about setting up Kerberos, see IBM Documentation.

- LDAP Authentication

  If you use LDAP Authentication, configure the HMC to use LDAP server. For more information about configuring the HMC for LDAP, see IBM Documentation.

  To indicate whether the HMC should automatically manage remotely authenticated LDAP users, use the Automanage Authentication option. Valid values are 0 to disable automatic management, or 1 to enable automatic management.

  When Automatic Management is enabled, an LDAP user can log in to the HMC. An HMC user is automatically created for the LDAP user if the HMC user does not exist when the LDAP user logs in. If the HMC user exists, it is updated with the current user definition that is retrieved from the LDAP server when the LDAP user logs in.

User Properties

User Properties has the following properties that you can set:

- Timeout Values

  These values specify values for various timeout situations:

  – Session timeout minutes

    Specifies the number of minutes, during a logon session, that a user is prompted for identity verification. If a password is not reentered within the amount of time that was specified in the Verify timeout minutes field, then the session is disconnected. A 0 is the default and indicates no expiration. You can specify up to a maximum value of 525600 minutes (equivalent to 1 year).

  – Verify timeout minutes

    Specifies the amount of time that is required for the user to reenter a password when prompted, if a value was specified in the Session timeout minutes field. If the password is not reentered within the specified time, the session is disconnected. A 0 indicates that there is no expiration. The default is 15 minutes. You can specify up to a maximum value of 525600 minutes (equivalent to 1 year).

  - Idle timeout minutes

    Specifies the number of minutes that the user's session can be idle. If the user does not interact with the session in the specified amount of time, the session disconnects. A 0 is the default and indicates no expiration. You can specify up to a maximum value of 525600 minutes (equivalent to 1 year).

  - Minimum time in days between password changes

    Specifies the minimum amount of time in days that must elapse between changes for the user's password. A 0 indicates that a user's password can be changed at any time.

- Inactivity Values

  These values define what actions to take due to various periods of inactivity:

  - Disable for inactivity in days

    This value defines the number of days of inactivity after which a user is temporarily disabled. A value of 0 means that the user will not be disabled regardless of the duration of inactivity.

  - Never disable for inactivity

    If you do not want to disable user access based on inactivity, select **Never disable for inactivity**.

  - Allow remote access using the web

    To enable remote web server access for the user that you are managing, select **Allow remote access via web**. If this option is not selected, the user has only local access to the HMC or access through the CLI by using an SSH session.

- User Lockout

  The user is locked out of the HMC after a specified number of invalid login attempts through any interface, including CLI, GUI, or REST. By default, the system is set to lock out the user after three failed attempts, with a lockout period of 5 minutes. You can configure these lockout parameters by using the following command:

  ```
  chhmcusr -t default -i “max_login_attempts=3,login_suspend_time=5”
  ```

Password policy

The HMC has default password policies that you can use to meet general corporate requirements. To meet specific requirements, users can create a custom password policy and apply it by using the HMC. Password policies are enforced for locally authenticated HMC users only.

To see what password policies are defined on the HMC, use the `lspwdpolicy`[1] command as follows:

- List all HMC password policies:

  ```
  lspwdpolicy -t p
  ```

- List only the names of all HMC password policies:

  ```
  lspwdpolicy -t p -F name
  ```

- List the HMC password policy status information:

  ```
  lspwdpolicy -t s
  ```

[1] https://www.ibm.com/docs/en/power10/7063-CR1?topic=commands-lspwdpolicy

The “HMC Medium Security Password Policy” is defined by default but not activated. It has the following settings:

- `min_pwage=1`
- `pwage=180`
- `min_length=8`
- `hist_size=10`
- `warn_pwage=7`
- `min_digits=0`
- `min_uppercase_chars=1`
- `min_lowercase_chars=6`
- `min_special_chars=0`
- `inactivity_expiration=180`

The policy can be activated by running the following `chpwpolicy`[2] command:

```
chpwdpolicy -o -n “HMC Medium Security Password Policy”
```

To deactivate the policy, run the following command:

```
chpwdpolicy -o d
```

If you deactivate a password policy, activate another policy to protect your system.

An additional defined policy, “HMC Standard Security Password Policy”, is also available and might be acceptable for use depending on your corporate requirements. Its settings are defined as follows:

- `min_lowercase_chars=1`
- `min_uppercase_chars=1`
- `min_digits=1`
- `min_special_chars=1`
- `pwage=90`
- `min_length=15`

If you want to create your own policy, use the `mkpwpolicy`[3] command. Example 3-1 shows an example of creating a password policy.

Example 3-1 Making a password policy example

```
mkpwdpolicy -i "name=xyzPolicy,description=Company xyz policy,
pwage=90,min_digits=2,min_uppercase_chars=0,min_lowercase_chars=0"
```

The `-i` flag uses the CLI input to define the parameters of the policy. The `-f` flag defines the parameters in a file to simplify the entry of the command and to provide consistency across your HMCs. When the policy is defined, it still must be activated before it is effective.

Deleting password policies is done by using the `rmpwdpolicy` command. The `-n` parameter specifies the name to delete. For example, `rmpwdpolicy -n xyzPolicy` deletes the policy “xyzPolicy”.

[2] https://www.ibm.com/docs/en/power10/7063-CR1?topic=commands-chpwdpolicy

[3] https://www.ibm.com/docs/en/power10/7063-CR1?topic=commands-mkpwdpolicy

WebUI session limit

The HMC provides a feature to limit the number of concurrent web user interface (WebUI) logins for each user. Because there is a maximum number of concurrent WebUI sessions that is supported by the HMC, limiting the number of WebUI logins per user helps avoid user lockouts due to exceeding the maximum WebUI session limit and potential denial of service (DoS) attacks.

You can use the following attributes to configure these session limits:

- Maximum WebUI sessions per user: Specify the maximum number of web user interface sessions that are allowed for a logged-in user. By default, 100 web user interface sessions are allowed for a user. The value for maximum WebUI sessions is 50 - 200.

  To set the session limit per user, use the following command:

  ```
  chhmcusr -t default -i “max_webui_sessions_per_user =50”
  ```

- Console maximum WebUI session: Specifies the maximum number of web user interface sessions that are allowed in the HMC. By default, 1000 web user interface sessions are allowed in the HMC. At the time of writing, this parameter is read-only and cannot be modified.

Enabling multi-factor authentication

Multi-factor authentication (MFA) is disabled on the HMC by default. For an HMC GUI login, when MFA is enabled and the user is configured on the PowerSC MFA server, enter the cache token credential (CTC) code in the password field. For SSH login, when MFA is enabled, all users that log in through SSH are prompted for a CTC code. If the user is configured on the PowerSC MFA server, then you can enter the CTC code at the prompt. If the user is not configured on the PowerSC MFA server, press Enter when prompted for the CTC code, and then enter the password of the user at the prompt. When HMC is enabled with PowerSC MFA, all the users, including local, LDAP, and Kerberos, are prompted by the PowerSC MFA authentication process. If for any reason you do not want MFA for a certain user, HMC is enabled with an PowerSC MFA allowlist for users that are exempt from MFA. The users that are on the allowlist are exempted from PowerSC MFA authentication on the HMC.

3.1.8 Auditing capabilities of the HMC

A secure system also requires strong auditing capabilities. This section describes some of the logging and auditing functions on the HMC.

Most tasks that are performed on the HMC (either locally or remotely) are logged by the HMC. These entries can be viewed by using the Console Events Log task, under **Serviceability** → **Console Events Log**, or by using the `lssvcevents` command from the restricted shell.

A log entry contains the timestamp, the username, and the task that is being performed. When a user logs in to the HMC locally or from a remote client, entries are also recorded. For remote login, the client hostname or IP address is also captured, as shown in Example 3-2.

Example 3-2 User login entry

```
lssvcevents -t console
time=11/11/2015 09:52:55,"text=User hscroot has logged on from location <ip
address> to session id 32. The user's maximum role is ""hmcsuperadmin""."
```

Standard log entries from `syslogd` can also be seen on the HMC by viewing the `/var/hsc/log` file. This file can be read by users with the hmcsuperadmin role. It is under logrotate control. A valid user can use the `cat` or `tail` commands to view the file.

A user with the hmcsuperadmin role can use the `scp` command to securely copy the file to another system. If you want to copy `syslogd` entries to a remote system, you may use the `chhmc` command to change the `/etc/syslog.conf` file on the HMC to specify a system to which to copy. For example, the following command causes the syslog entries to be sent to the `myremotesys.company.com` hostname:

```
chhmc -c syslog -s add -h myremotesys.company.com
```

The systems administrator must be sure that the `syslogd` daemon that is running on the target system is set up to receive messages from the network. On most Linux systems, this task can be done by adding the `-r` option to `SYSLOGD_OPTIONS` in the `/etc/sysconfig/syslog` file.

3.1.9 Secure Boot

The Secure Boot feature is enabled on the HMC hardware appliance (7063-CR2). Secure Boot is also supported when using the vHMC running on ESXi or KVM (on Ubuntu and RHEL). This feature helps ensure the integrity of the kernel of the system. The kernel images are signed by IBM private keys to ensure that the HMC starts only with IBM HMC supplied kernel images. The keys are validated as the first step in the boot process.

Due to the difference in architecture, different steps are required to enable Secure Boot on physical and vHMCs:

- For the documentation for the Secure Boot feature enablement steps for hardware HMC, see Enabling secure boot on 7063-CR2 HMC.
- For the dedicated steps to enable Secure Boot for vHMC based on VMware ESXi, see Installing the HMC virtual appliance enabled with secure boot by using VMware ESXi.
- For the steps to enable Secure Boot for vHMC for KVM Hypervisor on RHEL, see Installing the HMC virtual appliance enabled with secure boot by using KVM hypervisor on RHEL.
- For the steps to enable Secure Boot for vHMC for KVM Hypervisor on Ubuntu, see Installing the HMC virtual appliance enabled with secure boot by using KVM hypervisor on Ubuntu.

3.1.10 Enabling a PowerSC profile for the HMC

You can enable a PowerSC agent on the HMC the PowerSC server. You can enable PowerSC communication with the HMC by using firewall settings and by installing PowerSC server certificates on the HMC. The PowerSC server detects the HMC and starts managing it as an endpoint. Then, the HMC profile that is available from PowerSC can be applied on HMC and monitored for compliance.

For more information about enabling the PowerSC profile in the HMC, see Enabling PowerSC profile for the HMC: and HMC hardening profile.

3.1.11 Managing and understanding security vulnerabilities on the HMC

HMC users can subscribe to an email notification of corrective service at the IBM Fix Central website. Whenever a vulnerability is discovered on the HMC, a bulletin describing how to obtain the fix is sent to users. Usually because of the closed nature of the HMC and the presence of the restricted shell, some vulnerabilities that are found on non-HMC systems do not apply. Each time a new release of the HMC code is made available on the support website, a list of security fixes that are included in the release is also published. To find fixes for your HMC environment, see IBM Support Fix Central.

3.1.12 Electronic Service Agent setup wizard

The HMC includes a Call Home feature to notify IBM of any issues that occur. It can be configured to send call home data through direct LAN-based connections or indirect SSL connections to the internet. Also, internet support can be provided through a proxy server if needed.

For comprehensive documentation about configuring the HMC to send call home data, testing problem reporting, managing user authorization, and handling information transmission, see Configuring the local console to report errors to service and support.

For documentation to enable call home in a 7063-CR2 HMC, see How to configure a new 7063-CR2 HMC.

3.1.13 Inband communication setup on 7063-CR2

For the HMC to self-monitor for problem reporting, it must be able to communicate with the management controller that is built in to the HMC. The management controller interface is used to poll for platform events and checks the status of hardware components. For more information about how to configure the Baseboard Management Controller (BMC) on the 7063-CR2 HMC, see How to configure the BMC on HMC 7063-CR2.

To enable communication from the HMC to the BMC of the inband connection, configure credentials to allow the HMC to connect to the BMC for periodic monitoring of hardware problem events and other management console functions.

To communicate with OpenBMC, two things are needed:

- An inband or “pass-through” interface for the OS to connect to the BMC.

  The 7063-CR2 HMC uses an usb0-to-usb0 model of communication. There is an usb0 interface on the HMC OS, and an usb0 interface on the BMC. The two use a predefined set of IP addresses. The two interfaces are preconfigured, so no user intervention is needed for this step.

  The interfaces are defined as follows:

  - BMC usb0 IP: 169.254.95.120
  - HMC usb0 IP: 169.254.95.121

- Administrator privilege credentials are used to access the OpenBMC to enable discovery and retrieval of events.

  The administrator privilege credentials are necessary for the HMC to communicate with the API of the BMC. The default administrator privilege credentials for the BMC are as follows:

  – Username: root
  – Password: 0penBmc (zero for the O)

  > **Note:** The default password auto-expires on first access by the user and must be changed.

  It is a best practice that a local user, other than root, is configured with administrator privilege to be used for console inband communications. For more information, see How to add a user to the BMC on the 7063-CR2 HMC.

Managing HMC to BMC credentials

Once the HMC starts, a periodic event appears (after 20 minutes) to remind the user to set the inband credentials, unless the user sets them. If any users are using shell (ssh or rshterm) for access, they also receive a “wall” message. The notifications repeat every 24 hours while the credentials are not set.

How to configure the Console Inband Communication Credentials

This process is documented in IBM Documentation. Complete the following steps:

1. Select a local user on the BMC with administrator privileges.

   As a best practice, do not use root on the BMC, but instead create a user with administrator privileges. For more information, see How to add a user to the BMC on the 7063-CR2 HMC.

2. If you are running an HMC version earlier than Version 10r2.1030, on the HMC, select **Console Settings** → **Console Inband Communication Credentials**. For HMC Version 10r2.1030 or later, select **HMC Management** → **Inband BMC credentials**.

3. When the task loads, it checks whether credentials exist, and if so, it validates them. The user is informed whether the credentials are valid, failed, not set, or expired.

   There are two types of credentials-related tasks that are available: Set Credentials or Change Expired Password. The default task is Set Credentials, unless the previously provided password is expired, in which case it loads in the Change Expired Password task.

   > **Note:** The Change Expired Password task cannot be selected by the user. It is only available when the previously provided password has expired. This scenario can be common for first-time setups where the user has yet to configure the BMC and the default credentials of root/OpenBMC are still in place.

4. If the current credentials are valid, click **Close**, and then click **Close** again to end the task.

   If the credentials are failed or not set, then update the credentials by providing a valid username and password and clicking **Set Credentials**. If the credentials are accepted, click **Close** to exit.

5. If the credentials are expired, clicking **Close** switches to the Change Expired Password task. Provide a new password (twice to confirm) to update the new password for the user on the BMC. Click **Change Expired Password**.

3.1.14 Summary

The HMC is a part of the management of the IBM Power ecosystem. To provide its range of functions, it must connect to the servers it is managing, to IBM for call home, and to the users that are administering the environment. It is designed to make these connections in a secure manner. We described many areas where you can ensure that you are taking full advantage of the security options that are available.

3.2 Cloud Management Console security

As private and hybrid cloud deployments expand, enterprises require new management insights into these environments. Tools that offer consolidated information and analytics are crucial for smooth infrastructure operations. The IBM CMC for Power delivers a unified view of your Power cloud landscape, regardless of the number of systems or data centers that are involved.

3.2.1 Overview of Cloud Management Console

The CMC is an IBM cloud-based solution that manages your IBM Power environment across all data centers. It offers system inventories, including virtual components, and consolidated performance data to optimize utilization throughout your data centers. Also, the CMC provides aggregated logging information for deeper insights, and it has the Patch Planning feature to help identify and apply necessary updates.

Hosted in IBM Cloud, the CMC helps ensure secure, anytime access to enable system administrators to generate reports and gain insights into their Power cloud deployments. The CMC is required for IBM Power Enterprise Pools 2 (PEP2), which is a cloud solution that facilitates dynamic resource allocation and pay-as-you-go capacity management within the IBM Power environment.

CMC is not a single product, but a platform through which IBM delivers applications and microservices in a DevOps model. The solution is for mobile devices, tablets, and desktop browsers, which help ensure convenient access for cloud operators.

Cloud Connector is the service that runs on the IBM Power HMC and sends system resource usage data to the CMC service. The Cloud Connector and the CMC provide the applications that are shown in Table 3-2 for the IBM Power ecosystem.

Table 3-2 Cloud Management Console applications for IBM Power

Feature	Benefits
Inventory Aggregation	► Enterprise views of Power servers, HMCs, LPARs, and resources that are associated with these components, which provide insight into the health and status of the enterprise. ► Centralized hardware inventory. ► Grouping of resources by using customer-supplied names to customize views.
Performance Monitoring	► Enterprise views providing resource consumption and performance for Power servers, LPARs, and I/O components. ► Guest operating systems performance metrics that point out potential areas for improvement. ► Enterprise views of energy that is consumed by Power servers.

Feature	Benefits
Log Trends	System log aggregation across the Power enterprise, which provides a central point to view log trends from Live Partition Mobility, remote restart, and the lifecycle of LPARs.
Patch Planning	► Know all your patch planning needs at glance including firmware, VIOS, OS, and HMC. ► Identify all patch dependencies. ► Integrated, collaborative planning with stakeholders.
PEP2	Support management of resource sharing across systems that are defined in PEP2. PEP2 provides an advanced resource sharing option across multiple servers and multiple sites to allow an enterprise to run workloads on any server in the pool.

3.2.2 User and resource roles management on the CMC

Users can access the CMC portal by using a valid IBMid. IBM supports federated authentication with the IBMid so that organizations can use their own login pages and security controls to securely access the CMC.

User management is handled by the administrator of the organization that is registered for IBM CMC for Power services. Administrators can manage users from the Settings page. To access this page, click the navigation menu icon in the CMC portal header, and then select the **Settings** icon. On the Settings page, click the **Manage Users** tab to view all users that are configured for your organization. Users without administrator privileges have limited access to specific applications.

To be added to the CMC, users must have a valid IBMid. In addition to the IBMid, users must be added to the CMC application by the administrator within your company. The resource role assignment feature enables administrators to assign appropriate tasks to users.

Resource roles can be managed from the CMC Portal Settings page. On this page, click the **Manage Resource Roles** tab to view and manage resource roles. Administrators can add, modify, or delete resource roles for other users from this page.

3.2.3 Protection for sensitive data

The CMC provides techniques for protecting sensitive data. Although user data is not uploaded from the Cloud Collector, some of the information in the configuration data might be considered to be sensitive, and data from some systems might be considered sensitive. To provide flexibility in managing what data is uploaded, CMC provides a method for identifying systems that should not upload data to the CMC. CMC also provides a methodology for masking sensitive metadata within your configuration.

Blocklist, allowlist, and No List

Here are the three options for filtering systems for uploading data:

- Blocklist: To prevent the Cloud Connector from uploading data for specific managed systems, add these systems to the blocklist. The **Blocklist** tab in the Managed System Filter area shows the managed systems that are on the blocklist, including their model, type, and machine serial number (MTMS). Data from these systems is not uploaded to the cloud.
- Allowlist: To specifically permit the Cloud Connector to upload data from certain managed systems, add these systems to the allowlist. The **Allowlist** tab in the Managed System Filter area shows the managed systems that are on the allowlist, including their model, type, and MTMS. Only data from managed systems on the allowlist is uploaded to the cloud.

> **Attention:** Data filtering for the allowlist is supported only with HMC 1020 or later. If your version does not meet this requirement, data from systems that are not on the allowlist will still be uploaded to the CMC.

- No List: Selecting **No List** disables both filtering types and shows data from all managed systems.

To view the current managed systems on the blocklist or allowlist, click the **Blocklist** and **Allowlist** tabs in the Managed System Filter area.

> **Important:** Only one filter type (Blocklist, Allowlist, or No List) can be active at a time.

Managing the blocklist

To add a managed system to the blocklist, click the **Blocklist** tab and confirm your selection. In the Managed System Filter area, click **Edit Blocklis**t, enter the model, type, and serial number of the system that you want to block, and click **Add**. To remove a system from the blocklist, click the minus sign (-) next to the system’s name and confirm the removal.

Adding a system to the blocklist does not automatically remove its existing data from the cloud. To purge this data, ensure that Cloud Connector is running, and then run `chsvc -s cloudconn -o stop --purge` from the management console CLI.

> **Important:** Systems in Power Enterprise Pool 2.0 cannot be blocklisted.

Managing the allowlist

To add a managed system to the allowlist, click the **Allowlist** tab and confirm your selection. In the Managed System Filter area, click **Edit Allowlist**, specify the model, type, and serial number of the system that you want to allow, and click **Add**. To remove a system from the allowlist, click the minus sign (-) next to the system’s name and confirm the removal.

No List Option

To disable the blocklist and allowlist filters and display data from all managed systems, click **No List**.

Data Filter

To keep data in Cloud Connector from getting pushed to cloud storage, add the systems to this table. Selections are available to filter the System IP Address and Logical Partition/Virtual IO Server IP Address. These systems can be reenabled if you want.

After the selection is made, it takes 5 - 10 minutes to reflect the data in the CMC. The patch planning data is updated once a day, so you might see a delay in the changes in the Patch Planning app. The purge operation is supported if the HMC is at 8.8.6.0 SP 2 or later.

CMC Attribute Masking

By enabling the Attribute Masking feature, you can help ensure that sensitive data does not leave your data center. After enabling this feature, Cloud Connector masks sensitive data and sends the masked data to the CMC server, and the CMC UI shows these masked values of the resource attributes on all the CMC pages and apps. When Attribute Masking is enabled, the CMC APIs also contain masked data in their response.

To enable Attribute Masking, from the Cloud Console interface, select **Settings** → **Cloud Connector** → **Cloud Connector Management**, scroll down to the end of the page, and then set **Attribute Masking** to On.

The Attribute Masking feature is available with HMC 1040 and later only. Data from earlier HMC versions is not masked and remains unmasked even when attribute masking is enabled. For more information about what fields are masked, see Attribute Masking.

3.2.4 Cloud Connector connections

Cloud Connector must connect to multiple end points to send and receive the data that is used by the CMC. These end points are managed through outbound connection to the IP addresses and the ports for the different components. There are multiple sources requiring outbound connections:

- HMC Cloud Connector to CMC Cloud Portal Server
- HMC Cloud Connector to CMC Cloud Database (Configuration/Certificates)
- HMC Cloud Connector to CMC Cloud Data Ingestion Node

Cloud Connector connections from the HMC to the CMC Cloud Portal

Cloud Connector provides connections between the HMC in your data center and the CMC instance. It is a component that uploads data to the CMC cloud. Cloud Connector is preinstalled on the HMC, but is not started by default. The Cloud Connector can be started by using a key. To use the key, select **CMC Portal** → **Settings** → **Cloud Connector Settings**.

Cloud Connector supports either a direct outbound connection or a connection through a proxy server to the IBM CMC portal. The Cloud Connector supports a basic authentication protocol and other authentication protocols, such as Kerberos, LDAP, and Digest-MD5 to connect to the proxy server. This section provides details about various security aspects of the CMC.

IP connections and ports that are required

The IP addresses and ports that are required for connecting to the CMC Cloud Portal can be found on the CMC Portal (select **CMC Portal** → **Settings** → **Cloud Connector Settings**). The HMCs need either direct outbound connectivity to those endpoints or need a proxy connection. For the connection to the cloud portal, you can use an HTTP proxy.

> **Important:** Starting with HMC 9.1.941.0, the Cloud Connector supports an HTTP proxy. If you are using earlier versions of the HMC, the Cloud Connector requires a SOCKS5 proxy.

Cloud Connector supports Kerberos, LDAP, and Digest-MD5-based proxy server authentication, and Basic authentication. When you start Cloud Connector, you can specify an attribute that designates the authentication type to use for the proxy connection. The default authentication is Basic.

Starting the Cloud Connector

To start the Cloud Connector, complete the following steps on the HMC:

1. In the navigation area, click **HMC Management**.
2. In the content panel, click **Cloud Connector**. You see the State, Authentication Type, HTTP Proxy, and Sock Proxy of the Cloud Connector.
3. To start the Cloud Connector, click **Start Cloud Connector** and follow the steps in the wizard.

Cloud Connector uses a one-way push model where it initiates all outbound communication. For an automatic network-based configuration, where Cloud Connector pulls the configuration file from the Cloud Database, use HTTPS. For an application data flow (push) between Cloud Connector and the CMC data ingestion node, use TCP with SSL. All communication from the Cloud Connector to the CMC is secured by using the Transport Layer Security 1.2 (TLSv1.2) protocol.

Use the startup key for the HMC based Cloud Connector to establish a valid connection between the connector and the CMC Cloud Portal Server (cloud portal). This key is also used for a connection between the Cloud Connector and the configuration database. Once a valid connection is established to the cloud portal, credentials are returned to the Cloud Connector, which enables dynamic configuration and reconfiguration.

Figure 3-2 shows the Cloud Connector establishing trust with the cloud portal by pushing the user-provided key to a cloud portal key verification endpoint.

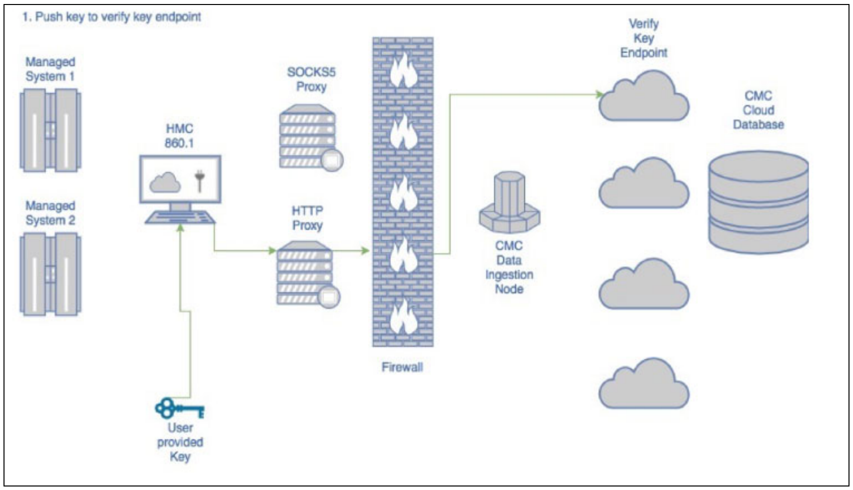


Figure 3-2 Cloud Connector pushes the key to the cloud portal[4]

Once the key verification is successful, the CMC cloud portal server returns credentials for pulling the Cloud Connector configuration file and SSL certificates, as shown in Figure 3-3 on page 69.

[4] Source: https://ibmcmc.zendesk.com/hc/en-us/article_attachments/360083545614/CloudConnectorSecurityWhitePaper.pdf

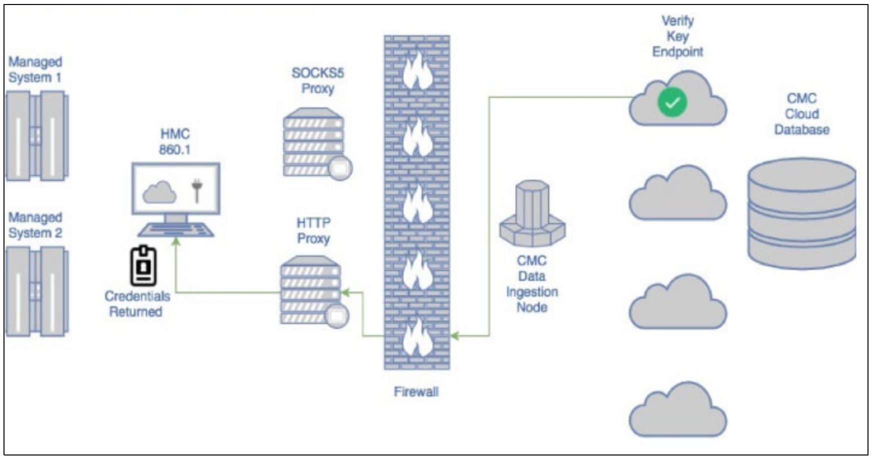


Figure 3-3 Cloud portal returns credentials to Cloud Connector[4]

A security test runs to assert that the startup key provided is valid. The test begins with a `GET` request from the connector to the cloud portal, which returns a cross-site request forgery (XSRF) header. With this XSRF header, along with a portion of the decoded key, a `POST` operation is performed to the same cloud portal endpoint. If the key is considered valid, the cloud portal responds with a set of encoded credentials that grant Cloud Connector access to a database containing the customer's Cloud Connector configuration file.

Cloud Connector connections from the HMC to the CMC Cloud Database

Once Cloud Connector establishes a successful connection with the cloud portal, a secure SSL connection is established between Cloud Connector and the Cloud Database to fetch the configuration file. This configuration file contains the following items:

- Cloud applications that are enabled by the customer.
- Data to push for those applications.
- Data to filter (block-listed managed systems, and selected system and partition IP addresses).
- IP address of the cloud data ingestion node.

Also, this file provides credentials for fetching the SSL certificate and key pair that are used in communication between the Cloud Connector and the cloud data ingestion node. The credentials are used to access a separate database from the one that is used to fetch the configuration file. However, the underlying network location and mechanism that are used to fetch the certificates are the same.

An SSL connection is established and the data is returned to the connector. Every minute, the Cloud Connector fetches a new configuration to ensure that changes are handled. All communication from the connector to the Cloud Database is secured by using the TLSv1.2 protocol. Using the received credentials that are shown in Figure 3-3 on page 69, the Cloud Connector pulls the customer-specific Cloud Connector configuration file from the CMC Cloud Configuration Database, as shown in Figure 3-4.

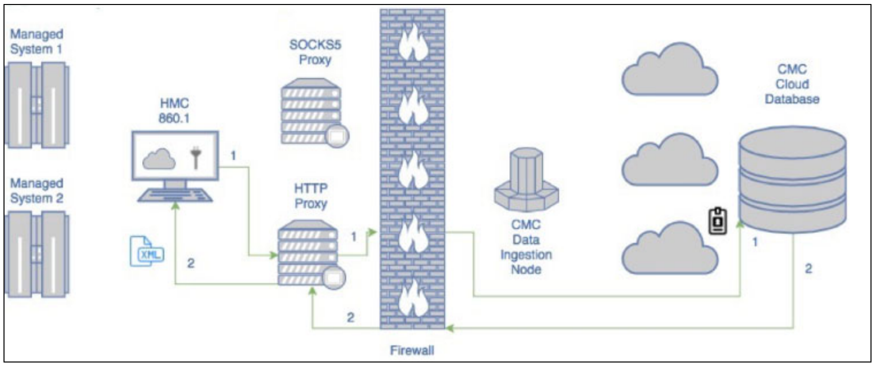


Figure 3-4 Cloud Connector pulls a configuration file from the CMC database[4]

Once the credentials from the configuration file are collected, as shown in Figure 3-3 on page 69, the Cloud Connector pulls the SSL certificates and key from the CMC Cloud Certificate Database, as shown in Figure 3-5.

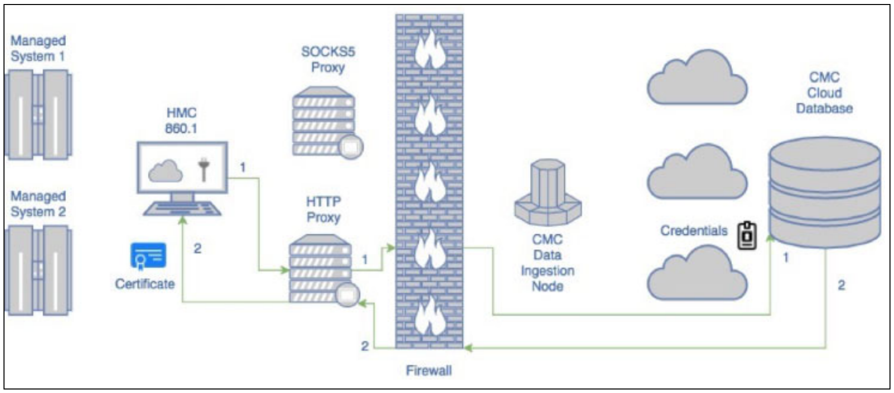


Figure 3-5 Cloud Connector pulls SSL keys from Cloud Database[4]

Cloud Connector connections from the HMC to the CMC cloud data ingestion node

Once the Cloud Connector is configured by the automated configuration process, The Cloud Collection begins collecting data and pushing that data to the data ingestion node. This channel is secured by using SSL with mutual authentication that uses the certificate and key that were noted in Figure 3-5. Using mutual authentication helps ensure that the connector sends data to only trusted data ingestion nodes. The certificate and key are stored on the HMC file system, but are accessible only by the root user.

Figure 3-6 on page 71 shows the connection between the HMC and the ingestion node through the SOCKS5 proxy by using the certificate that was obtained in Figure 3-5. With HMC 9.1.941.0, Cloud Connector can be started only with the HTTP proxy option.

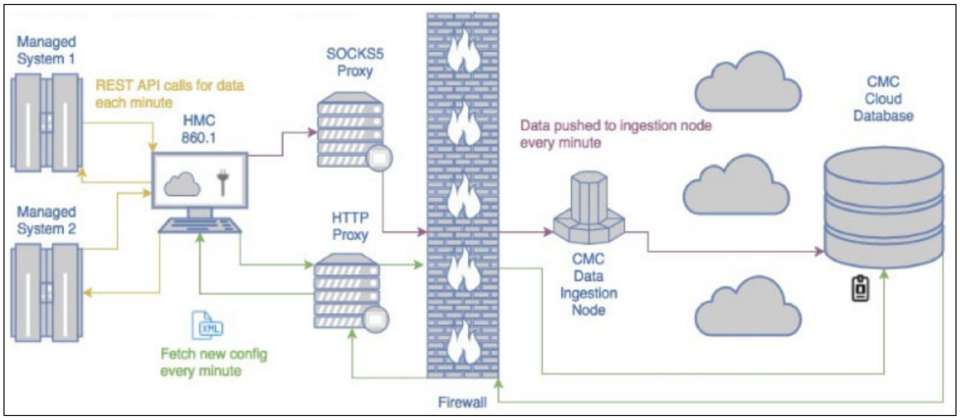


Figure 3-6 The Cloud Connector authentication to the CMC data ingestion node SOCK5 version[4]

If the Cloud Connector starts with only the HTTP proxy, then it uses the HTTP proxy to establish a connection between the HMC and the ingestion node, as shown in Figure 3-7. At the time of writing, the proxy options that are shown in Figure 3-6 are still supported in the HMCs, when the Cloud Connector starts with both HTTP and SOCKS5 proxies.

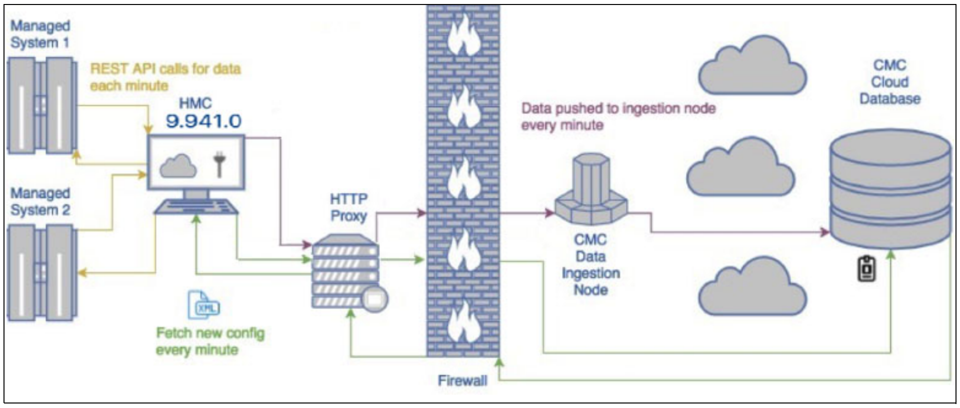

Managed System 1
REST API calls for data each minute
HMC 9.941.0
Data pushed to ingestion node every minute
CMC Cloud Database
Managed System 2
HTTP Proxy
CMC Data Ingestion Node
Fetch new config every minute
Firewall

Figure 3-7 The Cloud Connector authentication to the CMC data ingestion node in new HTTP proxy mode[4]

3.2.5 Hints for using the CMC

This section presents some best practices about on how to personalize the CMC to meet specific needs.

Enabling and disabling CMC applications

An administrator can enable or disable applications that are hosted by the CMC servers. In the navigation area, select **Settings**. In the contents area, click **Apps**. Set the switch next to each application to enable or disable that application.

IBM Power Enterprise Pools 2.0 provides enhanced multisystem resource sharing and by-the-minute consumption of on-premises compute resources to clients who deploy and manage a private cloud infrastructure. Power Enterprise Pool 2.0 is monitored and managed by the IBM CMC. The CMC Enterprise Pools 2.0 application helps you to monitor base and metered capacity across a Power Enterprise Pool 2.0 environment, with both summary views and sophisticated drill-down views of real-time and historical resource consumption by LPARs.

Important: To use the Power Enterprise Pools 2.0 app, first enable the Capacity Monitoring app to collect performance data.

Viewing logging dashboards

The CMC provides insight into various virtualization operations in the Dashboard. The Dashboard information includes Live Partition Mobility, remote restart, and other partition activities. The logging dashboard is in the navigation area. To access it, select **Logging**, and then in the Contents area, select the PowerVM Virtualization actions that you want to view.

Viewing Patch Planning information

You can get a comprehensive view of your inventory with information about the current patch state of the resources in your environment. You can also view a list of resources that must be updated, and the recommended service level for each resource. To see all the resources in your environment, including the operating systems, firmware, VIOSs, adapters, and HMCs, select **Patch Planning** → **Inventory** → **All**.

3.3 VIOS security

The VIOS is a specialized AIX-based appliance that can virtualize I/O adapters for Virtual I/O client LPARs running AIX, IBM i, and Linux. Typically, two VIOS instances are installed on the same managed system to help ensure HA and minimize risks from human errors, maintenance windows, and potential hardware failures.

Because all virtualized I/O traffic goes through VIOS, securing it is crucial. If an attacker compromises VIOS, they might gain access to all virtualized network and storage traffic on the system and potentially infiltrate client LPARs.

After you deploy VIOS, your first priority is to configure it securely. Many of the security settings that are applicable to AIX can also be applied to VIOS. Because of VIOS' appliance nature, if you are unsure about applying specific security configurations, contact IBM Support for help.

Although VIOS does not have its own published security benchmarks, you can refer to the Center for Internet Security (CIS) AIX benchmark to guide your VIOS security configuration. VIOS 3.1 is based on AIX 7.2, and VIOS 4.1 is based on AIX 7.3.

VIOS currency and security fixes

As with all other operating systems, keep your VIOS server updated and current. Regularly check for updates and be sure that you are using a supported version of VIOS. Unsupported versions do not get any security fixes and as such should not be considered to be secure.

The best way to get information about new VIOS releases, SPs, and security fixes is to subscribe to VIOS notifications at the IBM Support portal.

Occasionally, VIOS receives security fixes to address newly identified vulnerabilities. Apply these fixes in accordance with your company's security and compliance policies. Many VIOS administrators delay installing security updates, and opt to wait for the next SP instead. This approach might be acceptable if you evaluate the risks that are associated with a compromised virtualization infrastructure and your organization is prepared to accept those risks.

If your infrastructure is designed for HA with dual-VIOS instances per managed system and all client LPARs are correctly connected, you can install security fixes without disrupting operations, potentially even during normal business hours. To do this task, first install the fixes on one VIOS instance, and restart that instance to activate the updates. After the VIOS restarts, verify that all client LPAR storage paths and network connections are restored. Then, repeat the process on the second VIOS instance.

Decide which VIOS to update first based on the needs of your client LPARs. To prevent log messages during the update, you can disable the storage paths to the VIOS that is being updated.

3.3.1 Virtual I/O Server system security hardening

Beginning with VIOS 1.3, you can set security options that provide tighter security controls over your VIOS environment. These options enable you to select a level of system security hardening and specify the settings that are allowed within that level. The VIOS security feature also enables you to control network traffic by enabling the VIOS firewall. You can configure these options by using the `viosecure` command.

To help you set up system security when you initially install the VIOS, the VIOS provides the configuration assistance menu. You can access the configuration assistance menu by running the `cfgassist` command. By using the `viosecure` command, you can set, change, and view the security settings. By default, no VIOS security levels are set. Run the `viosecure` command to change the settings.

The system security hardening feature protects all elements of a system by tightening security or implementing a higher level of security. Although hundreds of security configurations are possible with the VIOS security settings, you can implement security controls by specifying a high, medium, or low security level.

By using the system security hardening features that are provided by VIOS, you can specify the following values:

- Password policy settings
- Actions such as `usrck`, `pwdck`, `grpck`, and `sysck`
- Default file-creation settings
- Settings that are included in the `crontab` command

Configuring a system at too high a security level might deny services that are needed. For example, `telnet` and `rlogin` are disabled for high-level security because the login password is sent over the network unencrypted. If a system is configured at too low a security level, the system might be vulnerable to security threats. Because each enterprise has its own unique set of security requirements, the predefined high, medium, and low security configuration settings are best suited as a starting point for security configuration rather than an exact match for the security requirements of a particular enterprise. As you become more familiar with the security settings, you can make adjustments by choosing the hardening rules that you want to apply. For more information about the hardening rules, run the `man` command.

Configuring Virtual I/O Server system security hardening

The section describes how you can set the security level to specify security hardening rules for your VIOS system.

To implement system security hardening rules, you can use the `viosecure` command to specify a security level of high, medium, or low. A default set of rules is defined for each level. You can also set a level of default, which returns the system to the system standard settings and removes any level settings that are applied.

The low-level security settings are a subset of the medium-level security settings, which are a subset of the high-level security settings. Therefore, high level is the most restrictive setting and provides the greatest level of control. You can apply all rules for a specified level or select which rules to activate for your environment. By default, no VIOS security levels are set, so you must run the `viosecure` command to modify the settings.

Setting the security level of a VIOS

To set a VIOS security level of high, medium, or low, use the following command:

```
viosecure -level low -apply
```

Changing the settings in a security level

To set a VIOS security level in which you specify which hardening rules to apply for the setting, run the `viosecure` command interactively. For example:

1. At the VIOS CLI, run `viosecure -level high`. All the security level options (hardening rules) at that level are displayed ten at a time (pressing Enter displays the next set in the sequence).
2. Review the options that are displayed and make your selection by entering the numbers, which are separated by a comma, that you want to apply, or type `ALL` to apply all the options or `NONE` to apply none of the options.
3. Press Enter to display the next set of options, and continue entering your selections.

> **Note:** To exit the command without making any changes, type “`q`”.

Viewing the current security setting

To display the current VIOS security level setting, run the following command:

```
viosecure -view
```

Removing security level settings

To unset any previously set system security levels and return the system to the standard system settings, run the command `viosecure -level default`.

To remove the security settings that are applied, run the command `viosecure -undo`.

3.3.2 Virtual I/O Server firewall

Using the VIOS firewall, you can enforce limitations on IP activity in your virtual environment. With this feature, you can specify which ports and network services may access the VIOS system. For example, if you must restrict login activity from an unauthorized port, you can specify the port name or number and specify `deny`, which removes it from the allowlist. You can also restrict a specific IP address.

Configuring the Virtual I/O Server firewall settings

This section describes how you can enable the VIOS firewall to control IP activity.

The VIOS firewall is not enabled by default. To enable the VIOS firewall, run the `viosecure` command with the `-firewall` option. When you enable the firewall, the default setting is activated, which enables access for the following IP services:

- `ftp`
- `ftp-data`
- `ssh`
- `web`
- `https`
- `rmc`
- `cimom`

> **Note:** The firewall settings are contained in the `viosecure.ctl` file in the `/home/ios/security` directory. If for some reason the `viosecure.ctl` file does not exist when you run the command to enable the firewall, you receive an error. You can use the `-force` option to enable the standard firewall default ports.

You can use the default settings or configure the firewall settings to meet the needs of your environment by specifying which ports or port services to allow. You can also turn off the firewall to deactivate the settings.

Configuring the VIOS firewall

To configure the VIOS firewall settings, complete the following steps on the VIOS CLI:

1. Enable the VIOS firewall by running the following command:

   ```
   viosecure -firewall on
   ```

2. Specify the ports to allow or deny by running the following command:

   ```
   viosecure -firewall allow | deny -port number
   ```

3. View the current firewall settings by running the following command:

   ```
   viosecure -firewall view
   ```

4. If you want to disable the firewall configuration, run the following command:

   ```
   viosecure -firewall off
   ```

> **Tip:** For more information about securing your VIOS, see IBM Documentation.

3.3.3 User management and authentication

Similar to AIX, you can have multiple user accounts on VIOS. The reason to have multiple accounts is to limit access to the padmin account, which is the main VIOS administrative account. You might want to create a view-only account for your monitoring software or a separate account for automating your VIOS operations with Ansible.

Creating an additional administrative account on VIOS

To create a user’s account on VIOS, run the command `mkuser`. An example is shown in Example 3-3.

Example 3-3 Creating an administrative user account on VIOS

```
$ mkuser admin
Changing password for “admin”
admin’s New password:
Enter the new password again:
```

The account “admin” has all the privileges to change a VIOS configuration but cannot switch into `oem_setup_env` mode and run root commands.

Creating a view-only user

To create a read-only account on VIOS, run the mkuser command with extra attributes, as shown in Example 3-4.

Example 3-4 Creating a view-only user account on VIOS

```
$ mkuser -attr pgrp=view monitor
Changing password for "monitor"
monitor's New password:
Enter the new password again:
```

The account “monitor” can log in to VIOS and see the current configuration of the VIOS, but cannot change the configuration.

Listing VIOS users

To list users on VIOS, run the `lsuser` command, as shown in Example 3-5.

Example 3-5 List the existing users on VIOS

```
$ lsuser
padmin roles=PAdmin,CacheAdm default_roles=PAdmin,CacheAdm account_locked=false
expires=0 histexpire=52 histsize=4 loginretries=0 maxage=13 maxexpired=4
maxrepeats=8 minage=4 minalpha=2 mindiff=0 minlen=10 minother=0 pwdwarntime=5
registry=files SYSTEM=compat
admin roles=Admin default_roles=Admin account_locked=false expires=0 histexpire=52
histsize=4 loginretries=0 maxage=13 maxexpired=4 maxrepeats=8 minage=4 minalpha=2
mindiff=0 minlen=10 minother=0 pwdwarntime=330 registry=files SYSTEM=compat
```

The super-administrators with access to `oem_setup_env` mode have the role `PAdmin`. All other administrators have the role `Admin`. View-only users have the role `ViewOnly`.

Changing user privileges

To change user privileges, run the `chuser` command. For example, if you want to make a user a read-only user, set its roles to `ViewOnly`.

```
chuser -attr roles=ViewOnly default_roles=ViewOnly pgrp=view admin
```

If you want to assign admin privileges to a user, run the following command:

```
chuser -attr roles=Admin default_roles=Admin pgrp=system admin
```

After you change the privileges, the user must relogin to the system to make the changes effective.

Changing the password policy for a user

You can set a password policy for users by changing the attributes that are shown in Table 3-3 with the command `chuser`

Table 3-3 Attributes for user passwords

Attribute	Meaning
`expires`	The expiration date of the account in the format MMDDhhmmyy. If the value is set to 0, the account does not expire.
`histexpire`	The period in weeks when the user cannot reuse an old password.
`histsize`	The number of previous passwords that the user cannot reuse.
`loginretries`	The number of unsuccessful logins attempts before the account is locked.
`maxage`	The maximum number of weeks that the password is valid.
`maxexpired`	The maximum number of weeks where the user can change their expired password.
`maxrepeats`	The maximum number of times that a character can be repeated in the password.
`minage`	The minimum number of weeks where the user cannot change the password after the new password is set.
`minalpha`	The minimum number of alphabetic characters in the password.
`mindiff`	The minimum number of characters in the new password that should differ from the old password.
`minlen`	The minimum length of the password.
`minother`	The minimum number of non-alphabetic characters in the password.
`pwdwarntime`	The number of days before a warning about password expiration is shown.

Locking and unlocking accounts

You can temporarily lock a user's account by running the following command:

```
chuser -attr account_locked=true admin
```

To unlock the account, set the attribute to `false`.

Removing unneeded accounts

If you do not need a user account, you can delete by running the following command:

```
rmuser -rmdir admin
```

The option `-rmidr` remove the user's home directory. The files on VIOS that are owned by the removed user do not change their ownership automatically,

3.3.4 Auditing of VIOS commands

All commands that are issued by administrative or view-only accounts are logged. The logs are saved in the file `/home/ios/logs/ioscli_global.log`. You can access the file only in `oem_setup_env` mode, which is why it is important to have separate accounts for administrators and not use the `padmin` user for daily tasks.

The local date, time, user account, and the issued command are saved in the file. An example is shown in Example 3-6.

Example 3-6 The ioscli_global.log file

```
Jun 24 2024, 18:55:20 monitor  vfcmap  -vadapter vfchost28 -fcp
Jun 24 2024, 18:56:10 admin    lsuser
Jun 24 2024, 18:58:01 padmin   chuser  -attr roles=ViewOnly default_roles=ViewOnly
pgrp=view admin
```

4

IBM AIX security

Security is an important topic. So much of our personally identifying data is stored online that a security incident most likely affected anyone reading this chapter. The penalties for breaches of the various standards (such as Health Insurance Portability and Accountability Act (HIPAA) or Payment Card Industry Data Security Standard (PCI DSS)) are significant. Good security requires a multi-layered approach that starts with people, then physical security, and then the various layers. Consider the whole environment and see how security can be applied at each level.

This chapter describes some of the basics of locking down your AIX logical partitions (LPARs). Some security hardening is implemented in AIX 7.3 by default. Usually, you must check and implement some additional security settings according to your environmental requirements, which include setting default permissions and umasks, using good usernames and passwords, hardening the security with AIX Expert, protecting the data at rest directly on the disk or at the logical volume (LV) layer through encryption, removing insecure daemons, and integrating with Lightweight Directory Access Protocol (LDAP) directory services or Microsoft Active Directory (AD)).

This chapter describes the following topics:

4.1 AIX security checklist

If you are setting up a new AIX system and need guidance about basic security measures and initial steps, IBM published a security checklist to help you. It is included here for your convenience.

Although this checklist is not exhaustive, it provides a solid foundation for developing a comprehensive security plan that is tailored to your environment. This section covers best practices and introduces other considerations in the following sections.

Here is a checklist of security actions to perform on a newly installed or existing system:

- Use secure base media to install AIX.
- Avoid installing desktop software, such as CDE, GNOME, or KDE, on servers.
- Apply necessary security fixes and any recommended maintenance and technology level updates. For the latest service bulletins, security advisories, and fix information, see the IBM Support Fix Central website.
- Back up the system after the initial installation and store the backup in a secure location.
- Set up access control lists (ACLs) for restricted files and directories.
- Disable unnecessary user and system accounts, such as daemon, bin, sys, adm, lp, and uucp. Deleting accounts is not advised because it removes account information, such as user IDs (UIDs) and user names, which might still be linked to data on system backups. If a user is re-created with a previously deleted ID and the backup is restored, the new user might gain unintended access.
- Regularly review and remove unnecessary daemons and services from the `/etc/inetd.conf`, `/etc/inittab`, `/etc/rc.nfs`, and `/etc/rc.tcpip` files.
- Verify that the permissions for the files that are listed in Example 4-1 are correctly set.

Example 4-1 Validating the permissions of the files

```
-rw-rw-r-- root     system  /etc/filesystems
-rw-rw-r-- root     system  /etc/hosts
-rw------- root     system  /etc/inittab
-rw-r--r-- root     system  /etc/vfs
-rw-r--r-- root     system  /etc/security/failedlogin
-rw-rw---- root     audit   /etc/security/audit/host
```

- Disable the root account's ability to remotely log in. The root account should be able to log in only from the system console.
- Enable system auditing. For more information, see Auditing overview.
- Enable a login control policy. For more information, see Login control.
- Disable user permissions for running the `xhost` command. For more information, see Managing X11 and CDE concerns.
- Prevent unauthorized changes to the PATH environment variable. For more information, see PATH environment variable.
- Disable telnet, rlogin, and rsh. For more information, see TCP/IP security.
- Establish user account controls. For more information, see User account control.
- Enforce a strict password policy. For more information, see Passwords.
- Establish disk quotas for user accounts. For more information, see Recovering from over-quota conditions.

- Allow only administrative accounts to use the su command. Monitor the su command's logs in the /var/adm/sulog file.
- Enable screen locking when using X Window System.
- Restrict access to the cron and at commands to only the accounts that need access to them.
- Use an alias for the ls command to show hidden files and characters in a file name.
- Use an alias for the rm command to avoid accidentally deleting files from the system.
- Disable unnecessary network services. For more information, see Network services.
- Perform frequent system backups and verify the integrity of the backups.
- Subscribe to security-related email distribution lists.

4.2 Encrypted File System

The Encrypted File System (EFS) enables individual users on the system to encrypt their data on the Journaled File System (JFS) by using their personal keystores. Each user is associated with a unique key. These keys are stored in a cryptographically protected keystore, and on successful login, the user's keys are loaded into the kernel and associated with the process's credentials.

Later, when the process needs to open an EFS-protected file, these credentials are checked. If a key matching the file protection is found, the process can decrypt the file key and the file content. Group-based key management is also supported.

> Note: EFS is part of an overall security strategy. It works with sound computer security practices and controls.

4.2.1 Encrypted File System usability

EFS key management, file encryption, and file decryption are transparent to users in normal operations.

EFS is part of the base AIX operating system. To enable EFS, root or any user with the role-based access control (RBAC) aix.security.efs authorization must use the efsenable command to activate EFS and create the EFS environment. For more information about who can manage EFS, see 4.2.3, “Root access to user keys” on page 82. This action is a one-time system enablement.

After the EFS is enabled, when a user logs in, their key and keystore are silently created and secured or encrypted with the user’s login password. Then, the user's keys are used automatically by the Enhanced Journaled File System (JFS2) for encrypting or decrypting EFS files. Each EFS file is protected with a unique file key, which is secured or encrypted with the file owner’s or group’s key, depending on the file permissions. By default, a JFS2 File System is not EFS-enabled.

When a file system is EFS-enabled, the JFS2 File System transparently handles encryption and decryption in the kernel for read/write requests. User and group administration commands (such as mkgroup, chuser, and chgroup) manage the keystores for the users and groups seamlessly.

The following EFS commands are provided so that users can manage their keys and file encryption:

`efskeymgr`	Manages and administers the keys.
`efsmgr`	Manages the encryption of files, directories, and the file system.

4.2.2 User keystores

The user keystore is managed automatically for basic operations. Users can perform advanced tasks and create encrypted files or directories by using the `efskeymgr` (files) and `efsmgr` (directories) commands. Access to group keystores is automatically granted when a user is added to a group. Similar to UNIX ACLs, file owners can control access to encrypted files.

Users can change their login password without affecting open keystores and the keystore password can be different from the login password. When the user password differs from the keystore password, you must manually load the keystore by using the `efskeymgr` command.

Keystore details

The keystore has the following characteristics:

- Protected with passwords and stored in the Public Key Cryptography Standards (PKCS) #12 format.
- Location:
  - User: `/var/efs/users/<username>/keystore`
  - Group: `/var/efs/groups/<groupname>/keystore`
  - `efs_admin`: `/var/efs/efs_admin/keystore`
- Users can choose the encryption algorithms and key lengths.
- Access is inherited by child processes.

Group key management

Keys are kept at both the user level and the group level. For group keys, consider the following items:

- Only group members can add and remove group keys in Root Guard mode.
- User keystores contain user private keys and passwords to access group keystores.
- Group keystores contain the group's private keys.

4.2.3 Root access to user keys

Root privileges can be configured to provide either unrestricted or limited access to user keys. Regardless of the configuration, root cannot directly assume a user's identity (by using `su`) to access encrypted files or keystores.

Unrestricted root access

In this mode, root can reset a user's keystore password, potentially gaining access to the keys within. This configuration offers greater flexibility for system administration.

Restricted root access

In this mode, root can reset only a user's login password, and not their keystore password. Root cannot impersonate a user (by using su) to access an open keystore. Although root can create, modify, and delete users and their associated keystores, accessing the keys within these keystores is prohibited. This configuration provides enhanced protection against malicious root activity.

4.2.4 EFS keystore management modes

There are two main modes for managing EFS keystores: root admin and root guard. Also, a special efs_admin key provides root-level access to all keystores.

Root admin mode

Root admin mode is the default setting. It offers full access to all keystores, including user and group keystores. In this mode, root can reset user keystore passwords, which might compromise data if a user forgets their password.

Root guard mode

Root guard mode provides strong security against unauthorized access, even for root users. However, losing the user’s keystore password results in data loss because there is no recovery method. Some keystore operations (adding or removing group access keys, and regenerating private keys) might require user intervention.

The efs_admin key

The efs_admin key is a special key that is stored in the root user's keystore that grants root-level access to all keystores in root admin mode. Permissions to access this key can be granted or revoked to specific users or groups by using the efskeymgr command. The user must have the aix.security.efs RBAC authorization to manage the EFS.

Note: The EFS keystore is opened automatically as part of the standard AIX login only when the user’s keystore password matches their login password. This approach is set up by default during the initial creation of the user’s keystore. Login methods other than the standard AIX login, such as loadable authentication modules and Pluggable Authentication Modules (PAMs), might not automatically open the keystore.

4.2.5 Encryption and inheritance

EFS is a feature of JFS2. The file system's efs option must be set to yes (see the mkfs and chfs commands). EFS automatically encrypts and decrypts user data. However, if a user has read access to an EFS activated file but does not have the right key, then the user cannot read the file in the normal manner. If the user does not have a valid key, it is impossible to decrypt the data.

All cryptographic functions come from the CLiC kernel services and CLiC user libraries.

By default, a JFS2 file system is not EFS-enabled. A JFS2 file system must be EFS-enabled before EFS inheritance can be activated or any EFS encryption of user data can take place. A file is created as an encrypted file either explicitly with the efsmgr command or implicitly through EFS inheritance. EFS inheritance can be activated either at the file system level, at the directory level, or both.

The ls command lists entries of an encrypted file with a preceding e.

The cp and mv commands can handle metadata and encrypted data seamlessly across EFS-to-EFS and EFS-to-non-EFS scenarios.

The backup, restore, and tar commands and other related commands can back up and restore encrypted data, including the EFS metadata that is used for encryption and decryption.

4.2.6 Backup and restore

Properly manage the archiving or backup of the keystores that are associated with the archived EFS files. Also manage and maintain the keystore passwords that are associated with the archived or backup keystores. Failure to do either of these tasks might result in data loss.

When backing up EFS encrypted files, you can use the –Z option with the backup command to back up the encrypted form of the file, along with the file's cryptographic metadata. Both the file data and metadata are protected with strong encryption. This approach has the security advantage of protecting the backed-up file through strong encryption. Back up the keystore of the file owner and group that are associated with the file that is being backed up. These keystores are in the following files:

Users keystores	/var/efs/users/user_login/*
Group keystore	/var/efs/groups//keystore
The efsadmin **keystore**	/var/efs/efs_admin/keystore

To restore an EFS backup that was made with the backup –Z command, use the restore command. The restore command helps ensure that the crypto metadata is also restored. During the restore process, it is not necessary to restore the backed-up keystores if the user has not changed the keys in their individual keystore. When a user changes their password to open their keystore, their keystore internal key is not changed. Use the efskeymgr command to change the keystore internal keys.

If the user's internal keystore key remains the same, the user can immediately open and decrypt the restored file by using their current keystore. However, if the key that is internal to the user's keystore changed, the user must open the keystore that was backed up in association with the backed-up file. This keystore can be opened with the efskeymgr –o command. The efskeymgr command prompts the user for a password to open the keystore. This password is the one that was used in association with the keystore at time of the backup.

For example, assume that user Bob's keystore was protected with the password *foo* (the password '*foo*' is not a secure password and only used in this example for simplicity sake) and a backup of Bob's encrypted files was performed in January along with Bob's keystore. In this example, Bob also uses *foo* for his AIX login password. In February, Bob changed his password to *bar*, which also changed his keystore access password to *bar*. If in March, Bob's EFS files were restored, then Bob would be able to open and view these files with his current keystore and password because he did not change the internal key of the keystore.

However, if it was necessary to change the internal key of Bob's keystore (with the efskeymgr command), then by default the old keystore internal key is deprecated and left in Bob's keystore. When the user accesses the file, EFS automatically recognizes that the restored file used the old internal key, and EFS uses the deprecated key to decrypt it. During this same access instance, EFS converts the file to using the new internal key. There is not a significant performance impact in the process because it is all handled through the keystore and the file's crypto metadata, and the file data does not need to be re-encrypted.

If the deprecated internal key is removed by using `efskeymgr`, then the old keystore containing the old internal key must be restored and used with the files that are encrypted with this internal key.

How do you securely maintain and archive old passwords? There are methods and tools to archive passwords, which involve using a file that contains a list of all old passwords, and then encrypting this file and protecting it with the current keystore, which is protected by the current password. However, IT environments and security policies vary from organization to organization, and consideration and thought should be given to the specific security needs of your organization to develop security policy and practices that are best suited to your environment.

4.2.7 The JFS2 EFS internal mechanism

Each JFS2 EFS encrypted file is associated with a special extended attribute (EA) that contains EFS metadata that is used to validate crypto authority; information that is used to encrypt and decrypt files such as keys; and a crypto algorithm

The EA content is not transparent to JFS2. Both user credentials and EFS metadata are required to determine the crypto authority (access control) for an EFS-activated file.

Note: Be careful in situations where a file or data might be lost, for example, removing the file's EA.

EFS protection inheritance

After a directory is EFS-activated, any newly created immediate children are automatically EFS-activated unless you manually override this action. The EFS attributes of the parent directory take precedence over the EFS attributes of the file system.

The scope of the inheritance of a directory is exactly one level. Any newly created child also inherits the EFS attributes of its parent if its parent directory is EFS-activated. Existing children maintain their current encrypted or non-encrypted state. The logical inheritance chain is broken if the parent changes its EFS attributes. These changes do not propagate down to the existing children of the directory and must be applied to those directories separately,

Workload partition considerations

Before enabling or using EFS within a workload partition, first enable EFS on the global system by running the `efsenable` command. This enablement is performed only once. Also, all file systems, including EFS-enabled file systems, must be created from the global system.

4.2.8 Setting up the Encrypted File System

This section describes how to set up an EFS. Complete the following steps:

1. Install the `clic.rte` file set. This file set contains the cryptographic libraries and kernel extension that are required by EFS. The `clic.rte` file set is part of the base AIX image from ESS.

   Enable EFS on the system by running the command `efsenable –a`. When prompted for a password, you may use the root password. Users keystores are created automatically when the user logs in or re-logs in after the `efsenable` command runs.

2. After `efsenable –a` runs on a system, the system is EFS-enabled and the command does not need to be run again.
3. Create an EFS-enabled file system by running the `crfs –a efs=yes` command. For example:

   ```
   crfs -v jfs2 -m /foo –A yes -a efs=yes -g rootvg -a size=20000
   ```

4. After mounting the file system, turn on the cryptographic inheritance on the EFS-enabled file system by running the `efsmgr` command. To continue the previous example where the file system `/foo` was created, run this command:

   ```
   efsmgr –s –E /foo
   ```

   Now, every file that is created and used in this file system can be an encrypted file.

If a file system exists, you can enable it for encryption by running the `chfs` command, for example:

```
chfs -a efs=yes /foo
```

It is impossible to disable EFS on a file system after it is enabled.

The following file systems cannot be converted to EFS-enabled file systems:

- `/`
- `/opt`
- `/usr`
- `/var`

From this point forward, when a user or process with an open keystore creates a file on this file system, the file will be encrypted. When the user or file reads the file, the file is automatically decrypted for users who are authorized to access the file.

4.2.9 Centralizing access to Encrypted File System keystores

In an enterprise environment, you can centralize your EFS keystores. When you store the databases that control the keystores on each system independently, it can be difficult to manage the keystores. AIX Centralized EFS Keystore enables you to store the user and group keystore databases in LDAP so that you can centrally manage the EFS keystore.

The LDAP defines a standard method for accessing and updating information in a directory (a database) either locally or remotely in a client/server model.

You can store all AIX EFS keystore databases in LDAP, which include the following EFS databases:

- User Keystore
- Group Keystore
- Admin Keystore
- Cookies

The AIX operating system provides utilities to help you perform the following management tasks:

- Export local keystore data to an LDAP server.
- Configure the client to use EFS keystore data in LDAP.
- Control access to EFS keystore data.
- Manage LDAP data from a client system.

All EFS keystore database management commands are enabled to use the LDAP keystore database. If the system-wide search order is not specified in the `/etc/nscontrol.conf` file, keystore operations depend on the user and group `efs_keystore_access` attribute. If you set `efs_keystore_access` to `ldap`, the EFS commands perform keystore operations on the LDAP keystore.

Table 4-1 describes the changes to EFS commands for LDAP.

Table 4-1 EFS command enablement for LDAP

Command	LDAP information
Any EFS command	When you set the `efs_keystore_access` attribute to `ldap`, you do not need to use the special option `-L domain` with any command to perform keystore operations on LDAP.
`efskeymgr`	Includes the `-L load_module` option so that you can perform explicit keystore operations on LDAP.
`efsenable`	Includes the `-d Basedn` option so that you can perform the initial setup on LDAP to accommodate the EFS keystore. The initial setup includes adding base distinguished names (DNs) for the EFS keystore and creating the local directory structure (`/var/efs/`).
`efskstoldif`	Generates the EFS keystore data for LDAP from the following databases on the local system: - `/var/efs/users/username/keystore` - `/var/efs/groups/groupname/keystore` - `/var/efs/efs_admin/keystore` - • Cookies, if they exist, for all the keystores

All keystore entries must be unique. Each keystore entry directly corresponds to the DN of the entry that contains the user and group name. The system queries the user IDs (uidNumber), group IDs (gidNumber), and the DNs. The query succeeds when the user and group names match the corresponding DNs. Before you create or migrate EFS keystore entries on LDAP, ensure that the user and group names and IDs on the system are unique.

Exporting Encrypted File System keystore data to LDAP

Populate the LDAP server with the keystore data to use LDAP as a centralized repository for the EFS keystore.

Before you create or migrate EFS keystore entries on LDAP, ensure that the user and group names and IDs on the system are unique.

To populate the LDAP server with the EFS keystore data, complete the following steps:

1. Install the EFS keystore schema for LDAP on to the LDAP server:
   a. Retrieve the EFS keystore schema for LDAP from the `/etc/security/ldap/sec.ldif` file on the AIX system.
   b. Run the `ldapmodify` command to update the schema of the LDAP server with the EFS keystore schema for LDAP.
2. Run the `efskstoldif` command to read the data in the local EFS keystore files and output the data in a format that is suitable for LDAP.

   To maintain unique keystore access, consider placing the EFS keystore data that is in LDAP under the same parent distinguished name (DN) as the user and group data.
3. Save the data to a file.
4. Run the `ldapadd -b` command to populate the LDAP server with the keystore data.

Configuring an LDAP client for an Encrypted File System keystore

To use EFS keystore data that is stored in LDAP, you must configure a system as an LDAP client. To configure an LDAP client for EFS keystore, complete the following steps:

1. To configure a system as an LDAP client, run the `/usr/sbin/mksecldap` command.

   The `mksecldap` command dynamically searches the specified LDAP server to determine the location of the EFS keystore data. Then, it saves the results to the `/etc/security/ldap/ldap.cfg` file. The `mksecldap` command determines the location for user, group, admin, and EFS cookies keystore data.
2. Complete one of the following steps to enable LDAP as a lookup domain for EFS keystore data:
   – Set the user and group `efs_keystore_access` attribute to `file` or `ldap`.
   – Define the search order for the keystore at the system level by using the `/etc/nscontrol.conf` file.

Table 4-2 shows an example.

Table 4-2 Example configuration for the /etc/nscontrol.conf file

Attribute	Description	Search order
`efsusrkeystore`	This search order is common for all users.	LDAP, files
`efsgrpkeystore`	This search order is common for all groups.	files, LDAP
`efsadmkeystore`	This search order locates the admin keystore for any target keystore.	LDAP, files

> **Attention:** The configuration that is defined in the `/etc/nscontrol.conf` file overrides any values that are set for the user and group `efs_keystore_access` attribute. The same is true for the user `efs_adminks_access` attribute.

After you configure a system as an LDAP client and enable LDAP as a lookup domain for the EFS keystore data, the `/usr/sbin/secldapclntd` client daemon retrieves the EFS keystore data from the LDAP server whenever you perform LDAP keystore operations.

4.3 Logical volume encryption

EFS provides data encryption at a file-system level. The EFS manages the data encryption key at a file level and protects the data encryption key for each user. If you want to avoid the complexity of fine-gained control of file-system encryption and selective file encryption, you can choose logical volume (LV) encryption. Starting with AIX 7.2 TL5, AIX added encryption at the LV level. This approach is one choice for data at rest encryption within AIX. By using this feature, you can encrypt the data to prevent data exposure due to lost or stolen hard disk drives or inappropriately decommissioned computers. *Data at rest* refers to data that is stored physically in any digital form.

Some organizations must show that data at rest is encrypted. A common example is the Payment Card Industry Data Security Standard (PCI DSS) requirement to encrypt sensitive data such as a direct link between a card holder name and card number.

Using LV encryption is similar to physical disk encryption. Once operational, the application environment does not know that the data is encrypted. The encryption is only noticeable when the (disk) storage is mounted somewhere else and the data is unreadable. Outside of the configured environment, information in the LV cannot be accessed.

Using LV encryption has the following advantages:

- The data owner controls the encryption keys.
- The data that is transmitted over the network (Fibre Channel or Ethernet) is encrypted and protected. These characteristics are important for virtual servers that are hosted in the cloud environment.

For more information about the LV encryption architecture, see AIX 72 TL5: Logical Volume Encryption.

LV encryption is simple to use and transparent to the applications. Once the system starts and an authorized process or user is active on the system, the data is accessible to authorized users based on classic access controls such as ACLs.

When enabled (by default, data encryption is not enabled in LVs), each LV is encrypted with a unique key. Data encryption must be enabled at the volume group level before you can enable the data encryption option at the LV level. The LV data is encrypted as the data is written to the physical volume (PV), and decrypted when it is read from the PV.

Enabling LV encryption creates one data encryption key for each LV. The data encryption key is protected by storing the keys separately in other data storage devices. The following types of key protection methods are supported:

- Paraphrase
- Key file
- Cryptographic key server
- Platform keystore (PKS), which is available in IBM PowerVM firmware starting at firmware level FW950

LV encryption enhancements in AIX 7.3

LV encryption was further enhanced in AIX 7.3. Starting from AIX 7.3, the following enhancements are added to the LV encryption function:

- You can encrypt LVs in the root volume group (rootvg) that are used in the start process. The LV encryption option must be selected during the installation of the base operating system. For more information, see BOS installation options.
- After you install the base operating system, you can use the `hdcryptmgr` conversion commands to change the encryption setting of an LV. However, the conversion of an LV in the rootvg is different from the conversion of an LV in a user volume group:
  - When you run the `hdcryptmgr` conversion command to change the encryption status of an LV in a rootvg, the `hdcryptmgr` command creates an LV to store the conversion recovery data.
  - When you run the `hdcryptmgr` conversion command to change the encryption status of an LV in a user volume group, the `hdcryptmgr` command stores the conversion recovery data in a file that is in the `/var/hdcrypt` directory.

    Therefore, the rootvg must have at least one available LPAR for successful conversion. When the conversion status of the encryption is successful, the LV that contains the conversion recovery data is deleted.
- When the rootvg is varied on, the network is not available. Hence, the PKS authentication method must be available for LVs that are used in the start process:
  - If the PKS authentication method is not available for an encrypted LV in the rootvg, the LV remains locked and not accessible until it is explicitly unlocked later.
  - You cannot delete a valid PKS authentication method from an LV in the rootvg that is used in the start process.
  - If you convert an unencrypted LV that is used in the start process to an encrypted LV, the PKS authentication method is automatically added to the LV.
  - If the PKS authentication method is not available or is corrupted for an encrypted LV that is used in the start process, you must start the operating system in maintenance mode and repair the PKS authentication method before you can resume the normal start operation.
- The following commands are enhanced to support LV encryption:
  - `cplv`
  - `splitvg`
  - `splitlvcopy`
  - `chlvcopy`
  - `snapshot`
  - `savevg`
  - `restvg`
- You can encrypt an LV in concurrent mode. If you change the encryption status of an LV in a node that is in concurrent mode, you cannot access the other nodes until the encryption conversion is complete.
- AIX 7.3 TL1 supports IBM Hyper Protect Crypto Services (HPCS) for AIX LV encryption. To use HPCS with AIX, provision Power Systems Virtual Server. The `keysvrmgr` command provides options to manage the integration.

4.3.1 LV encryption commands

Use the commands in this section to manage encryption keys and key server information.

The hdcryptmgr command

The hdcryptmgr utility manages the encrypted LVs by displaying LV and volume encryption information, controlling authentication, and many other functions. The utility and its help messages are built in a hierarchical and self-explanatory manner.

Example 4-2 shows a summary of the command usage. For a detailed man page, see the hdcryptmgr command.

Example 4-2 Help page for the hdcryptmgr command

```
# hdcryptmgr -h
Usage: hdcryptmgr <action> <..options..>

Display:
showlv       : Displays LV encryption status
showvg       : Displays VG encryption capability
showpv       : Displays PV encryption capability
showmd       : Displays encryption metadata that is related to the device
showconv     : Displays status of all active and stopped conversions

Authentication control:
authinit     : Initializes master key for data encryption
authunlock   : Authenticates to unlock the master key of the device
authadd      : Adds more authentication methods
authcheck    : Checks validity of an authentication method
authdelete   : Removes an authentication method
authsetrvgpwd : Adds "initpwd" passphrase method to all rootvg's LVs

PKS management:
pksimport    : Import the PKS keys
pksexport    : Export the PKS keys
pksclean     : Removes a PKS key
pksshow      : Displays PKS keys status

Conversion:
plain2crypt  : Converts an LV to encrypted
crypt2plain  : Converts an LV to not encrypted

PV encryption management:
pvenable     : Enables the Physical Volume Encryption
pvdisable    : Disables the Physical Volume Encryption
pvsavemd     : Save encrypted physical volume metadata to a file
pvrecovmd    : Recover encrypted physical volume metadata from a file
```

The keysvrmgr command

For the key server method, you can use the keysvrmgr utility to manage Object Data Manager (ODM) entries that are associated with the key server information, such as the key server hostname or IP address, the connection port, and certification location.

Example 4-3 shows a summary of the command usage. For a detailed man page, see the `keysvrmgr` command.

Example 4-3 Help page for the keysvrmgr command

```
# keysvrmgr -h
Usage: keysvrmgr <action> [-h] -t <server_type> <options> server_name
Manage ODM data for key server and HPCS.

<action> is one of the following items:
add     : Add a new key server or HPCS to ODM.
modify  : Modify a key server or HPCS ODM record.
remove  : Remove a key server or HPCS ODM record.
show    : Display key server or HPCS ODM records.
verify  : Verify an HPCS ODM record (HPCS only).
rekey   : Generate a new API key for an HPCS ODM record (HPCS only).

<server_type> is one of the following items:
keyserv : For (KMIP compliant) key management server.
hpcs    : For IBM Cloud Hyper Protect Crypto Services.

For more details on <options> run : keysvrmgr <action> -h
```

4.3.2 Prerequisites for using LV encryption

Before you implement LV encryption, make sure that you meet the following prerequisites:

- Use AIX 7.2.5 or later to encrypt an LV.
- The following file sets must be installed to encrypt the LV data. These file sets are included in the base operating system.
  - `bos.hdcrypt`
  - `bos.kmip_client`
  - `bos.rte.lvm`
  - `security.acf`
  - `openssl.base`
  - `oss.lib.libcurl`
  - `oss.lib.libjson-c`

Note: The `bos.hdcrypt` and `bos.kmip_client` file sets are not installed automatically when you run the `smit update_all` command or during an operating system migration operation. You must install it separately from your software source, such as a DVD or an ISO image.

4.3.3 Creating and authenticating an encrypted logical volume

To create an encrypted LV, complete the following steps:

1. Create an encryption-enabled volume group.
2. Create an encryption-enabled LV.
3. Authenticate the primary encryption key of the LV.

Creating an encryption-enabled volume group

To create an encryption-enabled volume group, complete the following steps:

1. Create a volume group in which the data encryption option is enabled by running this command:

   ```
   mkvg -f -y testvg -k y hdisk2
   ```

   `testvg` is the name of the new volume group and `hdisk2` is the PV that is used for the volume group.

2. Check the details of the new volume group by running the command that shown in Figure 4-4 on page 128. `ENCRYPTION` is set to `yes`.

Example 4-4 Showing the volume group settings

```
# lsvg testvg
VOLUME GROUP:       testvg                   VG IDENTIFIER:
00fb294400004c0000000176437c6663
VG STATE:           active                   PP SIZE:        8 megabytes
VG PERMISSION:      read/write               TOTAL PPs:      637 (5096 megabytes)
MAX LVs:            256                      FREE PPs:       637 (5096 megabytes)
LVs:                0                        USED PPs:       0 (0 megabytes)
OPEN LVs:           0                        QUORUM:         2 (Enabled)
TOTAL PVs:          1                        VG DESCRIPTORS: 2
STALE PVs:          0                        STALE PPs:      0
ACTIVE PVs:         1                        AUTO ON:        yes
MAX PPs per VG:     32512
MAX PPs per PV:     1016                     MAX PVs:        32
LTG size (Dynamic): 512 kilobytes            AUTO SYNC:      no
HOT SPARE:          no                       BB POLICY:      relocatable
PV RESTRICTION:     none                     INFINITE RETRY: no
DISK BLOCK SIZE:    512                      CRITICAL VG:    no
FS SYNC OPTION:     no                       CRITICAL PVs:   no
ENCRYPTION:         yes
```

3. Check the encryption state of varied-on volume groups by running the command that is shown in Example 4-5.

Example 4-5 Checking the encryption state of the volume groups

```
# hdcryptmgr showvg
VG NAME / ID         ENCRYPTION ENABLED
testvg                        yes
rootvg                        no
```

4. Check the volume group encryption metadata by running the command that is shown in Example 4-6.

Example 4-6 Validating volume group metadata

```
# hdcryptmgr showmd testvg
.....    Mon Dec  7 21:19:00 2020
.....    Device type : VG
.....    Device name : testvg
.....
=============== B: VG HEADER ================
Version                     : 0
Timestamp                   : Mon Dec  7 21:16:04 2020
```

```
Default data crypto algorithm: AES_XTS
Default MasterKey size       : 16 bytes
Auto-auth (during varyonvg)  : Enabled
=============== E: VG HEADER ================
=============== B: VG TRAILER ===============
Timestamp        : Mon Dec  7 21:16:04 2020
=============== E: VG TRAILER ===============
```

Creating an encryption-enabled logical volume

To create an encryption-enabled LV, complete the following steps:

1. Create an LV in which the data encryption option is enabled by running the command that is shown in Example 4-7.

Example 4-7 Creating an encrypted logical volume

```
# mklv -k y -y testlv testvg 10
testlv
mklv: Run :
hdcryptmgr authinit lvname [..] to define LV encryption options.
```

2. Check the details of the new volume group by running the command that is shown in Example 4-8.

Example 4-8 Validating the logical volume details

```
# lslv testlv
LOGICAL VOLUME:     testlv                             VOLUME GROUP:   testvg
LV IDENTIFIER:      00fb294400004c0000000176437c6663.1 PERMISSION:     read/write
VG STATE:           active/complete                    LV STATE:
closed/syncd
TYPE:               jfs                                WRITE VERIFY:   off
MAX LPs:            512                                PP SIZE:        8 megabytes
COPIES:             1                                  SCHED POLICY:   parallel
LPs:                10                                 PPs:            10
STALE PPs:          0                                  BB POLICY:      relocatable
INTER-POLICY:       minimum                            RELOCATABLE:    yes
INTRA-POLICY:       middle                             UPPER BOUND:    32
MOUNT POINT:        N/A                                LABEL:          None
MIRROR WRITE CONSISTENCY: on/ACTIVE
EACH LP COPY ON A SEPARATE PV?: yes
Serialize IO?:     NO
INFINITE RETRY:     no                                 PREFERRED READ: 0
ENCRYPTION:         yes
```

3. Check the authentication state of the LV by running the command that is shown in Example 4-9.

Example 4-9 Validating the authentication state of the logical volume

```
# hdcryptmgr showlv testlv
LV NAME   CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)  CONVERSION
testlv         yes              no             100            done
```

Authenticating the primary encryption key of the logical volume

To authenticate the primary encryption key of the LV, complete the following steps:

1. Initialize the primary key for an encrypted LV by running the command that is shown in Example 4-10. The LV is not accessible until the first passphrase method is initialized.

Example 4-10 Setting the authentication for the logical volume

```
# hdcryptmgr authinit testlv
Enter Passphrase:
Confirm Passphrase:
Passphrase authentication method with name "initpwd" added successfully.
```

2. Check the authentication status and authentication methods for the LV by running the command that is shown in Example 4-11.

Example 4-11 Checking the authentication for the LV

```
# hdcryptmgr showlv testlv -v
LV NAME    CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv          yes             yes             100              done

-- Authentication methods ------------
INDEX         TYPE               NAME
#0            Passphrase         initpwd
```

3. Vary off and vary on the volume group by running the following commands:

```
# varyoffvg testvg
# varyonvg testvg
```

4. Check the authentication status of the LV by running the command that is shown in Example 4-12.

Example 4-12 Checking the authentication of the LV after vary on

```
# hdcryptmgr showlv testlv
LV NAME      CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv            yes              no              100              done
```

The output shows that the LV `testlv` is not authenticated.

5. Unlock the authentication of the LV by running the command that is shown in Example 4-13.

Example 4-13 Unlocking the logical volume

```
# hdcryptmgr authunlock testlv
Enter Passphrase:
Passphrase authentication succeeded.
```

6. Check the authentication state of the LV again, as shown in Example 4-14.

Example 4-14 Validating the authentication status

```
# hdcryptmgr showlv testlv
LV NAME      CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv           yes              yes             100              done
```

4.3.4 Adding the platform keystore authentication method

To add the PKS authentication method, complete the following steps:

1. Check the LPAR PKS status by running the command that is shown in Example 4-15.

Example 4-15 Checking the status of the PKS

```
# hdcryptmgr pksshow
3020-0349 PKS is not supported or PKS is not activated.
3020-0218 hdcrypt driver service error. QUERY_PKS service failed with error 124:
An attempt was made to set an attribute to an unsupported value.
```

The output in this example shows that the PKS is not activated. The keystore size of an LPAR is set to 0 by default.

2. Shut down the LPAR and increase the keystore size in the associated Hardware Management Console (HMC). The keystore size is 4 KB - 64 KB. You cannot change the value of the keystore size when the LPAR is active.
3. Check the LPAR PKS status again by running the command that is shown in Example 4-16.

Example 4-16 Rechecking the PKS status

```
# hdcryptmgr pksshow
PKS uses 32 bytes on a maximum of 4096 bytes.
PKS_Label (LVid) Status
PKS_Label (objects)
```

4. Add the PKS authentication method to the LV by running the command that is shown in Example 4-17.

Example 4-17 Adding the PKS authentication method

```
# hdcryptmgr authadd -t pks -n pks1 testlv
PKS authentication method with name "pks1" added successfully.
```

5. Check the encryption status of the LV by running the command that is shown in Example 4-18.

Example 4-18 Checking the status of LV encryption

```
# hdcryptmgr showlv testlv -v
LV NAME        CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv         yes              yes              100              done
-- Authentication methods ------------
INDEX         TYPE                 NAME
#0            Passphrase           initpwd
#1            PKS                  pks1
```

6. Check the PKS status by running the command that is shown in Example 4-20 on page 97.

Example 4-19 Checking the PKS status

```
# hdcryptmgr pksshow
PKS uses 116 bytes on a maximum of 4096 bytes.
PKS_Label (LVid) Status
00fb294400004c0000000176437c6663.1 VALID KEY
PKS_Label (objects)
```

PKS is an automatic authentication method where the `varyonvg` command automatically unlocks the authentication of the LV.

7. Vary off the volume group by running the following command:

   ```
   # varyoffvg testvg
   ```

8. Check the PKS status by running the command that is shown in Example 4-20.

Example 4-20 Checking the PKS status again

```
# hdcryptmgr pksshow
PKS uses 116 bytes on a maximum of 4096 bytes.
PKS_Label (LVid) Status
00fb294400004c0000000176437c6663.1 UNKNOWN
PKS_Label (objects)
```

9. Vary on the volume group by running the following command:

   ```
   # varyonvg testvg
   ```

10. Check the encryption status of the LV by running the command that is shown in Example 4-21.

Example 4-21 Validating the LV encryption status

```
# hdcryptmgr showlv testlv
LV NAME         CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv          yes              yes              100              done
```

4.3.5 Adding the key server authentication method

You can use any Key Management Interoperability Protocol (KMIP) compliant key management server to use this type of authentication method. In this example, the AIX LPAR is installed and configured with IBM Security Key Lifecycle Manager 4.0 for AIX. The IBM Security Guardium® Key Lifecycle Manager key is used as an encryption key server.

To add the key server authentication method, complete the following steps:

1. Check the key servers in the LPAR by running the command that is shown in Example 4-22.

Example 4-22 Showing that no key server is defined

```
# keysvrmgr show
3020-0279 No key server in database
```

2. Add an encryption key server with the name `keyserver1` by running the command that is shown in Example 4-23.

Example 4-23 Adding a key server

```
# keysvrmgr add -i 9.X.X.X -s /tmp/sklm_cert.cer -c /tmp/ssl_client_cer.p12
keyserver1
Key server keyserver1 successfully added
```

3. Check the key servers in the LPAR again by running the command that is shown in Example 4-24.

Example 4-24 Validating that the key server is defined

```
# keysvrmgr show
List of key servers:
ID                    PWD            IP:PORT
keyserver1            N              9.X.X.X:5696
```

4. Check the encryption key server information that is saved in the ODM KeySvr object class by running the command that is shown in Example 4-25.

Example 4-25 Checking the key server definition

```
# odmget KeySvr
KeySvr:
        keysvr_id = "keyserver1"
        ip_addr = "9.X.X.X"
        port = 5696
        svr_cert_path = "/tmp/sklm_cert.cer"
        cli_cert_path = /tmp/ssl_client_cer.p12 "
        flags = 0
```

5. Add the key server authentication method to the LV by running the command that is shown in Example 4-26.

Example 4-26 Adding a key server authentication method to the LV

```
# hdcryptmgr authadd -t keyserv -n key1_testlv -m keyserver1 testlv
Keyserver authentication method with name "key1_testlv" added successfully.
```

6. Check the encryption status of the LV by running the command that is shown in Example 4-28.

Example 4-27 Checking the encryption status of the LV

```
#hdcryptmgr showlv -v testlv
LV NAME         CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv          yes              yes              100              done
-- Authentication methods ------------
INDEX                 TYPE              NAME
#0                    Passphrase        initpwd
#1                    PKS               pks1
#2                    Keyserver         key1_testlv
```

4.3.6 Adding a key file authentication method

To add a key file authentication method, complete the following steps:

1. Create a file that is named `testfile` that contains the passphrase text by running the command that is shown in Example 4-28.

Example 4-28 Creating a file

```
# cat /testfile
Add1ng Key f1le authent1cation meth0d
```

2. Add the key file authentication method to the LV by running the command that is shown in Example 4-29.

Example 4-29 Adding a key authentication method

```
# hdcryptmgr authadd -t keyfile -n key1_file -m /testfile testlv
Keyfile authentication method with name "key1_file" added successfully.
```

3. Check the contents of the `testfile` file by running the command that is shown in Example 4-30.

Example 4-30 Checking the contents of the file

```
# cat /testfile
Add1ng Key f1le authent1cation meth0d
00fb294400004c0000000176437c6663.1 xdxKjlJvZU+f9lFTgSM63kGoIoKW6Yxc+bKrk5GgCzc=
```

4. Check the encryption status of the LV by running the command that is shown in Example 4-31.

Example 4-31 Validating the encryption status

```
# hdcryptmgr showlv testlv -v
LV NAME          CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv           yes              yes              100              done
-- Authentication methods ------------
INDEX              TYPE               NAME
#0                 Passphrase         initpwd
#1                 PKS                pks1
#2                 Keyserver          key1_testlv
#3                 Keyfile            key1_file
```

4.3.7 Adding the passphrase authentication method

To add the passphrase authentication method, complete the following steps:

1. Add the passphrase authentication method to the LV by running the command that is shown in Example 4-32.

Example 4-32 Adding a passphrase

```
# hdcryptmgr authadd -t pwd -n test_pwd testlv
Enter Passphrase:
Confirm Passphrase:
Passphrase authentication method with name "test_pwd" added successfully.
```

2. Check the encryption status of the LV by running the command that is shown in Example 4-33.

Example 4-33 Checking the encryption status of the LV

```
# # hdcryptmgr showlv testlv -v
LV NAME            CRYPTO ENABLED   AUTHENTICATED   ENCRYPTION (%)   CONVERSION
testlv              yes             yes              100             done
-- Authentication methods ------------
INDEX         TYPE            NAME
#0            Passphrase      initpwd
#1            PKS             pks1
#2            Keyserver       key1_testlv
#3            Keyfile         key1_file
#4            Passphrase      test_pwd
```

4.3.8 Migrating the PKS to another LPAR before the volume group is migrated

To migrate the PKS to another LPAR, complete the following steps:

1. Export the PKS keys into another file by running the command that is shown in Example 4-34.

Example 4-34 Exporting the PKS keys

```
# hdcryptmgr pksexport -p /tmp/pksexp testvg
Enter Passphrase:
Confirm Passphrase:
1 PKS keys exported.
```

2. Import the volume group to another LPAR by running this command:

```
# importvg -y testvg hdisk2
```

3. Check the encryption status of the LV by running the command that is shown in Example 4-35.

Example 4-35 Checking the encryption status

```
# hdcryptmgr showlv testlv -v
LV NAME          CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)   CONVERSION
testlv           yes              yes              100              done
-- Authentication methods ------------
INDEX         TYPE             NAME
#0            Passphrase       initpwd
#1            PKS              pks1
#2            Keyserver        key1_testlv
#3            Keyfile          key1_file
#4            Passphrase       test_pwd
```

4. Check whether the authentication method is valid and accessible by running the command that is shown in Example 4-36.

Example 4-36 Testing the authentication

```
# hdcryptmgr authcheck -n pks1 testlv
3020-0199 Key does not exist in PKS storage.
3020-0127 hdcryptmgr authcheck failed for LV testlv.
```

5. Move the PKS key file to a new LPAR and run the command that is shown in Example 4-37.

Example 4-37 Importing the PKS key

```
#  hdcryptmgr pksimport -p /tmp/pksexp testvg
Enter Passphrase:
3020-0341 Key having LVid 00fb294400004c0000000176437c6663.1 is successfully
imported in LV testlv.
1 PKS keys imported.
```

6. Check whether the authentication method is valid and accessible by running the command that is shown in Example 4-38.

Example 4-38 Validating the authentication method

```
# hdcryptmgr authcheck -n pks1 testlv
PKS authentication check succeeded.
```

4.3.9 Changing the encryption policy of the volume group

Encryption metadata is saved at the end of each disk in the volume group. Enabling the volume group encryption requires available physical partitions on each disk in the volume group.

Complete the following steps:

1. Change the data encryption option of the volume group by running the command that is shown in Example 4-39.

Example 4-39 Changing the encryption option

```
# chvg -k y testvg
0516-1216 chvg: Physical partitions are being migrated for volume group
                descriptor area expansion. Wait.
```

2. Check the details of the volume group by running the command that is shown in Example 4-40.

Example 4-40 Listing the volume group details

```
# lsvg testvg
VOLUME GROUP:       testvg              VG IDENTIFIER:  00fb294400004c000000017648ff8d32
VG STATE:           active              PP SIZE:        8 megabytes
VG PERMISSION:      read/write          TOTAL PPs:      636 (5088 megabytes)
MAX LVs:            256                 FREE PPs:       506 (4048 megabytes)
LVs:                1                   USED PPs:       130 (1040 megabytes)
OPEN LVs:           0                   QUORUM:         2 (Enabled)
TOTAL PVs:          1                   VG DESCRIPTORS: 2
STALE PVs:          0                   STALE PPs:      0
ACTIVE PVs:         1                   AUTO ON:        yes
MAX PPs per VG:     32512
MAX PPs per PV:     1016                MAX PVs:        32
LTG size (Dynamic): 512 kilobytes     AUTO SYNC:      no
HOT SPARE:          no                  BB POLICY:      relocatable
PV RESTRICTION:     none                INFINITE RETRY: no
DISK BLOCK SIZE:    512                 CRITICAL VG:    no
FS SYNC OPTION:     no                  CRITICAL PVs:   no
ENCRYPTION:         yes
```

4.3.10 Changing the encryption policy of the logical volume

To change the encryption policy, complete the following steps.

Note: This capability is for experimental use only.

1. Enable the LV encryption by running the command that is shown in Example 4-41.

Example 4-41 Enabling the LV encryption

```
# hdcryptmgr plain2crypt testlv
Enter Passphrase:
Confirm Passphrase:
Passphrase authentication method with name "initpwd" added successfully.
Created recovery file: /var/hdcrypt/conv.004200021607542921
In case of error or if the conversion is canceled, this file may be
necessary to be able to recover the LV. If the conversion is fully
successful, then the file will be removed automatically
Successfully converted LV testlv to an encrypted LV.
```

This command performs the following operations:

– Enables the encryption policy of the LV.

– Initializes the master-key and encryption metadata for an encrypted LV.

– Encrypts the data in the LV.

2. Check the details of the LV by running the command that is shown in Example 4-42.

Example 4-42 Listing the LV details

```
# lsvg testvg
VOLUME GROUP:       testvg              VG IDENTIFIER:
00fb294400004c000000017648ff8d32
VG STATE:           active              PP SIZE:        8 megabytes
VG PERMISSION:      read/write          TOTAL PPs:      636 (5088 megabytes)
MAX LVs:            256                 FREE PPs:       506 (4048 megabytes)
LVs:                1                   USED PPs:       130 (1040 megabytes)
OPEN LVs:           0                   QUORUM:         2 (Enabled)
TOTAL PVs:          1                   VG DESCRIPTORS: 2
STALE PVs:          0                   STALE PPs:      0
ACTIVE PVs:         1                   AUTO ON:        yes
MAX PPs per VG:     32512
MAX PPs per PV:     1016                MAX PVs:        32
LTG size (Dynamic): 512 kilobytes     AUTO SYNC:      no
HOT SPARE:          no                  BB POLICY:      relocatable
PV RESTRICTION:     none                INFINITE RETRY: no
DISK BLOCK SIZE:    512                 CRITICAL VG:    no
FS SYNC OPTION:     no                  CRITICAL PVs:   no
ENCRYPTION:         yes
```

3. Check the encryption status of the LV by running the command that is shown in Example 4-43 on page 103.

Example 4-43 Checking the encryption status

```
# hdcryptmgr showlv testlv -v
LV NAME            CRYPTO ENABLED   AUTHENTICATED    ENCRYPTION (%)  CONVERSION
testlv              yes              yes             100             done
-- Authentication methods ------------
INDEX         TYPE                  NAME
#0                 Passphrase    initpwd
```

4.3.11 Best practices

When setting up and using LV encryption, follow these best practices:

- Use an inline log device for any file system that is created from an encrypted LV.
- If the file system is created with an external log device and the log device is shared across multiple file systems, unlock the authentication (run the `hdcryptmgr authunlock` command) for all encrypted LVs before you mount the file system.
- Use one of the non-PKS authentication methods to unlock the authentication of the snapshot volume group.
- To copy an encrypted LV by using the `cplv` command:
  a. Create an LV in which encryption is enabled.
  b. Use this newly created LV as the destination LV to copy the source LV.

4.3.12 Limitations of LV encryption

If an LV is encrypted, the following LV functions are not supported:

AIX Live Update The Live Update operation is not supported if LV encryption is enabled.

I/O serialization I/O serialization is not guaranteed while LV encryption conversion is in progress.

4.3.13 File system considerations for LV encryption

Consider the following items when you create or modify file systems that are associated with an encrypted LV:

- When you create or mount a file system on to an encrypted LV, ensure that the encrypted LV is unlocked and activated.
- If an encrypted LV that is hosting a file system by using the Network File System (NFS) `/etc/exports` file is not unlocked during system start, the mount operation of the file system fails and the table of physical file systems (PFSs) in the `/etc/exports` file is not updated. After the encrypted LV is unlocked and the file system is mounted, you can run the `exportfs -a` command to update the `/etc/exports f`ile.
- In JFS2, you can use a single log device across multiple file systems. If the log device is shared across multiple file systems and if the LV that is used by file systems is encrypted, the LV must be unlocked before mounting the file systems.

4.4 Physical volume encryption

PV encryption protects user data by encrypting data that is written to the PV. The base operating system performs PV data encryption and decryption during I/O operations. The data is encrypted before it is sent to an external Storage Area Network (SAN) device to protect data on the SAN. PV encryption also helps prevent data exposure because of lost or stolen hard disk drives or because of inappropriately decommissioned computers or storage devices. Applications that perform I/O operations can use the protected data without any modifications. The encrypted PVs can be used in the same way as unencrypted PVs. However, the `rootvg` volume group cannot contain any encrypted PVs.

With AIX 7.3 TL1, IBM continues to address clients' need to protect data by introducing encrypted PVs. This capability encrypts data at rest on disks, and because the data is encrypted in the OS, the disk data in transit is encrypted too.

Install the following file sets to encrypt the PV data. These file sets are included in the base operating system.

- `bos.hdcrypt`
- `bos.kmip_client`
- `security.acf`
- `openssl.base`

AIX historically supported encrypted files by using the EFS. More recently, AIX 7.2 TL 5 introduced support for LV encryption, as described in 4.3, "Logical volume encryption" on page 89.

Now, AIX offers a new level of security with PV encryption. With this feature, you can encrypt the entire PV, which provides enhanced protection for applications that do not rely on volume groups or LVs, such as certain database applications. However, it is also possible to create volume groups and LVs on encrypted disks.

PV encryption leverages the infrastructure that is developed for LV encryption. Therefore, many of the concepts and features that were described in 4.3, "Logical volume encryption" on page 89also apply to encrypted PVs. For example, both types support the same key management functions.

The `hdcryptmgr` command is used to manage encrypted PVs, and the `hdcrypt` driver handles the encryption process. Although the core function is similar, some new options and actions were added to the `hdcryptmgr` command specifically for PV encryption.

4.4.1 Configuring PV encryption

To use PV encryption, you must first format the disk for encryption. This operation erases any data on the disk. There is no support for directly encrypting existing data on a disk. To encrypt existing data, allocate a new PV, enable encryption on the new PV, and then copy the existing data to the new PV.

The size of the encrypted PV is smaller than the size of the PV before encryption because the encryption feature reserves some space on the PV for the encryption process.

The command to enable encryption on disk hdisk10 is `hdcryptmgr pvenable hdisk10`. This command prompts the user for a passphrase to use to unlock the disk and then reserves some space at the beginning of the disk for metadata. As with LV encryption, a data encryption key is created automatically when the disk is initialized for encryption. The `pvenable` action also prompts the user to add a passphrase wrapping key to encrypt the data encryption key. More wrapping keys may be added by using the `authadd` action of the `hdcryptmgr` command. Because space is reserved for metadata, the space that is available for user data on an encrypted PV is slightly smaller than the total size of the PV.

When the disk is initialized for encryption and unlocked, it may be used like any other disk in AIX, except that encrypted disks cannot be used as part of the `rootvg` volume group. As the OS writes data to the disk, the data is encrypted; when data is read from the disk, it is decrypted before being passed to the user.

4.4.2 Using encrypted physical volumes

Encrypted PV supports the same methods of key storage and retrieval as encrypted LVs. The key can be a typed passphrase; can be obtained from PKS; or can be obtained from a network key manager.

When the key is stored in a PKS or in a network key manager, the PV is unlocked automatically during the start process. The `authunlock` action parameter of the `hdcryptmgr` command can be used to manually unlock an encrypted PV. Any attempts to perform an I/O operation on a locked encrypted PV fails with a permission denied error until that PV is unlocked.

If the AIX LPAR restarts, encrypted disks that use only the passphrase wrapping key protection method must be manually unlocked by using the `hdcryptmgr authunlock` action. If one of the other methods, such as using a key server or PKS, was added to the disk by using the `authadd` action, AIX attempts to automatically unlock the disk during start. Any attempt to do I/O to an encrypted disk that is still locked fails.

Figure 4-1 illustrates the encryption process.

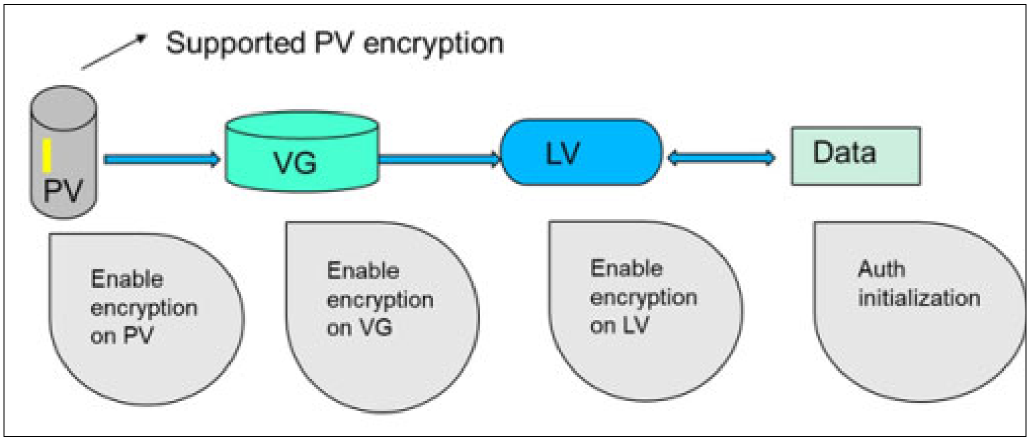
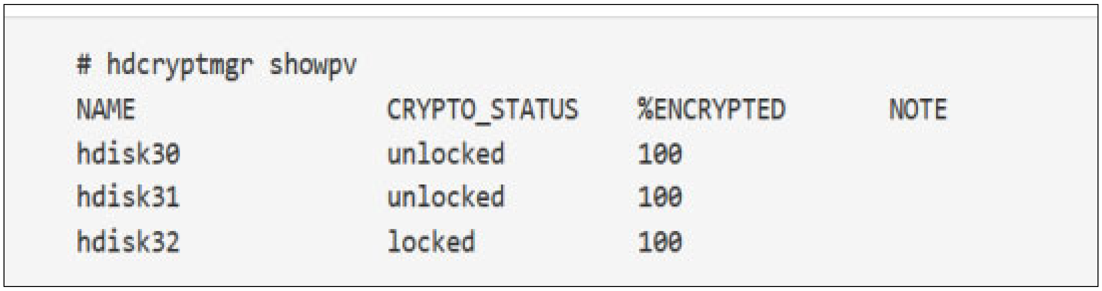


Figure 4-1 Physical volume encryption

Figure 4-2 shows the output of the `showpv` and `showmd` actions of the `hdcryptmgr` command. The `showpv` output displays three encrypted disks: two that are unlocked (able to be read from or written to) and one that is locked. Run `hdcrpytmgr authunlock hdisk32` on the locked disk before it is usable.



```
# hdcryptmgr showpv
NAME              CRYPTO_STATUS   %ENCRYPTED      NOTE
hdisk30           unlocked        100
hdisk31           unlocked        100
hdisk32           locked          100
```

Figure 4-2 Output from hdcryptmgr showpv

4.4.3 Limitations of encrypted PV

The encrypted PVs have the following limitations:

- The `rootvg` volume group must not contain any encrypted PV. If `rootvg` contains one or more encrypted PVs, the AIX start process fails. The `mkvg` and `extendvg` commands prevent the usage of encrypted PVs with `rootvg`.
- The existing PVs cannot be converted from unencrypted PVs to encrypted PVs, or vice versa. Enabling encryption on a PV deletes all the existing data on that disk.
- PV encryption requires extra disk attributes that are provided by the AIX operating system. If a disk is defined by using ODM definitions from another storage vendor, new ODM definitions from that vendor must be acquired to support PV encryption.
- Encrypted PVs can be shared with other AIX LPARs that are running AIX 7.3 Technology Level 1 or later. Sharing an encrypted PV with an older level of AIX corrupts data because the older level of AIX does not recognize that the PV is encrypted.
- PVs that are encrypted with PKS authentication can be used as traditional dump devices if they do not belong to the `rootvg` volume group.
- Encrypted PVs cannot be used as the destination disk when you use the `alt_disk_copy` and `alt_disk_mksysb` commands because the `rootvg` volume group does not support the encrypted PVs.
- Only SCSI PVs can be encrypted. You cannot encrypt NVMe or vPMEM disks.
- The same AIX operating system image cannot use geographical logical volume manager (GLVM) or AIX storage data caching (the `cache_mgt` command) with other SCSI disks while using encrypted PVs. GLVM or storage data caching can be used with NVMe disks or with vPMEM disks.

4.4.4 Disk backup considerations for encrypted PVs

The various methods of backing up data on a physical disk have different characteristics when encrypted PVs are used.

If the data backup operation is running in the operating system instance, the operating system reads data and decrypts that data before sending it to the backup software. The backup media contains the decrypted user data. The metadata that is related to encryption is not stored in the backup media. If this backup data is restored to another PV, data is encrypted only if encryption is enabled for that PV. If encryption is not enabled for the destination PV, the restored data is not encrypted and can be used directly even by older levels of AIX.

If data is backed up by using a storage device such as snapshot or IBM FlashCopy®, the data that is backed up is encrypted. The backup data in the storage device includes both the encryption metadata and the encrypted user data. The storage-based backup is a block-for-block copy of the encrypted data and the storage cannot determine that the data is encrypted by the operating system.

For more information about PV encryption in AIX, see the following resources:

- Understanding AIX Physical Volume Encryption
- Encrypted physical volumes
- Encrypting physical volumes

4.5 AIX access control lists

A standard UNIX environment uses only three attributes to determine which user has access to a file or a directory:

- Owner of the file or the directory.
- Group of the file or the directory.
- Permissions mode to define what the owner, the group, and others can do with the file (read, write, or execute).

In addition to the standard UNIX discretionary access control (DAC) AIX has ACLs. ACLs define access to files and directories more granularly. Typically, an ACL consists of series of entries that is called an Access Control Entry (ACE). Each ACE defines the access rights for a user in relationship to the object.

When access is attempted, the operating system uses the ACL that is associated with the object to see whether the user has the rights to do so. These ACLs and the related access checks form the core of the DAC mechanism that is supported by AIX.

The operating system supports several types of system objects that enable user processes to store or communicate information. The most important types of access-controlled objects are as follows:

- Files and directories
- Named pipes
- IPC objects such as message queues, shared memory segments, and semaphores

All access permission checks for these objects are made at the system call level when the object is first accessed. Because System V Interprocess Communication (SVIPC) objects are accessed statelessly, checks are made for every access. For objects with file system names, it is necessary to be able to resolve the name of the actual object. Names are resolved either relatively (to the process' working directory) or absolutely (to the process' root directory). All name resolution begins by searching one of these directories.

The DAC mechanism enables effective access control of information resources and provides for separate protection of the confidentiality and integrity of the information. Owner-controlled access control mechanisms are only as effective as users make them. All users must understand how access permissions are granted and denied, and how they are set.

For example, an ACL that is associated with a file system object (file or directory) can enforce the access rights for various users regarding access to an object. It is possible that such an ACL might enforce different levels of access rights, such as read/write, for different users.

Typically, each object has a defined owner, and sometimes are associated with a primary group. The owner of a specific object controls its discretionary access attributes. The owner's attributes are set to the creating process's effective UID.

The following list contains direct-access control attributes for the different types of objects:

- Owner

  For SVIPC objects, the creator or owner can change the object's ownership. SVIPC objects have an associated creator that has all the rights of the owner (including access authorization). The creator cannot be changed, even with root authority.

  SVIPC objects are initialized to the effective group ID (GID) of the creating process. For file system objects, the direct-access control attributes are initialized to either the effective GID of the creating process or the GID of the parent directory (determined by the group inheritance flag of the parent directory).

- Group

  The owner of an object can change the group. The new group must be either the effective GID of the creating process or the GID of the parent directory. (SVIPC objects have an associated creating group that cannot be changed, and they share the access authorization of the object group.)

- Mode

  The `chmod` command (in numeric mode with octal notations) can set base permissions and attributes. The `chmod` subroutine that is called by the command disables extended permissions. The extended permissions are disabled if you use the numeric mode of the `chmod` command on a file that has an ACL. The symbolic mode of the `chmod` command disables extended ACLs for the NSF4 ACL type but does not disable extended permissions for AIXC type ACLs. For more information about numeric and symbolic mode, see the `chmod` man page.

Many objects in the operating system, such as sockets and file system objects, have ACLs that are associated for different subjects. The details of these ACLs for these object types can vary from one to another.

Traditionally, AIX supported mode bits for controlling access to the file system objects. It also supported a unique form of ACL for mode bits. This ACL consisted of base mode bits, and can define multiple ACE entries, which each ACE entry defining access rights for a user or group for the mode bits. This classic type of ACL behavior is still supported as the AIXC ACL type.

The support of an ACL on file system objects depends on the underlying PFS. The PFS must understand the ACL data and be able to store, retrieve, and enforce the accesses for various users. It is possible that some of the PFSs do not support any ACLs at all (they might support only the base mode bits) compared to a PFS that support multiple types of ACLs. Few of the file systems under AIX are enhanced to support multiple ACL types. JFS2 and GPFS can support an NFS version 4 protocol-based ACL type too. This ACL is named NFS4 ACL type on AIX. This ACL type adheres to most of the ACL definition in the NFS version 4 protocol specifications. It also supports more granular access controls compared to the AIXC ACL type, and provides for capabilities such as inheritance.

4.6 Role-based access control

System administration is an important aspect of daily operations, and security is an inherent part of most system administration functions. Also, in addition to securing the operating environment, it is necessary to closely monitor daily system activities.

Most environments require that different users manage different system administration duties. It is necessary to maintain separation of these duties so that no single system management user can accidentally or maliciously bypass system security. Although traditional UNIX system administration cannot achieve these goals, RBAC can.

4.6.1 AIX role-based access control

AIX provided a limited RBAC implementation before AIX 6.1.

Beginning with AIX 6.1, a new implementation of RBAC provides for a fine-grained mechanism to segment system administration tasks. Because these two RBAC implementations differ greatly in function, the following terms are used:

- Legacy RBAC Mode: The historic behavior of AIX roles that apply to versions before AIX 6.1.
- Enhanced RBAC Mode: The new implementation that was introduced with AIX 6.1.

Both modes of operation are supported. However, Enhanced RBAC Mode is the default on newly installed AIX systems after AIX 6.1. The following sections provide a brief description of the two modes and their differences. They also include information about configuring the system to operate in the correct RBAC mode.

Legacy RBAC Mode

Before AIX 6.1, AIX provided limited RBAC functions that enabled non-root users to perform certain system administration tasks.

In this RBAC implementation, when an administrative command is started by a non-root user, the code in the command determines whether the user is assigned a role with the required authorization. If a match is found, the command execution continues. If not, the command fails with an error. Often, the command that is controlled by an authorization must be `setuid` to the root user for an authorized invoker to have the necessary privilege to accomplish the operation.

This RBAC implementation also introduced a predefined but user-expandable set of authorizations that can be used to determine access to administrative commands. Also, a framework of administrative commands and interfaces to create roles, assign authorizations to roles, and assign roles to users, is provided.

Although this implementation can partially segment system administration responsibilities, it functions with the following constraints:

- The framework requires changes to commands and applications to be RBAC-enabled.
- Predefined authorizations are not granular and the mechanisms to create authorizations are not robust.
- Membership in a certain group is often required, and a user must have a role with a given authorization to run a command.
- Separation of duties is difficult to implement. If a user is assigned multiple roles, there is no way to act under a single role. The user always has the authorizations for their roles.
- The least privilege principle is not adopted in the operating system. Commands must typically be SUID to the root user.

Legacy RBAC Mode is supported for compatibility, but Enhanced RBAC Mode is the default RBAC mode. Enhanced RBAC Mode is preferred on AIX.

Enhanced RBAC Mode

A more powerful implementation of RBAC is provided starting with AIX 6.1. Applications that require administrative privileges for certain operations have new integration options with the enhanced AIX RBAC infrastructure.

These integration options center on the use of granular privileges and authorizations and the ability to configure any command on the system as a privileged command. Features of the enhanced RBAC mode are installed and enabled by default on all installations of AIX beginning with AIX 6.1.

The enhanced RBAC mode provides a configurable set of authorizations, roles, privileged commands, devices, and files through the following RBAC databases. With enhanced RBAC, the databases can be either in the local file system or managed remotely through LDAP.

- Authorization database
- Role database
- Privileged command database
- Privileged device database
- Privileged file database

Enhanced RBAC mode introduces a new naming convention for authorizations that allows a hierarchy of authorizations to be created. AIX provides a granular set of system-defined authorizations, and an administrator may create more user-defined authorizations as necessary.

The behavior of roles was enhanced to provide separation of duty functions. Enhanced RBAC introduces the concept of role sessions. A *role session* is a process with one or more associated roles. A user can create a role session for any roles that they are assigned, thus activating a single role or several selected roles at a time. By default, a new system process does not have any associated roles. Roles are further enhanced to support the requirement that the user must authenticate before activating the role to protect against an attacker taking over a user session because the attacker would need to authenticate to activate the user's roles.

The introduction of the privileged command database implements the least privilege principle. The granularity of system privileges is increased, and explicit privileges can be granted to a command, and the execution of the command can be governed by an authorization. This approach provides the function to enforce authorization checks for command execution without requiring a code change to the command itself. Using the privileged command database eliminates the need for SUID and SGID applications because the capability of assigning only required privileges is possible.

The privileged device database provides access to devices to be governed by privileges, while the privileged file database provides unprivileged users access to restricted files based on authorizations. These databases increase the granularity of system administrative tasks that can be assigned to users who are otherwise unprivileged.

The information in the RBAC databases is gathered and verified and then sent to an area of the kernel that is designated as the Kernel Security Tables (KSTs). The state of the data in the KST determines the security policy for the system. Entries that are modified in the user-level RBAC databases are not used for security decisions until this information is sent to the KST by using the `setkst` command.

Note: For more information about RBAC on AIX, see Role-based access control.

4.6.2 AIX Toolbox for Open Source Software

Many of the most prominent tools that are used in AIX and Linux are open source and should be considered when designing and maintaining virtual machines (VMs) on IBM Power platforms. This list does not serve as an endorsement of these tools, but that they might be useful in your specific environment.

Note: For AIX users, these commands are available in the IBM AIX Toolbox for Open Source Software.

In addition to the GNU Public License (GPL), each of these packages includes its own licensing information, so review the individual tools for their licensing information.

Important: The freeware packages that are provided in the AIX Toolbox for Open Source Software are made available as a convenience to IBM customers. IBM does not own these tools; did not develop or exhaustively test them; or provide support for these tools. IBM compiled these tools so that they run with AIX.

The following tools are available:

- RBAC: `sudo` grants specific users or groups the ability to run commands as root or other users, which enhances security by limiting the need for users to have full root access.
- System monitoring and management:
  - Nagios: Monitors system metrics, services, and network protocols. Nagios Core is the open-source version.
  - Zabbix: An enterprise-level monitoring solution that provides real-time monitoring and alerting of various metrics.
  - Prometheus: A monitoring and alerting toolkit with a flexible query language and powerful visualization capabilities when paired with Grafana.

- Intrusion detection:
  - OSSEC: A host-based intrusion detection system (HIDS) that performs log analysis, file integrity checking, and more.
  - Snort: An open-source network intrusion detection system (NIDS) that can be configured as an intrusion prevention system (IPS).
- Network analysis:
  - Wireshark: A widely used network protocol analyzer that helps capture and analyze network traffic.
  - `tcpdump`: An ACLI packet analyzer that is used for network diagnostics and monitoring.
- Log management and analysis:
  - The Elasticsearch, Logstash, and Kibana (ELK) stack: A powerful suite for log management and analysis, which helps in searching, analyzing, and visualizing log data.
  - Graylog: A log management tool that provides real-time analysis, alerting, and visualization capabilities.
- File synchronization: `rsync` is a file synchronization and backup tool that is used for incremental backups and mirroring data.
- Package management: YUM and DNF are package managers for RPM-based distributions that handle package installations, updates, and dependencies.

4.7 AIX Security Expert compliance

AIX Security Expert is a comprehensive tool for managing system security settings, which include TCP, NET, Internet Protocol Security (IPsec), system configurations, and auditing. A As part of the `bos.aixpert` file set, AIX Security Expert facilitates system security hardening through a simple interface. The tool offers predefined security levels of High, Medium, Low, and AIX Standard Settings, which cover over 300 security configurations. It also enables advanced administrators to customize each security setting as needed.

With AIX Security Expert, you can easily apply a chosen security level without the need for extensive research and manual implementation of individual security elements. Also, the tool enables you to create a security configuration snapshot, which you can use to replicate the same settings across multiple systems, which streamline security management and ensure consistency across an enterprise environment.

AIX Security Expert can be accessed either through SMIT or by using the `aixpert` command.

AIX Security Expert settings

The following coarse-grain security settings are available:

- High-Level Security: Applies the high-level security settings definition.
- Medium-Level Security: Applies the medium-level security settings definition.
- Low-Level Security: Applies the low-level security settings definition.
- Advanced Security: Applies custom user-specified security settings.
- AIX Standard Settings: Uses the original system default security settings.
- Undo Security: Allows some AIX Security Expert configuration settings to be undone.
- Check Security: Provides a detailed report of the security settings.

4.7.1 AIX Security Expert security hardening

Security hardening protects all elements of a system by tightening security or implementing a higher level of security. It helps ensure that all security configuration decisions and settings are adequate and appropriate. Hundreds of security configuration settings might need to be changed to harden the security of an AIX system.

AIX Security Expert provides a menu to centralize effective and common security configuration settings. These settings are based on extensive research on properly securing UNIX systems. Default security settings are provided for broad security environment needs (High-Level Security, Medium-Level Security, and Low-Level Security), and advanced administrators can set each security configuration setting independently.

Configuring a system at too high a security level might deny necessary services. For example, telnet and rlogin are disabled for High-Level Security because the login password is sent over the network unencrypted. Conversely, if a system is configured at too low a security level, it can be vulnerable to security threats. Because each enterprise has its own unique set of security requirements, the predefined High-Level Security, Medium-Level Security, and Low-Level Security configuration settings are best used as a starting point rather than an exact match for the security requirements of an enterprise.

The practical approach to using AIX Security Expert is to establish a test system (in a realistic test environment) similar to the production environment in which it will be deployed. Install the necessary business applications and run AIX Security Expert by using the GUI. The tool analyzes this running system in its trusted state. Depending on the security options that you choose, AIX Security Expert enables port scan protection, turns on auditing, blocks network ports that are not used by business applications or other services, and applies many other security settings. After re-testing with these security configurations in place, the system is ready to be deployed in a production environment. Also, the AIX Security Expert XML file that defines the security policy or configuration of this system can be used to implement the same configuration on similar systems in your enterprise.

Note: For more information about security hardening, see NIST Special Publication 800-70, NIST Security Configurations Checklist Program for IT Products.

For a full description of AIX Security Expert on AIX 7.3, see AIX Security Expert.

4.8 File Permission Manager

The AIX File Permission Manager manages the permissions of commands and daemons that are owned by privileged users with `setuid` or `setgid` permissions.

The `fpm` command enables administrators to harden their system by setting permissions for important binary files and dropping the `setuid` and `setgid` bits on many commands in the operating system. This command is intended to remove the `setuid` permissions from commands and daemons that are owned by privileged users, but you can also customize it to address the specific needs of unique computer environments.

The `setuid` programs on the base AIX operating system are grouped to enable levels of hardening. This grouping enables administrators to choose the level of hardening according to their system environment. Also, you can use the `fpm` command to customize the list of programs that must be disabled in your environment. Review the levels of disablement and choose the right level for your environment.

Changing the execution permissions of commands and daemons with the fpm command affects non-privileged users by denying their access to these commands and daemons or functions of the commands and daemons. Also, other commands that call or depend on these commands and daemons can be affected. Any user-created scripts that depend on commands and daemons with permissions that were altered by the fpm command cannot operate as expected when run by non-privileged users. Give full consideration to the effect and potential impact of modifying default permissions of commands and daemons.

Perform appropriate testing before using this command to change the execution permissions of commands and daemons in any critical computer environment. If you encounter problems in an environment where execution permissions were modified, restore the default permissions and re-create the problem in this default environment to ensure that the issue is not due to lack of appropriate execution permissions.

The fpm command provides the capability to restore the original AIX installation default permissions by using the -l default flag.

Also, the fpm command logs the permission state of the files before changing them. The fpm log files are created in the /var/security/fpm/log/<date_time> file. If necessary, you can use these log files to restore the system's file permissions that are recorded in a previously saved log file.

When the fpm command is used on files that have extended permissions, it disables the extended permissions, although any extended permission data that existed before the fpm invocation is retained in the extended ACL.

Customized configuration files can be created and enacted as part of the high, medium, low, and default settings. File lists can be specified in the /usr/lib/security/fpm/custom/high/* directory, the /usr/lib/security/fpm/custom/medium/* directory, and the /usr/lib/security/fpm/custom/default/* directory. To leverage this feature, create a file containing a list of files that you want to be automatically processed in addition to the fpm command's internal list. When the fpm command runs, it also processes the lists in the corresponding customized directories. To see an example of the format for a customized file, view the /usr/lib/security/fpm/data/high_fpm_list file. The default format can be viewed in the /usr/lib/security/fpm/data/default_fpm_list.example file. For the customization of the -l low flag, the fpm command reads the same files in the /usr/lib/security/fpm/custom/medium directory, but removes the setgid permissions, but the -l medium flag removes both the setuid and setgid permissions.

The fpm command cannot run on Trusted Computing Base (TCB)-enabled hosts.

AIX File Permission Manager examples

Here are some AIX File Permission Manager examples:

- To apply the fpm command's low-level security settings, run the following command:

  ```
  fpm –l low
  ```

  This command also processes any file list in the /usr/lib/security/fpm/custom/med/ directory.

- To check whether the system commands are presently set to fpm low-level permissions, run the following command:

  ```
  fpm –c –l low
  ```

  This command reports any file with permissions out of conformance.

- To restore the traditional, default permissions, run the following command:

  ```
  fpm –l default
  ```

  This command also processes any file list in the `/usr/lib/security/fpm/custom/default/` directory.
- To list or preview what permission changes must be done to make the system compliant with the `fpm` command's high-level security without changing any file permissions, run the following command:

  ```
  fpm -l high –p
  ```

  This command also previews any file list in the `/usr/lib/security/fpm/custom/high/` directory.
- To apply the `fpm` command's high-level security settings, run the following command:

  ```
  fpm –l high
  ```

  This command also processes any file list in the `/usr/lib/security/fpm/custom/high/` directory.
- To list the status of the system that was changed by the `fpm` command, run the following command:

  ```
  fpm –s
  ```
- If the `fpm -l` level command was run on 7 January 2024 at 8:00 AM, then the permission state of the affected files was captured by the `fpm` command before it made any changes. To restore the file permissions to their state of 7 January 2007 at 8:00 AM, run the following command:

  ```
  fpm –l default –f /var/security/fpm/log/01072024_08:00:00
  ```

4.9 Trusted Execution

Trusted Execution (TE) refers to a collection of features that are used to verify the integrity of the system and implement advanced security policies, which together can be used to enhance the trust level of the complete system.

The usual way for a malicious user to negatively impact the system typically involves gaining unauthorized access and then installing harmful programs like trojans or rootkits, or modifying sensitive security files, which render the system vulnerable and prone to exploitation. TE aims to prevent such activities or in cases where incidents do occur, quickly identify them.

Using the functions that are provided by TE, the system administrator can define the exact set of executable files that are permitted to run or specify the kernel extensions that may load. Also, you can use TE to examine the security status of the system and identify files that were updated, which raises the trustworthiness of the system and makes it harder for an attacker to cause damage.

The set of features under TE can be grouped into the following categories:

- Managing a Trusted Signature Database
- Auditing the integrity of the Trusted Signature Database
- Configuring security policies
- Trusted Execution Path and Trusted Library Path

Important: There is a TCB function in the AIX operating system. However, TCB features can be enabled during the Base Operating System installation process *only* by expressly turning on its corresponding menu option or by performing a Preservation installation on an installed AIX system.

TE is a powerful and enhanced mechanism that overlaps some of the TCB functions and provides advance security policies to better control the integrity of the system. Although the TCB is still available, TE introduces a new and more advanced concept of verifying and guarding the system integrity.

AIX Trusted Execution uses whitelisting to prevent or detect malware that runs on your AIX system. It provides the following features:

- Provides cryptographic checking that you can use to determine whether a hacker replaced an IBM published file with his trojan horse.
- San for root kits.
- Detect whether various attributes of a file were altered.
- Corrects certain file attribute errors.
- Provides whitelisting.
- Provides a numerous configuration options.
- Detects and prevents malicious scripts, executable files, kernel extensions, and libraries.
- Protects files from being altered by a hacker that gains root access.
- Provides a function to protect the TE's configuration from a hacker that gains root access.
- Provides a function for using digital signatures to verify that IBM and non IBM published files were not altered by an attacker.

TE is available in AIX 6 and later.

Trusted Signature Database management

Similar to TCB, there is a database that is used to store the critical security parameters of trusted files that are on the system. This database is the Trusted Signature Database (TSD), and is in the `/etc/security/tsd/tsd.dat` file.

Auditing the integrity of the Trusted Signature Database

To audit the integrity state of the file definitions in the TSD against the actual files, use the `trustchk` command.

Security policies configuration

The TE feature provides a runtime file integrity verification mechanism. Use this mechanism to configure the system to check the integrity of the trusted files before every request to access those file so that only the trusted files that pass the integrity check can be accessed on the system.

Quick reference for security checks

To enable or disable the Trusted Library Path or Trusted Execution Library and set the colon-separated path list for both, use the `trustchk` command.

Table 4-3 compares the TCB function with the TE function.

Table 4-3 Comparison of TE and TCB function

Function	TE	TCB
Integrity Checking reference	System and runtime checking	System checking only
System Enablement	Enabled at any time	Installation time option
Security Database Files	`/etc/security/tsd/tsd.dat`	`/etc/security/sysck.cfg`
Management Commands	`trustchk`	`tcbck`

Trusted Execution Policy Management

To enable or disable different security policies that are used with the TE mechanism, use the `trustchk` command. You can specify the policies that are shown in Table 4-4.

Table 4-4 Policies

Policy	Policy description
CHKEXEC:	Checks the integrity of commands before running the commands.
CHKSHLIBS:	Checks the integrity of shared libraries before loading the libraries.
CHKSCRIPTS:	Checks the integrity of shell scripts before running the scripts.
CHKKERNEXT:	Checks the integrity of kernel extensions before loading the extensions.
LOCK_TSD:	Disables modification of the TSD.
LOCK_TSD_	FILES: Disables modification of TSD files.
STOP_UNTRUSTD:	Does not load files unless they are in the TSD.
STOP_ON_CHKFAIL:	If an integrity check fails, the policy does not load a file.
TEP:	Allows execution of commands from a defined list of directories.
TLP:	Allows library loads from a defined list of directories.

For TE to work, the CryptoLight for C library (CLiC) and kernel extension must be installed. To see whether it is installed and loaded into the kernel, run the commands that are shown in Example 4-44.

Example 4-44 Checking for the CLiC installation

```
# lslpp -l "clic*"
File set Level State Description
----------------------------------------------------------------------------
Path: /usr/lib/objrepos
clic.rte.kernext 4.3.0.0 COMMITTED CryptoLite for C Kernel
clic.rte.lib 4.3.0.0 COMMITTED CryptoLite for C Library
Path: /etc/objrepos
clic.rte.kernext 4.3.0.0 COMMITTED CryptoLite for C Kernel
# genkex|grep clic
4562000 37748 /usr/lib/drivers/crypto/clickext
```

If the file set is not installed, install it on your system and load it into the kernel when installation completes successfully, by running the following command:

```
# /usr/lib/methods/loadkclic
```

The TSD database stores the critical security parameters of trusted files that are on the system. This database is in at `/etc/security/tsd/tsd.dat` and comes with AIX media. In TE’s context, trusted files are files that are critical from the security perspective of a system and, if compromised, can jeopardize the security of the entire system. Typically, the files that match this definition are as follows:

- Kernel (operating system)
- All SUID root programs
- All SGID root programs
- Any program that is exclusively run by root or by a member of the system group
- Any program that must be run by the administrator while on the trusted communication path (for example, the `ls` command)
- The configuration files that control system operation
- Any program that is run with the privilege or access rights to alter the kernel
- The system configuration files

Every trusted file must ideally have an associated stanza or a file definition that is stored in the TSD. You can mark a file as *trusted* by adding its definition to the TSD by using the `trustchk` command. You can use this command to add, delete, or list entries from the TSD. You can lock The TSD so that even root cannot write to it. Locking the TSD becomes immediately effective.

Example 4-45 shows how the `ksh` command appears in the TSD database file.

Example 4-45 The ksh entry in the TSD database file

```
/usr/bin/ksh:
Owner = bin
Group = bin
Mode = TCB,555
Type = FILE
Hardlinks = /usr/bin/sh,/usr/bin/psh,/usr/bin/tsh,/usr/bin/rksh
Symlinks =
Size = 294254
Cert_tag = 00af4b62b878aa47f7
Signature =
8e8118ec793fd4899ccc38c0f4ab88571b0488024aff80f83d0bde2380f3ae44137a26607cd5d4c5e5
8e02ad1f872ca1c398f8702ad38f3a0f0a584c2061bb09de3e5218405f1b07d80efe0be192d3333b8c
d49a4ff980ce5e1f15f6b64d3b38f75d0cc6fb5ef9e7d8b410547c40181847c5ae980979abf3279f25
c6b512178a
hash_value = f3a2e9b92e2cfc10ffb2274680c97f29742ff2dd12dda04de85544fd8c039fd8
t_accessauths = aix.mls.system.access.dir
t_innateprivs = PV_DAC_R,PV_DAC_X,PV_MAC_R


/usr/lib/drivers/igcts:
Owner = root
Group = system
Mode = 555
Type = HLINK
```

```
Size = 7714
Cert_tag = 00af4b62b878aa47f7
Signature =
b47d75587bbd4005c3fe98015d9c0776fd8d40f976fb0f529796ffe1b2f9028500ffd2383ca31cd2f3
9712f70e36c522dc1ba52c44334781a389ea06cdabd82c72d705fd94
bffe59817b5a4d45651e2d5457cb83ebdb3b705a3b5c981c51eae79facfe271fbde0e396b7ea64d4db
d6ab753a3fa7a9578b7f5e6458b83d8f08df
Hash_value = 6d13bbd588ecfdd06cbb2dc3a17eabad6b51a42bd1fd62e7ae5402a75116e8bd
```

To enable TSD protection, run the commands that are shown in Example 4-46.

Example 4-46 Enabling TSD protection

```
# trustchk -p tsd_lock=on
# trustchk -p te=on
```

The TSD is immediately protected against any modification by either the `trustchk` command or by manually editing the file, as shown in Example 4-47.

Example 4-47 TSD is protected

```
# trustchk -d /usr/bin/ps
Error writing to database file
# echo >> /etc/security/tsd/tsd.dat
Operation is not permitted.
```

To enable the TSD for write access again, turn off TE or set `tsd_lock` to `off` by using the `trustchk` command.

When the system is blocking any untrusted shell scripts by using the CHKSCRIPT policy, as shown in Example 4-48m make sure that all scripts that are needed by your services are included in the TSD.

Example 4-48 Commands to set chkscript on and stop untrusted execution

```
# trustchk -p stop_untrustd=on
# trustchk -p chkscript=on
```

For example, if you are using OpenSSH, make sure that the `sshd` and `ksshd` start and stop scripts in `/etc/rc.d/rc2.d` are in the TSD. Otherwise, `sshd` does not start when the system is restarted, and it will not shut down on system shutdown.

When you try to start a script with `chkscript=on` and that script is not included in the TSD, its execution is denied regardless of its permissions, even when root is starting it. This situation is shown in Example 4-49.

Example 4-49 Showing permission denied for access

```
# ./foo
ksh: ./foo: 0403-006 permission denied.

# ls -l foo
-rwx------- root system 17 May 10 11:51 foo
```

The Trusted Execution Path defines a list of directories that contain the trusted commands. When Trusted Execution Path verification is enabled, the system loader allows commands in the specified paths to run.

The Trusted Library Path has the same function as Trusted Execution Path, with the only difference is that it is used to define the directories that contain trusted libraries of the system. When the Trusted Library Path is enabled, the system loader allows the libraries from this path to be linked to the commands.

You can use the `trustchk` command to enable or disable the Trusted Execution Path or Trusted Execution Library or to set the colon-separated path list for both by using the Trusted Execution Path and Trusted Library Path command-line interface (CLI) attributes of `trustchk`.

> **Important:** Be careful when you are changing either the Trusted Execution Path or the Trusted Library Path. Do not remove the paths from their default settings, which are as follows:
>
> ```
> TEP=/usr/bin:/usr/sbin:/etc:/bin:/sbin:/sbin/helpers/jfs2:/usr/lib/instl:/usr/c
> cs/bin
> TLP=/usr/lib:/usr/ccs/lib:/lib:/var/lib
> ```
>
> If you remove the paths, the system will not restart and function properly because it cannot access the necessary files and data.

Here are some common command usages of the `trustchk` command:

- To perform a system check comparison with the TSD and report errors, run the following command:

  ```
  # trustchk -n ALL
  ```

- To delete the entry for `/usr/bin/ls` in the TSD, run the following command:

  ```
  # trustchk -d /usr/bin/ls
  ```

- To enable a policy that checks the commands that are listed in TSD on every load, run the following command:

  ```
  # trustchk -p CHKEXEC=ON
  ```

- To turn on runtime TSD checking, run the following command:

  ```
  # trustchk -p TE=ON
  ```

- To check the runtime policy in effect, run the following command:

  ```
  # trustchk -p
  ```

Here are some other examples:

- Adding the `STOP_ON_CHKFAIL` option stops commands when they fail the test.

  The trustchk flags are as follows:

`TE=[ON|OFF]`	Turns on runtime checks.
`CHKEXEC=[ON|OFF]`	Turns on command checks.
`STOP_ON_CHKFAIL= [ON|OFF]`	Stops commands if they fail the test.
`STOP_UNTRUSTD= [ON|OFF]`	Stops commands that are not listed in `/etc/security/tsd/tsd.dat`.

  Example 4-50 on page 121 shows an example of the command.

Example 4-50 Demonstrating STOP_ON_CHKFAIL

```
# openssl dgst -sha256 /usr/bin/ls | awk '{print $2}'
8f3505509771df3915b6f8c7e45fc6a56ec68d4c082bfb640f89c2251bf9550c

# trustchk -q /usr/bin/ls | grep hash_value | awk '{print $3}'
8f3505509771df3915b6f8c7e45fc6a56ec68d4c082bfb640f89c2251bf9550c

# cp /usr/bin/ls /usr/bin/.goodls
- Hash value of "/usr/bin/ls" command changed
# trustchk -p TE=ON CHKEXEC=ON STOP_ON_CHKFAIL=ON

# ls
ksh: ls: 0403-006 permission denied.

# cp /usr/bin/ls /usr/bin/.badls
# cp /usr/bin/.goodls /usr/bin/ls
# chown bin:bin /usr/bin/ls

# ls
file1 file2 dir1
```

- Adding the `STOP_UNTRUSTD=ON` option stops executable files that are not listed in `/etc/security/tsd/tsd.dat`, as shown in Example 4-51.

Example 4-51 Demonstrating STOP_UNTRUSTD=ON

```
# trustchk -p TE=ON CHKEXEC=ON STOP_UNTRUSTD=ON
# ls
file1 file2 dir1
# /usr/bin/.goodls
ksh: /usr/bin/.goodls: 0403-006 permission denied.
# ls -l /usr/bin/.goodls
-r-xr-xr-x 1 bin bin 26732 May 28 17:39 /usr/bin/.ls
```

4.10 Internal Protocol Security network filtering

This section describes IBM AIX Internal Protocol Security (IPsec) network filters and tunneling.

With the constant threat of security breaches, companies are under pressure to lock down every aspect of their applications, infrastructure, and data.

One method of securing IBM AIX network transactions is to establish networks that are based on the IPsec protocol. IPsec is an IBM AIX network-based protocol that defines how to secure a computer network at the IP layer. When determining how to secure your IPsec connections, you might need to consider these items:

- The connectivity architecture, whether it is an internal or external connection
- Encryption mechanisms or by using authentication services

IBM AIX IPsec uses the `mkfilt` and `genfilt` binary files to activate and add the filter rules. You can also use it to control the filter logging functions, which work on IP version 4 and IP version 6. With the IPsec feature enabled, you can also create IP filtering rules to block the IP address from accessing hosts or exact ports.

One of the interesting features of IPsec is dynamic tunnels. These tunnels use the Internet Key Exchange (IKE) protocol to protect IP traffic by authenticating and encrypting IP data. The `ike` command performs several functions, such as activate, remove, or list IKE and IPsec tunnels.

For more information, see the following resources:

- IBM AIX: Using IPsec Rules to Filter Network Traffic
- The mkfilt command
- The genfilt command
- The ike command
- The command-line interface for IKE tunnel configuration

4.11 AIX Event Monitoring

AIX has built-in functions for monitoring and auditing events that occur within the environment. These functions are described in this section.

4.11.1 Auditing

Auditing is the process of examining systems, accounts, or activities to verify accuracy, compliance, and efficiency. An auditing subsystem records system events to monitor and analyze transactions, and help ensure transparency and traceability.

By default, auditing is disabled in AIX. When activated, the auditing subsystem begins collecting information based on your configuration settings. The frequency of auditing depends on your environment and usage patterns. Although auditing is a best practice for enhanced security and troubleshooting, the decision to enable auditing and its frequency is yours.

Any security-relevant occurrence on the system is considered an auditable event. Auditing involves detecting, recording, and analyzing these events to maintain system security and integrity. The set of auditable events determines which occurrences can be audited and the level of detail that is provided. By examining audit trails, organizations can identify patterns, trends, and anomalies that might indicate potential security threats or issues.

Understanding the AIX auditing subsystem

The auditing subsystem provides a mechanism to record security-related information and alert system administrators of potential or actual security policy violations. Collected information includes the auditable event name, status (success or failure), and event-specific security details.

The auditing subsystem contains detection, collection, and processing functions:

- Event detection: Distributed throughout the TCB in the kernel and trusted programs. Detects security-relevant occurrences and reports them to the system audit logger.
- Event information collection: The kernel audit logger records selected auditable events, which construct a complete audit record that consists of a header and trail. The audit trail can be written in BIN or STREAM mode.
- Audit trail information processing: The operating system offers various options for processing the audit trail, which includes compression, filtering, and formatting.

The system administrator can configure each of these functions.

Auditing event detection

Event detection is integrated throughout the TCB, and it encompasses both the kernel (supervisor state code) and trusted programs (user state code). An auditable event is any occurrence that is relevant to system security, such as changes to the system's security state, attempted or actual violations of access control or accountability policies, or both. Programs and kernel modules that are responsible for detecting these events must report them to the system audit logger, which operates as part of the kernel. The logger can be accessed through a subroutine (for trusted program auditing) or through a kernel procedure call (for supervisor state auditing). The reported information includes the event name, its success or failure status, and any extra details that are pertinent to security auditing.

Event detection configuration

Configuring event detection involves enabling or disabling the audit subsystem and specifying which events to audit for particular users. To manage event detection, use the `audit` command to activate or deactivate the subsystem. The `/etc/security/audit/config` file contains the configuration details for the events and users that are processed by the audit subsystem.

Event information collection

Information collection encompasses logging the selected auditable events. This function is performed by the kernel audit logger, which provides both a system call and an intra-kernel procedure call interface that records auditable events.

The audit logger is responsible for constructing the complete audit record, which contains the audit header, which contains information that is common to all events (such as the name of the event, the user responsible, and the time and return status of the event), and the audit trail, which contains event-specific information. The audit logger appends each successive record to the kernel audit trail, which can be written in either (or both) of two modes:

- BIN mode

  The trail is written into alternating files, which provide safety and long-term storage.

- STREAM mode

  The trail is written to a circular buffer that is read synchronously through an audit pseudo-device. STREAM mode offers immediate response.

Information collection can be configured at both the front end (event recording) and at the back end (trail processing). Event recording is selectable on a per-user basis. Each user has a defined set of audit events that are logged in the audit trail when they occur. At the back end, the modes are individually configurable so that the administrator can employ the back-end processing that is best suited for a particular environment. In addition, BIN mode auditing can be configured to generate an alert if the file system space that is available for the trail is getting too low.

Audit trail information processing

The operating system offers various options for processing the kernel audit trail. In BIN mode, the audit trail can be compressed, filtered, or formatted for output, or a combination of these methods can be used before archiving the audit trail.

- Compression: Performed by using Huffman encoding.
- Filtering: Achieved through SQL-like audit record selection (by using the `auditselect` command), which allows for selective viewing and retention of records.
- Formatting: Enables examination of the audit trail, generation of periodic security reports, and printing of a paper audit trail.

These processing options help manage and analyze audit data effectively.

The STREAM mode audit trail can be monitored in real time to provide an immediate threat-monitoring capability. Configuring these options is handled by separate programs that can be started as daemon processes to filter either BIN or STREAM mode trails, although some of the filter programs are more naturally suited to one mode or the other.

To help ensure that the AIX audit subsystem can retrieve information from the AIX security audit, set the following files of the AIX server to be monitored:

- `streamcmds`
- `config`
- `events`
- `objects`

For more information about how to configure the AIX audit subsystem for collecting, recording, and auditing the events, see the following resources:

- AIX AUDIT: The Audit Subsystem in AIX
- The config file
- The events file
- Configuring the AIX Audit subsystem
- Auditing overview
- Enhanced auditing

4.11.2 Accounting

The accounting subsystem provides features for monitoring system resource usage and billing users for the usage of resources. Accounting data can be collected on various system resources: processors, memory, disks, and such.

Other data that is collected by the accounting system is connect-time usage accounting, which lets you know how many users are connected to a system and for how long. You can use the connect time data to detect unused accounts, which must be invalidated (for security reasons) or even erased to save resources. Also, you can use the connect-time usage data to discover suspect activities (such as too many unsuccessful logon attempts) that signal that security measures should be adopted.

The data that is collected by the accounting subsystem is used to automatically generate reports, such as daily and weekly reports. The reports can be generated at any time by using accounting-specific commands. The accounting subsystem provides tools that you can use to observe how the system reacts at a particular moment in time (for example, when running a specific command or task).

Accounting data provides valuable information to accomplish the following goals:

- Develop effective charge-back policies.
- Assess the adequacy of the current resources.
- Effectively balance and control resource allocation.
- Forecast future needs.

For more information about how to set up accounting subsystem and accounting internals, see Administering system accounting.

4.11.3 AIX Event Infrastructure for AIX and AIX clusters

The AIX Event Infrastructure is an event monitoring framework for monitoring predefined and user-defined events.

An event in the AIX Event Infrastructure refers to any detectable change in a system's state or values by the kernel or its extensions at the moment that the modification takes place. These events are stored as files within a specialized file system that is known as the pseudo file system. The AIX Event Infrastructure offers several benefits:

- There is no need for constant polling. Users monitoring the events are notified when those events occur.
- Detailed information about an event (such as stack trace, and user and process information) is provided to the user monitoring the event.
- Existing file system interfaces are used so that there is no need for a new application programming interface (API).
- Control is handed to the AIX Event Infrastructure at the time that the event occurs.

Autonomic Health Advisor File System (AHAFS) architecture

The AIX Event Infrastructure is made up of the following four components:

1. The kernel extension that implements the pseudo file system.
2. The event consumers that consume the events.
3. The event producers that produce events.
4. The kernel component that serves as an interface between the kernel extension and the event producers.

Figure 4-3 illustrates the architecture.

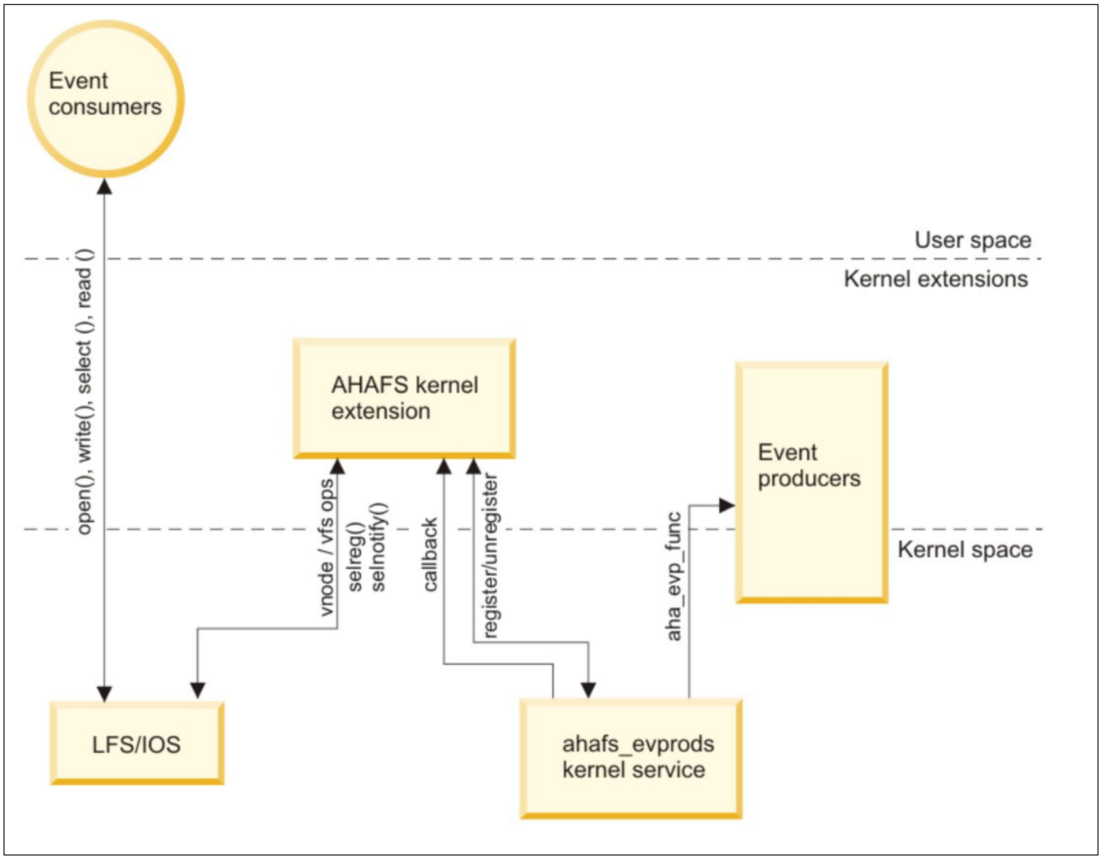


Figure 4-3 Autonomic Health Advisor File System architecture[1]

For more information, see the AIX Event Infrastructure documentation.

4.12 In-core cryptography support

Using encryption to protect data requires a significant amount of processing to encrypt and decrypt data. To reduce the performance impact of these activities, IBM POWER processors provide on-chip capabilities to off-load these processor-intensive cryptographic operations. These benefits are provided by on-chip accelerators and by using new crypto processor instructions (in-core) to accelerate these functions. These hardware accelerators can reduce CPU usage and improve the performance of AIX applications that use crypto features.

IBM POWER7+ was the first IBM POWER processor to include Nest Accelerator (NX) for symmetric (shared key) cryptography. The accelerators are shared among the LPARs under the control of the PowerVM Hypervisor, and accessed through a hypervisor call. The internal NX crypto API calls require extra pages of memory to perform the relevant hypervisor calls. The impact of NX calls makes them suitable for large data only. A tuning parameter (`min_sz`) for data size is implemented to set the minimal data size for NX accelerator operations.

[1] Source: https://www.ibm.com/docs/en/aix/7.3?topic=ahafs-aix-event-infrastructure-components

The IBM Power8® processor provided a new set of VMX and VSX in-core symmetric cryptographic instructions that are aimed at improving performance of various crypto operations. In most circumstances, the in-core crypto instructions provide better performance with lower latency and no extra page requirements. To use the in-core crypto instructions in the kernel, you must have a few resources to save and restore the vector register content.

More improvements were made in the IBM Power9 and IBM Power10 processors to increase the encryption capabilities and improve system performance.

Advance Crypto Facility in AIX

The Advance Crypto Facility (ACF) is the AIX cryptographic framework that provides crypto services (APIs) for kernel and user space applications. It implements the supported crypto algorithms in software that can be replaced by other crypto providers, such as crypto cards and hardware accelerations when the respective hardware acceleration is enabled. The leverage of hardware acceleration is done in a manner transparent to the callers.

The ACF kernel services are implemented in the PKCS11 device driver (kernel extension), which provides services for other kernel subsystems like EFS, IPsec, and LV-Encryption. User space applications can also use ACF kernel services by calling the AIX PKCS #11 subsystem library (`/usr/lib/pkcs11/ibm_pkcs11.so`).

Here are some details about ACF:

- The purpose of this feature is to improve the performance in the PKCS11 kernel extension.
- Using the AES instruction set can reduce CPU usage and improve the performance of AIX applications that use AES crypto features, such as EFS, IPsec, and TE.
- This feature enables the in-core vector AES crypto instructions under the CLiC interfaces of the PKCS11 kernel extension.
- Customers can enable or disable the in-core support in the ACF kernel extension though a CLI.
- ODM support is provided to enable or disable the feature after restarts.
- Displays the status of the in-core crypto enablement.
- Supports IBM Power8 and later processors.
- Prerequisites:
  - OS level: 7.2 TL5 and later
  - Virtual I/O Server (VIOS): 3.1
  - Hardware: Power8 or later processor
  - Firmware: Any
- Enablement:

  Two flags were introduced:

  - `in_core_capable`
  - `in_core_enabled` (`acfo -t in_core_enabled=1`)
- Recovery (how to turn off or disable ACF if something goes wrong):

  ```
  acfo -t in_core_enabled=0
  ```
- Default settings: For Power8 and later processors, the default values are `in_core_Capable=1` and `in_core_enabled=0`.

For more information, see Exploitation of In-Core Acceleration of POWER processors for AIX.

4.13 Lightweight Directory Access Protocol

Lightweight Directory Access Protocol (LDAP) is a distributed hierarchical, directory service access protocol that you use to access the repositories of users and other network-related entities. LDAP is an open standard for managing directory data.

LDAP defines a message protocol that is used by directory clients and directory servers. LDAP originated from the X.500 Directory Access Protocol, which is considered heavyweight. X.500 needs the entire OSI protocol stack, and LDAP is built on the TCP/IP stack. LDAP is considered lightweight because it omits many X.500 operations that are rarely used.

An application-specific directory stores the information that does not have general search capabilities. Keeping multiple copies of information up-to-date and synchronized is difficult. What is needed is a common, application-independent directory. You can achieve a single, common directory by using LDAP. Clients can interact independent of the platform. Also, clients can be set up without any dependencies.

LDAP works with most vendor directory services, such as Microsoft Active Directory (AD). With LDAP, sharing information about users, services, systems, networks, and applications from a directory service to other applications and services is simpler to implement. When using LDAP, client access is independent of the platform. Because LDAP is a standard protocol, you can set up clients without any dependency on the specific LDAP server that you use.

For example, if you have a Microsoft AD (LDAP server), you can configure an LDAP client with IBM Tivoli Directory Server file sets and access the data from the server.

Example 4-52 shows a sample of an LDAP entry for multiple applications.

Example 4-52 Sample LDAP entries

```
Application 1: FirstName, LastName
Application 2: name
Application 3: firstname, middlename, lastname
```

When you implement a setup to use LDAP, multiple applications such as IBM Verse, intranet page, BestQuest, RQM, and ClearQuestcan may be connected to a user entry in the same directory. If a user changes their password once, it is reflected in all applications.

Figure 4-4 illustrates this concept.

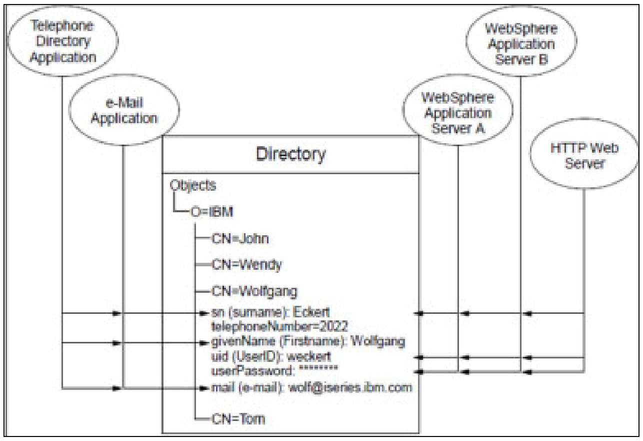


Figure 4-4 Several applications that use the attributes of a single entry

4.13.1 How LDAP works

An application that wants to read/write information in a directory does not access the directory directly. It calls a function or API that sends a message to another process. This second process accesses the information in the directory on behalf of the requesting application through TCP/IP.

Here are some considerations for this process:

- The default TCP/IP ports are 636 for secure communications and 389 for unencrypted communications.
- Use the `mksecdlap` command in AIX to set up LDAP.
- IBM Db2® is the database.

4.13.2 Where LDAP is used

Most of the applications need a function to authenticate the user and store some user or device information. Without LDAP, you create database schema, create tables, and maintain the database. Also, you must create an application program that accesses the database. With LDAP, the LDAP server does all this work. The user uses the services by using the available APIs.

Here is a list of possible LDAP configurations:

- Server configurations:
  - Master/replica.
  - Peer-peer.
- Client configurations:
  - LDAP Anonymous bind: Automated per ISST Auto bucket.
  - LDAP Automount: Automated per ISST auto bucket.
  - LDAP Netgroups: Automated per ISST auto bucket.
  - LDAP with Auth only mode: Automated per ISST auto bucket.
  - IPsec with LDAP: The automation code is ready, and you must need to verify it.
  - LDAP with Secure Sockets Layer (SSL) communication.

4.13.3 LDAP usage in AIX

IBM Security Verify Directory is a highly scalable and robust LDAP directory server. The AIX security subsystem's usage of IBM Security Verify Directory enables centralized security authentication, and access to user and group information. This function can be used in a host clustering environment to keep authentication, user, and group information in common. The `mksecldap` command sets up an AIX cluster that consists of one or more servers, and one or more clients that use the IBM Security Verify Directory (LDAP) for security authentication, and user and group management.

LDAP usage in AIX provides the following functions:

- Seamless Integration:
  - Automated client and server configuration.
  - Centralized AIX user, group, and network management.
  - User, group, and network export tools.
- Features:
  - RFC 2307-based implementation: Heterogeneous LDAP environment compatibility.
  - A client-side daemon manages requests and connections.
  - Fault-tolerant and priority-based server failover and reconnection.

For more information about how to set up an LDAP server and to configure clients in AIX, see the following resources:

- *Integrating AIX into Heterogeneous LDAP Environments*, SG24-7165
- LDAP in AIX
- How to Install LDAP On AIX 7.1 and Configure As LDAP Server
- Setting up an LDAP client

4.13.4 LDAP-based user and group management on AIX

LDAP for the security subsystem is implemented as the LDAP authentication load module. It is conceptually similar to the other load modules, such as NIS, DCE, and KRB5. Load modules are defined in the `/usr/lib/security/methods.cfg` file. The LDAP load module provides user authentication and centralized user and group management functions through the LDAP protocol. A user that is defined on an LDAP server can be configured to log in to an LDAP client even if that user is not defined locally.

The AIX LDAP load module is fully integrated within the AIX operating system. After the LDAP authentication load module is enabled to serve user and group information, high-level APIs, commands, and system management tools work in their usual manner. An `-R` flag is introduced for most high-level commands to work through different load modules.

AIX supports LDAP-based user and group management by integrating with IBM Security Verify Directory servers, non-IBM RFC 2307-compliant servers, and Microsoft AD. As a best practice, use IBM Security Verify Directory to define AIX users and groups. For more information about setting up the server, see Setting up an IBM Security Verify Directory Server.

AIX supports non IBM directory servers. A directory server that is RFC 2307 compliant is supported, and AIX treats these servers similarly to IBM Security Verify Directory Servers. Directory servers that are not RFC 2307 compliant can be used, but they require extra manual configuration to map the data schema. There might be some limitations due to the subset of user and group attributes in RFC 2307 compared to the AIX implementation. LDAP Version 3 protocol support is required.

AIX also supports Microsoft AD as an LDAP server for user and group management. For this setup, the UNIX supporting schema must be installed (it is included in Microsoft Service for UNIX). AIX supports AD running on Windows 2000, 2003, and 2003 R2 with specific Microsoft Service for UNIX schema versions.

Some AIX commands might not function with LDAP users if an AD server is used because of the differences in user and group management between UNIX and Windows systems. Most user and group management commands (such as `lsuser`, `chuser`, `rmuser`, `lsgroup`, `chgroup`, `rmgroup`, `id`, `groups`, `passwd`, and `chpasswd`) should work, depending on access rights.

4.13.5 Setting up IBM Security Verify Directory

IBM Security Verify Directory is a highly scalable and robust LDAP directory server. The AIX security subsystem’s usage of IBM Security Verify Directory allows for centralized security authentication, and access to user and group information. You can use this function in a host clustering environment to keep authentication, user, and group information in common. The `mksecldap` command sets up an AIX cluster that consists of one or more servers, and one or more clients that use the IBM Security Verify Directory (LDAP) for security authentication, and user and group management.

To set up the AIX security subsystem to use IBM Security Verify Directory (LDAP), complete the following steps:

1. Set up a IBM Security Verify Directory Server that serves as a centralized repository for user and group information when authenticating.
2. Set up the host systems (clients) to use the IBM Security Verify Director server for authentication and to retrieve user and group information.

For more information about installing the IBM Security Verify Directory, see LDAP on AIX: Step by step instructions for installing the LDAP client file sets on AIX.

5

IBM i security

The IBM Power family covers many users. Security on the IBM i platform is flexible enough to meet the requirements of these users and many situations. This chapter describes the features and options that are available.

This chapter describes the following topics:

5.1 Introducing IBM i security

To create a security policy and plan security measures for your system, you must understand the following security concepts, some of which are general concepts and some of which are specific to the hardware type.

A small system might have 3 - 5 users and a large system might have several thousand users. Some installations have all their workstations in a single, relatively secure area. Others have widely distributed users, which include users who connect by dialing in and indirect users that are connected through personal computers or system networks. Security on IBM i is flexible enough to meet the requirements of this wide range of users and situations.

System security has some important objectives. Each security control or mechanism should satisfy one or more of the following security goals:

- Confidentiality
- Integrity
- Availability
- Authentication
- Authorization
- Auditing or logging

Confidentiality

Confidentiality concerns include the following items:

- Protecting against disclosing information to unauthorized people
- Restricting access to confidential information
- Protecting against unauthorized system users and outsiders

Integrity

Integrity is an important aspect when applied to data within your enterprise. Integrity goals include the following items:

- Protecting against unauthorized changes to data
- Restricting manipulation of data to authorized programs
- Providing assurance that data is trustworthy

Availability

Systems are often critical to keep an enterprise running. Availability includes the following items:

- Preventing accidental changes or destruction of data
- Protecting against attempts by outsiders to abuse or destroy system resources

Authentication

Ensuring that your data is accessible only by entities that are authorized is one of the basic tenets of data security. Proper authentication methodologies are important for the following reasons:

- Determine whether users are who they claim to be. The most common technique to authenticate is by user profile name and password.
- Provide extra methods of authentication, such as using Kerberos as an authentication protocol in a single sign-on (SSO) environment.

Authorization

Once a user is authenticated, you must ensure that they access only the data and tasks that are relevant to their job. Proper authorization is important for the following reasons:

- Permit a user to access resources and perform actions on them.
- Define access permissions (public or private rights) to objects to help ensure that they are not accessed except by users that have authorization.

Auditing or logging

Auditing and logging are important in discovering and stopping access threats before the system becomes compromised.

- When your security plan is implemented, you must monitor the system for any out-of-policy security activity and resolve any discrepancies that are created by the activity.
- Depending on your organization and security policy, you might need to issue a security warning to the person who performed the out-of-policy security activity so that they know not to perform this action in the future.

System security is often associated with external threats, such as hackers or business rivals. However, protection against system accidents by authorized system users is often the greatest benefit of a well-designed security system. In a system without good security features, pressing the wrong key might result in deleting important information. System security can prevent this type of accident.

The best security system functions cannot produce good results without good planning. Security that is set up in small pieces without planning can be confusing and difficult to maintain and audit. Planning does not imply designing the security for every file, program, and device in advance. It does imply establishing an overall approach to security on the system and communicating that approach to application designers, programmers, and system users.

As you plan security on your system and decide how much security that you need, consider these questions:

- Is there a company policy or standard that requires a certain level of security?
- Do the company auditors require some level of security?
- How important is your system and the data on it to your business?
- How important is the error protection that is provided by the security features?
- What are your company security requirements for the future?

To facilitate installation, many of the security capabilities on your system are not activated when your system is shipped. There are best practices this chapter to bring your system to a reasonable level of security. Always consider the security requirements of your own installation as you evaluate any best practices.

5.2 Maintaining IBM i

One of the most important tasks in maintaining a system's security is to help ensure that all available operating system updates are installed promptly. Fixes are essential for system maintenance to help ensure optimal availability, reduced downtime, and added functions.

IBM periodically releases fixes to address issues that are discovered in IBM i programs. These fixes are bundled into cumulative Product Temporary Fix (PTF) packages, which contain fixes for specific periods. Consider installing cumulative PTF packages twice a year in dynamic environments and less frequently in stable ones. Also, apply them when making major hardware or software changes.

By prioritizing fixes, fix groups, cumulative packages, and high-impact pervasive (HIPER) fixes, you can prevent security issues that are caused by failing to implement operating system fixes to correct known issues.

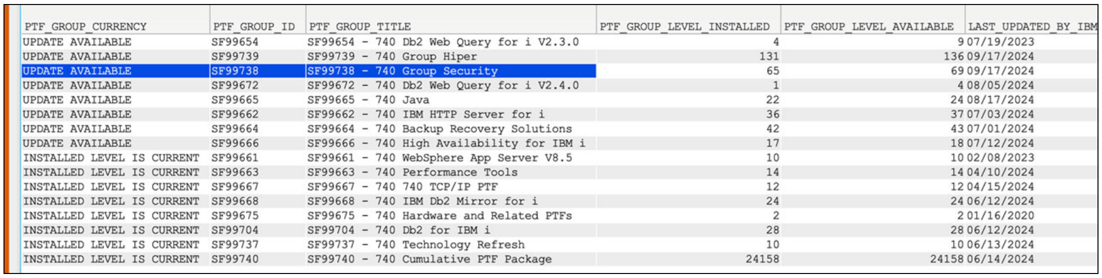
The IBM Navigator for i web-based tool contains technology for doing system management tasks across one or more systems concurrently. IBM Navigator for i provides wizards that simplify fix management. You can use the wizards to send, install, or uninstall fixes. You can also use the compare and update wizard to compare a model system to multiple target systems to find missing or extra fixes. Also, you can use tools like IBM Administration Runtime Expert for i and Ansible to compare and automate this process.

Another option for managing fixes is to use an SQL query to identify any issues, as documented in this IBM document. The query is shown in Example 5-1.

Example 5-1 Group PTF currency query

```
--
-- Derive the IBM i operating system level and then
-- determine the level of currency of PTF Groups
--
With iLevel(iVersion, iRelease) AS
(
select OS_VERSION, OS_RELEASE from sysibmadm.env_sys_info
)
  SELECT P.*
     FROM iLevel, systools.group_ptf_currency P
     WHERE ptf_group_release =
           'R' CONCAT iVersion CONCAT iRelease concat '0'
     ORDER BY ptf_group_level_available -
        ptf_group_level_installed DESC;
```

An example output is shown in Figure 5-1.



PTF_GROUP_CURRENCY	PTF_GROUP_ID	PTF_GROUP_TITLE	PTF_GROUP_LEVEL_INSTALLED	PTF_GROUP_LEVEL_AVAILABLE	LAST_UPDATED_BY_IBM
UPDATE AVAILABLE	SF99654	SF99654 - 740 Db2 Web Query for i V2.3.0	4	9	07/19/2023
UPDATE AVAILABLE	SF99739	SF99739 - 740 Group Hiper	131	136	09/17/2024
UPDATE AVAILABLE	SF99738	SF99738 - 740 Group Security	65	69	09/17/2024
UPDATE AVAILABLE	SF99672	SF99672 - 740 Db2 Web Query for i V2.4.0	1	4	08/05/2024
UPDATE AVAILABLE	SF99665	SF99665 - 740 Java	22	24	08/17/2024
UPDATE AVAILABLE	SF99662	SF99662 - 740 IBM HTTP Server for i	36	37	07/03/2024
UPDATE AVAILABLE	SF99664	SF99664 - 740 Backup Recovery Solutions	42	43	07/01/2024
UPDATE AVAILABLE	SF99666	SF99666 - 740 High Availability for IBM i	17	18	07/12/2024
INSTALLED LEVEL IS CURRENT	SF99661	SF99661 - 740 WebSphere App Server V8.5	10	10	02/08/2023
INSTALLED LEVEL IS CURRENT	SF99663	SF99663 - 740 Performance Tools	14	14	04/10/2024
INSTALLED LEVEL IS CURRENT	SF99667	SF99667 - 740 740 TCP/IP PTF	12	12	04/15/2024
INSTALLED LEVEL IS CURRENT	SF99668	SF99668 - 740 IBM Db2 Mirror for i	24	24	06/12/2024
INSTALLED LEVEL IS CURRENT	SF99675	SF99675 - 740 Hardware and Related PTFs	2	2	01/16/2020
INSTALLED LEVEL IS CURRENT	SF99704	SF99704 - 740 Db2 for IBM i	28	28	06/12/2024
INSTALLED LEVEL IS CURRENT	SF99737	SF99737 - 740 Technology Refresh	10	10	06/13/2024
INSTALLED LEVEL IS CURRENT	SF99740	SF99740 - 740 Cumulative PTF Package	24158	24158	06/14/2024

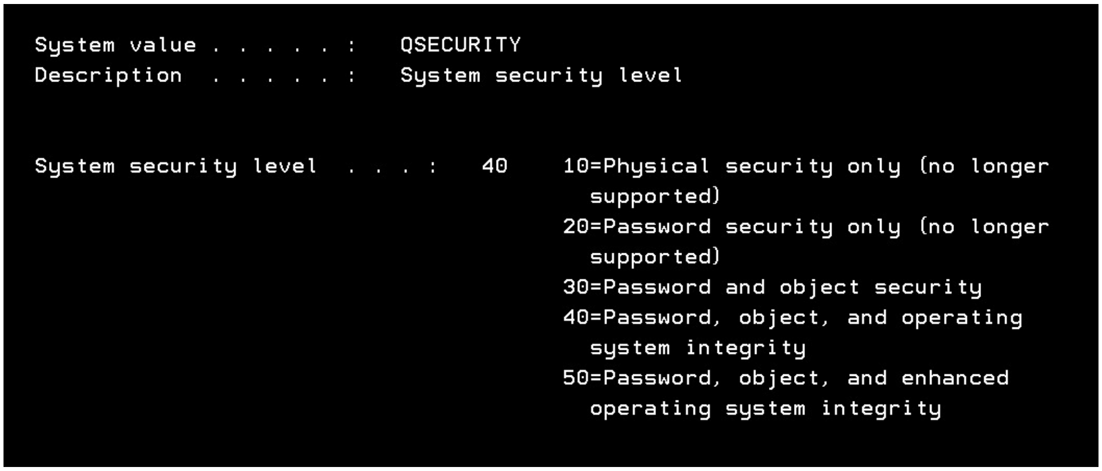
Figure 5-1 PTF currency query results.

For more information about maintaining IBM i, see Using fixes.

5.3 Security levels

Security on IBM I systems is a series of levels with each level offering a greater degree of security and protection of your data than the previous level. You can choose how much security that you want the system to enforce by setting the security level (`QSECURITY`) system value. IBM i supports these fully integrated system security levels. IBM i platform offers five levels of security.

Figure 5-2 shows the QSECURITY panel where the level can be set.



```
System value . . . . . :   QSECURITY
Description  . . . . . :   System security level


System security level  . . . :   40    10=Physical security only (no longer
                                         supported)
                                       20=Password security only (no longer
                                         supported)
                                       30=Password and object security
                                       40=Password, object, and operating
                                         system integrity
                                       50=Password, object, and enhanced
                                         operating system integrity
```

Figure 5-2 The QSECURITY system value and the various security levels on IBM i

Level 10: Password security (changing to level 10 is no longer supported)

At security level 10, you have no security protection. It is not supported. Running at this security level is both a security and integrity risk.

Level 20: Password security (changing to level 20 is no longer supported)

At security level 20, users require both a unique user ID (UID) and a password that are created by the system administrator to access the system. A significant issue with this level is that it grants users `*ALLOBJ` special authority, allowing them unrestricted access to all system data, files, and objects. There is no way to restrict individual user privileges, which is a severe security risk. Because of the potential for unauthorized access and data breaches, running at security level 20 is no longer supported. Higher security levels (40 and 50) offer enhanced protection mechanisms that can reduce risks.

Level 30: Password and resource security

Level 30 provides more security functions in addition to what is provided at level 20. Users must have specific authority to use resources on the system. Users do not have automatic access to everything on the system, and the system administrator must define a valid UID and password for a user. User access is limited by the security policies of the business. Level 30 is *not* considered a secure level because the integrity protection features that are available on security level 40 and 50 are not activated at security level 30. Running at this security level is both a security and integrity risk because you do not have the protection of the higher security levels activated.

Level 40: Integrity protection

At this security level, resource security and integrity protection are enforced, and the system itself is protected against users. Integrity protection functions, such as the validation of parameters for interfaces to the operating system, help protect your system and the objects on it from tampering by experienced system users. For example, user-written programs cannot directly access the internal control blocks through pointer manipulation. Level 40 is the default security level for every new installation and is the recommended security level for most installations.

Level 50: Advanced integrity protection

At this security level, advanced integrity protection is added to the resource security and level 40 integrity protection enforcement. Advanced integrity protection includes further restrictions, such as the restriction of message-handling between system state programs and user state programs. The system is protected against user-written programs, and users have access only to data on the system, rather than information about the system itself. This level offers greater security against anyone attempting to learn about your system. Level 50 is the recommended level of security for most businesses because it offers the highest level of security possible at the time of writing. Security level 50 is intended for IBM i platforms with high security requirements, and it meets Common Criteria (CC) security requirements.

5.4 System values

System values are part of the global settings of your system. They customize many characteristics of your system. The security system values are used to control the security settings on your system and are broken into four groups:

- System values that control passwords
- System values that control auditing
- General security system values
- Other system values that are related to security

System values also provide customization of many characteristics of your IBM i platform. You can use system values to define system-wide security settings. To access the jobs category of system values from IBM Navigator for i, select **Configuration and Services** and then select **System Values**, as shown in Figure 5-3 on page 139.

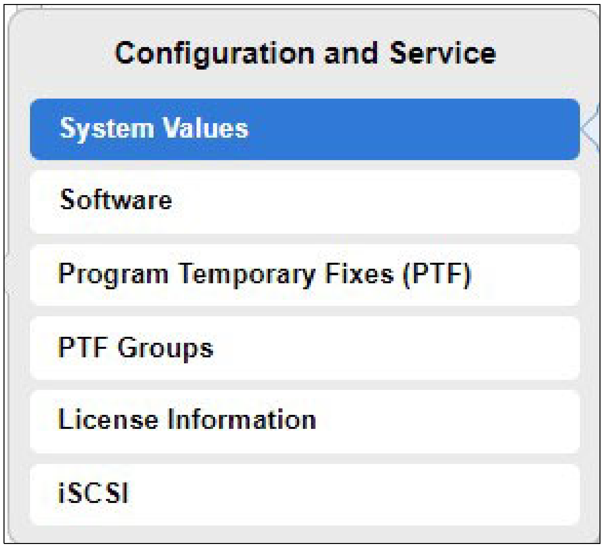


Figure 5-3 System Values option under the Configuration and Service menu of IBM Navigator for i

For example, you can specify the following settings:

- How many sign-on attempts you allow at a device.
- Whether the system automatically signs off an inactive workstation.
- How often passwords must be changed.
- The length and composition of passwords.

You can restrict users from changing the security-related system values. The Change SST Security Attributes (`CHGSSTSECA`) command, System Service Tools (SST), and Dedicated Service Tools (DST) are options to lock these system values. By locking the system values, you can prevent even a user with `*SECADM` and `*ALLOBJ` authority from changing these system values with the `CHGSYSVAL` command. In addition to restricting changes to these system values, you can also restrict adding digital certificates to a digital certificate store with the Add Verifier application programming interface (API) and restrict password resetting on the digital certificate store.

To see a list of all the security-related system values, got IBM Navigator for I and select **Security** → **Security Config. info**. The values are typically related to your security environment requirement and might differ slightly for every organization.

5.5 Authentication

Authentication is the set of methods that are used by organizations to help ensure that only authorized personnel, services, and applications with the correct permissions can access company resources. There are bad actors who want to gain access to your systems with ill intentions, thus making authentication a critical part of cybersecurity. These bad actors try to steal credentials from users who have access to your environment.

Your authentication process should primarily include these three steps:

1. Identification: Ensure that the user that requests access is who they claim to be, usually through a username or other type of login ID.
2. Authentication: Users usually provide a password (a random word or phrase or sequence of characters that the user is the only one who is supposed to know) to prove that they are who they claim to be, but if you want to strengthen security, organizations may also require the user to provide something they have (a phone or token device) to further prove their identity, or a unique characteristic that is part of their person (a face or fingerprint scan).
3. Authorization: The system verifies that the user is indeed who they claim to be and allows them access to the system or application that they are trying to gain access to.

Authentication helps your organization protect your applications, systems, data, websites, and networks from internal and external attacks. It also aids in keeping an individual's personal data confidential so that they can conduct their everyday business online with less risk. When authentication systems are weak, attackers can compromise a user account either by guessing a password or tricking a person into handing over their credentials. This situation might lead to any of the following risks:

- Exfiltration or a data breach
- Installation of various types of malware (in enterprise environments, the most prevalent of them is ransomware)
- Non-compliance with different data privacy regulations

Because compromising a user's access to a system is a common method for attackers to obtain unauthorized access to an organization's resources, it is of utmost importance that strong authentication security is enforced.

5.5.1 Single sign-on enablement

SSO is an authentication process where a user can access more than one system by entering a single UID and password. In today's heterogeneous networks with partitioned systems and multiple platforms, administrators must cope with the complexities of managing identification and authentication for network users.

To enable an SSO environment, IBM provides two technologies that work together to enable users to sign in with their Windows username and password and authenticate to IBM i platforms in the network. Network Authentication Services and Enterprise Identity Mapping (EIM) are the two technologies that an administrator must configure to enable an SSO environment. Windows operating systems, AIX, and z/OS use the Kerberos protocol to authenticate users to the network. A secure, centralized system that is called a key distribution center authenticates principals (Kerberos users) to the network.

Network Authentication Services allows an IBM i platform to participate in the Kerberos realm, and EIM provides a mechanism for associating these Kerberos principals to a single EIM identifier that represents that user within the entire enterprise. Other UIDs, such as an IBM i username, can also be associated with this EIM identifier. When a user signs on to the network and accesses an IBM i platform, that user is not prompted for a UID and password. If the Kerberos authentication is successful, applications can look up the association to the EIM identifier to find the IBM i username. The user no longer needs a password to sign on to the IBM i platform because the user is already authenticated through the Kerberos protocol. Administrators can centrally manage UIDs with EIM, and network users need only to manage one password. You can enable SSO by configuring Network Authentication Services and EIM on your system.

Note: For more information about SSO for IBM i 7.5, see IBM i 7.5: Security Single sign-on.

5.5.2 User profiles

On the IBM i operating system, every system user has a user profile. Create a user profile before a user can sign on.

The user profile is a powerful and flexible tool. It controls what the user can do and customizes the way that the system appears to the user. The following sections describe some of the important security features of the user profile.

Special authority

Special authorities determine whether the user may perform system functions, such as creating user profiles or changing the jobs of other users. The special authorities that are available are shown in Table 5-1.

Table 5-1 Security-specific special authorities

Special Authority	Description
*ALLOBJ	All-object (*ALLOBJ) special authority grants access to any resource on the system, even if private authority exists for the user.
*SECADM	Security administrator (*SECADM) special authority allows a user to create, change, and delete user profiles.
*JOBCTL	The Job control (*JOBCTL) special authority allows a user to change the priority of jobs and of printing, end a job before it has finished, or delete output before it has printed. *JOBCTL special authority can also give a user access to confidential spooled output if output queues are specified a as OPRCTL(*YES).
*SPLCTL	Spool control (*SPLCTL) special authority allows the user to perform all spool control functions, such as changing, deleting, displaying, holding, and releasing spooled files.
*SAVSYS	Save system (*SAVSYS) special authority gives the user the authority to save, restore, and free storage for all objects on the system, regardless of whether the user has object existence authority to the objects.
*SERVICE	Service (*SERVICE) special authority allows the user to start SST by using the Start SST (STRSST) command. This special authority allows the user to debug a program with only *USE authority to the program and perform the display and alter service functions. It also allows the user to perform trace functions.
*AUDIT	Audit (*AUDIT) special authority gives the user the ability to view and change auditing characteristics.
*IOSYSCFG	System configuration (*IOSYSCFG) special authority gives the user the ability to change how the system is configured. Users with this special authority can add or remove communications configuration information, work with TCP/IP servers, and configure the internet connection server (ICS). Most commands for configuring communications require *IOSYSCFG special authority.

Initial menu and initial program

The initial menu and initial program determine what the user sees after signing on the system. You can limit a user to a specific set of tasks by restricting the user to an initial menu.

Limit capabilities

The Limit capabilities field in the user profile determines whether the user can enter commands and change the initial menu or initial program when signing on. The Limit capabilities field in the user profile and the `ALWLMTUSR` parameter on commands apply only to commands that are run from the CLI, the Command Entry display, FTP, REXEC, the QCAPCMD API, or an option from a command grouping menu. Users are not restricted to performing the following actions:

- Run commands in control language (CL) programs that are running a command as a result of taking an option from a menu.
- Run remote commands through applications.

5.5.3 Signing for objects

You can reinforce integrity by signing software objects that you use.

A key component of security is integrity: You must trust that objects on the system were not tampered with or altered. Your IBM i software is protected by digital signatures.

Signing your software object is important if the object has been transmitted across the internet or stored on media that you feel might have been modified. The digital signature can be used to detect whether the object has been altered.

Digital signatures and their usage for verification of software integrity can be managed according to your security policies by using the Verify Object Restore (`QVFYOBJRST`) system value, the Check Object Integrity (`CHKOBJITG`) command, and the Digital Certificate Manager (DCM) tool. Also, you can choose to sign your own programs (all licensed programs that are included with the system are signed).

You can restrict adding digital certificates to a digital certificate store by using the Add Verifier API and restrict resetting passwords on the digital certificate store. SST provides a new menu option that is called “Work with system security” where you can restrict adding digital certificates.

5.5.4 Group profiles

A *group profile* is a special type of user profile. Rather than granting authority to each user individually, you can use a group profile to define authority for a group of users.

A group profile can own objects on the system. You can also use a group profile as a pattern when creating individual user profiles by using the copy profile function.

5.6 Transport Layer Security

Transport Layer Security (TLS), originally know as Secure Socket Layer or Secure Sockets Layer (SSL), uses certificates to establish an encrypted link between a server and a client. With this link, sensitive information like credit card details can be transmitted securely over the internet. The certificate contains a public key that authenticates the identity of a website and allows for encrypted data transfer through asymmetric, or public-key cryptography. The matching private key is kept secret on the server.

There are two types of encryption keys that are used in SSL/TLS:

- Asymmetric keys: The public and private key pair are used to identify the server and initiate the encrypted session. The private key is known only to the server, and the public key is shared through a certificate.
- Symmetric session keys: Disposable keys are generated for each connection and used to encrypt and decrypt transmitted data. The symmetric keys are securely exchanged by using asymmetric encryption.

SSL and TLS support multiple symmetric ciphers and asymmetric public key algorithms. For example, AES with 128-bit keys is a common symmetric cipher, and RSA and Elliptic Curve Cryptography commonly use asymmetric algorithms.

5.6.1 Secure Sockets Layer and Transport Layer Security on IBM i

This section explores the SSL/TLS implementations that are available on IBM i. These protocols provide secure communication channels for data transmission among applications.

Overview

The IBM i system offers multiple SSL/TLS implementations, each adhering to industry-defined protocols and specifications that are set by the Internet Engineering Task Force (IETF). These implementations cater to different application needs and offer varying functions. The specific implementation that is used by an application depends on the chosen API set.

For Java applications, the configured Java Secure Socket Extension (JSSE) provider determines the implementation because Java interfaces are standardized. Alternatively, an application can embed its own implementation for exclusive usage.

Available implementations

Here is a breakdown of the available SSL/TLS implementations on IBM i:

- System SSL/TLS:
  - Primarily used by ILE applications.
  - Certificate management is handled by the DCM. The certificates are stored in Certificate Management Services (CMS) format (.KDB files).
  - Although Java applications can use system SSL/TLS it is not the typical approach. Even rarer is a Java application that concurrently uses both System SSL/TLS and a Java Keystore.

- IBMJSSE2 (IBMJSSEProvider2):
  - A pure Java implementation of SSL/TLS protocols is available on various platforms. It is known as `com.ibm.jsse2.IBMJSSEProvider2` in the `java.security` provider list.
  - It is the default JSSE provider for all Java SDK versions on IBM i, which makes it the most commonly used option for Java applications.
  - Certificates typically are in Java Keystore files (.JKS) and managed through a Java keytool or IBM Key Management (iKeyman).
  - For more information about JSSE, see Java Secure Socket Extension (JSSE).
  - Specific details for IBMJSSE2 can be found in the platform-independent documentation for your corresponding Java SDK version. For Java SDK 8, see Security Reference for IBM SDK, Java Technology Edition, Version 8.
- OpenSSL:
  - An open-source toolkit that offers an SSL/TLS protocol implementation and a comprehensive cryptography library.
  - It has limited availability, and is only accessible within the IBM Portable Application Solutions Environment for i (PASE for i).
  - Certificates are typically stored in PEM files and managed by using OpenSSL commands.
  - It is primarily used by applications like the Common Information Model Object Manager (CIMOM).
  - For more information, see the Common Information Model.

System SSL/TLS

System SSL/TLS is a set of generic services that are provided in the IBM i Licensed Internal Code (LIC) to protect TCP/IP communications by using the SSL/TLS protocol. System SSL/TLS is tightly coupled with the operating system, and the LIC sockets code specifically provides extra performance and security.

System TLS has the infrastructure to support multiple protocols. The following protocols are supported by System TLS:

- Transport Layer Security 1.3 (TLSv1.3)
- Transport Layer Security 1.2 (TLSv1.2)
- Transport Layer Security 1.1 (TLSv1.1)
- Transport Layer Security 1.0 (TLSv1.0)
- Secure Sockets Layer 3.0 (SSLv3)

System TLS also supports the following cipher suites:

- AES_128_GCM_SHA256
- AES_256_GCM_SHA384
- CHACHA20_POLY1305_SHA256
- ECDHE_ECDSA_AES_128_GCM_SHA256
- ECDHE_ECDSA_AES_256_GCM_SHA384
- ECDHE_RSA_AES_128_GCM_SHA256
- ECDHE_RSA_AES_256_GCM_SHA384
- ECDHE_ECDSA_CHACHA20_POLY1305_SHA256
- ECDHE_RSA_CHACHA20_POLY1305_SHA256

System values for setting protocols and cipher suites

The QSSLPCL system value setting identifies the specific protocols that are enabled on the system. Applications can negotiate secure sessions with only protocols that are listed in QSSLPCL. For example, to restrict the System TLS implementation to use only TLSv1.3 and not allow any older protocol versions, set QSSLPCL to contain only *TLSV1.3.

The QSSLPCL special value *OPSYS allows the operating system to change the protocols that are enabled on the system. The value of QSSLPCL remains the same when the system upgrades to a newer operating system release. If the value of QSSLPCL is not *OPSYS, then the administrator must manually add newer protocol versions to QSSLPCL after the system moves to a new release.

For more information about System SSL/TLS support for protocols and cipher suites, see System SSL/TLS.

> **Important:** It is a best practice to always run your IBM i server with the following network protocols *disabled*. Using configuration options that are provided by IBM to enable weak protocols results in your IBM i server being configured to allow usage of weak protocols. This configuration results in your IBM i server potentially being at risk of a network security breach.
>
> - TLSv1.1
> - TLSv1.0
> - SSLv3
> - SSLv2

The QSSLCSL system value setting identifies the specific cipher suites that are enabled on the system. Applications can negotiate secure sessions with only a cipher suite that is listed in QSSLCSL. No matter what an application does with code or configuration, it cannot negotiate secure sessions with a cipher suite if it is not listed in QSSLCSL. Individual application configuration determines which of the enabled cipher suites are used for that application.

To restrict the System TLS implementation from using a particular cipher suite, complete the following steps:

1. Change the QSSLCSLCTL system value to the special value *USRDFN so that you can edit the QSSLCSL system value.
2. Remove all cipher suites what you want to restrict from the list in QSSLCSL.

The QSSLCSLCTL system value special value *OPSYS allows the operating system to change the cipher suites that are enabled on the system. The value of QSSLCSLCTL remains the same when the system upgrades to a newer operating system release. If the value of QSSLCSLCTL is *USRDFN, then the administrator must manually add newer cipher suites to QSSLCSL after the system moves to a new release. Setting QSSLCSLCTL to *OPSYS also adds the new values to QSSLCSL.

A cipher suite cannot be added to QSSLCSL if the TLS protocol that is required by the cipher suite is not set in QSSLPCL.

5.7 Service tools

Service tools are used to configure, manage, and service all models of IBM i.

Service tools can be accessed from DSTs or SST. Service tools UIDs are required if you want to access DST or SST, and to use the IBM Navigator for i functions for disk unit management.

Service tools UIDs are referred to as DST user profiles, DST UIDs, service tools user profiles, or a variation of these names. Within this topic collection, the term “service tools user IDs” is used.

> **Note:** For more information about Service Tools for IBM i 7.5, see IBM i 7.5: Security Service Tools.

5.7.1 System Service Tools

If your user profile has the required authorizations, you can use SST to access service tools.

The service tools UID that you use to access SST must have the functional privilege to use SST. The IBM i user profile must have the following authorizations:

- Authorization to use the `STRSST` CL command.
- Service special authority (*SERVICE).

To access service tools by using SST, complete the following steps:

1. Enter `STRSST` on an IBM i CLI. The Start SST Sign On display opens.
2. Enter the following information:
   - Service Tools User ID: The service tools UID that you sign on with.
   - Password: The password that is associated with this UID.
3. Press Enter. The SST display opens.

To exit from SST, press F3 (`Exit`) until you get to the Exit SST display, and then press Enter to exit SST.

5.7.2 Dedicated Service Tool

To access service tools, you can use the *DST* from the system console.

The service tools UID that you use to access service tools with DST must have the functional privilege to use DST. You can start the DST by using function 21 from the system control panel or by using a manual initial program load (IPL).

Accessing service tools by using the DST from the system control panel

To access service tools by using the DST from the control panel, complete the following steps:

1. Put the control panel into manual mode.
2. Use the control panel to select function 21 and press Enter. The DST Sign On display appears on the console.

3. Sign on to the DST by using your service tools UID and password. The Use Dedicated Service Tools (DST) display appears.
4. Select the appropriate option from the following list and press Enter:
   - Select option 5 (Work with DST environment) to get to more options for working with service tools UIDs.
   - Select option 7 (Start a service tool) to start any of the service tools that are available from the DST.
   - Select any other option.

Accessing service tools by using the DST from a manual IPL

To access service tools by using the DST from a manual IPL, complete the following steps:

1. Put the control panel in manual mode.
2. Take either of the following actions:
   - If the system is powered off, turn on the system.
   - If the system is powered on, enter the Power Down System (`PWRDWNSYS`) command, `PWRDWNSYS *IMMED RESTART(*YES)`, on a CLI to turn off the system and restart it.
3. Sign on to the DST by using your service tools UID and password. The **IPL or Install the System** menu is shown.
4. At the **IPL or Install the System** menu:
   a. Select option 3 (`Use Dedicated Service Tools (DST)`).
   b. Select the appropriate option from the list and press Enter.
   c. Select option 5 (`Work with DST environment`) to get more options for working with service tools UIDs.
   d. Select option 7 (`Start a service tool`) to start any of the service tools that are available from the DST.
   e. Select any of the other options as needed.

Exiting the DST

To exit the DST, press F3 (`Exit`) until you return to the Exit Dedicated Service Tools (DST) display, and then select option 1 (`Exit Dedicated Service Tools (DST)`).

5.8 Digital Certificate Manager

With *DCM*, you can manage digital certificates for your network and use TLS to enable secure communications for many applications.

A digital certificate is an electronic credential that you can use to establish proof of identity in an electronic transaction. There are an increasing number of uses for digital certificates to provide enhanced network security measures. For example, digital certificates are essential to configuring and to using the TLS. Using TLS enables you to create secure connections between users and server applications across an untrusted network, such as the internet. TLS provides one of the best solutions for protecting the privacy of sensitive data, such as usernames and passwords, over the internet. Many IBM i platforms and applications, such as FTP, Telnet, and HTTP Server, provide TLS support to help ensure data privacy.

IBM i provides extensive digital certificate support so that you can use digital certificates as credentials in many security applications. In addition to using certificates to configure TLS, you can use them as credentials for client authentication in both TLS and virtual private network (VPN) transactions. Also, you can use digital certificates and their associated security keys to sign objects. Signing objects enable you to detect changes or possible tampering to object contents by verifying signatures on the objects to help ensure their integrity.

Capitalizing on the IBM i support for certificates is simple when you use DCM to centrally manage certificates for your applications. DCM enables you to manage certificates that you obtain from any certificate authority (CA). Also, you can use DCM to create and operate your own local CA to issue private certificates to applications and users in your organization.

Planning and evaluation are the keys to using certificates effectively for their added security benefits.

5.9 Encryption and cryptography

Cryptography is the study and implementation of processes that manipulate data for hiding and authenticating information. A comprehensive cryptography solution is an important part of a successful security strategy. The usage of cryptography for encryption of data as it is processed within the IBM i partition provides enhanced security for memory and stored data as part of the IBM Power10 infrastructure.

As described in 2.1, “Encryption technologies and their applications” on page 30, Power10 provides Transparent Memory Encryption (TME), which transparently encrypts and protects memory within the system by using the encryption acceleration processors that are built in to the Power10 processing chip, which provides protection without performance penalties.

IBM i offers various levels of encryption for databases and attached storage devices. By using Field Procedures within IBM Db2, IBM i provides field-level encryption to directly protect sensitive data fields within the database. Also, IBM i supports encryption for directly attached storage devices to safeguard data at rest within the system.

IBM i includes both software cryptography and a range of cryptographic hardware options for data protection and secure transaction processing. Users can leverage the built-in encryption acceleration processors on the Power10 chip or integrate specialized cryptographic coprocessors. Both options provide robust security without compromising performance.

IBM i cryptographic services help ensure data privacy, maintain data integrity, authenticate communicating parties, and prevent repudiation when a party denies sending a message.

Cryptographic Services Key Management

Use *Cryptographic Services Key Management* for the IBM i to store and manage master keys and keystores. Because you exchange sensitive data to manage master keys and keystores, it is a best practice to use a secure session.

Cryptographic services support a hierarchical key system. At the top of the hierarchy is a set of master keys. These keys are the only key values that are stored in the clear (unencrypted). Cryptographic services securely store the master keys within the IBM i LIC.

Eight general-purpose master keys are used to encrypt other keys, which can be stored in keystore files. Keystore files are database files. Any type of key that is supported by cryptographic services can be stored in a keystore file, for example, AES, RC2, RSA, or SHA1-HMAC.

In addition to the eight general-purpose master keys, cryptographic services supports two special-purpose master keys:

- The auxiliary storage pool (ASP) master key is used for protecting data in the independent auxiliary storage pool (IASP) (in the Disk Management GUI, the IASP is known as an Independent Disk Pool).
- The save/restore master key is used to encrypt the other master keys when they are saved to media by using a Save System (`SAVSYS`) operation.

You can work with Cryptographic Services Key Management by using the IBM Navigator for i interface, as shown in Figure 5-4.

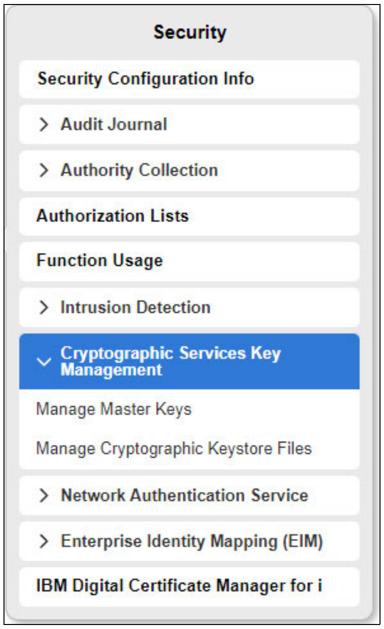

Security
Security Configuration Info
Audit Journal
Authority Collection
Authorization Lists
Function Usage
Intrusion Detection
Cryptographic Services Key Management
Manage Master Keys
Manage Cryptographic Keystore Files
Network Authentication Service
Enterprise Identity Mapping (EIM)
IBM Digital Certificate Manager for i

Figure 5-4 Security menu within IBM Navigator for i

After you connect to IBM Navigator for i, select **Security** → **Cryptographic Services Key Management**. Then, you can manage master keys and cryptographic keystore files.

You can also use the cryptographic services APIs or CL commands to work with the master keys and keystore files.

Note: Use TLS to reduce the risk of exposing key values while performing key management functions.

4769 Cryptographic Coprocessor

IBM offers Cryptographic Coprocessors, which are available on various system models. Cryptographic Coprocessors contain hardware engines, which perform cryptographic operations that are used by IBM i application programs and IBM i TLS transactions. The 4769 Cryptographic Coprocessor (requires IBM i 7.3 or later) appears as hardware feature #EJ35 or #EJ37 on Power10 systems. For more information about IBM Cryptographic cards, see “IBM PCIe Cryptographic Coprocessor cards” on page 34.

Note: The IBM 4767 Cryptographic Coprocessor is no longer available but is still supported. For more information, see IBM i 7.5: Security Cryptography.

5.10 Resource security

The ability to access an object is called *authority*. Resource security on IBM i enables you to control object authorities by defining who can use which objects and how those objects can be used.

You can specify detailed authorities, such as adding records or changing records, or you can use the system-defined subsets of authorities: `*ALL`, `*CHANGE`, `*USE`, and `*EXCLUDE`.

Files, programs, and libraries are the most common objects that require security protection, but you can specify authority for any object on the system. The following list describes the features of resource security:

- Group profiles

  A group of similar users can share the authority to use objects. For more information, see 5.5.4, "Group profiles" on page 142.

- Authorization lists

  Objects with similar security needs can be grouped into one list. Authority can be granted to the list rather than to the individual objects.

- Object ownership

  Every object on the system has an owner. Objects can be owned by an individual user profile or by a group profile. The correct assignment of object ownership helps you manage applications and delegate responsibility for the security of your information.

- Primary group

  You can specify a primary group for an object. The primary group's authority is stored with the object. Using primary groups might simplify your authority management and improve your authority checking performance.

- Library authority

  You can put files and programs that have similar protection requirements into a library and restrict access to that library, which is often simpler than restricting access to each individual object.

- Directory authority

  You can use directory authority the same way that you use library authority. You can group objects in a directory and secure the directory rather than the individual objects.

- Object authority

  In cases where restricting access to a library or directory is not specific enough, you can restrict authority access to individual objects.

- Public authority

  For each object, you can define what access is available for any system user who does not have any other authority to the object. Public authority is an effective means for securing information and provides good performance.

- Adopted authority

  Adopted authority adds the authority of a program owner to the authority of the user running the program. Adopted authority is a useful tool when a user needs different authority for an object, depending on the situation.

- Authority holder

  An authority holder stores the authority information for a program-described database file. The authority information remains, even when the file is deleted. Authority holders are commonly used when converting from the System/36 because System/36 applications often delete files and create them again.

- Field-level authority

  Field-level authorities are given to individual fields in a database file. You can use SQL statements to manage this authority.

5.11 Security audit journal

You can use security audit journals to audit the effectiveness of security on your system.

The IBM i can log selected security-related events in a security audit journal. Several system values, user profile values, and object values control which events are logged.

The security audit journal is the primary source of auditing information about the system. This section describes how to plan, set up, and manage security auditing, what information is recorded, and how to view that information.

A security auditor inside or outside your organization can use the auditing function that is provided by the system to gather information about security-related events that occur on the system.

You can define auditing on your system at three different levels:

- System-wide auditing that occurs for all users.
- Auditing that occurs for specific objects.
- Auditing that occurs for specific users.

When a security-related event that might be audited occurs, the system checks whether you selected that event for audit. If you have, the system writes a journal entry in the current receiver for the security auditing journal (`QAUDJRN` in library `QSYS`).

When you want to analyze the audit information that you collected in the journal, use IBM Navigator for i or use these SQL commands.

5.12 Independent disk pool

An independent disk pool, also known as an IASP, can group storage that then can be taken offline or brought online independently of system data or other unrelated data. The terms IASP and independent disk pool are synonymous.

Auxiliary storage is the permanent disk space that you assign to the system in the form of disk units, which are either physical or virtual. The disk pool can contain objects, libraries, directories, and many other objects, such as object attributes. The concept of IASP also forms a base for high availability and disaster recovery (HADR) solutions like PowerHA.

The concept of IASP is straightforward, and there are many solutions that are built around it. IASPs provide an attractive solution for clients who are looking at server consolidation and continuous availability with a minimum amount of downtime. Using the IASP provides both technical and business advantages on IBM i.

The key difference between the system ASP and an IASP is that the system ASP is always accessible when a system is running, and IASP can be brought online or offline independently of the system activity on any other pools.

An IASP must be brought online or “varied on” to make it visible to the system before attempting access data on it. If you want to make the IASP inaccessible to the system, “vary off” the IASP. The varyon process is not instantaneous and can take several minutes. The amount of time that is required depends on several factors.

Figure 5-5 shows a system with its SYSBAS or ASP and an IASP that is defined.

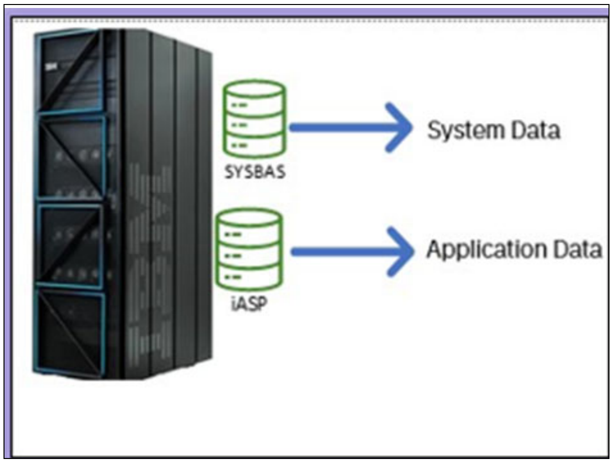
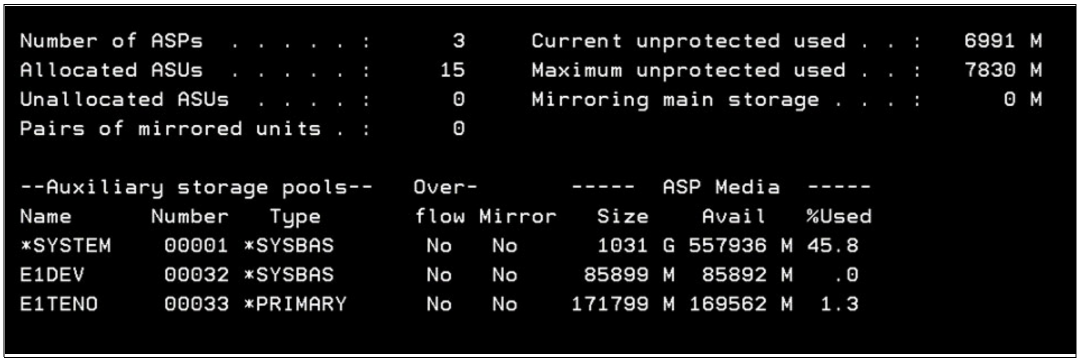


Figure 5-5 Single system with an IASP where application data is

IASPs are numbered 33 - 255. Basic ASPs are numbered 2 - 32. All basic ASPs are automatically available when the system is online and cannot be independently varied on or off.

Figure 5-6 on page 153 shows a system with a system pool, a user ASP, and an IASP that is defined.



```
Number of ASPs  . . . . . :       3     Current unprotected used . . :   6991 M
Allocated ASUs  . . . . . :      15     Maximum unprotected used . . :   7830 M
Unallocated ASUs  . . . . :       0     Mirroring main storage . . . :      0 M
Pairs of mirrored units . :       0

--Auxiliary storage pools--   Over-       -----  ASP Media  -----
Name      Number   Type       flow Mirror   Size    Avail   %Used
*SYSTEM    00001 *SYSBAS       No   No      1031 G 557936 M 45.8
E1DEV      00032 *SYSBAS       No   No     85899 M  85892 M   .0
E1TENO     00033 *PRIMARY      No   No    171799 M 169562 M  1.3
```

Figure 5-6 System with a user ASP and IASP

An IASP can be deployed by using external storage (Storage Area Network (SAN) storage) and defined on internal disks. However, there are many benefits of using a SAN instead of internal disks. SAN storage that is combined with an IASP enables cluster configurations where the replication technologies that are available on the SAN storage can provide HADR options. Many clients across many different sectors, such as manufacturing, insurance, aviation, banking, and retail, run with this configuration.

An independent disk pool can be either one of the following pools:

Switchable An IASP that is used across two or more IBM i partitions in a clustered environment. When the switch occurs and the independent disk pool is “varied on”, all the contents on the IASP are available.

Non-Switchable An IASP that is created locally to the system and not shared with any other system.

When considering IASP implementation, business needs should be considered first and a plan should be made to implement them in the client environment. At the application level, you should have a good understanding about where objects are, who are the users, and how the program and data is accessed. There are certain types of objects that are supported in IASP but should remain in the system ASP only to maintain the expected or normal behavior of the system. Some work management-related changes must be made when you introduce an IASP. In general there are two environments in which IASP can be used:

- Single system environment

  In this case, you have an independent disk pool on the local system. It can be brought online or offline without impacting other storage pools on the system or doing an IPL. This environment is often used in a local system that contains multiple databases on IASPs. The IASP can be made available while the system is active without performing an IPL, and the independent disk pool may remain offline until it is needed. This setup is common if you want to separate application data; keep historical and archived data on the same system; maintain multiple application versions; or meet data compliance rules where you must have data in different pools and keep it offline unless it is needed by the business.

- Multi-system environment

  In this case, you have one or more IASPs that are shared between multiple IBM i partitions (on the same system or on different systems, possibly even in another locations) that are members of the cluster. In this setup, the IASP can be switched between these systems without an IPL for any of the partitions. This environment is an advantage because it allows continuous availability of the data. There can be various reasons for implementing IASPs in multi-system environments. For example, if you are implementing a new HADR solution, then you normally choose a switchable IASP setup for the most flexible implementation.

Some practical implementations of Independent Disk pool are as follows.

- HADR scenarios

  When planning a HADR solution, one of the key prerequisites for PowerHA, which provides geographical mirroring, IBM HyperSwap®, and metro distance or global replication, is that you must use an IASP. With an IASP, you can have an active partition at the target site for minimal interruption while moving workloads between systems. PowerHA is an integrated extension of the IBM i and offers environmental, application, and data resiliency solutions for managing data access and replicated storage for planned or unplanned outages. This solution provides near continuous availability of data. The time to recover user access to their applications depends on many factors in your environment, which include the distance between your data centers and the bandwidth that is available between them. By using a tool like PowerHA, you do not need to use a software replication tool.

  You can use IASP with IBM Db2 Mirror for i to implement continuous availability when you use an IASP for the Integrated File System (IFS). There is minimal maintenance that is required when this solution is implemented. The data can remain available to an application even if there is a system outage. Based on your implementation choices, the switchover is automatic.

- Data isolation or protection

  You can use an IASP when you want to divide data between multiple databases. You can use the IASP to store data that is accessed only occasionally because of business requirements. Using an IASP this way means that the data is brought online only when needed, which provides more protection in terms of data access. The data remains offline until needed, so no one can access this portion of storage until it is varied on. When it is varied on, you must have the correct privileges to access it. Customers can use this IASP to store and retain any historical data that they might need to use occasionally and as a part of security compliance because the IASP can remain offline and inaccessible. The most common usage is to maintain different application versions on the same system and efficiently use them.

- IBM FlashCopy

  You can use IASP to do a full system FlashCopy copy. The copy is taken for the IBM i environment. This process delivers an almost instantaneous copy of the data in the shortest possible time. This copy can be varied on to separate partitions by using the IASP. When the IASP is varied on, the data is available for access to save it to physical or virtual tapes. This process can be automated by using Backup Recovery and Media Services (BRMS), which is fully integrated with IASP to automate the full process and provide reporting.

5.13 Exit points

An *exit point* signifies the point in a system function or program where control is turned over to one or more exit programs to perform a function.

The *registration facility* provides a central point to store and retrieve information about IBM i and other exit points and their associated exit programs. This information is stored in the registration facility repository, and it can be retrieved to determine which exit points and exit programs exist.

You can use the registration facility APIs to register and unregister exit points, add and remove exit programs to retrieve information about exit points and exit programs. You can also perform some of these functions by using the Work with Registration Information (WRKREGINF) command.

The *exit point provider* is responsible for defining the exit point information; defining the format in which the exit program receives data; and calling the exit program. There are four areas where exit points provide another layer of security.

Securing network access

When different network protocols on IBM i, FTP, NetServer, JDBC, ODBC, DDM, DRDA, and others are used extensively by users to connect to back-end databases on IBM i. IBM provides dozens of exit points that cover most network access protocols, which means that exit programs can be created and assigned to these exit points to monitor and log activity, and to control access by using various criteria.

Securing communication port access

There are a few network protocols that do not have their own exit points and cannot be protected in the same way as protocols that have specific exit points. These network protocols include Secure Shell (SSH), SFTP, SMTP, and others. In addition, organizations might need to control communication access in a way that networks or other types of exit points cannot because it is not possible to specify a port number in these other types of exit points. For example, a specific type of network connection might need to use only one or more secured ports.

IBM provides socket exit points that make it possible to develop exit programs for securing connections to your IBM i by specific ports and IP addresses.

Securing database access

One powerful exit point is called Open Database File, and it allows development of exit programs that protect sensitive data from any access. The added layer of security that this exit point provides is significant because of its ability to invoke an exit program whenever a specified file on the system is opened, whether it is a physical file, logical file, SQL table, or SQL view. As with other exit points, your exit program can be defined to audit all activity, such as the user, the method of access, the date and time, and the operation (read, update, add, or delete). Also, the exit program can contain a granular set of rules that control under what conditions the file can be accessed and by whom.

Securing command access

IBM provides exit points that enable the usage of commands, which make it possible to develop exit programs that allow or disallow access to any command within specific circumstances, regardless of whether the access attempt comes from a user performing a CLI function directly within the IBM i, through network access, or otherwise. Because command exit programs supersede normal object-level security, they add an extra, useful layer of security that can control the usage of commands, even for users with powerful authorities such as *ALLOBJ or *SECADM. As with other types of exit points, you can define command exit programs so that each command can have its own specific rules of usage while providing logging of any activity.

5.14 Function usage

Function usage enables you to define who can use an application, the parts of an application, or the functions within a program.

This support is not a replacement for resource security. Function usage does not prevent a user from accessing a resource (such as a file or program) from another interface. Function usage support provides APIs to perform the following tasks:

- Register a function.
- Retrieve information about the function.
- Define who can or cannot use the function.
- Check to see whether the user may use the function.

To use a function within an application, the application provider must register the functions when the application is installed. The registered function corresponds to a code block for specific functions in the application. When the user runs the application, before the application invokes the code block, it calls the check usage API to verify that the user has the authority to use the function that is associated with the code block. If the user may use the registered function, the code block runs. If the user may not use the function, the user is prevented from running the code block.

The system administrator specifies who may access a function. The administrator can either use the Work with Function Usage Information (`WRKFCNUSG`) command to manage the access to program function, or select **Security** → **Function Usage** in IBM Navigator for i.

Separation of duties

Separation of duties helps businesses comply with government regulations and simplifies the management of authorities. It can divide administrative functions across individuals without overlapping responsibilities so that one user does not possess unlimited authority, such as with the `*ALLOBJ` authority. The `QIBM_DB_SECADM` function can grant authority, revoke authority, change ownership, or change the primary group, but without providing access to the object or, in the case of a database table, to the data that is in the table or allowing other operations on the table.

`QIBM_DB_SECADM` function usage can be granted only by a user with the `*SECADM` special authority, and it can be granted to a user or a group.

You can use `QIBM_DB_SECADM` to administer Row and Column Access Control (RCAC). RCAC can restrict which rows a user may access in a table and whether a user may see information in certain columns of a table. For more information, see Row and column access control (RCAC).

> **Note:** For more information about IBM i security, see IBM i 7.5 Security Reference, IBM i 7.5 Security - Planning and setting up system security, and Security.

5.15 IFS security on IBM i

IFS is a component of the IBM i that facilitates stream input/output and storage management. much like personal computers and UNIX systems, while offering a cohesive structure for all information that is stored within the system. Ensuring IFS security is a primary concern when developing security strategies for IBM i. One of the challenges that are faced by security administrators is managing security that is related to the root directory ('/'). This root directory is in most IBM i products, third-party applications, configuration files, code, and data.

A major concern is that the root directory "/" is publicly accessible because the default setting enables full access for public users. After you install the IBM i, the default permission for root is set to *RWX, which is a considerable risk and should be modified. The IFS enables users to store and manage various types of data, such as documents, images, program source code, and more. It offers a unified interface for accessing and managing files across different platforms, which simplifies the integration of IBM i systems with other systems in a diverse IT environment. Often, the data that is stored in IFS is sensitive and requires robust security measures.

Figure 5-7 shows a conceptual view of the IFS in IBM i.

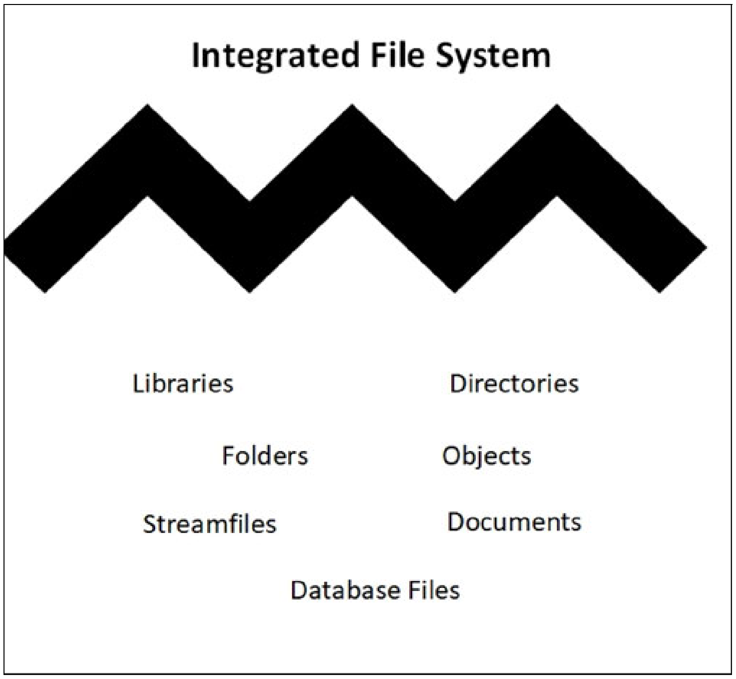


Figure 5-7 A structure for all information that is stored in the IBM i

Virus scanning for the IFS

Viruses target specific computer architectures. The unique architecture of the IBM i system reduces the likelihood of a virus being developed to exploit it. Viruses that are intended for PC environments cannot operate on IBM i, and IBM does not offer any dedicated anti-virus, anti-spyware, anti-malware, or anti-ransomware solutions for this platform. However, if a file that is infected on a PC is transferred to the IFS and then shared with another PC, it can potentially spread the virus to that new machine. Similarly, if a network drive is connected to the IFS, a virus from a PC that can affect files on a network drive might also compromise files that are stored on the IFS.

IBM i does support scanning for malicious activities through third-party software. Users can scan objects within the IFS, which provide them with the flexibility to determine the timing of scans and the actions to take based on the outcomes. There are two exit points that related to this support:

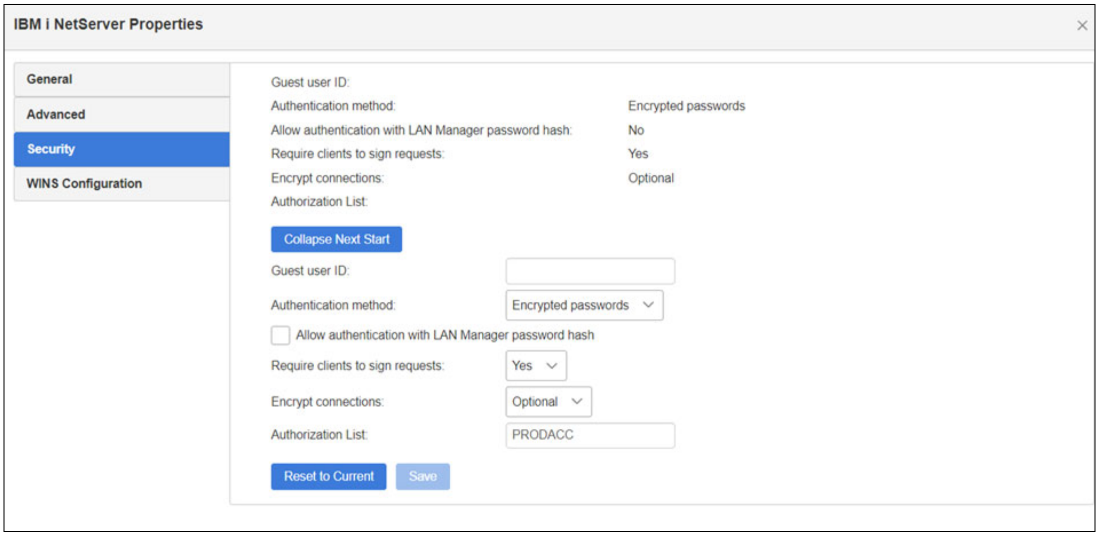
- "QIBM_QPOL_SCAN_OPEN: Integrated File System Scan on Open Exit Program. For this exit point, the IFS scan on an open exit program is called to do scan processing when an IFS object is opened under certain conditions.
- "QIBM_QPOL_SCAN_CLOSE: Integrated File System Scan on Close Exit Program. For this exit point, the IFS scan on the close exit program is called to do scan processing when an IFS object is closed under certain conditions.

> **Note:** Only objects in file systems that are fully converted to *TYPE2 directories are scanned.

Example for securing a file share directory

Standard IBM i security processes are used to set the security of files in the IFS.

Figure 5-8 shows setting a file share directory (/ptf) that is secured by limiting access to members of the authorization list PRODACC.

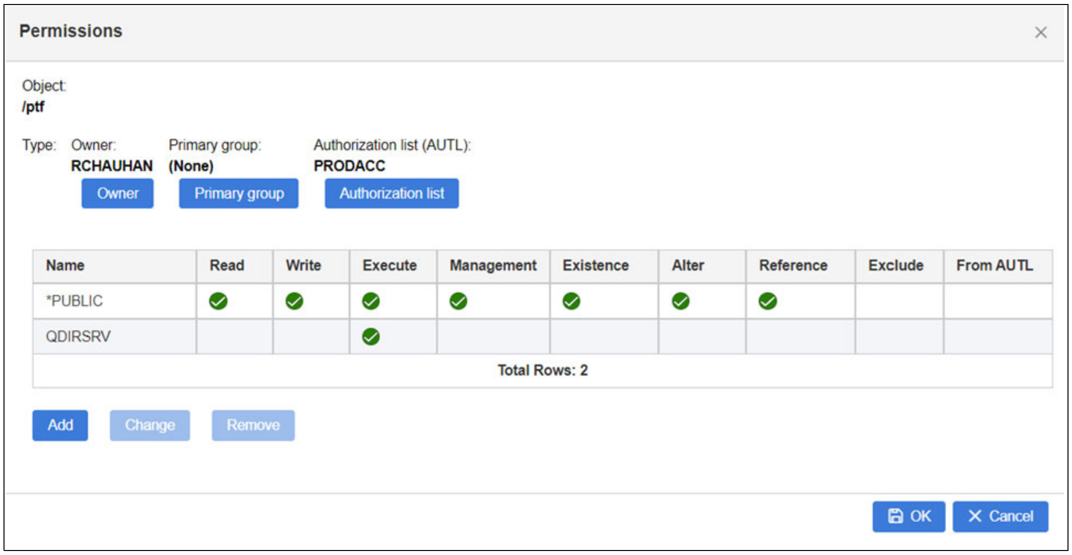


Guest user ID:	
Authentication method:	Encrypted passwords
Allow authentication with LAN Manager password hash:	No
Require clients to sign requests:	Yes
Encrypt connections:	Optional
Authorization List:	

Guest user ID:	
Authentication method:	Encrypted passwords
□ Allow authentication with LAN Manager password hash	
Require clients to sign requests:	Yes
Encrypt connections:	Optional
Authorization List:	PRODACC

Figure 5-8 Limiting access of /ptf to an access list

Figure 5-9 on page 159 shows the interface to display the current access permissions for directory /ptf. Additional access can be set from this window.



Permissions

Object:
/ptf

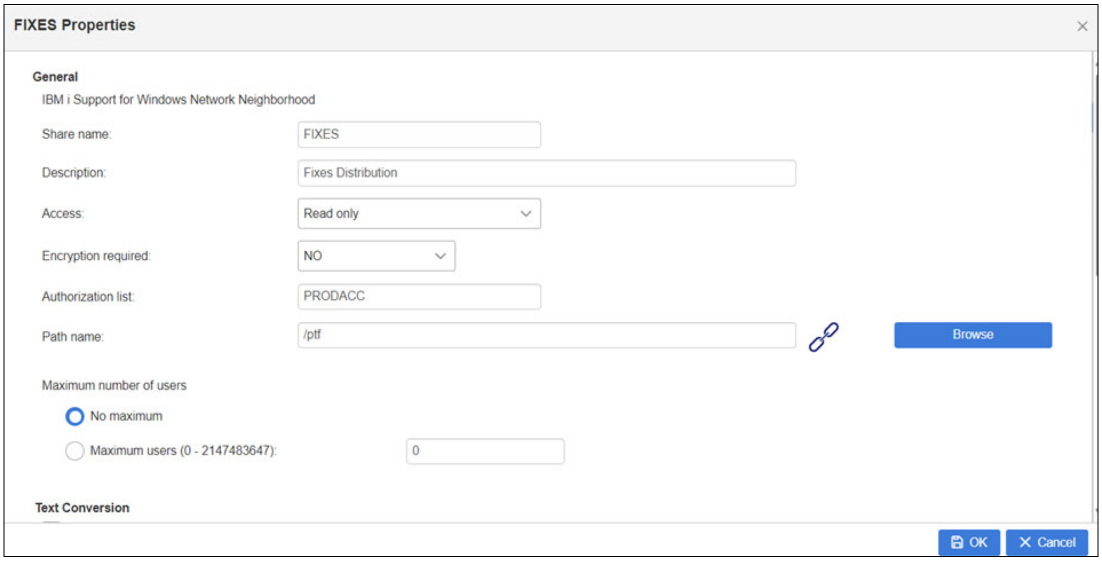
Type: Owner: **RCHAUHAN**
Primary group: **(None)**
Authorization list (AUTL): **PRODACC**

Name	Read	Write	Execute	Management	Existence	Alter	Reference	Exclude	From AUTL
*PUBLIC	✓	✓	✓	✓	✓	✓	✓		
QDIRSRV			✓						

Total Rows: 2

Figure 5-9 Permissions display for /ptf

Figure 5-10 shows the definition of a shared directory (`FIXES`) that points to the `/ptf` directory and defines the access as limited to members in the PRODACC group.



FIXES Properties

General

IBM i Support for Windows Network Neighborhood

Share name:	FIXES
Description:	Fixes Distribution
Access:	Read only
Encryption required:	NO
Authorization list:	PRODACC
Path name:	/ptf

Maximum number of users

● No maximum
○ Maximum users (0 - 2147483647): 0

Text Conversion

Figure 5-10 Share definition for FIXES (connects to /ptf)

Tips to enhance security measures

To enhance security measures for the IFS, consider the following tips:

1. Review file shares.

   Do a systematic review of the file shares regularly. This evaluation is important because it can greatly lower risk by helping ensure that any unnecessary shares in the system are removed.

2. Set shares to read-only wherever possible.

   Because of the potential risks that are associated with write access, particularly in the context of ransomware or malware, implement access review processes. The primary goal of these exercises is to lower the risk of security breaches by routinely examining access rights. This process involves making educated decisions about who has access to sensitive data and critical resources, and then determining whether such access is warranted for each user. This includes root and should be changed to `*PUBLIC` use only.

3. Do not share the root or `/QSYS.lib`.

   Sharing the root directory ("/") or `/QSYS.lib` poses a considerable security threat, especially in the absence of effective access controls. If someone maps to root, they can see the entire structure of then system. This practice reveals all files and directories under the root, which is not advisable. Instead, create specific shares as needed, and no higher.

4. Use Authority Collection if there is an unclear policy about the usage of shares by users.

   The Authority Collection feature is included as part of the core operating system. It operates by collecting data that is linked to the runtime authority verification processes that are embedded within the IBM i system. The primary goal of this feature is to support security administrators and application providers in safeguarding application objects with the least amount of authority that is required for effective operation. Using authority collection to reduce or prevent excessive authority contributes to strengthening the overall security of the objects that are employed by an application.

5. Restrict server and share access by using an authorization list.

   Starting in IBM i 7.5, user access to IBM i NetServer and specific shares can be restricted by assigning an authorization list. IBM i NetServer now allows assigning an authorization list object to the server and individual shares. The authorization list is used as an extra layer of protection for shared resources. You can update the configuration through Navigator by changing the IBM i NetServer properties or Share properties or by using the IBM i NetServer APIs. A user must have at least `*USE` authority to the authorization list to access the server when an authorization list is assigned.

> **Important:** Authorization lists do not restrict access to users with the `*ALLOBJ` special authority. Any user profile with the `*ALLOBJ` special authority may access IBM i NetServer as though there is no authorization list restriction in place. You can use this special authority to create administrative shares that can be accessed only by IBM i administrative profiles by specifying an authorization list that lists only public `*EXCLUDE`.

For more information about IBM i NetServer security, see IBM i NetServer security.

For more information about the IBM i 7.5 IFS, see IBM i 7.5: Files and file systems Integrated File System.

5.16 Security Compliance Tools for IBM i

IBM Technology Expert Labs for Security is your accelerator to maximum value with its unmatched IBM product expertise. Its mission is to successfully deploy, optimize, and expand IBM Security platforms by using the most knowledgeable IBM Security Software experts.

The IBM Technology Expert Labs team for IBM i Security is an IBM team that specializes in IBM i security services, such as security assessments, system hardening, and developing IBM i utilities. This family of utilities is known as Security Compliance Tools for IBM i.

For more information about these offerings, see “Security assessment for IBM Power from IBM Technology Expert Labs” on page 262.

6

Linux security and compliance on IBM Power

IBM Power servers enable the operation of Linux operating systems that leverage the superior capabilities of IBM Power hardware, which include high performance, dependability, and resilience. With the increasing adoption of Linux workloads on IBM Power servers, ensuring stringent and accurate security measures across those Linux landscapes assumes a paramount importance.

This chapter provides an overview of Linux on Power by highlighting its unique features and challenges. It describes various supported Linux environments and offers guidance about implementing robust security measures to establish a secure and high-performing Linux system. By combining the strengths of Linux and IBM Power technology, organizations can benefit from a powerful and flexible infrastructure.

This chapter describes the following topics:

- 6.1, “Overview” on page 164
- 6.3, “Linux on Power” on page 166
- 6.4, “Hardening Linux systems” on page 168
- 6.5, “Best practices” on page 196
- 6.6, “Developing an incident response plan” on page 198

6.1 Overview

Many clients run multiple different operating systems on their IBM Power servers, which include AIX, IBM i, and Linux on Power. For those mission-critical environments that are implemented on Linux, it is important to ensure the security and compliance of those systems.

Linux is an open-source based system. In contrast to AIX or IBM i, which experienced fewer than ten reported vulnerability reports in 2023, the Linux kernel suffered more than a hundred documented flaws during the same period. Because of its open nature and extensive user base, this outcome was predictable and it makes the task of protecting Linux workloads even more critical.

This chapter provides a comprehensive guide to understanding the various threats, implementing effective security measures, and adopting best practices to help ensure data integrity and confidentiality on Linux systems running on the IBM Power Architecture®, regardless of the distribution that is chosen. This chapter addresses pertinent security and Linux compliance concerns.

Regarding security, the chapter describes practices, processes, and tools that safeguard Linux on Power from cyberthreats to help ensure the confidentiality, integrity, and availability of these systems.

Ensuring robust security in Linux environments requires a comprehensive strategy that integrates multiple layers of defense. This strategy should encompass meticulous configuration management, proactive vulnerability assessments, and strict adherence to regulatory compliance frameworks.

The intricate nature of Linux systems demands a diverse set of tools and methodologies to effectively reduce the attack surface and bolster defenses against both established and emerging threats.

> **Note:** In our laboratory setting, we use various distributions, including Red Hat, SUSE, and Ubuntu, Debian, CentOS, Fedora, Alma, Rocky, and OpenSUSE, which offer robust support for the ppc64le architecture.

6.1.1 Implementation and best practices

Your method of implementing each security measure and maximizing the utility of the available tools varies depending on the distribution and version that is chosen. Therefore, this guide delineates general principles without delving into the minutiae of specific configurations, which can vary and subject to constant change. However, this chapter provides straightforward examples and guidelines about how to apply this knowledge by using open source software that has been tested on the ppc64le architecture. This chapter contains solutions that are best practices in achieving your objective of making your Linux on Power servers as secure as possible.

6.2 Threats

Linux is not immune to security threats. Its vulnerabilities can expose systems to various attacks, including malware, unauthorized access, and data breaches, even in Power servers due to their widespread deployment. To safeguard these systems, a comprehensive, cross-functional approach is required to identify, assess, and mitigate these threats.

To enhance system security, Linux offers a wide range of potent tools, many of which are available, open-source, compatible with the ppc64le architecture. By implementing these tools and adhering to best practices, you can bolster the overall security posture of Power servers, starting from a basic logical partition (LPAR) that is equipped with Linux and suitable software configurations.

6.2.1 Malware

Malware, including viruses, worms, trojans, and ransomware, poses significant risks to Linux on Power servers. These malicious programs can disrupt operations, steal sensitive information, and cause substantial financial and reputational damage.

Implementing a comprehensive security strategy that includes anti-virus solutions, regular system scans, security patches, strong authentication measures, and user awareness training is crucial for safeguarding Linux systems from malware threats.

6.2.2 Unauthorized access

Unauthorized access can occur through multiple channels, such as exploiting compromised credentials, unaddressed weaknesses, or manipulative tactics (social engineering). Counteracting unauthorized entry requires establishing robust authentication protocols, such as multi-factor authentication (MFA), frequently updating systems, and training users in cybersecurity fundamentals.

6.2.3 Data leaks

Data leaks, whether accidental or malicious, can lead to severe consequences, which include regulatory penalties and loss of customer trust. Effective measures to prevent data leaks include data encryption, stringent access controls, and regular security audits.

6.2.4 Misconfiguration and human errors

Incorrectly configured systems can leave them open for exploitation, which includes weak passwords, open ports, and incorrect permissions. Applying security profiles such as Center for Internet Security (CIS) benchmarks and custom security policies is essential to mitigate the impact of errors and misconfigurations.

6.3 Linux on Power

IBM's long history and strong commitment to open source is the best kept secret in open source. Although open-source communities have long appreciated IBM's role in the movement, until the recent acquisition of Red Hat, not many people outside of those communities would have associated IBM with open source.

IBM was one of the earliest champions of open source. IBM backed influential communities like Linux, Apache, and Eclipse, and pushed for open licenses, open governance, and open standards. Beginning in the late 1990s, IBM supported Linux with patent pledges, a $1 billion investment of technical and other resources, and helped to establish the Linux Foundation in 2000. Since then, IBM has consistently supported open-source initiatives in general, and Linux and accompanying technologies in particular. Proof of this is IBM's support of Linux on IBM hardware, including IBM Power.

IBM Power supports many f Linux distributions, each offering unique features and capabilities. This section provides an overview of the supported distributions and their main security features and utilities.

6.3.1 Linux distributions on Power

When we talk about Linux on Power, we often refer to Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server, but there are other alternatives. IBM has been a long supporter of Linux across all hardware platforms. Linux is supported in both IBM (natively or running under IBM z/VM®) and IBM Power. IBM has been supporting Linux on Power as early as the POWER4-based processors.

Red Hat-based distributions

RHEL and its derivatives, CentOS, Fedora, Alma and Rocky Linux, are widely used distributions that are known for their stability, security, and enterprise-grade support for different use cases, such as classic workloads (such as application servers and large databases) to containers, machine learning, and others.

For more information about RHEL, see this Red Hat website.

Debian-based distributions

Debian is a widely-used operating system that is primarily known for its stability, reliability, security, and extensive software repositories. It is a Linux distribution that consists entirely of no-charge software. Debian is the foundation for many other distributions, most notably Ubuntu, which is also supported on Power.

Ubuntu is optimized for workloads in the mobile, social, cloud, big data, analytics, and machine learning spaces. With its unique deployment tools (including Juju and MAAS), Ubuntu makes the management of those workloads simpler. Starting with Ubuntu 22.04 LTS, IBM Power9 and IBM Power10 processors are supported. For more information about the Ubuntu Server, see Scale out with Ubuntu Server.

SUSE-based distributions

SUSE Linux Enterprise Server, which is traditionally used for SAP HANA on Power environments, is an alternative for classic workloads. Also, OpenSUSE Leap is a community-driven, open-source Linux distribution that was developed by the OpenSUSE Project. It shares its core with SUSE Linux Enterprise Server, and provides a highly stable and well-tested base. It receives the same security fixes when they are released to SUSE Linux Enterprise Server customers.

SUSE Linux Enterprise Server for IBM POWER is an enterprise-grade Linux distribution that is optimized for IBM Power servers. It delivers increased reliability and provides a high-performance platform to meet increasing business demands and accelerate innovation while improving deployment times.

SUSE Linux Enterprise Server for POWER at a glance

SUSE Linux Enterprise Server for POWER was the first Linux distribution that was optimized for IBM Power, and the first to run in Power9 base mode. It has the following benefits:

- Increase reliability and reduce costs for mission-critical applications with advanced reliability, availability, and serviceability (RAS) capabilities that are optimized to support IBM Power features.
- Deliver a high-performance platform to meet increasing business demands with improved application performance and instant access to data.
- Accelerate innovation and improve deployment times with support for Power9 Little Endian Mode (ppc64le) for a broad choice of open source and partner solutions.

For more information, see SUSE Linux Enterprise Server for IBM Power.

Supported distributions

Table 6-1 is a table of Linux distributions that are supported by IBM on IBM Power10 servers. Also listed are the Ubuntu distributions where the support comes directly from Canonical.

Table 6-1 A list of supported Linux distributions on IBM Power10 servers

IBM Power10 processor-based systems	PowerVM LPARs
► 9043-MRX (IBM Power E1050) ► 9105-22A (IBM Power S1022) ► 9105-22B (IBM Power S1022s) ► 9105-41B (IBM Power S1014) ► 9105-42A (IBM Power S1024) ► 9786-22H (IBM Power L1022) ► 9786-42H (IBM Power L1024)	► Red Hat Enterprise Linux 9.0, any subsequent RHEL 9.x releases ► Red Hat Enterprise Linux 8.4, any subsequent RHEL 8.x releases ► SUSE Linux Enterprise Server 15 SP3, any subsequent SUSE Linux Enterprise Server 15 updates ► Red Hat OpenShift Container Platform 4.9 or later ► Ubuntu 22.04 or later[a]
► 9080-HEX (IBM Power E1080)	► Red Hat Enterprise Linux 9.0, any subsequent RHEL 9.x releases ► Red Hat Enterprise Linux 8.4, any subsequent RHEL 8.x releases ► Red Hat Enterprise Linux 8.2 (Power9 Compatibility mode only)[b] ► SUSE Linux Enterprise Server 15 SP3, any subsequent SUSE Linux Enterprise Server 15 updates ► SUSE Linux Enterprise Server 12 SP5 (Power9 Compatibility mode only) ► Red Hat OpenShift Container Platform 4.9, or later ► Ubuntu 22.04 or later[a]

IBM Power10 processor-based systems	PowerVM LPARs
▶ 9028-21B (IBM Power S1012)	▶ Red Hat Enterprise Linux 9.2, for PowerLE, or later ▶ Red Hat OpenShift Container Platform 4.15, or later ▶ Ubuntu 22.04 or later[a]

[a] Ubuntu on Power support is available directly from Canonical.

[b] Red Hat Business Unit approval is required for using RHEL 8.2 on IBM Power10 processor-based systems.

IBM Power10 processor-based systems support the following configurations per LPAR:

- SUSE Linux Enterprise Server 15 SP4: Up to 64 TB of memory and 240 processor cores
- SUSE Linux Enterprise Server 15 SP3: Up to 32 TB of memory and 240 processor cores
- Red Hat Enterprise Linux 8.6, or later: Up to 64 TB of memory and 240 processor cores
- Red Hat Enterprise Linux 8.4 and 9.0: Up to 32 TB of memory and 240 processor cores
- SUSE Linux Enterprise Server 12 SP5 and RHEL 8.2: Up to 8 TB of memory and 120 processor cores

The recommended Linux distribution for a particular server is always the latest level distribution that is optimized for the server. The listed distributions are the operating system versions that are supported for the specific hardware. For more information about the product lifecycles for Linux distributions, see the support site for each distribution.

- SUSE Linux Enterprise Server: SUSE Product Support Lifecycle
- Red Hat Enterprise Linux: Red Hat Enterprise Linux Life Cycle
- Ubuntu: Ubuntu Release Life Cycle

For libraries and tools that can help leverage the capabilities of Linux on Power, see IBM Software Development Kit for Linux on Power tools. Other information about packages and migration assistance can be found in Find packages built for POWER in the Linux on Power developer portal.

Cores are supported in the Red Hat OpenShift Container Platform. For more information about Red Hat OpenShift Container Platform, see Getting started with Red Hat OpenShift on IBM Cloud and Architecture and dependencies of the service.

6.4 Hardening Linux systems

System hardening in Linux is an ongoing, dynamic process that involves the careful application of security principles and practices to safeguard the system against threats.

Given the complexity of Linux systems, various tools and methodologies are necessary to effectively minimize the attack surface and strengthen defenses against both established and emerging threats.

The security measures and available tools that you use depend on the chosen distribution and version. This section outlines general principles without focusing on specific configurations, which might change over time.

This section covers essential aspects of hardening a GNU/Linux OS on IBM Power from a distribution-neutral perspective. We provide practical examples and guidelines by using open-source software that is tested on ppc64le, specifically in Debian and Fedora to ensure that our Linux on Power servers are as secure as possible through an open-source first approach.

Although Linux offers the advantage of open-source software, challenges remain in building and deploying applications on ppc64le due to lack of access to some proprietary programs and tools, and missing dependencies and build processes. However, as data centers embrace multi-architecture environments, these gaps are closing. When selecting tools, prioritize ones with native ppc64le support.

6.4.1 Compliance

Compliance helps ensure that Linux deployments meet the minimum required standards in terms of configuration, patching, security, and regulatory compliance.

CIS Benchmarks and DISA Security Technical Implementation Guides

CIS Benchmarks and Defense Information Systems Agency (DISA) Security Technical Implementation Guides (STIGs) provide detailed guidelines for securing IT systems, including Linux OS servers.

CIS Benchmarks are developed by the CIS and offer best practices for securing a wide range of systems and applications, including various Linux distributions. They are community-driven and cover a broad spectrum of security configurations.

DISA STIGs are developed by the DISA and tailored to the stringent security requirements of the US Department of Defense (DoD). These guides provide highly detailed security configurations and are mandatory for DoD-related systems. DISA STIGs offer comprehensive security measures that address potential threats that are specific to defense environments. Implementing these guidelines helps ensure that systems meet federal security standards and are protected against sophisticated threats.

For our purpose of providing a good basis for Linux security in Power, we use CIS as a reference, but other standards such as Payment Card Industry Data Security Standard (PCI DSS) might be more appropriate depending on the environment.

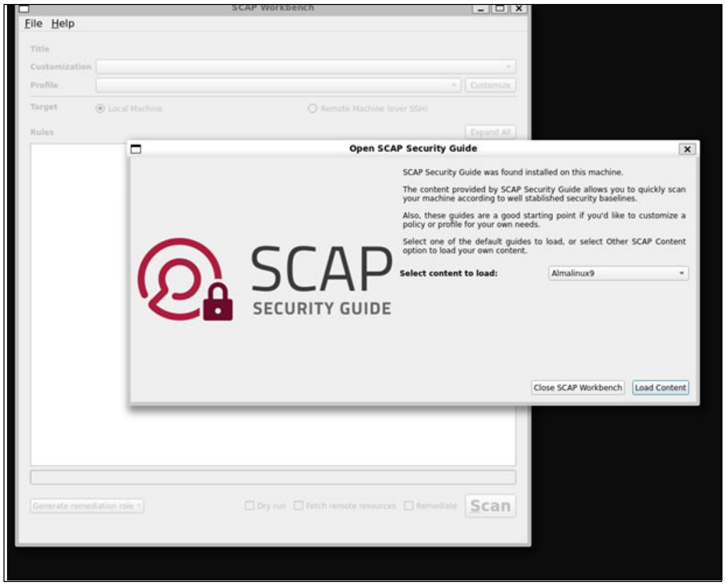
SCAP Security Guide

The SCAP Security Guide (SSG) is a comprehensive collection of security policies, baselines, and configuration guides that are developed by the open-source community and maintained by the National Institute of Standards and Technology (NIST). It provides a standardized approach to configuring and managing security settings for various operating systems and applications. The SSG has three key components:

1. Security Baselines: Predefined sets of security policies that align with various regulatory standards, such as CIS, DISA STIG, PCI DSS, and others. These baselines help organizations comply with industry-specific security requirements.
2. SCAP Content: Machine-readable files that are written in XML that describe the security policies and configurations. These files include benchmarks, rules, checks, and remediation scripts.
3. Automation Tools: SSG provides tools and scripts for automating the process of scanning, evaluating, and remediating security configurations. These tools can generate reports, fix scripts, and apply configurations across multiple systems.

OpenSCAP, often referred to by its CLI tool `oscap`, is an open-source framework that implements SCAP standards. It provides tools to audit and verify system configurations, vulnerabilities, and compliance against the content that is provided by SCAP, including the content from the SSG.

OpenSCAP can verify that the ppc64le system adheres to various security benchmarks and standards, such as CIS Benchmarks, NIST guidelines, custom security policies, or vulnerability lists. It also has a GUI, `scap-workbench`, which is available on RHEL-based distributions on Power, such as Alma Linux 9, which is shown in Figure 6-1.



Open SCAP Security Guide

SCAP Security Guide was found installed on this machine.

The content provided by SCAP Security Guide allows you to quickly scan your machine according to well stablished security baselines.

Also, these guides are a good starting point if you'd like to customize a policy or profile for your own needs.

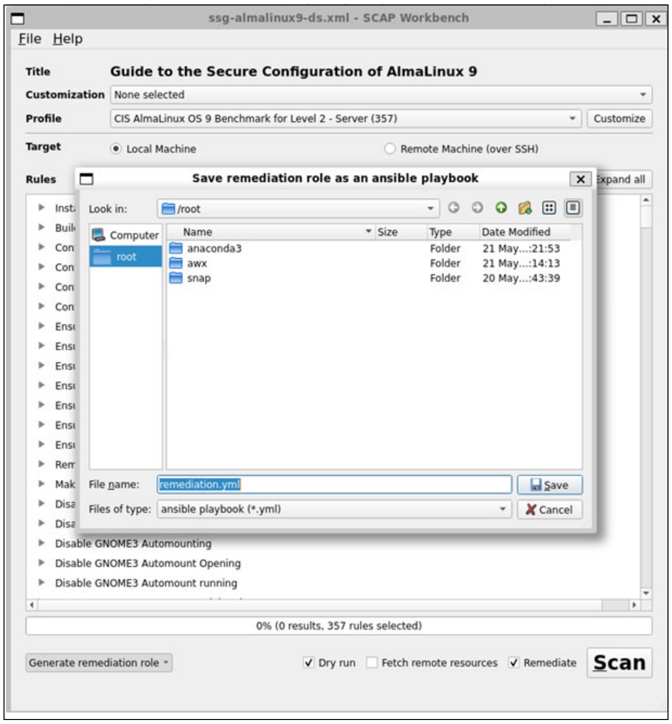
Select one of the default guides to load, or select Other SCAP Content option to load your own content.

Select content to load: Almalinux9

Close SCAP Workbench | Load Content

Figure 6-1 SCAP workbench GUI

Figure 6-2 shows one of the management windows in the GUI.



ssg-almalinux9-ds.xml - SCAP Workbench

Title: Guide to the Secure Configuration of AlmaLinux 9

Customization: None selected

Profile: CIS AlmaLinux OS 9 Benchmark for Level 2 - Server (357)

Target: ● Local Machine ○ Remote Machine (over SSH)

Save remediation role as an ansible playbook

Look in: /root

Name	Size	Type	Date Modified
anaconda3		Folder	21 May...:21:53
awx		Folder	21 May...:14:13
snap		Folder	20 May...:43:39

File name: remediation.yml

Files of type: ansible playbook (*.yml)

Disable GNOME3 Automounting

Disable GNOME3 Automount Opening

Disable GNOME3 Automount running

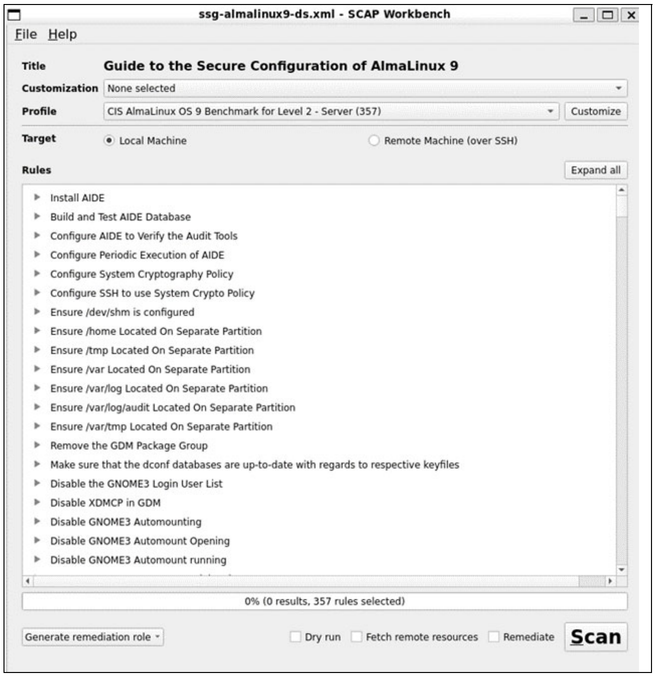
0% (0 results, 357 rules selected)

Generate remediation role ☑ Dry run □ Fetch remote resources ☑ Remediate Scan

Figure 6-2 Management window in the GUI

To comply with the security regulations and policies, complete the following steps:

1. Install the Linux ISO of your choice.
2. Decide which set of rules to use (always start with a dry run). Figure 6-3 shows choosing a CIS Level 2 benchmark by using scap-workbench (GUI).



ssg-almalinux9-ds.xml - SCAP Workbench

Title: Guide to the Secure Configuration of AlmaLinux 9

Customization: None selected

Profile: CIS AlmaLinux OS 9 Benchmark for Level 2 - Server (357)

Target: ● Local Machine ○ Remote Machine (over SSH)

Rules

- Install AIDE
- Build and Test AIDE Database
- Configure AIDE to Verify the Audit Tools
- Configure Periodic Execution of AIDE
- Configure System Cryptography Policy
- Configure SSH to use System Crypto Policy
- Ensure /dev/shm is configured
- Ensure /home Located On Separate Partition
- Ensure /tmp Located On Separate Partition
- Ensure /var Located On Separate Partition
- Ensure /var/log Located On Separate Partition
- Ensure /var/log/audit Located On Separate Partition
- Ensure /var/tmp Located On Separate Partition
- Remove the GDM Package Group
- Make sure that the dconf databases are up-to-date with regards to respective keyfiles
- Disable the GNOME3 Login User List
- Disable XDMCP in GDM
- Disable GNOME3 Automounting
- Disable GNOME3 Automount Opening
- Disable GNOME3 Automount running

0% (0 results, 357 rules selected)

□ Dry run □ Fetch remote resources □ Remediate

Figure 6-3 Choosing benchmarks through the GUI

3. Automatically address these compliance gaps when technically feasible by using Bash scripts and Ansible playbooks by using the CLI:

```
oscap xccdf generate fix --profile [PROFILE_ID] --output remediation_script.sh
\  usr/share/xml/scap/ssg/content/ssg-[OS].xml
```

Automated remediation might yield unexpected results on systems that already are modified. Therefore, thoroughly evaluate the potential impact of remediation actions on your specific systems. You might want to make a snapshot or backup before continuing.

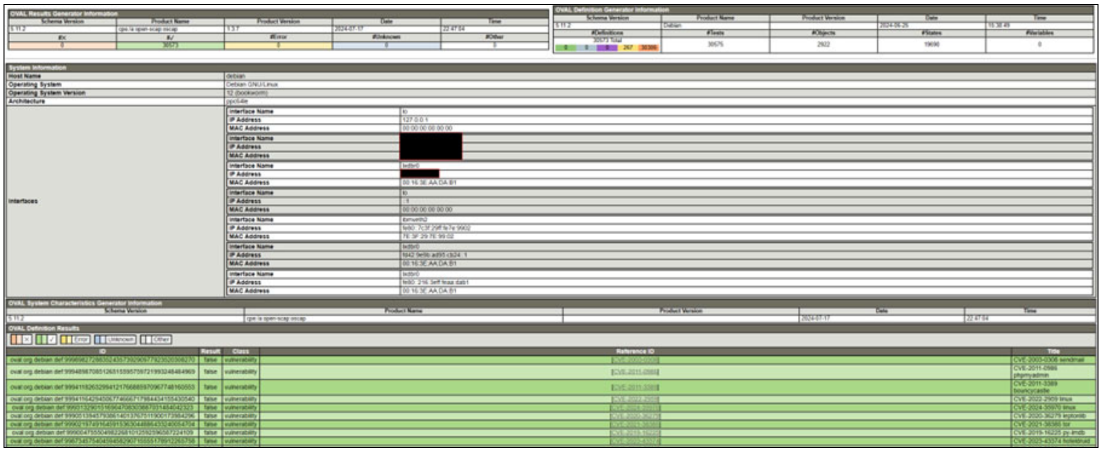
> **Tip:** Under normal conditions, the remediation of compliance issues is the result of several iterations and some backtracking by recovering snapshots or backups until you reach a level of security that is adequate for your purposes. Compliance should be balanced against the usability of the system.

OpenSCAP can also help you check whether there are any vulnerabilities in your current OS version by running Open Vulnerability and Assessment Language (OVAL) and generating a report. Example 6-1 shows how you can generate the report by using Debian.

Example 6-1 Generating a vulnerability report

```
wget https://www.debian.org/security/oval/oval-definitions-$(lsb_release -cs).xml.bz2
bzip2recover oval-definitions-$(lsb_release -cs).xml.bz2
oscap oval eval --report report.html oval-definitions-$(lsb_release -cs).xml
```

Figure 6-4 shows the generated HTML report with no vulnerabilities found.



OVAL Results Generator Information				
Schema Version	Product Name	Product Version	Date	Time
5.11.2	cpe:/a:open-scap:oscap	1.3.7	2024-07-17	22:47:04
#X	#✓	#Error	#Unknown	#Other
0	30573	0	0	0

OVAL Definition Generator Information				
Schema Version	Product Name	Product Version	Date	Time
5.11.2	Debian		2024-06-25	15:38:49
#Definitions	#Tests	#Objects	#States	#Variables
30573 Total 0 0 0 267 30306	30575	2922	19690	0

System Information		
Host Name	debian	
Operating System	Debian GNU/Linux	
Operating System Version	12 (bookworm)	
Architecture	ppc64le	
Interfaces	Interface Name	lo
	IP Address	127.0.0.1
	MAC Address	00:00:00:00:00:00
	Interface Name	[redacted]
	IP Address	[redacted]
	MAC Address	[redacted]
	Interface Name	lxdbr0
	IP Address	[redacted]
	MAC Address	00:16:3E:AA:DA:B1
	Interface Name	lo
	IP Address	::1
	MAC Address	00:00:00:00:00:00
	Interface Name	ibmveth2
	IP Address	fe80::7c3f:29ff:fe7e:9902
	MAC Address	7E:3F:29:7E:99:02
	Interface Name	lxdbr0
	IP Address	fd42:9e9b:ad95:cb24::1
	MAC Address	00:16:3E:AA:DA:B1
	Interface Name	lxdbr0
	IP Address	fe80::216:3eff:feaa:dab1
	MAC Address	00:16:3E:AA:DA:B1

OVAL System Characteristics Generator Information				
Schema Version	Product Name	Product Version	Date	Time
5.11.2	cpe:/a:open-scap:oscap		2024-07-17	22:47:04

OVAL Definition Results

X ✓ Error Unknown Other

ID	Result	Class	Reference ID	Title
oval:org.debian:def:99969827286352435739290977923520308270	false	vulnerability	[CVE-2003-0308]	CVE-2003-0308 sendmail
oval:org.debian:def:99948987085126515595759721993248484969	false	vulnerability	[CVE-2011-0986]	CVE-2011-0986 phpmyadmin
oval:org.debian:def:99941182632994121766885970967748160553	false	vulnerability	[CVE-2011-3389]	CVE-2011-3389 bouncycastle
oval:org.debian:def:99941164294506774666717984434155430540	false	vulnerability	[CVE-2022-2909]	CVE-2022-2909 linux
oval:org.debian:def:99901329015169047083098670314840042323	false	vulnerability	[CVE-2024-35970]	CVE-2024-35970 linux
oval:org.debian:def:99905139457938614013767511900173984296	false	vulnerability	[CVE-2020-36279]	CVE-2020-36279 leptonlib
oval:org.debian:def:99902197491645915363044886433240054704	false	vulnerability	[CVE-2021-38385]	CVE-2021-38385 tor
oval:org.debian:def:99900475550498226810125925965872241109	false	vulnerability	[CVE-2019-16225]	CVE-2019-16225 py-lmdb
oval:org.debian:def:99673457540459458290710555178912263758	false	vulnerability	[CVE-2023-43374]	CVE-2023-43374 hoteldruid

Figure 6-4 Vulnerability report

Tip: If you are satisfied with the image that you evaluated and you use IBM PowerVC (the IBM Power virtualization solution that is based on OpenStack), it is a best practice to capture this system as a template or create an OVA. Be careful when using “default” installations because they might be missing important security protection settings. Always use the appropriate compliance policies to help ensure that your Linux systems running on IBM Power are all configured and protected.

To summarize, although compliance signifies that organizations adhere to the minimal required standards, it does not automatically imply full security. A firm might fulfill every criterion for a PCI DSS assessment without ensuring effective employee training, which can result in inadequate execution and a higher likelihood of security breaches.

6.4.2 Network security

Network security measures protect Linux systems from external and internal threats. This protection includes implementing intrusion detection and prevention systems (IDPS) and encrypting data in transit.

Firewall technologies

Firewalls are a critical component of network security. They are essential for controlling the flow of incoming and outgoing traffic based on predefined security rules. Effective firewall management on Linux systems involves various tools, each offering different levels of control, efficiency, and simplicity of use. This section explores the primary tools that are used in Linux firewall implementations, their relationships, and practical guidance on their use.

Linux firewalls have evolved over time. They started with simple packet filtering mechanisms, and now use more sophisticated management tools. The primary tools that are used in Linux firewall implementations include `iptables`, `nftables`, `firewalld`, and Uncomplicated Firewall (UFW). Understanding the background and functions of these tools helps you choose the correct one for your specific needs (including the distribution that you chose).

Front-end tools

Here are the front-end tools for setting up firewalls in Linux:

- *Netfilter* is a framework within the Linux kernel that provides various networking-related operations, such as packet filtering, network address translation (NAT), and packet mangling. It is the core infrastructure that enables these operations, with hooks in the kernel where modules can register callback functions to handle network packets. Both `iptables` and `nftables` are user-space utilities that interact with the Netfilter framework,
- `iptables` is a CLI utility that administrators can use to configure the Linux kernel firewall. It enables granular control over packet filtering and manipulation, offering precise management of data flow and security policies. Despite its powerful features, `iptables` can be complex to configure due to its detailed syntax and structure.

  This command enables the use of Secure Shell (SSH):

  ```
  sudo iptables -A INPUT -p tcp --dport 22 -j ACCEPT
  ```
- `nftables` is the successor to `iptables`, and is designed to provide a more efficient and streamlined framework for packet filtering and NAT. Introduced in the Linux kernel in Version 3.13, `nftables` offers a simplified syntax and enhanced performance. It has been gradually adopted by many distributions as the default back end for firewall configurations, aiming to overcome part of the complexity and performance limitations of `iptables`. The command to permit SSH in `nftables` is as follows:

  ```
  sudo nft add rule inet filter input tcp dport 22 accept
  ```
- `bpfilter` is a newer and primary kernel-space framework that is designed to improve firewall management in Linux by leveraging Extended Berkeley Packet Filter (eBPF). `bpfilter` aims to replace the older `iptables` and `nftables` with a more flexible and efficient packet filtering mechanism, but it is still under heavy development and has not yet seen widespread adoption.

Firewall tools

Within the front-end tools for creating and maintaining firewall rules in Linux (by using `iptables` or `nftables`), you have two main options:

- `firewalld` is a dynamic firewall management tool that has been part of RHEL since Version 7. It simplifies firewall management by using the concept of network zones, which define the trust level of network connections and interfaces. `firewalld` enables real-time changes without restarting the firewall, which provides a flexible and dynamic approach compared to traditional static tools like `iptables`. `firewalld` uses `nftables` as its back end by default on modern systems, and `firewall-cmd` as the CLI tool.

  The following command is used to set the firewall:

  ```
  sudo firewall-cmd --zone=public --add-port=22/tcp --permanent
  ```
- UFW is designed to provide an interface for managing firewall settings. With its straightforward command-line interface (CLI), UFW enables users to implement basic firewall rules with minimal effort. It is useful for users who might not have extensive networking or firewall management experience but still must ensure system security. UFW is included by default in Ubuntu and can be installed on Debian systems.

  Here is a command to enable SSH to UFW:

  ```
  sudo ufw allow 22/tcp
  ```

Recommendations for Linux firewalls

The CIS advises setting a default deny policy for both incoming and outgoing traffic to ensure that only explicitly allowed traffic is permitted. This action involves allowing essential services such as SSH from trusted networks, and loop-back traffic, while restricting other services to mitigate unauthorized access.

Regular reviews and updates of firewall rules are a best practice to maintain compliance with security policies and adapt to emerging threats. These measures collectively aim to fortify Linux systems against various network-based threats.

A best practice is to forward Linux firewall logs to a solution that automates their analysis, alerts, and responses. For an example that uses IBM QRadar, see this document.

Example 6-2 show a simple firewall configuration on Linux on Power by using `firewall-cmd`.

Example 6-2 Configuring a firewall with firewall-cmd

```
# Install firewalld (if not already installed)
sudo dnf install firewalld -y

# Start and enable firewalld
sudo systemctl start firewalld
sudo systemctl enable firewalld

# Allow SSH traffic
sudo firewall-cmd --permanent --add-service=ssh

# Set default deny policies
# Firewalld uses zones to manage rules. The default zone is "public".
# By default, firewalld allows all outgoing traffic. To mimic iptables behavior,
you can configure it to deny by default.
sudo firewall-cmd --permanent --set-target=DROP

# Allow loopback traffic
sudo firewall-cmd --permanent --add-interface=lo --zone=trusted
sudo firewall-cmd --permanent --zone=trusted --add-source=127.0.0.1

# Enable logging (optional)
sudo firewall-cmd --set-log-denied=all

# Allow specific outgoing traffic (optional)
sudo firewall-cmd --permanent --add-port=80/tcp
sudo firewall-cmd --permanent --add-port=443/tcp

# Reload firewall to apply changes
sudo firewall-cmd --reload

# Review firewalld rules
sudo firewall-cmd --list-all
```

Example 6-3 shows this same configuration by using UFW.

Example 6-3 Configuring a firewall with UFW

```
# Install UFW
sudo apt-get install ufw

# Allow SSH traffic
sudo ufw allow ssh

# Set default deny policies
sudo ufw default deny incoming
sudo ufw default deny outgoing

# Enable UFW
sudo ufw enable

# Allow specific outgoing traffic (optional)
sudo ufw allow out to any port 80
sudo ufw allow out to any port 443

# Review UFW status
sudo ufw status verbose
```

Tip: Be careful not to lock yourself out. Although the new rules do not apply to existing connections, make sure that you either have a script to disable the firewall automatically after a few seconds or direct console access in an emergency.

Also, CIS emphasizes the importance of logging and auditing firewall activity to detect and respond to suspicious behavior, and suggests using stateful inspection and rate limiting to prevent attacks like denial of service (DoS).

Intrusion detection and prevention

Intrusion detection systems (IDS) and intrusion prevention systems (IPS) monitor network traffic for suspicious activities. Tools such as Snort, OSSEC, Zeek, and Suricata can be deployed to detect and prevent potential threats.

This use case uses *Suricata* because it has strong support for the ppc64le architecture. Suricata is a versatile and high-performance Network Security Monitoring (NSM) tool that can detect and block network attacks. By default, Suricata operates as a passive IDS by scanning for suspicious traffic on a server or network and generating logs and alerts for further analysis. Also, you can configure it as an active IPS to log, alert, and block network traffic that matches specific rules. Suricata is open source and managed by the Open Information Security Foundation.

Example 6-4 shows setting up Suricata.

Example 6-4 Intrusion detection by using Suricata

```
root@debian:~/snort3# systemctl status suricata

suricata.service - Suricata IDS/IDP daemon
Loaded: loaded (/lib/systemd/system/suricata.service; enabled; preset: enabled)
Active: active (running) since Wed 2024-07-17 19:07:00 BST; 2s ago
Docs: man:suricata(8)
      man:suricatasc(8)
      https://suricata-ids.org/docs/
    Process: 2965553 ExecStart=/usr/bin/suricata -D --af-packet -c
/etc/suricata/suricata.yaml --pidfile /run/suricata.pid (code=exited,
status=0/SUCCESS)
   Main PID: 2965554 (Suricata-Main)
      Tasks: 38 (limit: 9635)
     Memory: 255.8M
        CPU: 335ms
     CGroup: /system.slice/suricata.service
             ··2965554 /usr/bin/suricata -D --af-packet -c
/etc/suricata/suricata.yaml --pidfile /run/suricata.pid

##
## Step 3: Configure common capture settings
##
## See "Advanced Capture Options" below for more options, including Netmap
## and PF_RING.
##

# Linux high-speed capture support
af-packet:
  - interface: ibmveth2 <- modify this
    # Number of receive threads. "auto" uses the number of cores
    #threads: auto
```

Suricata includes a tool that called `suricata-update` that can fetch rule sets from external providers. Example 6-5 shows how to download the latest rule set for your Suricata server.

Example 6-5 Installing suricata-update

```
root@debian:~# sudo suricata-update -o /etc/suricata/rules
20/7/2024 -- 21:38:39 - <Info> -- Using data-directory /var/lib/suricata.
20/7/2024 -- 21:38:39 - <Info> -- Using Suricata configuration /etc/suricata/suricata.yaml
20/7/2024 -- 21:38:39 - <Info> -- Using /etc/suricata/rules for Suricata provided rules.
20/7/2024 -- 21:38:39 - <Info> -- Found Suricata version 6.0.10 at /usr/bin/suricata.
20/7/2024 -- 21:38:39 - <Info> -- Loading /etc/suricata/suricata.yaml
20/7/2024 -- 21:38:39 - <Info> -- Disabling rules for protocol http2
20/7/2024 -- 21:38:39 - <Info> -- Disabling rules for protocol modbus
20/7/2024 -- 21:38:39 - <Info> -- Disabling rules for protocol dnp3
20/7/2024 -- 21:38:39 - <Info> -- Disabling rules for protocol enip
20/7/2024 -- 21:38:39 - <Info> -- No sources configured, will use Emerging Threats Open
20/7/2024 -- 21:38:39 - <Info> -- Fetching
https://rules.emergingthreats.net/open/suricata-6.0.10/emerging.rules.tar.gz.
20/7/2024 -- 21:40:43 - <Info> -- Loaded 367 rules.
20/7/2024 -- 21:40:43 - <Info> -- Disabled 14 rules.
20/7/2024 -- 21:40:43 - <Info> -- Enabled 0 rules.
20/7/2024 -- 21:40:43 - <Info> -- Modified 0 rules.
20/7/2024 -- 21:40:43 - <Info> -- Dropped 0 rules.
20/7/2024 -- 21:40:43 - <Info> -- Enabled 0 rules for flowbit dependencies.
20/7/2024 -- 21:40:43 - <Info> -- Backing up current rules.
20/7/2024 -- 21:40:43 - <Info> -- Writing rules to /etc/suricata/rules/suricata.rules:
total: 367; enabled: 311; added: 367; removed 0; modified: 0
20/7/2024 -- 21:40:43 - <Info> -- Writing /etc/suricata/rules/classification.config
20/7/2024 -- 21:40:43 - <Info> -- Testing with suricata -T.
20/7/2024 -- 21:40:43 - <Info> -- Done.
```

For more information about Suricata, including installation instructions, see the Suricata documentation.

Encryption in transit

Encrypting data in transit protects it from being intercepted and read by unauthorized parties. Protocols such as Secure Sockets Layer (SSL)/Transport Layer Security (TLS) are used to secure communications over networks.

- SSL/TLS are secure protocols for encrypting web traffic, email, and other communications.

  To secure your web server with SSL/TLS, obtain a digital certificate. Certbot is an automated tool that is designed to streamline the process of acquiring and installing SSL/TLS certificates. It is one of many technology projects that are developed by the Electronic Frontier Foundation (EFF) to promote online freedom.

  Certbot is available in different Linux repositories, including the ppc64le versions, making installation straightforward, It has plug-ins for both Apache and nginx, among other deployments, and includes a tool to automatically renew these certificates.

  Example 6-6 shows how to install Certbot in a Debian Linux system. Other Linux versions might differ slightly.

Example 6-6 Installing Certbot

```
sudo apt-get install certbot python3-certbot-apache
sudo apt-get install certbot python3-certbot-nginx
```

  To obtain and automatically install the certificate for your web server, run either sudo `certbot –apache` or `sudo certbot --nginx`, depending on which tool you use.

  Certbot prompts you to enter your email address and agree to the terms of service. It interacts with your web server to perform the domain verification process and install the SSL/TLS certificate.

  For more information and how-to guides, see the Certbot website.

- Virtual private networks (VPNs) create encrypted tunnels to provide secure remote access. Setting up a VPN is essential for protecting communications and helping ensure the privacy of transmitted data. VPNs are highly recommended by the CIS as a best practice for securing remote connections.

  To install and configure OpenVPN on Linux on Power, follow the specific guides for different Linux on Power distributions:

  – Debian-based

  – RHEL-based

  – SUSE-based

Encryption at rest

Encryption at rest is a form of encryption that is designed to prevent an attacker from accessing data by ensuring that the data is encrypted when stored on a persistent device. You can do this task at different layers, from physical storage systems to the OS. If you choose to encrypt at the OS level, employ full disk encryption by using Linux Unified Key Setup (LUKS) with LVM (Debian / RHEL-Based) or BTRFS (SUSE).

LUKS offers a suite of tools that are designed to simplify the management of encrypted devices. LUKS enables encryption of block devices and supports multiple user keys that can decrypt a master key. This master key is used for the bulk encryption of the partition.

You can configure disk encryption at the installation time or later by using `cryptsetup`, which is a CLI tool that you can use to set up disk encryption based on the `dm-crypt` kernel module. It offers a range of functions, including creating, opening, and managing encrypted volumes.

Here are the prerequisites for installing LUKS:

- A Linux system with disk attached.
- `cryptsetup` is installed.
- Root or sudo privileges

Example 6-7 provides an example of activating encryption at rest for a logical volume (LV) by using `lvm`.

Example 6-7 Setting up encryption at rest for an LVM volume

```
sudo apt-get install lvm2 cryptsetup

Initialize the physical volume
sudo pvcreate /dev/sdX

Create a volume group
sudo vgcreate acme_vg /dev/sdX

Create a logical volume
sudo lvcreate -n acme_lv01 -L 10G acme_vg

Encrypt the new LV
sudo cryptsetup luksFormat /dev/acme_vg/acme_lv01

WARNING!
========
This will overwrite data on /dev/sdX irrevocably.

Are you sure? (Type uppercase yes): YES
Enter LUKS passphrase:
Verify passphrase:
Command successful.
Create a mapping
sudo cryptsetup open /dev/acme_vg/acme_lv01 encrypted_lv01

Format the encrypted logical volume
sudo mkfs.ext4 /dev/mapper/encrypted_lv01

Mount it (and use it)
sudo mount /dev/mapper/encrypted_lv01 /media
```

Automating encryption at system start

To ensure that the encrypted LVM is available at system start, configure `/etc/crypttab` and `/etc/fstab`.

To configure `/etc/crypttab`, add the following line:

```
encrypted_lv01 /dev/acme_vg/acme_lv01 none luks
```

To configure `/etc/fstab`, add the following line:

```
/dev/mapper/encrypted_lv01 /mnt ext4 defaults 0 2
```

For more information, see GitLab.

6.4.3 User policies and access controls

User policies and administration are vital for maintaining a secure environment. They include defining and enforcing password policies and managing user access. The CIS provides some recommendations that you can apply to Linux on Power.

Linux uses Pluggable Authentication Modules (PAMs) in the authentication process, serving as an intermediary layer between users and applications. PAM modules are accessible on a system-wide basis, allowing any application to request their services. The PAM modules implement most of the user security measures that are defined in various files within the `/etc` directory, including Lightweight Directory Access Protocol (LDAP), Kerberos and Active Directory (AD) connections, or MFA options.

Access control mechanisms help ensure that only authorized users can access specific resources, which include configuring SUDO, managing user groups, and maintaining access logs.

Password policies

Enforcing strong password policies is crucial to prevent unauthorized access. Policies should mandate complex passwords, regular password changes, and account lockout mechanisms after multiple failed login attempts.

A password policy can specify the minimum length that a password must have and the maximum duration that it can be used before it must be changed. All users under this policy must create passwords that are long enough and update them regularly. Implementing password policies helps mitigate the risk of passwords being discovered and misused.

A minimum password policy for Linux on Power should contain at least the following characteristics:

- Complex passwords: Require a mix of uppercase, lowercase, numbers, and special characters. This policy can be enforced with the `pam_pwquality` PAM module. CIS recommends that passwords should be at least 14 characters long with no limit on the enforced maximum number of characters, among other requirements.

Example 6-8 is an excerpt of a sample configuration of `/etc/security/pwquality.conf` (Ubuntu).

Example 6-8 Example security configuration

```
# Minimum acceptable size for the new password (plus one if
# credits are not disabled which is the default). (See pam_cracklib manual. Cannot be set
to a lower value than 6.
minlen = 15

# The maximum credit for having digits in the new password. If less than 0 it is the
minimum number of digits in the new password.
dcredit = -1

# The maximum credit for having uppercase characters in the new password. # If less than 0
it is the minimum number of uppercase characters in the new
# password.
ucredit = -1

..
```

- Regular changes: CIS recommends specific password change policies for Linux systems to enhance security. These policies include setting a maximum password age of 90 days or less to ensure regular password updates; a minimum password age of 7 days to prevent rapid password changes that might cycle back to previous passwords; and a password expiration warning of 7 days to notify users before impending password expiry. These guidelines help maintain robust security by helping ensure that passwords are regularly updated and users are adequately informed.

  Enforced password expiration policies are defined in `login.defs`. Example 6-9 shows an excerpt of a sample configuration of `/etc/login.defs` (Ubuntu) to log both successful logins and `su` activity.

Example 6-9 Sample login definition

```
#
# Enable logging of successful logins
#
LOG_OK_LOGINS yes


#
# If defined, all su activity is logged to this file.
#
SULOG_FILE var/log/sulog
```

- Account lockout: Typically, it is recommended by CIS to lock the account after five unsuccessful attempts and to unlock it automatically after a specified period, such as 15 minutes. These settings help mitigate the risk of unauthorized access by deterring repeated login attempts and ensuring that legitimate users can regain access after a brief lockout period.

  This policy is configured in the `/etc/pam.d/common-auth` file (Debian, Ubuntu, and SUSE) or `/etc/pam.d/system-auth` file (RHEL). The `pam_faillock` module performs a function similar to the legacy `pam_tally` and `pam_tally2` but with more options and flexibility. Check which is the recommended method in your chosen distribution and version.

  Here is an excerpt of a sample legacy configuration (SUSE):

  ```
  auth   required  pam_tally2.so  onerr=fail deny=3 unlock_time=1800
  ```

Groups

Grouping users based on their roles and responsibilities helps manage permissions. Assigning users to appropriate groups helps ensure that they have access only to the necessary resources. For example, a file with permissions `rw-rw----` (660) allows the owner and the group to read/write the file, but others cannot access it. These permissions reduce the risk of accidental or malicious modifications to sensitive files.

This way, developers can be part of a dev group with access to development files and the production team is part of a *prod* group with access to production files.

CIS advises regular audits of group memberships to help ensure that users have the appropriate permissions and to remove any unnecessary or outdated group assignments. Also, the creation of custom groups for specific tasks or roles is recommended to further refine access control and minimize potential security risks.

Access control lists

Access control lists (ACLs) provide more fine-grained control over permissions for files and directories than user and group permissions. Example 6-10 shows the ACL for a file.

Example 6-10 Viewing an ACL

```
touch testfile
ls -l testfile
-rw-r--r-- 1 user user 0 Jul 17 14:05 testfile
```

Example 6-11 shows the process to grant read/write permissions to another user (in this case, `john`). After setting the ACL for the user, use the `getfacl` command to display the ACL as shown in the example.

Example 6-11 Setting an ACL for a specific user

```
sudo setfacl -m u:john:rw testfile
getfacl testfile
# file: testfile
# owner: user
# group: user
user::rw-
user:john:rw-
group::r--
mask::rw-
other::r--
```

Multi-factor authentication

MFA is a security process that requires users to verify their identity through multiple methods before gaining access to a system or application. Unlike single-factor authentication (for example, a password), MFA enhances security by combining two or more independent credentials from the following categories:

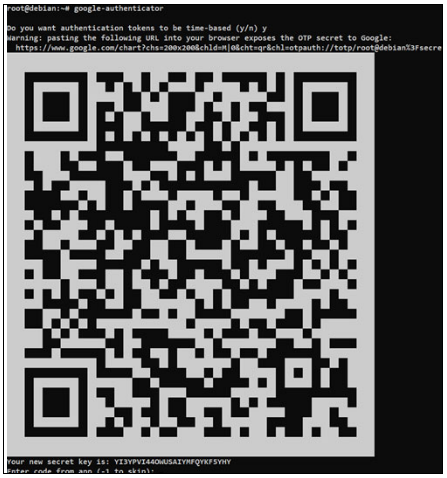
- Something that you know: Typically, a password or personal identification number (PIN)
- Something that you have: A physical device, such as a smartphone
- Something that you are: Biometric verification like fingerprints, facial recognition, or iris scans

When attempting to log in to a system that is secured by MFA, users must supply extra credentials beyond their standard username and password. In the context of Linux systems, SSH serves as a common method for remotely accessing the system. To enhance security further, incorporate MFA when you use SSH.

One method of implementing MFA is using IBM PowerSC. However, MFA can also be implemented by using native tools like Google authenticator by using the following commands:

- `google-authenticator libpam-google-authenticator` for Debian-based systems
- `google-authenticator-libpam` in SUSE-based systems
- `google-authenticator` in Extra Packages for Enterprise Linux (EPEL) for RHEL-based systems

Google authenticator has a standard setup script for configuration. It uses the Google Authenticator app that is available for Android and iOS to generate authentication codes. The authentication code is shown in Figure 6-5.



```
root@debian:~# google-authenticator

Do you want authentication tokens to be time-based (y/n) y
Warning: pasting the following URL into your browser exposes the OTP secret to Google:
  https://www.google.com/chart?chs=200x200&chld=M|0&cht=qr&chl=otpauth://totp/root@debian%3Fsecre
```

```
Your new secret key is: YI3YPVI44OWUSAIYMFQYKF5YHY
Enter code from app (-1 to skip):
```

Figure 6-5 Google authenticator code

For more information about adding MFA to other distributions, see the following resources:

- Configure SSH to use two-factor authentication
- Set up two-factor authentication for SSH on Fedora

Role-based access control

Although organizations have various user provisioning methodologies to choose from, role-based access control (RBAC) is among the most prevalent. RBAC is a method of restricting system access to authorized users based on their roles within an organization. In RBAC, permissions are not assigned directly to users but to roles, and users are assigned to these roles. This abstraction simplifies the management of permissions.

RBAC offers a more detailed approach to identity and access management (IAM) compared to ACLs, yet it is simpler and more straightforward to implement than attribute-based access control (ABAC). Other IAM methods, such as mandatory access control (MAC) or discretionary access control (DAC), can be effective for particular scenarios.

The following list provides some methods to help with setting the appropriate access controls within your system:

- *SELinux* (RHEL/SUSE-based) defines roles and the assignment of domains (or types) to these roles. Users are assigned roles, and the roles define the allowable operations on objects within the system, which makes it a RBAC-like solution. SELinux uses security policies that are label-based, identifying applications through their file system labels. SELinux might be complex to configure and manage. For more information, see GitHub.
- *AppArmor* (Debian-based) employs security profiles that are path-based, identifying applications by their executable paths. AppArmor does not have a traditional RBAC approach, but can define profiles for applications, which can be seen as a form of access control. For more information, see AppArmor.

- *FreeIPA* helps provide a centrally managed Identity, Policy, and Audit (IPA) system. FreeIPA, which is the upstream open-source project for Red Hat Identity Management, is an integrated security information management solution combining Fedora Linux, 389 Directory Server, Kerberos, Network Time Protocol (NTP), DNS, and Dogtag (Certificate System). It provides centralized identity management and includes support for RBAC so that administrators can define roles and associate permissions and policies with these roles across a network of Linux systems. For more information, see FreeIPA.
- *RHEL System Roles* is a collection of Ansible roles and modules that provide a stable and consistent configuration interface to automate and manage multiple releases of RHEL. RHEL System Roles are provided in the following formats:

  – As an RPM package in the RHEL 9 or RHEL 8 Application Streams repositories
  – As a supported collection in the Red Hat Automation Hub

  For more information, see RHEL System Roles.

Each solution offers different features and complexities so that administrators can choose the most appropriate tool based on their specific security requirements and environment. Red Hat based distributions come preconfigured with several SELinux policies, but the configuration might be more complex than FreeIPA or AppArmour. RHEL roles are typically part of automation policies.

The sudo command

The widespread reliance on `sudo` in most Linux distributions is attributed to the granular control that it provides over user permissions. `sudo` simplifies the delegation of limited root access, specifies commands that are allowed through the `sudoers` file, and maintains an audit trail, which makes it highly practical for routine administrative tasks. Other tools involve complex management and a level of detail that is typically unnecessary for everyday operations, making them more suitable for specialized use cases.

The fundamental approach to configuring `sudo` is consistent across Debian-based, Red Hat-based, and SUSE-based distributions, but the specifics of default configurations and group usage vary:

- Debian-based: Focus on the `sudo` group. The root user is disabled by default (Ubuntu).
- Red Hat-based: Use the wheel group. The root user is enabled by default.
- SUSE-based: Flexible with `sudo` or users groups. The root user is enabled by default.

To implement group access control by using sudo, complete the following steps:

1. Determine the different roles in the organization and the specific permissions or commands that each role needs.
2. Create UNIX groups corresponding to each role. For example, admin, developer, auditor, and others. Example 6-12 shows adding groups.

Example 6-12 Creating groups for sudo

```
sudo groupadd admin
sudo groupadd developer
sudo groupadd auditor
```

3. Add users to the appropriate groups based on their roles, as shown in Example 6-13.

Example 6-13 Adding users to sudo

```
sudo usermod -aG admin alice
sudo usermod -aG developer bob
sudo usermod -aG auditor charlie
```

4. Edit the `sudoers` file to grant permissions to groups. To do this task, use the `visudo` command to ensure proper syntax and prevent mistakes:

   `sudo visudo`

   In the `sudoers` file, define the commands that each group can run. Example 6-14 shows the group permissions.

Example 6-14 Group permissions

```
    %admin ALL=(ALL) ALL
    # Admins can run any command
    %developer ALL=(ALL) /usr/bin/git, /usr/bin/make, /usr/bin/gcc
    # Developers can run git, make, gcc
   %auditor ALL=(ALL) /usr/bin/less, /bin/cat, /usr/bin/tail               #
Auditors can run less, cat, tail
```

You can now use the sudo command to run commands based on user roles, as shown in Example 6-15.

Example 6-15 Role definitions in sudo

```
sudo git commit -m "example commit"   # For a developer
sudo less /var/log/syslog             # For an auditor
sudo systemctl restart apache2        # For an admin
```

In this example, the admin role has full control over the system; the developer role grants access to development tools like `git`, `make`, and `gcc`; and the auditor role has read-only access to logs and configuration files.

For more information about `sudo`, see What is sudo?

6.4.4 Logging, audits, and file integrity monitoring

Access logging provides a record of user activities, which are crucial for auditing and identifying suspicious behavior. They offer insights into system activities and help ensure compliance with security policies. You can leverage several tools.

The commands and their typical use cases are as follows:

`rsyslog` Centralizes logs by using syslog to monitor and analyze security events.

`auditd` A powerful auditing tool that logs system calls and user activities.

`ausearch` A tool that can query the audit daemon logs for events based on different search criteria.

We show how to deploy and combine these tools in practical examples.

The rsyslog tool

syslog is a standard for message logging that allows separation of the software that generates messages from the system that stores them and the software that reports and analyzes them. rsyslog is an enhanced version of syslog. It builds on the foundation of syslog, providing advanced features and greater flexibility.

To use rsyslog, complete the following steps:

1. Install rsyslog (Debian-based) by running the following command:

   ```
   sudo apt-get install rsyslog
   ```

2. Edit /etc/rsyslog.conf to configure the log levels and destinations, as shown in Example 6-16.

Example 6-16 Configuring the log levels

```
# Log all user messages to /var/log/user.log user.
* /var/log/user.log
# Log all auth messages to /var/log/auth.log auth.
* /var/log/auth.log
```

3. Restart the syslog service by running the following command:

   ```
   sudo systemctl restart rsyslog
   ```

The auditd and ausearch tools

auditd is the user space component of the Linux Auditing System, which is used to collect, filter, and store audit records that are generated by the kernel. These records can include information about system calls, file accesses, user logins, and other security events. The audit daemon (auditd) is responsible for writing these records to disk and managing the log files.

To use auditd, complete the following steps:

1. Install auditd by running the following command:

   ```
   sudo apt-get install auditd audispd-plugin
   ```

2. Edit the audit.rules file by running the following command:

   ```
   vi /etc/audit/rules.d/audit.rules
   ```

3. Append the lines that are shown in Example 6-17.

Example 6-17 Lines to append to audit.rules

```
# Monitor changes to /etc/passwd
-w /etc/passwd -p wa -k passwd_changes

# Monitor changes to /etc/shadow
-w /etc/shadow -p wa -k shadow_changes

# Monitor use of privileged commands
-a always,exit -F arch=b64 -S execve -C uid!=euid -F euid=0 -k privileged
```

4. Restart auditd to make the changes take effect:

   ```
   sudo systemctl restart auditd
   ```

5. To validate that the changes took effect, change one password:

   ```
   #passwd hugo
   ```

6. Search the audit log for password changes:

   ```
   sudo ausearch -k passwd_changes
   ```

The results are shown in Example 6-18.

Example 6-18 Audit display of password changes

```
----
time->Wed Jul 17 18:35:16 2024
type=PROCTITLE msg=audit(1721237716.900:98064):
proctitle=2F7362696E2F617564697463746C002D52002F6574632F61756469742F61756469742E72756C6573
type=SYSCALL msg=audit(1721237716.900:98064): arch=c0000015 syscall=335 success=yes
exit=1084 a0=3 a1=7fffd8d50114 a2=43c a3=0 items=0 ppid=2962757 pid=2962773 auid=4294967295
uid=0 gid=0 euid=0 suid=0 fsuid=0 egid=0 sgid=0 fsgid=0 tty=(none) ses=4294967295
comm="auditctl" exe="/usr/sbin/auditctl" subj=unconfined key=(null)
type=CONFIG_CHANGE msg=audit(1721237716.900:98064): auid=4294967295 ses=4294967295
subj=unconfined op=add_rule key="passwd_changes" list=4 res=1
```

You can also use `ausearch` with `aureport` for detailed reports, as shown in Example 6-19. The command is as follows:

```
sudo aureport -k
```

Example 6-19 Audit report

```
Key Report
===============================================
# date time key success exe auid event
===============================================
17/07/24 18:35:16 passwd_changes yes /usr/sbin/auditctl -1 98064
17/07/24 18:35:16 shadow_changes yes /usr/sbin/auditctl -1 98065
17/07/24 18:35:16 privileged yes /usr/sbin/auditctl -1 98066
17/07/24 18:35:33 shadow_changes yes /usr/bin/passwd 1000 98076
```

6.4.5 File system security

With detailing and auditing, you can know who does what, when, and how. The challenge is determining that among all the millions of events that can happen in a system, one of them modifies a critical file, or even worse, thousands of them at the same time (ransomware attack).

AIDE helps to monitor and verify the integrity of files and directories on a system. It helps detect unauthorized changes, such as modifications, deletions, or additions, by creating a database of file attributes and comparing the current state to the baseline. It is a File Integrity Monitor that initially was developed as a no-cost and open source replacement for Tripwire that is licensed under the terms of the GNU General Public License.

To use AIDE, complete the following steps:

1. To install AIDE (Debian), run the following command:

   ```
   sudo apt-get install aide
   ```

2. To begin using AIDE, you must make sure that the database is present. To do so, run the following command:

   ```
   ls /var/lib/aide
   ```

3. If you see the file `aide.db` in the output of the `ls` command, then proceed to step 4. If you see the file `aide.db.new`, then you rename it to `aide.db` by running the following command:

   ```
   sudo mv /var/lib/aide/aide.db.new /var/lib/aide/aide.db
   ```

4. Once the AIDE database is in place, you can initialize the database by running the following command. This process takes a while.

   ```
   aide --config /etc/aide/aide.conf --init
   ```

5. To perform an initial check of the directories and files that are specified in `/etc/aide/aide.conf`, run the following command:

   ```
   sudo aide -check
   ```

   If everything in the monitored directories and files is correct, you will see the following message when the check completes:

   ```
   All files match the AIDE database. Looks okay!
   ```

AIDE also runs daily through the `/etc/cron.daily/aide` crontab, and the output is emailed to the user that is specified in the `MAILTO=` directive of the `/etc/default/aide` configuration file. Set up a cron job for regular checks by running the following command:

```
sudo crobtab -e 0 0 * * * /usr/bin/aide --check
```

AIDE can determine what changes were made to a system, but it cannot determine who made the change, when the change occurred, and what command was used to make the change. To discover this information, use `auditd` and `ausearch`.

By combining these tools, you establish a robust system for logging, integrity checking, and auditing. This multi-layered approach enhances the security and integrity of your Linux installation on the ppc64le architecture, providing early detection of potential security incidents and unauthorized changes.

Tip: Forwarding these events to a Security Information and Event Management (SIEM) or remote log solution (including PowerSC Trusted Logging on Virtual I/O Server (VIOS)) is a best practice to help ensure that these logs are tamper-proof and cannot be modified or deleted. This situation applies to any other log or audit file of security interest.

6.4.6 SIEM and endpoint detection and response integration

CIS recommends configuring Linux systems to send logs to a centralized log server. This approach enhances security by protecting logs from being tampered with on the local machine, and simplifies log management.

Integrating Linux on Power with SIEM tools such as IBM QRadar requires several steps to help ensure that logs from the Linux systems are collected, transmitted, and ingested by the SIEM platform. The same steps apply if you use a remote log collector or other observability tool to centralize the logs from different environments for their secure storage and analysis.

Figure 6-6 shows the concept.

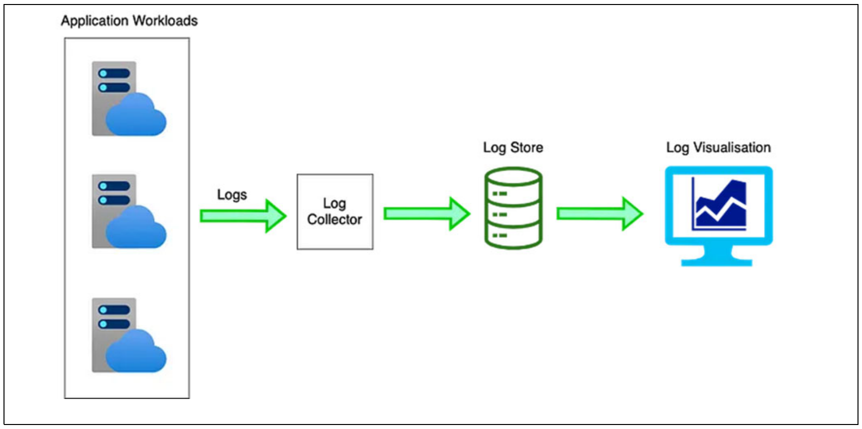


Figure 6-6 Log management

syslog-based approach

`syslog` is a protocol that was created in the 1980s. It remains the default on OpenBSD. There are two options that are available:

- `syslog-ng`

  `syslog-ng` was developed in the late 1990s as a robust replacement for `syslog`. It introduced support for TCP, encryption, and numerous other features. `syslog-ng` became the standard and was included in distributions such as SUSE, Debian, and Fedora for many years.

  Example 6-20 shows a configuration example to forward auth logs to a remote SIEM on IP 1.2.3.4 (Debian-based) by using `syslog-ng`.

Example 6-20 Configuration to forward logs to a remote SIEM

```
# Source configuration
source s_src {
    system();
    internal();
};

# Destination configuration for QRadar
destination d_remotesiem {
    udp("1.2.3.4" port(514));
};

# Filter for authentication logs
filter f_auth { facility(auth, authpriv); };

# Log path for authentication logs to QRadar
log {
    source(s_src);
    filter(f_auth);
    destination(d_remotesiem);
};
```

- `rsyslog`

  Launched in 2004 as a competitor to `syslog-ng`, `rsyslog` has become the default `syslog` daemon on Ubuntu, RHEL, and many other distributions. If you are using a common, up-to-date Linux distribution, you are likely using `rsyslog` by default.

  Example 6-21 shows forwarding the auth log by using `rsyslog` (RHEL-based).

Example 6-21 Forwarding the auth log

```
# Load necessary modules
module(load="imfile")  # File reader module
module(load="omfwd")   # Forwarding module

# Input configuration for auth logs
input(type="imfile" File="/var/log/secure" Tag="auth-log"
StateFile="state-auth-log")

# Forward auth logs to remote SIEM
if $syslogtag == 'auth-log' then {
    action(type="omfwd" target="1.2.3.4" port="514" protocol="udp"
template="RSYSLOG_ForwardFormat")
```

Field-based log approach (JSON)

The problem with `syslog` messages is that they must be parsed to extract every field by using regular expressions, which is time-consuming.

Another approach is to send JSON or field-based logs to IBM QRadar without using traditional `syslog` daemons or after storing messages in a database. You can accomplish this task by using a tool like Fluentd or even your own scripts in Python.

Fluentd is an extensively deployed open-source log collector that is written in Ruby. It stands out for its versatile pluggable architecture. This design enables it to seamlessly connect to a broad array of log sources and storage solutions, including Elasticsearch, Loki, `rsyslog`, MongoDB, Amazon Web Services (AWS) S3 object storage, and Apache Kafka, among others.

Figure 6-7 shows how Fluentd can help with log management.

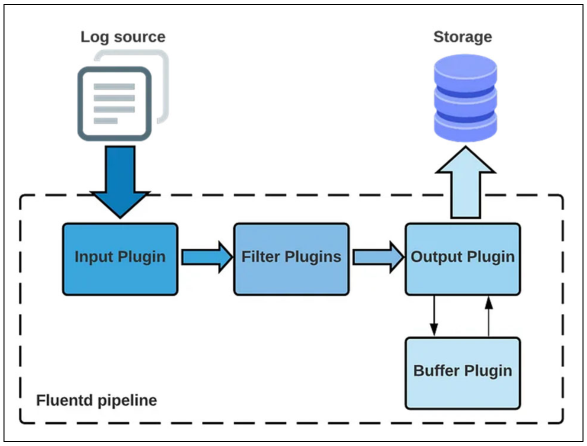


Figure 6-7 Fluentd used for log management

IBM leverages Fluentd to streamline its log management processes across diverse environments, including sending logs from Kubernetes based deployments on IBM Cloud or Red Hat OpenShift and deploying in Power by using a Docker image.

6.4.7 Malware protection

Malware, which encompasses any malicious software that harms or exploits programmable devices, services, or networks, includes ransomware, which is a specific type of malware that encrypts victims' files and demands a ransom for their release, and viruses, which attach to programs or files, spreading from one computer to another one.

To prevent these threats from infecting Linux systems, you can use Comprehensive Malware Detection and Removal (ClamAV) to detect and remove various forms of malware through regular scans and real-time protection, and `chkrootkit` can identify and report signs of rootkits. Both tools enhance security by helping to ensure that the system remains free from unauthorized access and malicious activity.

Both tools are open source and available for the ppc64le architecture.

Virus detection

There are a couple of options for virus detection on IBM Power.

ClamAV

ClamAV is a versatile and powerful open-source antivirus engine that is designed to detect trojans, viruses, malware, and other malicious threats. It offers several features that make it a valuable tool for enhancing Linux system security:

- Regular scans: You can configure ClamAV to perform regular scans of the system, which helps ensure that any new or existing malware is detected and addressed.
- Real-time protection: With the ClamAV daemon, you can enable real-time scanning to monitor file activity continuously, providing immediate detection and response to potential threats.
- Automatic updates: ClamAV includes an automatic update mechanism for its virus definitions, which helps ensure that the system is protected against the latest threats.
- Cross-platform support: ClamAV supports multiple platforms, making it a flexible solution for various environments working on Linux on Power but also in AIX and IBM i (IBM Portable Application Solutions Environment for i (PASE for i)).

To install and configure ClamAV on a Debian-based Linux system, complete the following steps:

1. Install ClamAV by running the following command:

   ```
   sudo apt-get install clamav clamav-daemon
   ```

2. Update the ClamAV database by running the following command:

   ```
   sudo freshclam
   ```

3. Start the ClamAV daemon by running the following command:

   ```
   sudo systemctl start clamav-daemon
   ```

4. Schedule a daily scan (add this line to your crontab) and send a report by email by running the following command:

   ```
   MAILTO=admin@acme.com 0 1 * * * /usr/bin/clamscan -ri --no- summary /
   ```

Tip: ClamAV is also available for AIX and IBM i.

Powertech Antivirus

Offered by Fortra (previously Help Systems), Powertech Antivirus is a commercially available antivirus solution providing native scanning for IBM systems, including IBM i, AIX, and Linux on Power (it also supports Linux on IBM Z, LinuxONE, and Linux on x86).

Powertech Antivirus offers both on-demand and scheduled scanning, enabling you to balance security and system performance. Compatible with IBM, Fortra, and third-party scheduling solutions, you can customize the scan frequency and target directories. Powertech Antivirus can be run independently on each endpoint or it can be centrally managed.

Here are some key features:

- Native scanning for IBM i, AIX, Linux x86, LinuxONE, Linux on IBM Z, PowerLinux, and Solaris systems
- Advanced detection, cleaning, and quarantining
- Object integrity scanning
- Advanced heuristic analysis
- Scalable multi-threaded daemon
- On-demand or scheduled scans
- Scanning SMTP mail sent from the server
- Partition scanning
- Automatic DAT file downloads
- Alert enabled
- Syslog integration

For more information, see Fortra Powertech Antivirus.

Rootkit detection

The tool `chkrootkit` is a rootkit detector that checks for signs of rootkits on UNIX-based systems. It scans for common signatures of known rootkits and helps ensure that the system remains uncompromised.

You can scan for many types of rootkits and detect certain log deletions by using `chkrootkit`. Although it does not remove any infected files, it does specifically tell you which ones are infected so that you can remove, reinstall, or repair the file or package.

To install `chkrootkit`, run the following command:

```
sudo apt-get install chkrootkit
```

To run `chkrootkit`, run the following command:

```
sudo chkrootkit
```

You can also schedule a daily scan through `cron`.

6.4.8 Backup strategy

Regular backups are essential for data recovery if there is an attack or system failure. Implementing automated backup solutions is a must. There are open-source projects with support for ppc64le like Bacula and enterprise-grade solutions such as IBM Storage Protect and others.

Bacula is an enterprise-grade, open-source backup solution that automates the process of backing up, recovering, and verifying data across a network of computers. It is highly flexible and scalable, making it suitable for various environments, from small businesses to large enterprises. Bacula is known for its robustness, extensive feature set, and support for multiple operating systems, including Linux on ppc64le (Red Hat Fedora)

IBM Storage Protect is a comprehensive, enterprise-grade data protection solution that can safeguard critical data across diverse environments. Designed to automate the processes of data backup, recovery, and archiving, IBM Storage Protect helps ensure that business-critical information remains secure, available, and verifiable. Its robust architecture and extensive feature set make it an ideal choice for organizations of all sizes, from small businesses to large enterprises. Both the clients and the server itself are supported on Linux on the ppc64le architecture. For more information, see the IBM Storage Protect documentation.

6.4.9 Consistent update strategy

Updating and maintaining Linux environments is crucial for helping ensure security, performance, and compliance. For enterprises that use IBM Power (ppc64le) systems, integrating Foreman and Katello and their supported products, such as Red Hat Satellite™ and Orcahino, offers a robust solution for consistent and automated updates. To date, server packages are not available on the ppc64le architecture, but client packages are.

Using Foreman and Katello together with Red Hat Satellite provides one option for update management.

Foreman is an open-source lifecycle management tool that system administrators can use to manage servers throughout their lifecycle, from provisioning and configuration to monitoring and management. It provides an integrated, comprehensive solution for managing large-scale infrastructures.

Katello is a plug-in for Foreman that adds content management and subscription management capabilities. Administrators can use to manage software repositories, handle updates, and ensure compliance with subscription policies.

Figure 6-8 shows how Red Hat Satellite works.

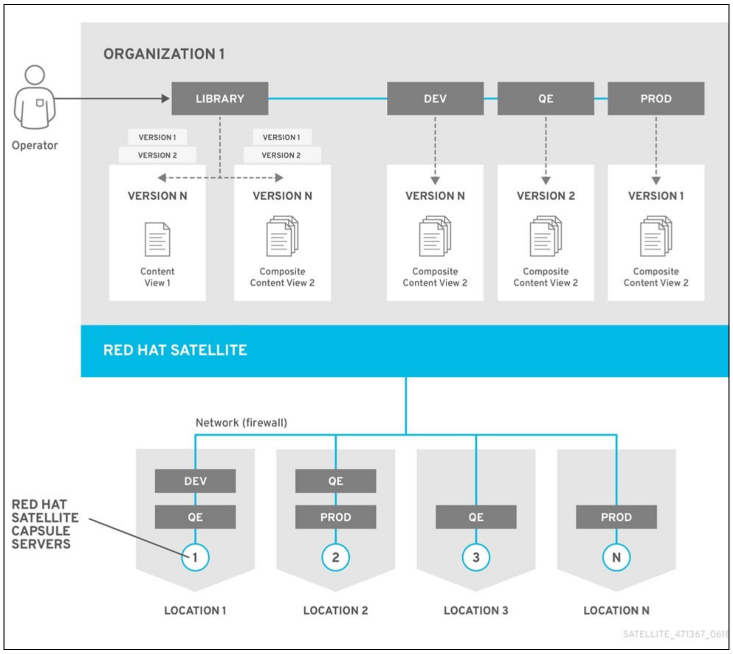


Figure 6-8 Red Hat Satellite

Figure 6-9 shows the Red Hat Satellite GUI.

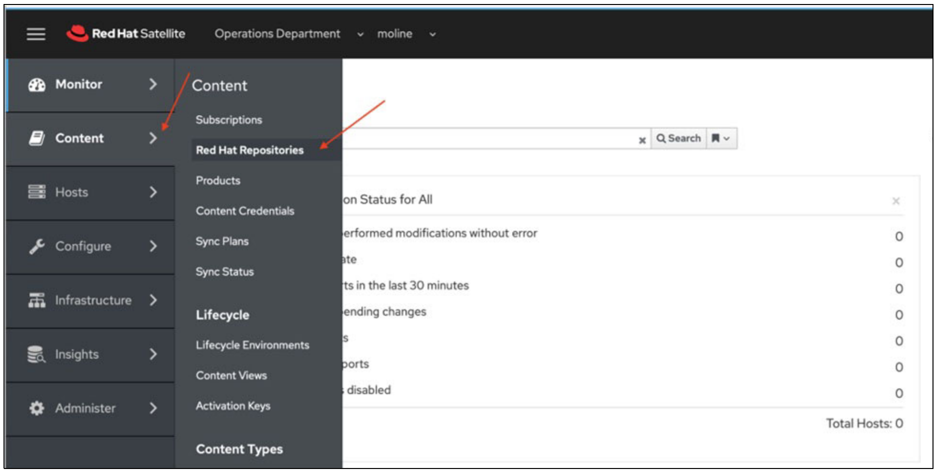


Figure 6-9 Red Hat Satellite GUI

Using Red Hat Satellite along with Foreman and Katello manages package and patch lifecycles, including update distribution. Using Red Hat Satellite in this environment can initiate updates. However, Ansible offers a more comprehensive automation solution for keeping systems up to date:

- Performs prechecks, backups, and snapshots.
- Initiates patch updates.
- Restarts systems.
- Conducts post-checks for complete patch automation.

Combining Satellite and Ansible is optimal. Satellite handles lifecycle management and package provisioning, and Ansible automates the entire patching process. This integration helps ensure efficient and consistent updates across your environment.

6.4.10 Monitoring

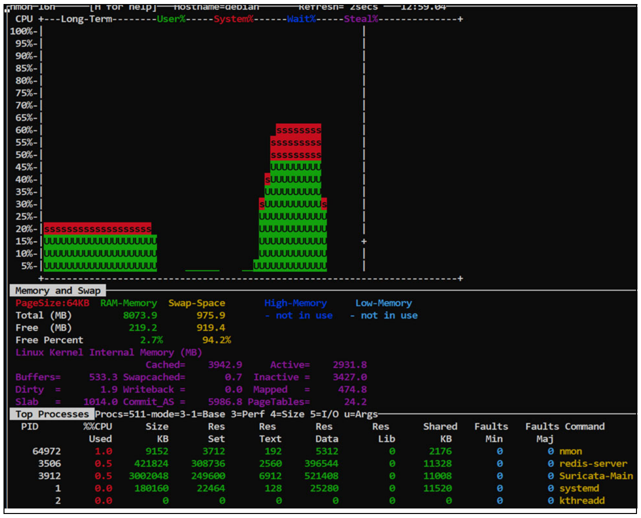
Monitoring Linux on Power servers is vital to help ensure their security, supplementing the specialized tools that are described in this chapter and providing more insights.

Specific examples include detecting abnormal CPU or network consumption by unfamiliar applications, which might indicate underlying issues. Proper sizing of workloads and appropriate distribution of system resources contribute to their accessibility.

There are many options for monitoring Linux on Power servers. Most commercial and community solutions have ppc64le agents. For example, consider Pandora FMS. There are also solutions that support monitoring Linux on Power and also fit into a complete monitoring infrastructure across all your IBM Power workloads running on Linux, AIX, and IBM i, where you can visualize the status any partition, and generate alerts that can be redirected to a centralized monitoring environment.

One of the simplest options for this multiple architecture monitoring tool is nmon. `nmon` was originally written for AIX, and is now an integrated tool within AIX. A version for Linux was written by IBM and later released as open source for Linux across multiple platforms, including x86, IBM Power, IBM Z, and even ARM. There are multiple integrations for using and analyzing `nmon` data, including charts and spreadsheet integrations. There is even a newer version (`nmonj`) that saves the performance data in JSON format for better integration into databases and for web browser graphing.

Figure 6-10 show `nmon` reporting on utilization.

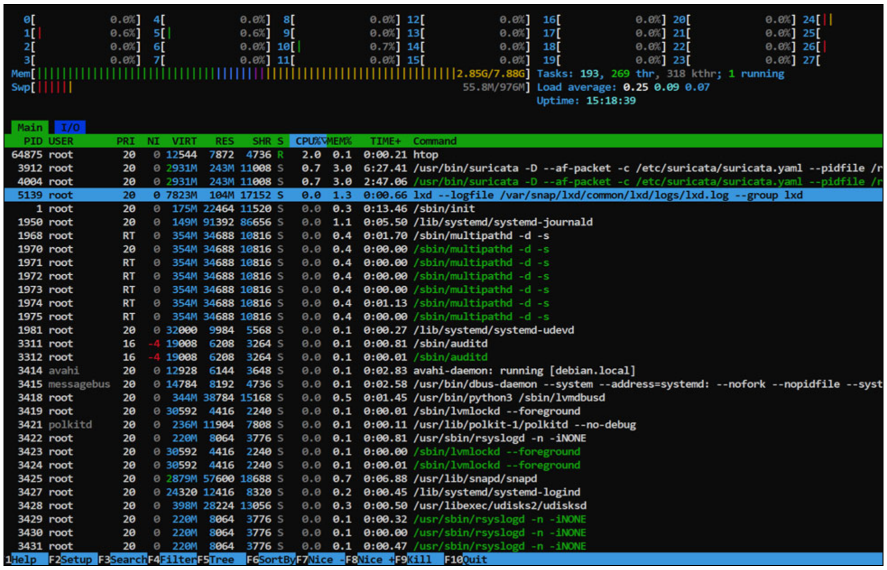


```
nmon-16h        [H for help]   Hostname=debian         Refresh= 2secs   12:59.04
 CPU +---Long-Term--------User%-----System%------Wait%-----Steal%-------------+
100%-|
 95%-|
 90%-|
 85%-|
 80%-|
 75%-|
 70%-|
 65%-|
 60%-|
 55%-|
 50%-|
 45%-|
 40%-|
 35%-|
 30%-|
 25%-|
 20%-|
 15%-|
 10%-|
  5%-|
     +---------------------------------------------------------------------+
 Memory and Swap
 PageSize:64KB  RAM-Memory  Swap-Space       High-Memory     Low-Memory
 Total (MB)         8073.9       975.9       - not in use   - not in use
 Free  (MB)          219.2       919.4
 Free Percent          2.7%       94.2%
 Linux Kernel Internal Memory (MB)
                       Cached=     3942.9     Active=     2931.8
 Buffers=     533.3 Swapcached=       0.7  Inactive =    3427.0
 Dirty  =       1.9 Writeback =       0.0  Mapped   =     474.8
 Slab   =    1014.0 Commit_AS =    5986.8 PageTables=      24.2
 Top Processes Procs=511-mode=3-1=Base 3=Perf 4=Size 5=I/O u=Args
  PID       %%CPU    Size      Res      Res      Res      Res    Shared   Faults   Faults Command
             Used      KB      Set     Text     Data      Lib       KB      Min      Maj
   64972      1.0    9152     3712      192     5312        0     2176        0        0 nmon
    3506      0.5  421824   308736     2560   396544        0    11328        0        0 redis-server
    3912      0.5 3002048   249600     6912   521408        0    11008        0        0 Suricata-Main
       1      0.0  180160    22464      128    25280        0    11520        0        0 systemd
       2      0.0       0        0        0        0        0        0        0        0 kthreadd
```

Figure 6-10   A nmon display

Another tool is htop, an interactive process viewer that offers several enhanced functions that make it simple and versatile. For example, users may scroll through and select processes for detailed information, change priorities, and terminate processes directly from the interface.

Figure 6-11 shows an example panel from `htop`.



```
    0[           0.0%]   4[           0.0%]   8[           0.0%]  12[           0.0%]  16[           0.0%]  20[           0.0%]  24[||
    1[|          0.6%]   5[|          0.6%]   9[           0.0%]  13[           0.0%]  17[           0.0%]  21[           0.0%]  25[
    2[           0.0%]   6[           0.0%]  10[|          0.7%]  14[           0.0%]  18[           0.0%]  22[           0.0%]  26[|
    3[           0.0%]   7[           0.0%]  11[           0.0%]  15[           0.0%]  19[           0.0%]  23[           0.0%]  27[
  Mem[||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||2.85G/7.88G] Tasks: 193, 269 thr, 318 kthr; 1 running
  Swp[||||||                                                                  55.8M/976M] Load average: 0.25 0.09 0.07
                                                                                          Uptime: 15:18:39

  Main  I/O
    PID USER       PRI  NI  VIRT   RES   SHR S  CPU%▽MEM%   TIME+  Command
  64875 root        20   0 12544  7872  4736 R   2.0  0.1  0:00.21 htop
   3912 root        20   0 2931M  243M 11008 S   0.7  3.0  6:27.41 /usr/bin/suricata -D --af-packet -c /etc/suricata/suricata.yaml --pidfile /r
   4004 root        20   0 2931M  243M 11008 S   0.7  3.0  2:47.06 /usr/bin/suricata -D --af-packet -c /etc/suricata/suricata.yaml --pidfile /r
   5139 root        20   0 7823M  104M 17152 S   0.0  1.3  0:00.66 lxd --logfile /var/snap/lxd/common/lxd/logs/lxd.log --group lxd
      1 root        20   0  175M 22464 11520 S   0.0  0.3  0:13.46 /sbin/init
   1950 root        20   0  149M 91392 86656 S   0.0  1.1  0:05.50 /lib/systemd/systemd-journald
   1968 root        RT   0  354M 34688 10816 S   0.0  0.4  0:01.70 /sbin/multipathd -d -s
   1970 root        20   0  354M 34688 10816 S   0.0  0.4  0:00.00 /sbin/multipathd -d -s
   1971 root        RT   0  354M 34688 10816 S   0.0  0.4  0:00.00 /sbin/multipathd -d -s
   1972 root        RT   0  354M 34688 10816 S   0.0  0.4  0:00.00 /sbin/multipathd -d -s
   1973 root        RT   0  354M 34688 10816 S   0.0  0.4  0:00.00 /sbin/multipathd -d -s
   1974 root        RT   0  354M 34688 10816 S   0.0  0.4  0:01.13 /sbin/multipathd -d -s
   1975 root        RT   0  354M 34688 10816 S   0.0  0.4  0:00.00 /sbin/multipathd -d -s
   1981 root        20   0 32000  9984  5568 S   0.0  0.1  0:00.27 /lib/systemd/systemd-udevd
   3311 root        16  -4 19008  6208  3264 S   0.0  0.1  0:00.81 /sbin/auditd
   3312 root        16  -4 19008  6208  3264 S   0.0  0.1  0:00.01 /sbin/auditd
   3414 avahi       20   0 12928  6144  3648 S   0.0  0.1  0:02.83 avahi-daemon: running [debian.local]
   3415 messagebus  20   0 14784  8192  4736 S   0.0  0.1  0:02.58 /usr/bin/dbus-daemon --system --address=systemd: --nofork --nopidfile --syst
   3418 root        20   0  344M 38784 15168 S   0.0  0.5  0:01.45 /usr/bin/python3 /sbin/lvmdbusd
   3419 root        20   0 30592  4416  2240 S   0.0  0.1  0:00.01 /sbin/lvmlockd --foreground
   3421 polkitd     20   0  236M 11904  7808 S   0.0  0.1  0:00.11 /usr/lib/polkit-1/polkitd --no-debug
   3422 root        20   0  220M  8064  3776 S   0.0  0.1  0:00.81 /usr/sbin/rsyslogd -n -iNONE
   3423 root        20   0 30592  4416  2240 S   0.0  0.1  0:00.00 /sbin/lvmlockd --foreground
   3424 root        20   0 30592  4416  2240 S   0.0  0.1  0:00.01 /sbin/lvmlockd --foreground
   3425 root        20   0 2879M 57600 18688 S   0.0  0.7  0:06.88 /usr/lib/snapd/snapd
   3427 root        20   0 24320 12416  8320 S   0.0  0.2  0:00.45 /lib/systemd/systemd-logind
   3428 root        20   0  398M 28224 13056 S   0.0  0.3  0:00.50 /usr/libexec/udisks2/udisksd
   3429 root        20   0  220M  8064  3776 S   0.0  0.1  0:00.32 /usr/sbin/rsyslogd -n -iNONE
   3430 root        20   0  220M  8064  3776 S   0.0  0.1  0:00.00 /usr/sbin/rsyslogd -n -iNONE
   3431 root        20   0  220M  8064  3776 S   0.0  0.1  0:00.47 /usr/sbin/rsyslogd -n -iNONE
F1Help  F2Setup F3SearchF4FilterF5Tree  F6SortByF7Nice -F8Nice +F9Kill  F10Quit
```

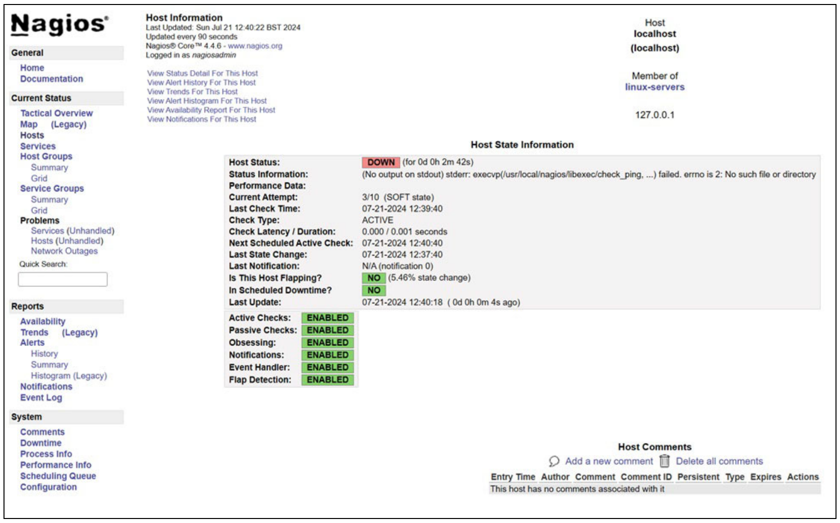
Figure 6-11 Screen capture of htop

There are more projects being developed by using Python and other languages that are portable between architectures. Some of them have good export capabilities to InfluxDB, Cassandra, OpenTSDB, StatsD, Elasticsearch, or RabbitMQ.

Nagios

The next level is deploying complete monitoring environments, such as Nagios or Zabbix. These frameworks support extensive customization and scalability. Their source code can be downloaded and compiled on IBM Power, with agents and integrations for both AIX and IBM i. Some of them have enterprise support options or require licenses from some instances.

Figure 6-12 is an example of Nagios (core version) running on Debian on Power.



Host Information
Last Updated: Sun Jul 21 12:40:22 BST 2024
Updated every 90 seconds
Nagios® Core™ 4.4.6 - www.nagios.org
Logged in as *nagiosadmin*

Host
localhost
(localhost)

Member of
linux-servers

127.0.0.1

Host State Information

Host Status:	DOWN (for 0d 0h 2m 42s)
Status Information:	(No output on stdout) stderr: execvp(/usr/local/nagios/libexec/check_ping, ...) failed. errno is 2: No such file or directory
Performance Data:	
Current Attempt:	3/10 (SOFT state)
Last Check Time:	07-21-2024 12:39:40
Check Type:	ACTIVE
Check Latency / Duration:	0.000 / 0.001 seconds
Next Scheduled Active Check:	07-21-2024 12:40:40
Last State Change:	07-21-2024 12:37:40
Last Notification:	N/A (notification 0)
Is This Host Flapping?	NO (5.46% state change)
In Scheduled Downtime?	NO
Last Update:	07-21-2024 12:40:18 (0d 0h 0m 4s ago)

Active Checks:	ENABLED
Passive Checks:	ENABLED
Obsessing:	ENABLED
Notifications:	ENABLED
Event Handler:	ENABLED
Flap Detection:	ENABLED

Host Comments

Entry Time	Author	Comment	Comment ID	Persistent	Type	Expires	Actions
This host has no comments associated with it							

Figure 6-12 Nagios running on Linux on Power

IBM Instana

In the field of commercial monitoring solutions, we highlight IBM Instana™®. It leverages various open-source projects to provide advanced monitoring and observability capabilities, making it an excellent enterprise-supported solution for monitoring Linux on Power (ppc64le), AIX, and IBM i.

IBM Instana integrates with technologies such as Apache Kafka for real-time data processing; Prometheus for metrics collection; Grafana for data visualization; OpenTelemetry for tracing and metrics; Elasticsearch, Logstash, and Kibana (ELK) for log management; Kubernetes for container orchestration; and Jenkins for continuous integration and delivery.

With support for Debian, Red Hat, and SUSE on ppc64le, Instana helps ensure comprehensive, real-time visibility into the performance and health of applications and systems that is backed by IBM's robust enterprise support.

For more information, see IBM Instana.

6.5 Best practices

To help ensure the security and reliability of Linux on Power (ppc64le), several best practices should be implemented. These practices include system hardening, regular updates, access control, and data protection strategies.

System hardening

Implement the following best practices:

- Minimal installation: Begin with a minimal base installation to reduce the attack surface. Install only the necessary software and services, which reduce the surface attack by limiting the installed software and services. It is simpler to add software than to remove it.
- Compliance: Use tools such as OpenSCAP or PowerSC to help ensure minimum levels of compliance in all systems. This task can be accomplished by generating a base image and being rigorous with change control by using tools such as Ansible for configuration management. This approach enforces consistent security policies across all systems.

- Patch management: Regularly apply security patches and updates. Use automated tools like Red Hat Satellite to keep the system current and consistent. Automation tools help you do the work.
- File system security: Implement secure file system permissions and use encryption for sensitive data. Regularly audit file permissions and access controls. Use tools like AIDE to monitor and verify the integrity of files and directories to protect sensitive data.

Access control

Implement the following best practices:

- User authentication: Implement strong authentication mechanisms, including MFA. Use SSH key pairs instead of passwords for remote access, with a second method of authentication.
- RBAC: Assign permissions based on roles rather than individual users. `sudo` is a powerful tool to do it locally.
- Password policies: Enforce strong password policies, including complexity requirements, expiration, and account lockout mechanisms.

Data protection

Implement the following best practices:

- Encryption: Use encryption for data at rest and in transit. Implement SSL/TLS for network communications and encrypt sensitive files on disk.
- Backup strategies: Regularly back up critical data and test restore procedures. Use tools like Bacula or IBM Storage Protect for automated backups.

Regular monitoring and logging

Implement the following best practices:

- Effective monitoring and logging are essential for detecting and responding to security analysis and compliance requirements. Search and analytics engines like Elasticsearch are a great choice for different use cases.
- SIEM integration: Integrate logs with SIEM systems for real-time analysis and alerting.
- Intrusion detection and prevention: Deploy IDS and IPS tools like Suricata to monitor network traffic and detect suspicious activities.
- Regular audits: Perform regular audits of log files and security configurations to identify and address potential issues. Use tools like `auditd` and `ausearch` for detailed audit logs. They can also be forwarded and analyzed by external tools from traditional SIEMs to Elasticsearch, Logstash, and Kibana (ELK) stacks.

Summary

In summary, layered security provides enhanced safety. However, even the most robust defenses have weaknesses, which make their effectiveness dependent on the least secure component. Achieving the right balance between security and usability is essential. Although technologies advance and operating systems change, core problems remain and new ones emerge.

6.6 Developing an incident response plan

An incident response plan (IRP) is a comprehensive, systematic approach to handling and managing security breaches or cyberattacks. It outlines the procedures and actions that an organization must take to detect, respond to, and recover from security incidents to minimize the impact on business operations, data integrity, and overall security posture.

A well-defined IRP is crucial for minimizing the impact of security incidents. Here are the key components of an effective IRP:

- Preparation: Define roles and responsibilities for an incident response that are specific to Linux on Power environments. Ensure that all team members are trained and familiar with the response procedures for this architecture. Make sure that you have a clearly defined architecture where the people who specialize in each technology (PowerVM, SUSE, Red Hat, Ubuntu, databases, applications, storage, and communications) understand the Linux on Power environment.
- Identification: Implement monitoring and alerting mechanisms to quickly identify potential security incidents in the Power environment. Use log analysis and SIEM tools that are tailored for the ppc64le architecture to detect anomalies. Ensure compatibility and optimization of these tools for the Power architecture.
- Containment: Develop strategies for containing incidents to prevent further damage. This component might involve isolating affected Power servers or networks. Consider the specific containment techniques that are suitable for Power hardware, such as leveraging virtualization features to isolate affected LPARs, VLANs, or shared storage.
- Eradication: Identify and remove the root cause of the incident in the Power environment. This component might involve applying patches, removing malware, or addressing configuration issues that are specific to ppc64le systems. Ensure that the incident response team is familiar with patch management and malware removal tools that are compatible with Linux on Power.
- Recovery: Restore affected Power servers to normal operation. This component might involve restoring data from backups, rebuilding compromised LPARs, or reconfiguring network settings that are specific to the Power architecture. Ensure that recovery procedures are tested and validated for ppc64le environments.
- Lessons learned: After resolving an incident, conduct a postmortem analysis to identify lessons that are learned and improve future response efforts. Update IRPs and security policies based on findings, considering any unique aspects of the Power environment. Document any architecture-specific issues and resolutions to enhance future readiness.

Tip: Regularly test and update your IRP to ensure that it remains effective and relevant to the threat landscape.

7

Red Hat OpenShift Security

This chapter describes the Red Hat OpenShift platform by providing a basic overview of its history, key features, the role that it plays in the modern IT landscape, and how it compares to similar platforms. This chapter also describes Kubernetes, which is the foundation of Red Hat OpenShift, and how Red Hat OpenShift improves on the fundamental features of Kubernetes. You learn about the advantages of using Red Hat OpenShift for orchestrating containers in your environment. The chapter concludes by providing a brief description of Red Hat OpenShift when it is implemented on an IBM Power infrastructure.

This chapter describes the following topics:

7.1 Red Hat OpenShift fundamentals

This section introduces Red Hat OpenShift, which is a leading enterprise cloud platform that you can use to design and build applications to run in a hybrid cloud environment. Red Hat OpenShift can run on your on-premises infrastructure, across many cloud vendors, and in your edge environments, enabling seamless migration of services across the hybrid cloud environment.

7.1.1 What is Red Hat OpenShift

Red Hat OpenShift is a leading, enterprise Kubernetes platform that provides a robust foundation for developing, deploying, and scaling cloud-native applications. It extends Kubernetes with additional features and tools to enhance productivity and security, making it an ideal choice for businesses looking to leverage container technology at scale.

Red Hat OpenShift is a unified platform to build, modernize, and deploy applications at scale. Work smarter and faster with a complete set of services for bringing apps to market on your choice of infrastructure. Red Hat OpenShift delivers a consistent experience across public cloud, on-premises, hybrid cloud, or edge architectures.

Red Hat OpenShift offers you a unified, flexible platform to address several business needs, such as an enterprise-ready Kubernetes orchestrator to a comprehensive cloud-native application development platform that can be self-managed or used as a fully managed cloud service.

Figure 7-1 shows how Kubernetes is only one component (albeit a critical one) in Red Hat OpenShift.

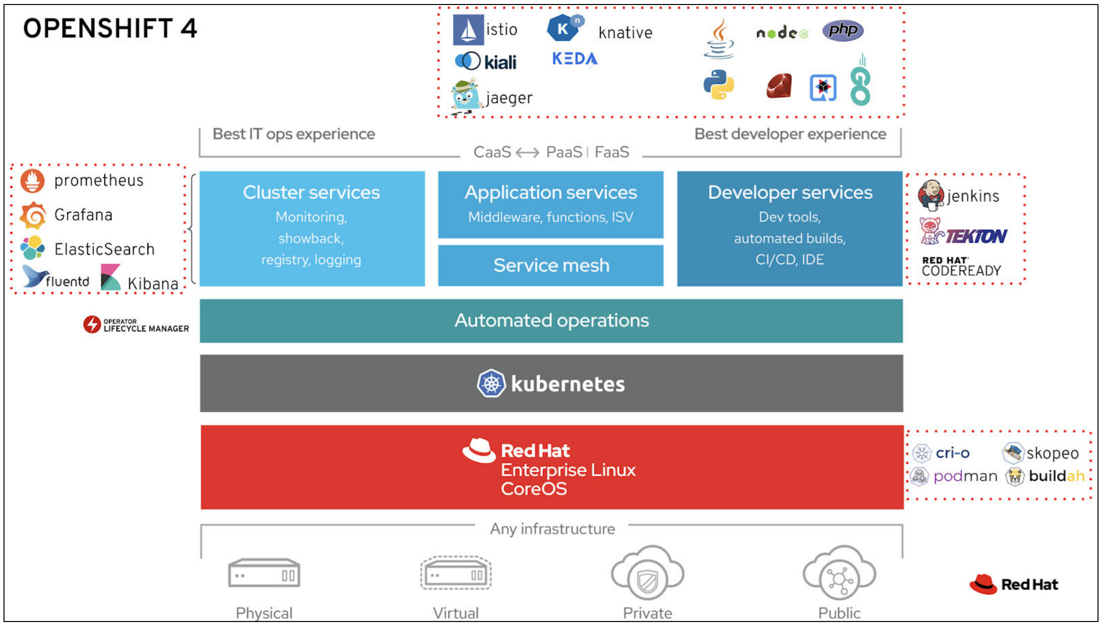


Figure 7-1 Red Hat OpenShift components

Red Hat OpenShift provides the following features:

- The ability to deploy and run in any environment, and the flexibility to build new applications, modernize existing applications, run third-party ISV applications, or use public cloud services under a single platform.
- The tools that are necessary to help customers integrate data analytics, and artificial intelligence (AI) and machine learning (ML) capabilities into cloud-native applications to deliver more insight and value.
- Consistency and portability to deploy and manage containerized workloads; make infrastructure and investments future-ready; and deliver speed and flexibility on-premises, across cloud environments, and to the edge of the network.
- Advanced security and compliance capabilities, enabling end-to-end management and observability across the entire architecture.

Built by open source leaders, Red Hat OpenShift includes an enterprise-ready Kubernetes solution with a choice of deployment and usage options to meet the needs of your organization. From self-managed to fully managed cloud services, you can deploy the platform in the data center, in cloud environments, and at the edge of the network. With Red Hat OpenShift, you can get advanced security and compliance capabilities, end-to-end management and observability, and cluster data management and cloud-native data services. Red Hat Advanced Cluster Security for Kubernetes modernizes container and Kubernetes security, enabling developers to add security controls early in the software lifecycle. With Red Hat Advanced Cluster Management for Kubernetes, you can manage your entire application lifecycle and deploy applications on specific clusters based on labels. Red Hat OpenShift Data Foundation supports performance at scale for data-intensive workloads.

7.1.2 Distinguishing features

There are many cloud and container management options that are available. Red Hat OpenShift has integrated these features to enhance the cloud and container management experience:

- With a focus on developer-centric tools, Red Hat OpenShift enhances Kubernetes with developer-friendly tools, including Red Hat OpenShift Console (a developer-centric view for application management) and Source-to-Image (S2I) technology. These tools simplify the process of building reproducible container images from source code.
- Designed with advanced security in mind, built-in security at every layer of the application stack (from the operating system (Red Hat Enterprise Linux CoreOS) to the application services) helps ensure that compliance features and security best practices are built in from the start.
- Focused on hybrid cloud capabilities, Red Hat OpenShift is designed to operate across on-premises, public cloud, and hybrid cloud environments, which provide consistent application portability and flexibility in deployment options.

Red Hat OpenShift is an enterprise-level production product that provides enterprise-level support that is based on Kubernetes and Kubernetes management. Red Hat OpenShift provides the following benefits:

- Red Hat OpenShift offers automated installation, upgrades, and lifecycle management throughout the container stack (the operating system, Kubernetes, cluster services, and applications) on any cloud.
- Red Hat OpenShift helps teams build with speed, agility, confidence, and choice. Get back to doing work that matters.
- Red Hat OpenShift is focused on security at every level of the container stack and throughout the application lifecycle. It includes long-term, enterprise support from one of the leading Kubernetes contributors and open-source software companies.

7.1.3 The role of Red Hat OpenShift in modern IT

Red Hat OpenShift plays a crucial role in modern IT by facilitating the DevOps approach, improving software delivery speed, and enabling a more agile development environment. It offers a scalable platform that supports both microservices and traditional application models, accommodating a wide range of programming languages and frameworks. Through its comprehensive tool set, Red Hat OpenShift addresses the needs of developers, system administrators, and IT managers, making it a pivotal tool in enterprise digital transformation strategies.

7.1.4 Key takeaways

When considering which cloud management platform to use, consider the following takeaways:

- Red Hat OpenShift provides extensive enterprise features, including advanced security features, integrated developer tools, and extensive automation capabilities that might not be as comprehensive in other Kubernetes services.
- The Red Hat OpenShift ability to deploy across multiple environments (cloud, on-premises, and hybrid) with consistency makes it a strong choice for organizations with complex infrastructure needs.
- Red Hat OpenShift distinctly benefits developers with features like S2I, a comprehensive web console, and application templates, which facilitate a smoother and more productive development experience compared to basic Kubernetes services.

Red Hat OpenShift is a strong leader in the cloud landscape of Kubernetes platforms, and is chosen for its strengths in enterprise environments, multi-environment consistency, and developer-centric features.

7.1.5 Kubernetes fundamentals

Kubernetes serves as the backbone of Red Hat OpenShift, providing the essential framework for orchestrating containerized applications. Understanding these core concepts is crucial for leveraging Red Hat OpenShift effectively. This section provides a detailed exploration of the fundamental components and mechanisms of Kubernetes as implemented in Red Hat OpenShift.

Basic components

The basic components of Kubernetes can be described as follows:

Pods The smallest deployable units that are created and managed by Kubernetes. A pod is a group of one or more containers that share storage, network, and specifications about how to run the containers. Pods are ephemeral by nature; they are created and destroyed to match the state that is specified by users.

Nodes The physical or virtual machines (VMs) where Kubernetes runs the pods. A node can be a worker node or a master node, although with the latest Kubernetes (and by extension Red Hat OpenShift) practices, the distinction is often abstracted away, especially in managed environments.

Clusters A cluster consists of at least one worker node and at least one master node. The master node manages the state of the cluster, including scheduling workloads, and handling scaling and health monitoring.

Figure 7-2 shows a basic cluster architecture. Although a cluster can technically be created with one master node and two worker nodes, a best practice is to use at least three master nodes, which can share functions and provide failover, and three or more worker nodes to provide failover and scalability.

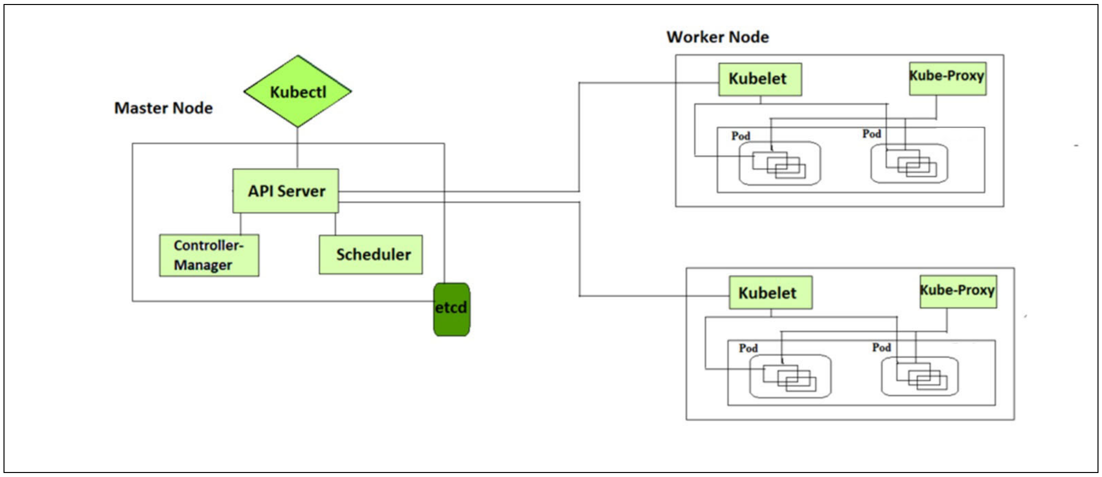


Figure 7-2 Base Kubernetes cluster

Control plane components

There are two distinct types of nodes in a Kubernetes cluster: a master node and a worker node. The worker nodes run containers with the applications that run your business. The master nodes run the services that control the cluster and manage the pods that are running the application code. Collectively, the master nodes create what is called the control plane.

The major services that are running in the control plane are as follows:

Apiserver Acts as the front end for Kubernetes. The application programming interface (API) server is the component that clients and external tools interact with.

etcd A highly available key-value store that is used as Kubernetes' backing store for all cluster data. It maintains the state of the cluster.

Scheduler Watches for newly created pods with no assigned node, and selects a node for them to run on based on resource availability, policies, and specifications.

Controller Manager Runs controller processes, which are background tasks in Kubernetes that handle routine tasks, such as helping to ensure the correct number of pods for replicated applications.

Workload resources

The control plane is in charge of setting up and managing the worker nodes that are running the application code. Workload components can be described as follows:

Deployments A deployment specifies a state for a group of pods. You describe a state in a deployment, and the Deployment Controller changes the actual state to the wanted state at a controlled rate. You can define deployments to create ReplicaSets or to remove existing deployments and adopt all their resources into new deployments.

ReplicaSet A ReplicaSet maintains a stable set of replica pods running at any given time. As such, it is often used to help ensure the availability of a specified number of identical pods. This approach maintains a stable set of replica pods running at any given time.

StatefulSets Used for applications that require persistent storage and a unique identity for each pod, making them ideal for databases and other stateful applications.

DaemonSets Helps ensure that each node in the cluster runs a copy of a pod. Useful for deploying system services that need to run on all or certain nodes.

Networking

Networking connectivity between pods, and between pods and outside services, is managed within a Kubernetes cluster. The following functions are maintained by the cluster:

Service An abstraction that defines a logical set of pods and a policy by which to access them. Services enable communication between different pods and external traffic routing into the cluster.

Ingress Manages external access to the services in a cluster, typically HTTP. Ingress can provide load-balancing, Secure Sockets Layer (SSL) termination, and name-based virtual hosting.

Storage

Containers are by definition ethereal, as is any data that is stored in the container. To enable persistent storage, Kubernetes uses the following concepts:

Persistent volumes Persistent volumes are resources in the cluster that can be connected to containers to provide persistent storage.

Persistent volume claims (PVCs) PVCs are requests for storage by users. These requests are satisfied by allocating persistent volumes.

Configurations and secrets

The following components relate to configurations and secrets in Kubernetes:

ConfigMaps Enables you to decouple configuration artifacts from image content to keep containerized applications portable.

Secrets Used to store and manage sensitive information such as passwords, OAuth tokens, and SSH keys.

Security

Role-based access control (RBAC) controls authorization, which determines what operations a user can perform on cluster resources. RBAC is crucial for maintaining the security of the cluster.

7.1.6 Red Hat OpenShift enhancements to Kubernetes

Kubernetes offers a robust platform for container orchestration, and Red Hat OpenShift enhances this foundation with more features and tools to meet the needs of enterprise environments and developer workflows. These enhancements improve usability, security, and operational efficiency:

- A web console for management and monitoring.
- Enhanced security features that are tailored to strict compliance requirements.
- Built-in tools for continuous integration and continuous deployment (CI/CD) to streamline the development process.

Here is a detailed look at how Red Hat OpenShift builds on the core Kubernetes architecture:

- Enhanced developer productivity:
  - Red Hat OpenShift includes a sophisticated web-based console that provides a simpler interface than the standard Kubernetes dashboard. This console enables developers to manage their projects, visualize the state of their applications, and access a broad range of development tools directly.
  - Code-ready containers simplify the setup of local Red Hat OpenShift clusters for development purposes, providing a minimal, preconfigured environment that can run on a developer s workstation. They are useful for simplifying the “getting started” experience.
  - The S2I tool is a powerful feature for building reproducible container images from source code. This tool automates the process of downloading code, injecting it into a container image, and assembling an image. The new image incorporates runtime artifacts that are necessary to run the code, which streamlines the workflow from source code to deployed application.
- Advanced security features:
  - Red Hat OpenShift enhances Kubernetes security by implementing Security Context Constraints (SCCs). SCCs are like Pod Security Policies but provide more granular security controls over the deployment of pods. They enable administrators to define a set of conditions that a pod must run with to be accepted into the system, such as forbidding running containers as root.
  - Red Hat OpenShift integrates an OAuth server that can connect to external identity providers, which enables a streamlined authentication and authorization process. This integration enables users to log in to Red Hat OpenShift by using their corporate credentials, simplifying access management and enhancing security.

  - Red Hat OpenShift provides extensive support for Kubernetes network policies, which dictate how pods communicate with each other and other network endpoints. Red Hat OpenShift takes this support further with egress firewall capabilities, which enables administrators to control outbound traffic from pods to external networks.

- Operational efficiency:

  - Red Hat OpenShift fully embraces the Kubernetes Operator pattern, which extends Kubernetes capabilities by automating the deployment, scaling, and management of complex applications. Red Hat OpenShift includes the Operator Hub, which is a marketplace where users can find and deploy operators for software stacks.

  - Red Hat OpenShift offers a streamlined and highly automated installation process that simplifies the setup of production-grade Kubernetes clusters. This process extends to updates, which can be applied automatically across the cluster, reducing downtime and manual intervention.

  - Red Hat OpenShift includes built-in monitoring and telemetry capabilities that are preconfigured to collect metrics from all parts of the cluster. This feature provides insights into the performance and health of applications and infrastructure, enabling proactive management and troubleshooting.

- Enterprise Integration and support:

  - Red Hat OpenShift integrates Istio-based service mesh capabilities directly into the platform, facilitating a microservices architecture by providing service discovery, load-balancing, failure recovery, metrics, and monitoring, along with complex operational requirements like A/B testing, canary releases, and more.

  - Red Hat OpenShift integrates with various CI/CD tools, offering built-in Jenkins support and integrations with other major CI/CD platforms. This integration supports the automation of the build, test, and deployment lifecycle within the same platform.

7.1.7 Key features of Red Hat OpenShift

Red Hat OpenShift is focused on the developer's experience and has integrated many features that are designed to make development of applications efficient and productive. This section provides an overview of some of those enhancements.

Developer productivity

Red Hat OpenShift is designed to enhance developer productivity by streamlining processes and reducing the complexities that are typically associated with deploying and managing applications. Here is a detailed look at how Red Hat OpenShift achieves this goal through its key features:

- Developer-focused user interface:

  - The Red Hat OpenShift Console is a powerful interface that provides developers with an overview of all projects and resources within the cluster. It offers a perspective that is tailored to developers' needs, enabling them to create, configure, and manage applications directly from the browser. Features like the Topology view enable developers to visualize their applications and services in a GUI, making it simpler to understand and manage the relationships between components.

  - Red Hat OpenShift includes a Developer Catalog that offers many build and deploy solutions, such as databases, middleware, and frameworks, which can be deployed on the cluster with a few clicks. This self-service portal accelerates the setup process for developers, which enables them to focus more on coding and less on configuration.

- Code-Ready Workspaces

  Red Hat OpenShift integrates with Code-Ready Workspaces, a Kubernetes-native IDE that developers can use within their browser. This IDE provides a fully featured development environment that is complete with source code management, run times, and dependencies that are managed and kept consistent across the development team. This approach helps ensure that the team works within a controlled and replicable environment, reducing “works on my machine” problems.

- Application templates and S2I:
  - Red Hat OpenShift application templates are predefined configurations for creating applications based on specific languages, frameworks, or technologies. These templates include everything that you need to build and deploy an application quickly, such as build configurations, deployment strategies, and required services.
  - S2I is a tool for building reproducible Docker images from source code. S2I enables developers to build containerized applications without writing Dockerfiles or becoming experts in Docker. It combines source code with a base Docker image that contains the appropriate runtime environment for the application. The result is a ready-to-run Docker image that is built according to best practices.
- Automated build and deployment pipelines:
  - Red Hat OpenShift has robust support for CI/CD processes, integrating tools like Jenkins, GitLab CI, and others directly into the platform. It automates the build, test, and deployment pipeline, enabling developers to commit code changes frequently without manual steps.
  - Red Hat OpenShift can automatically trigger builds and deployments when code changes are pushed to a source code repository or when other specified events occur. This feature helps ensure that applications are always up to date with the latest code changes.
- Live application development:
  - Red Hat OpenShift supports hot deployments, where changes to application code can be made active without restarting the entire application. This capability is crucial for environments where uptime is critical, and it enables developers to see changes instantly.
  - Developers can access real-time logs and debugging tools directly through the Red Hat OpenShift console, making it simpler to diagnose and resolve issues in development and production environments.

By focusing on these aspects of developer productivity, Red Hat OpenShift lowers the barrier to entry for deploying applications in a Kubernetes environment, simplifies the management of these applications, and accelerates the development cycle. This approach enables developers to spend more time coding and less time dealing with deployment complexities, leading to faster innovation and deployment cycles in a cloud-native landscape.

7.1.8 Enterprise-grade security

Red Hat OpenShift is designed with security as a foundational aspect, integrating robust security features that support the demanding requirements of enterprise environments. This approach includes everything from strict access controls to helping ensure container and platform integrity.

Here is a deeper look into how Red Hat OpenShift delivers enterprise-grade security:

- SCCs:
  - Red Hat OpenShift enhances the security of container environments by using SCCs to define a set of conditions that a container must comply with to run on the platform. These role-based constraints can limit the actions that a pod can perform and the resources that it can access, reducing the risk of unauthorized actions.
  - Fine-grained permissions: Administrators can use SCCs to manage permissions at a granular level, controlling whether pods can run as privileged containers, access sensitive volumes, or use host networking and ports, among other security settings.
- Integrated authentication and authorization:
  - Red Hat OpenShift integrates with existing enterprise authentication systems, such as Lightweight Directory Access Protocol (LDAP), Active Directory (AD), and public OAuth providers to provide a robust user authentication process seamlessly across the organization.
  - RBAC in Red Hat OpenShift enables administrators to regulate access to resources based on the roles of individual users within the enterprise. This approach helps ensure that only authorized users have access to control critical operations, which secures the environment against internal and external threats.
- Network policies and encryption:
  - Red Hat OpenShift enables administrators to define network policies that govern how pods communicate with each other and with other network endpoints. This approach helps ensure that applications are isolated and protected from network-based attacks.
  - Data in transit and at rest can be encrypted, providing an extra layer of security. Red Hat OpenShift supports Transport Layer Security (TLS) for all data in transit and can integrate with enterprise key management solutions to manage encryption keys for data at rest.
- Security enhancements and compliance:
  - Red Hat OpenShift provides automated mechanisms to apply security patches and updates to the container host, run time, and the application containers themselves. This approach helps maintain security compliance and reduces the vulnerability window.
  - Red Hat OpenShift includes features to support compliance with various regulatory requirements, such as Payment Card Industry Data Security Standard (PCI DSS), Health Insurance Portability and Accountability Act (HIPAA), and General Data Protection Regulation (GDPR). It provides extensive logging and auditing capabilities that help track all user actions and system changes, which are crucial for forensic analysis and compliance reporting.
- Container security and image assurance:
  - Red Hat OpenShift integrates with tools like Quay.io to provide automated container image scanning. This feature scans images for vulnerabilities before they are deployed, and image signing helps ensure that only approved and verified images are used in the environment.
  - Running on Red Hat Enterprise Linux (RHEL), Red Hat OpenShift leverages SELinux to enforce mandatory access control (MAC) policies that isolate containers from each other and from the host system. This approach prevents a compromised container from affecting others or gaining undue access to host resources.

- Secure default settings and practices:
  - Red Hat OpenShift encourages the usage of minimal base images that contain only the essential packages that are needed to run applications, reducing the potential attack surface.
  - Red Hat OpenShift is preconfigured with security best practices and regularly updated security benchmarks that guide users in setting up and maintaining a secure environment.

By providing these comprehensive security features, Red Hat OpenShift addresses the complex security challenges that are faced by enterprises today, which helps ensure that their deployments are secure by design, compliant with industry standards, and capable of withstanding modern cybersecurity threats. This security-first approach is integral to maintaining trust and integrity in enterprise applications and data.

7.2 Designing for security

Red Hat OpenShift provides a methodology to build applications as a set of independent containerized microservices, containing their dependencies and running on a shared host in a multi-tenant fashion. In this regard, the paradigm of application development is experiencing a shift from monolithic applications to modern containerized applications, with the latter enabling a faster release cycle and the possibility to upscale and downscale microservices according to business needs, such as a sudden increase in customer demand. The operational advantages that Red Hat OpenShift adoption brings to enterprises are scalability, flexibility, and maintainability.

As the application modernization effort is increasingly embraced by organizations, the decision of the preferred container platform inevitably must consider security as a main requirement. A microservices architecture introduces security challenges that must be considered by security teams in the enterprise workforce. This section highlights the unique security proposition of Red Hat, which positions the container platform as leading in the enterprise dimension.

The major security challenges that are linked with distributed environments are as follows:

Complexity and visibility Microservices challenge security due to the distributed nature of its building blocks. Because containers are independent and built on various frameworks (such as different languages libraries), the security challenges require a preventive strategy to monitor containers.

Communication Microservices communicate with each other through APIs, increasing the attack surface. Encryption of data and authentication are measures that must address the issue effectively.

Access control Access must be monitored granularly, and specific policies are required to help ensure a balance between a smooth development workflow and a highly secure environment.

Beginning with the operating system layer, this section then explores the Compute layer, specifically focusing on the IBM Power server to emphasize the security features that are integrated into its hardware design. Before delving into a more detailed description, we also introduce the Network and Storage layers, highlighting how the Red Hat OpenShift platform provides strategies to address challenges.

7.2.1 Operating system layer

Red Hat OpenShift Container Platform is secure by design, and a significant contribution is its operating system. The host OS addresses the security challenges (complexity and visibility, communication, and access control) by securing the host from container vulnerabilities, and isolating containers from each other.

Red Hat OpenShift Container Platform leverages Red Hat CoreOS, a container-oriented operating system that implements the Security Enhanced Linux (SELinux) kernel to achieve container isolation and supports access control policies. CoreOS includes the following features:

- Ignition: A boot system configuration that is responsible for starting and configuring machines.
- CRI-O: A container run time integrating with the OS. It is responsible for running, stopping, and restarting containers (it replaces the Docker Container Engine).
- Kubelet: A node agent that is responsible for monitoring containers.

An extra security measure is implemented by namespaces, which enable you to abstract the resources that are consumed, including the OS, so that the running container appears as though it is running its own OS to limit the attack surface and prevent vulnerabilities that contaminate other containers. Compromised containers are a vector for the host OS and for other containers that are not running SELinux, so with the control groups, the administrator can set a limitation on the resources that a collection of containers can consume from the host. Secure computing profiles can be defined to limit the system calls that are available to a collection of containers.

Ultimately, SELinux isolates namespaces, control groups, and secure computing nodes.

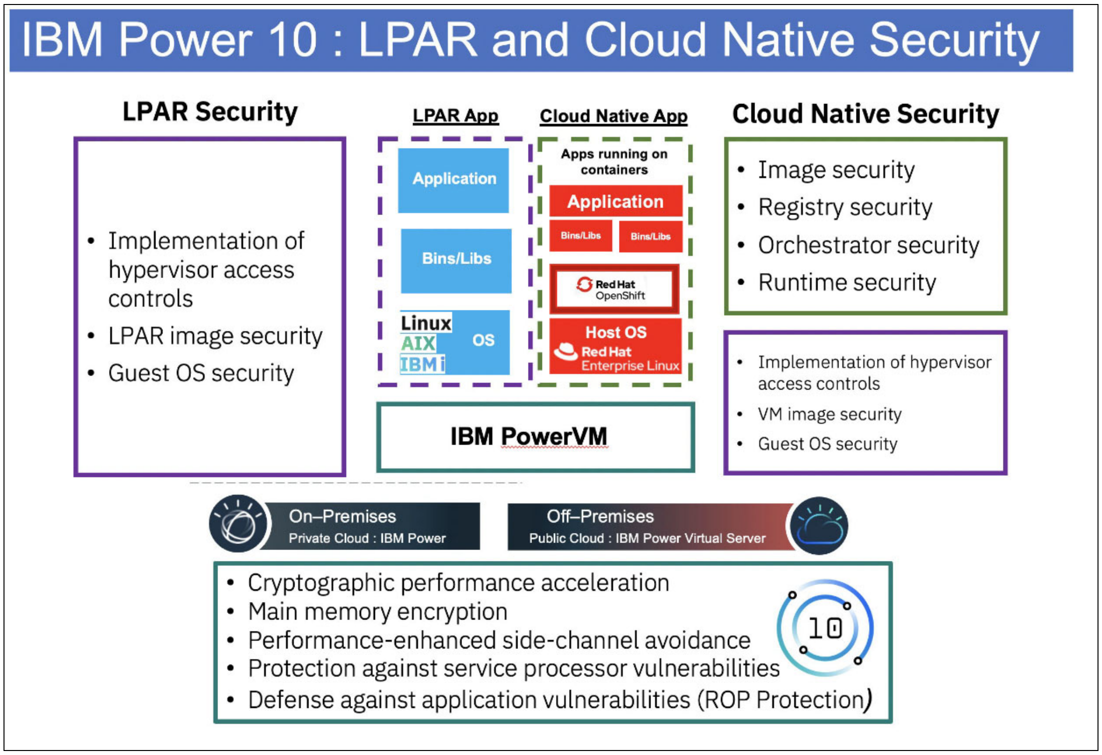
A best practice to secure a multi-tenant environment is to design a container with the least privileges (see 7.4.1, “Privileges” on page 218) possible. As described in 7.3, “Securing your container environment” on page 213, the administrator can (and should) apply a MAC for every user and application while making sure that control groups limit the resources that containers may consume from the host.

7.2.2 Compute layer

The combination of Red Hat OpenShift and IBM Power servers is synergistic from a scalability point of view because it can achieve more than 3x container density compared to x86 based servers. Even more important is the advantage that IBM Power provides from a security standpoint. IBM Power10 encryption of data at rest and in transit are clear assets to securing cloud-native applications in a hybrid-cloud environment, as shown in Figure 7-3 on page 211.

IBM Power10 has in-core hardware that protects against return-oriented programming (ROP) cyberattacks, with limited impact (1 - 2%). ROP attacks are difficult to identify and contain because they collect and reuse existing code from memory (also known as “gadgets”) rather than injecting new code into the system. Hackers chain the commands in memory to perform malicious actions.

IBM Power10 isolates the Baseboard Management Controller (BMC), which is the micro-controller that is embedded on the system board that is responsible for controlling remote management capabilities, and implements allowlist and blocklist approaches to limit the CPU resources that the BMC can access.


IBM Power 10 : LPAR and Cloud Native Security

LPAR Security

- Implementation of hypervisor access controls
- LPAR image security
- Guest OS security

Cloud Native Security

- Image security
- Registry security
- Orchestrator security
- Runtime security

- Implementation of hypervisor access controls
- VM image security
- Guest OS security

On–Premises
Private Cloud : IBM Power
Off–Premises
Public Cloud : IBM Power Virtual Server
10

- Cryptographic performance acceleration
- Main memory encryption
- Performance-enhanced side-channel avoidance
- Protection against service processor vulnerabilities
- Defense against application vulnerabilities (ROP Protection)

Figure 7-3 IBM Power10 security for logical partitions and cloud-native applications

For more information, see 1.4, “Architecture and implementation layers” on page 10.

7.2.3 Network layer

When working with containerized applications that are distributed across multiple hosts and nodes, the network becomes crucial in securing communications. This section introduces the role of networking in Red Hat OpenShift Container Platform so that components in “Network isolation and API endpoint security” are put into context.

Red Hat OpenShift comes with Red Hat Single Sign-On, which acts as an API authentication and authorization measure to secure platform endpoints.

Kubernetes clusters are composed of at least one master node (preferably more for redundancy purposes) and multiple worker nodes, which are virtual machines (VMs) or physical machines that the containers run on. Each node has an IP address, and containerized applications are deployed on these nodes as pods. Each pod is identified with a unique IP address, which helps network management ease because the pod can be treated as a physical host or VM in terms of port allocation, naming, and load-balancing.

The apparent complexity of communication in this distributed infrastructure architecture is solved by the implementation of virtual networking. In fact, each container in a pod shares network ports, facilitating the port allocation task. Although containers in a pod communicate with each other through localhost, communication with containers belonging to different pods requires the coordination of networking resources. The connectivity is implemented by Red Hat Software-Defined Networking (SDN).

Red Hat SDN uses Open vSwitch to manage network traffic and resources as software, enabling policy-based management. SDN controllers satisfy application requests by managing networking devices and routing the data packages to their destination.

The network components in a cluster are managed by a Cluster Network Operator (CNO), which runs on an Red Hat OpenShift cluster.

Leveraging Single Root I/O Virtualization (SR-IOV) on IBM Power servers, the network design becomes more flexible.

Before moving to storage, another functional aspect of Red Hat OpenShift is Network File System (NFS), which is the method that is used to share files across clusters over the network. Although NFS is an excellent solution for many environments, understanding the workload requirements of an application is important when selecting NFS-based storage solutions.

7.2.4 Storage layer

Storage considerations regarding Red Hat OpenShift workloads are derived from a preliminary intrinsic condition of containers. Monolithic applications reserve storage resources in a static fashion, but containers require more flexibility to be agile, simple to manage, and quickly movable between environments.

This section addresses complexity and visibility.

When a container is created, there is a transient layer that handles all read/write data. However, when the container stops running, this ephemeral layer is lost. According to the nature of the container, administrators decide to assign either volumes (bound to the lifetime of the pod) or persistent volumes (persisting longer than the lifetime of the pod).

The dynamic provisioning requirements that benefit from a microservices architecture are facilitated by the Container Storage Interface (CSI), which enables a vendor-neutral management of file and block storage. Using a CSI API enables the following functions:

- Provision or deprovision a determined volume.
- Attach or detach a volume from a node.
- Mount or unmount a volume from a node.
- Consume block and mountable volumes.
- Create or delete a snapshot.
- Provision a volume from a snapshot.

With the Red Hat OpenShift Platform Plus plan, the enterprise can leverage Red Hat Data Foundation, which is a software-defined storage orchestration platform for container environments. The data fabric capabilities of the Red Hat OpenShift Data Platform are derived from the combination of Red Hat Ceph (a software-defined storage platform), Rook.io (a storage operator), and NooBaa (a storage gateway). Red Hat OpenShift Data Platform can be deployed as an internal storage cluster or external storage cluster. It uses CSI to serve storage to the Red Hat OpenShift Container Platform pods. With these capabilities, you can manage block, file, and object storage to serve databases, CI/CD tools, and S3 API endpoints to the nodes.

With the contextual framework of the storage layer in Red Hat OpenShift clarified, here are the security measures that Red Hat Ceph enforces to address threat and vulnerability management, encryption, and identity and access management (IAM):

- Maintaining upstream relationships and community involvement to help focus on security from the start.
- Selecting and configuring packages based on their security and performance track records.
- Building binary files from associated source code (instead of accepting upstream builds).

- Applying a suite of inspection and quality assurance tools to prevent an extensive array of potential security issues and regressions.
- Digitally signing all released packages and distributing them through cryptographically authenticated distribution channels.
- Providing a single, unified mechanism for distributing patches and updates.

For more information, see this Red Hat documentation article.

7.3 Securing your container environment

Red Hat OpenShift provides a container environment that has capabilities that are missing in a Kubernetes distribution. The following sections of this chapter describe the security features that position Red Hat OpenShift as the enterprise choice for containerized workloads.

Source trusting

When pulling code from a GitHub repository, consider whether you can trust the third-party developer. Developers might overlook the vulnerabilities of libraries or other dependencies that are used in the code, so conduct due diligence before deploying a container in your enterprise environment.

To mitigate this risk, Red Hat provides Quay, which is a security focused container image registry that is included in Red Hat OpenShift Platform Plus.

If you prefer to scan for vulnerabilities with different tools, you can integrate Red Hat OpenShift scanners such as OpenSCAP, BlackDuck Hub, JFrog Xray, and Twislock.

Protecting the software build process

S2I provides a framework to integrate code with the runtime libraries and other dependencies. A best practice is to integrate automated security scanning tools into the CI/CD pipeline of enterprises. RESTful APIs, in this context, enable the integration of the workflow with Static Application Security Testing (SAST) or Dynamic Application Security Testing (DAST) tools, such as IBM AppScan or HCL AppScan.

Deployments on a cluster

As a best practice, leverage automated policy-based tools to deploy containers in production environments. SCCs, which are packaged in Red Hat OpenShift Container Platform, help administrators secure sensitive information by allowing/denying access to volumes, accepting/denying privileges, and extending/limiting capabilities that a container requires.

Orchestrating securely

Red Hat OpenShift extends Kubernetes capabilities in terms of secure containers orchestration as follows:

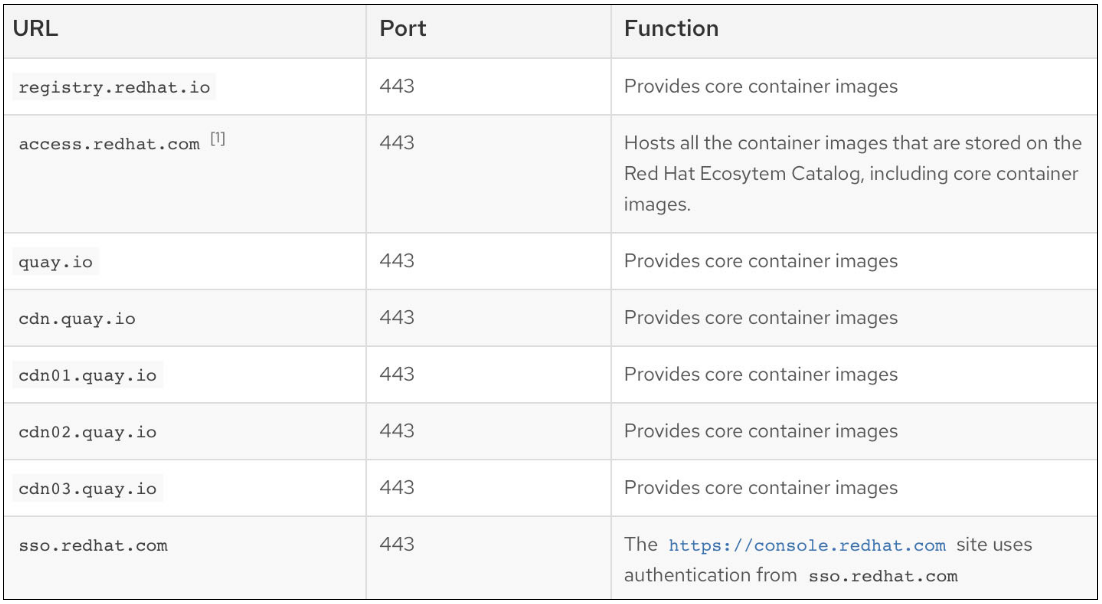
- Handling access to the master node through TLS, which helps ensure that data over the internet is encrypted.
- Helping ensure that the apiserver access is based on X.509 certificates or OAuth access tokens.
- Avoiding exposing etcd (an open source key-value store database for critical data) to the cluster.
- Using SELinux.

Network isolation and API endpoint security

The SDN facilitates the management and visibility of the complex distributed workload. In fact, it is possible to control the outbound traffic of data out of the cluster (with further network control that is implemented through a router or firewall) to see which IP addresses allow/deny access to all others.

Red Hat Single Sign-On, which is an API authentication and authorization service, features client adapters for Red Hat JBoss, which is a Node.js and LDAP-based directory service. An API management tool to use in this context is Red Hat 3scale API management.

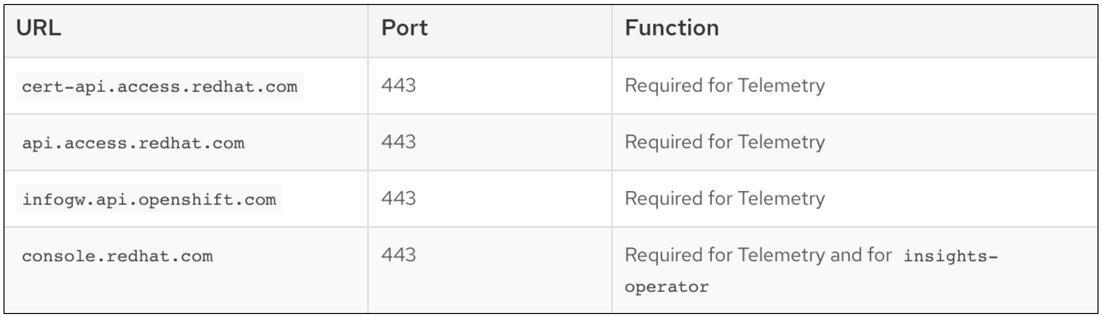
To configure a firewall for Red Hat OpenShift Container Platform 4.12, define the sites that Red Hat OpenShift Container Platform requires so that the firewall grants access to those sites. Create an allowlist that contains the URLs that are shown in Figure 7-4. If a specific framework requires more resources, include them now.



URL	Port	Function
`registry.redhat.io`	443	Provides core container images
`access.redhat.com` [1]	443	Hosts all the container images that are stored on the Red Hat Ecosytem Catalog, including core container images.
`quay.io`	443	Provides core container images
`cdn.quay.io`	443	Provides core container images
`cdn01.quay.io`	443	Provides core container images
`cdn02.quay.io`	443	Provides core container images
`cdn03.quay.io`	443	Provides core container images
`sso.redhat.com`	443	The `https://console.redhat.com` site uses authentication from `sso.redhat.com`

Figure 7-4 Allowlist for the Red Hat OpenShift Container Platform firewall

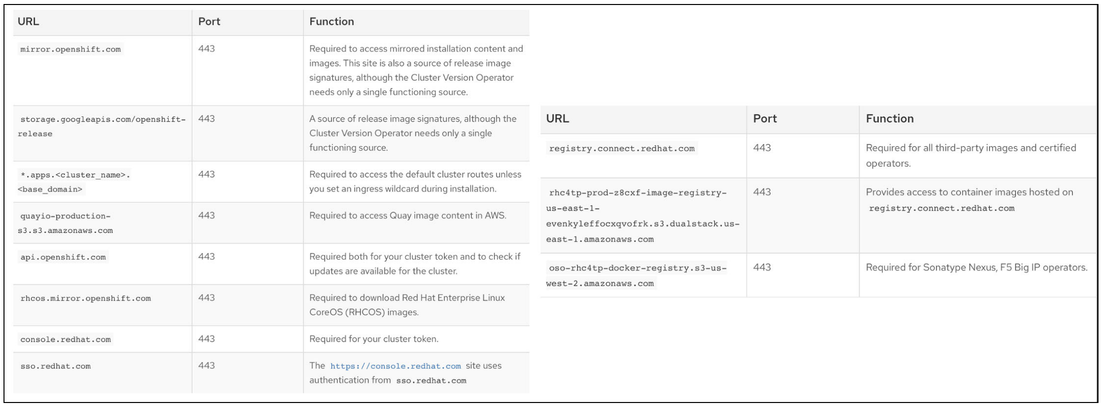
If you want to use Telemetry to monitor the health, security, and performance of application components, use the URLs that are shown in Figure 7-5 to access Red Hat Insights.



URL	Port	Function
`cert-api.access.redhat.com`	443	Required for Telemetry
`api.access.redhat.com`	443	Required for Telemetry
`infogw.api.openshift.com`	443	Required for Telemetry
`console.redhat.com`	443	Required for Telemetry and for `insights-operator`

Figure 7-5 Telemetry URLs to use

If the environment extends to Alibaba, Amazon Web Services (AWS), GCP, or Azure to host the cluster, grant access to the provider API and DNS for the specific cloud, as shown in Figure 7-6 on page 215.



URL	Port	Function
mirror.openshift.com	443	Required to access mirrored installation content and images. This site is also a source of release image signatures, although the Cluster Version Operator needs only a single functioning source.
storage.googleapis.com/openshift-release	443	A source of release image signatures, although the Cluster Version Operator needs only a single functioning source.
*.apps.<cluster_name>.<base_domain>	443	Required to access the default cluster routes unless you set an ingress wildcard during installation.
quayio-production-s3.s3.amazonaws.com	443	Required to access Quay image content in AWS.
api.openshift.com	443	Required both for your cluster token and to check if updates are available for the cluster.
rhcos.mirror.openshift.com	443	Required to download Red Hat Enterprise Linux CoreOS (RHCOS) images.
console.redhat.com	443	Required for your cluster token.
sso.redhat.com	443	The https://console.redhat.com site uses authentication from sso.redhat.com
registry.connect.redhat.com	443	Required for all third-party images and certified operators.
rhc4tp-prod-z8cxf-image-registry-us-east-1-evenkyleffocxqvofrk.s3.dualstack.us-east-1.amazonaws.com	443	Provides access to container images hosted on registry.connect.redhat.com
oso-rhc4tp-docker-registry.s3-us-west-2.amazonaws.com	443	Required for Sonatype Nexus, F5 Big IP operators.

Figure 7-6 Cloud connections to use

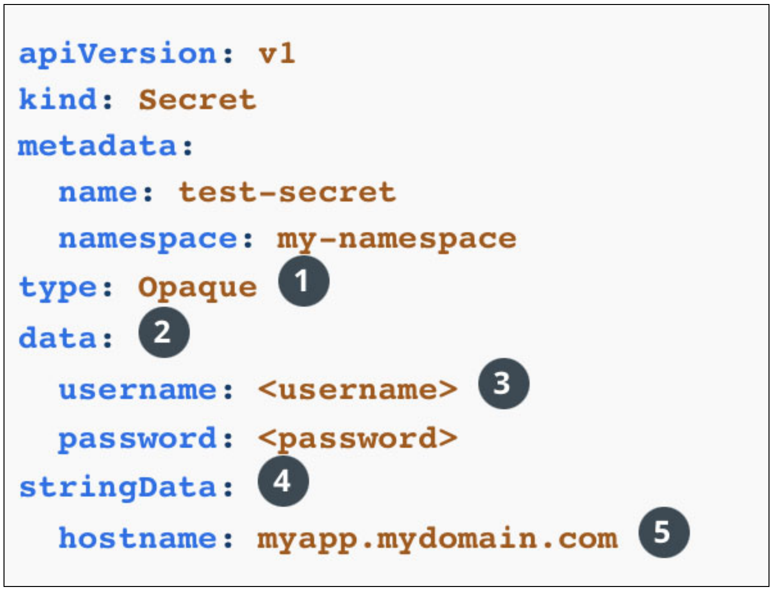
If the preferred option is the default Red Hat Network Time Protocol (NTP) server, use `rhel.pool.ntp.org`.

For more information, see Configuring your firewall for Red Hat OpenShift Container Platform.

Federation of containerized applications

Federations are deployment models that are composed of different meshes that are managed by different administrative domains. Federated namespaces create namespaces in the federation control plane so that the pods have consistent port ranges and IP addresses that are assigned. Federations can share services and workloads while helping ensure extensive security through "secrets". A secret is an object that holds sensitive information (such as passwords) that is decoupled from the pod. Secret data volumes can be shared within a namespace and are never at rest on a node.

Figure 7-7 shows an example of a YAML definition of a secret object type and describes some of the contents.



```
apiVersion: v1
kind: Secret
metadata:
  name: test-secret
  namespace: my-namespace
type: Opaque
data:
  username: <username>
  password: <password>
stringData:
  hostname: myapp.mydomain.com
```

Figure 7-7 YAML file that describes a secret

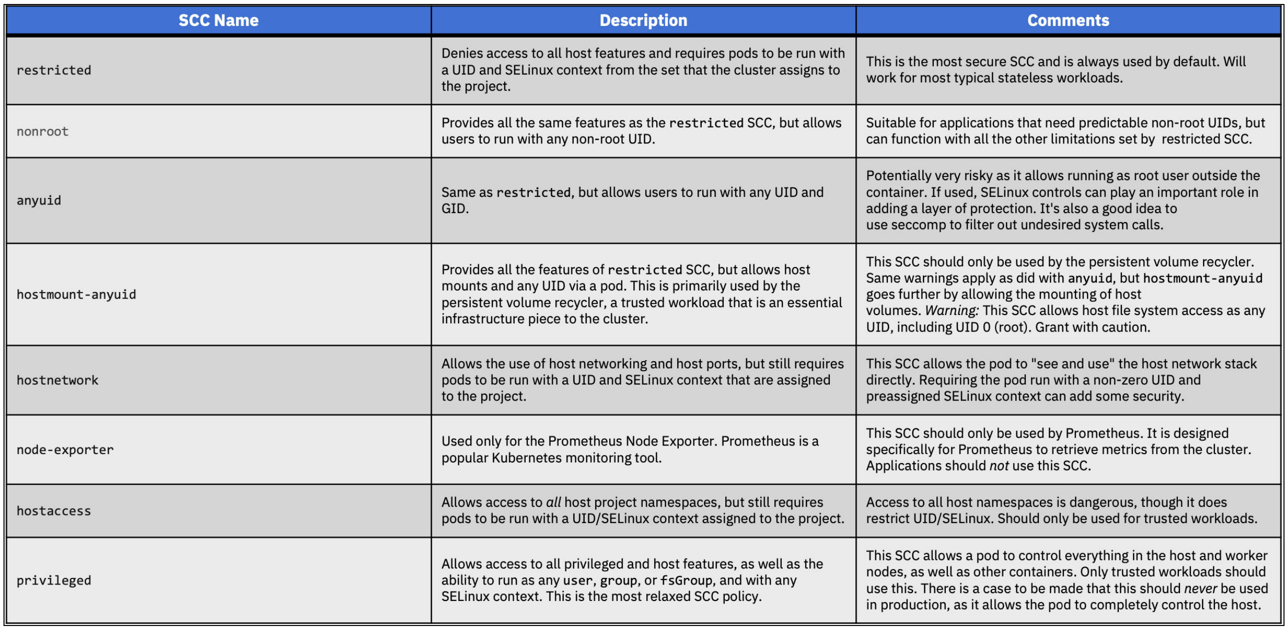
1. Indicates the structure of the secret (in this case, opaque identifies a key-value pair).
2. The format for the keys in `data` must meet the guidelines for DNS_SUBDOMAIN of the Kubernetes glossary. For more information, see Identifiers and Names in Kubernetes.
3. Values that are associated with the keys in `data` must be base64-converted.
4. Entries in `stringdata` are converted to base64 and moved to `data` automatically.
5. Plain text strings are associated with the `stringdata` key.

7.4 Security contexts and Security Context Constraints

Access control to any shared computing environment, such as VMs and containers, is an essential task for the Chief Security Officer’s (CSO) team. Red Hat OpenShift provides security contexts (SCs) and SCCs to help manage security in the container environment.

SCs and SCCs are required for a container to configure access to protected Linux operating system functions on an Red Hat OpenShift Container Platform cluster. SCs are defined by the development team, and SCCs are defined by cluster administrators. An application's SCs specify the permissions that the application needs, and the cluster's SCCs specify the permissions that the cluster allows. An SC with an SCC enables an application to request access while limiting the access that the cluster grants.

An example of SCC and SC implementation is shown in Figure 7-8.



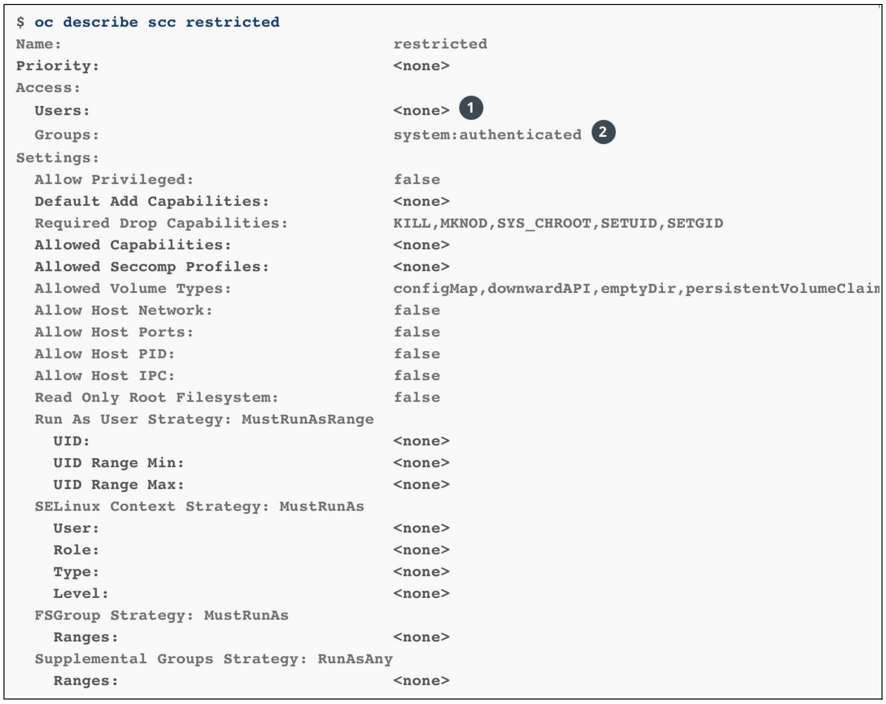
SCC Name	Description	Comments
`restricted`	Denies access to all host features and requires pods to be run with a UID and SELinux context from the set that the cluster assigns to the project.	This is the most secure SCC and is always used by default. Will work for most typical stateless workloads.
`nonroot`	Provides all the same features as the `restricted` SCC, but allows users to run with any non-root UID.	Suitable for applications that need predictable non-root UIDs, but can function with all the other limitations set by restricted SCC.
`anyuid`	Same as `restricted`, but allows users to run with any UID and GID.	Potentially very risky as it allows running as root user outside the container. If used, SELinux controls can play an important role in adding a layer of protection. It's also a good idea to use seccomp to filter out undesired system calls.
`hostmount-anyuid`	Provides all the features of `restricted` SCC, but allows host mounts and any UID via a pod. This is primarily used by the persistent volume recycler, a trusted workload that is an essential infrastructure piece to the cluster.	This SCC should only be used by the persistent volume recycler. Same warnings apply as did with `anyuid`, but `hostmount-anyuid` goes further by allowing the mounting of host volumes. *Warning:* This SCC allows host file system access as any UID, including UID 0 (root). Grant with caution.
`hostnetwork`	Allows the use of host networking and host ports, but still requires pods to be run with a UID and SELinux context that are assigned to the project.	This SCC allows the pod to "see and use" the host network stack directly. Requiring the pod run with a non-zero UID and preassigned SELinux context can add some security.
`node-exporter`	Used only for the Prometheus Node Exporter. Prometheus is a popular Kubernetes monitoring tool.	This SCC should only be used by Prometheus. It is designed specifically for Prometheus to retrieve metrics from the cluster. Applications should *not* use this SCC.
`hostaccess`	Allows access to *all* host project namespaces, but still requires pods to be run with a UID/SELinux context assigned to the project.	Access to all host namespaces is dangerous, though it does restrict UID/SELinux. Should only be used for trusted workloads.
`privileged`	Allows access to all privileged and host features, as well as the ability to run as any `user`, `group`, or `fsGroup`, and with any SELinux context. This is the most relaxed SCC policy.	This SCC allows a pod to control everything in the host and worker nodes, as well as other containers. Only trusted workloads should use this. There is a case to be made that this should *never* be used in production, as it allows the pod to completely control the host.

Figure 7-8 SCC and SC implementation

By default, Red Hat OpenShift prevents the containers running in a cluster from accessing protected functions. These functions (Linux features such as shared file systems, root access, and some core capabilities, such as the `kill` command) can affect other containers running in the same Linux kernel, so the cluster limits access to them. Most cloud-native applications work with these limitations, but some (especially stateful workloads) need greater access. Applications that need these functions can still use them, but they need the cluster's permission.

SCs are defined as a YAML file within the pod that attempts to deploy the application into production. SCCs determine which Linux functions a pod can request for its application. The pod requesting access to specific functions through SCs fails to start unless SCCs give permission to proceed.

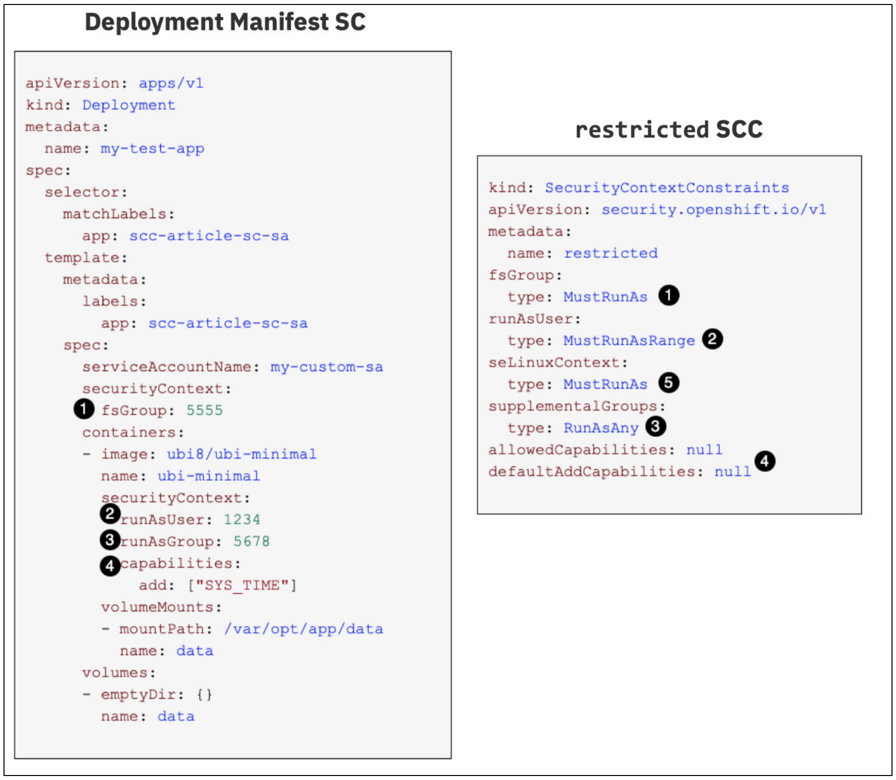
A list of predefined (default) SCCs is shown in Figure 7-9.



```
$ oc describe scc restricted
Name:                                   restricted
Priority:                               <none>
Access:
  Users:                                <none> ❶
  Groups:                               system:authenticated ❷
Settings:
  Allow Privileged:                     false
  Default Add Capabilities:             <none>
  Required Drop Capabilities:           KILL,MKNOD,SYS_CHROOT,SETUID,SETGID
  Allowed Capabilities:                 <none>
  Allowed Seccomp Profiles:             <none>
  Allowed Volume Types:                 configMap,downwardAPI,emptyDir,persistentVolumeClaim
  Allow Host Network:                   false
  Allow Host Ports:                     false
  Allow Host PID:                       false
  Allow Host IPC:                       false
  Read Only Root Filesystem:            false
  Run As User Strategy: MustRunAsRange
    UID:                                <none>
    UID Range Min:                      <none>
    UID Range Max:                      <none>
  SELinux Context Strategy: MustRunAs
    User:                               <none>
    Role:                               <none>
    Type:                               <none>
    Level:                              <none>
  FSGroup Strategy: MustRunAs
    Ranges:                             <none>
  Supplemental Groups Strategy: RunAsAny
    Ranges:                             <none>
```

Figure 7-9 Predefinefd SCCs

Taking a closer look to one of the SCCs, the object looks like the one that is represented in Figure 7-10.



Deployment Manifest SC

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-test-app
spec:
  selector:
    matchLabels:
      app: scc-article-sc-sa
  template:
    metadata:
      labels:
        app: scc-article-sc-sa
    spec:
      serviceAccountName: my-custom-sa
      securityContext:
     ❶ fsGroup: 5555
      containers:
      - image: ubi8/ubi-minimal
        name: ubi-minimal
        securityContext:
       ❷runAsUser: 1234
       ❸runAsGroup: 5678
       ❹capabilities:
            add: ["SYS_TIME"]
        volumeMounts:
        - mountPath: /var/opt/app/data
          name: data
      volumes:
      - emptyDir: {}
        name: data
```

restricted SCC

```
kind: SecurityContextConstraints
apiVersion: security.openshift.io/v1
metadata:
  name: restricted
fsGroup:
  type: MustRunAs ❶
runAsUser:
  type: MustRunAsRange ❷
seLinuxContext:
  type: MustRunAs ❺
supplementalGroups:
  type: RunAsAny ❸
allowedCapabilities: null
defaultAddCapabilities: null ❹
```

Figure 7-10 Description of the “restricted” SCC

The field that is designated by the number 1 in Figure 7-10 represents the users to which the SCC “restricted” applies to. Field 2 identifies the group to which the SCC “restricted” applies.

It is a best practice to avoid modifying default SCCs, although administrators can create customized SCCs that better fit specific requirements and policies in the organizational processes. An example of how a new SCC can be created is described in 7.4.2, “Access controls” on page 219 and in Example 7-3 on page 219.

The following sections describe protected Linux functions, such as privileges, access controls, and capabilities.

7.4.1 Privileges

Privileges describe the authority of a pod and the containerized applications running within it. There are two places that privileges can be assigned: either in the SC when the privilege is set equal to `true` in the SC request, or set in the SCC where the privilege is set to `true`, as shown in Example 7-1.

Example 7-1 Privileged container settings in an SCC

```
allowPrivilegedContainer
allowPrivilegedEscalation
```

In Example 7-1, the first line indicates that the container runs with specified privileges, and the second line grants the possibility for a pod that is derived by the parent pod to run with more privileges than the parent pod.

A request for privileges in an SC is shown in Example 7-2. When privileges are requested from a SCs perspective, the developer needs to only request privileges, but from the SCCs' perspective, the administrator must be specific about the set of privileges that are allowed.

Example 7-2 Request for privileges in an SC

```
securityContext.privileged: true
```

It is a best practice to remember that privileged pods might endanger the host and other containers, so only well-trusted processes should be allowed privileges.

7.4.2 Access controls

Administrators define *access control* within SCCs to manage user IDs (UIDs) and group IDs (GIDs) that access pods. Here are some examples:

`runAsUser` Specifies the range of UIDs that may run within a pod and access all its respective containers.

`supplementalGroups` Determines the GIDs that may run all containers within a pod.

`fsGroup` Enables a range of GIDs to control the pod storage volumes.

`seLinuxContext` Specifies the values for setting SELinux user, role, type, and level.

`mustRunAs` Assigned to fields 1:4. Enforces the range of UIDs that a container can request.

`mustRunAsRange` Assigned to fields 1:4. Enforces the range of UIDs that a container can request.

`RunaAsAny` An UID may be requested by a pod even if its ID is not within a specified range in the SCCs (also referred to as the root ID). Use it carefully because it creates a privileged access that must be cautiously assigned.

`MustRunAsNonRoot` Helps ensure that any non-root IDs must be specified.

Here is the correct syntax for the development team to use to include these requests:

```
securityContext.field
```

`field` is any of the fields in Example 7-3.

Once the request is made, it is processed and validated against the cluster SCCs.

Example 7-3 shows how a new SCC looks by integrating the fields that are listed in 7.4.2, "Access controls" on page 219.

Example 7-3 Example SCC

```
kind: SecurityContextConstraints
apiVersion: v1
metadata:
  name: scc-admin
allowPrivilegedContainer: true
runAsUser:
  type: RunAsAny
seLinuxContext:
  type: RunAsAny
fsGroup:
  type: RunAsAny
```

```
supplementalGroups:
  type: RunAsAny
users:
- my-admin-user
groups:
- my-admin-group
```

7.4.3 Capabilities

Some capabilities, specifically Linux OS capabilities, take precedence over the pod's settings. A list of these capabilities can be found in this document. For completeness, Example 7-4 shows some of the most popular ones.

Example 7-4 Capabilities that are override a pod's settings

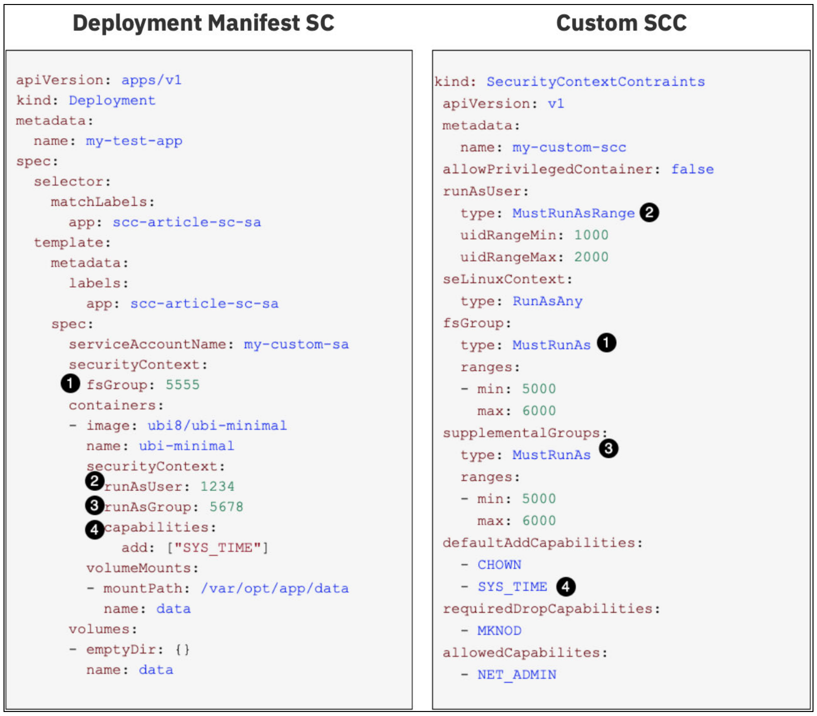
```
chown
kill
setcap
```

7.4.4 Deployment examples

The purpose of this final section on SCs and SCCs is to combine the pieces that have been described so far, illustrating a deployment scenario in which an SC manifest is validated, first against a default "restricted" SCC (Figure 7-11 on page 221) and then against a customized SCC (Figure 7-12 on page 221).

In Figure 7-11 on page 221, the SC fails to pass due to three critical issues, which are shown as points 1, 2, and 4.

- In an attempt to control the pod storage volumes, the SC requests `fsGroup 5555`. The reason that this action fails is that SCC `restricted` does not specify a range for `fsGroup`, so the default range (1000000000 - 1000009999) is used, which excludes `fsGroup 5555`. (1)
- The SC requests `runAsUser 1234`. However, the SCC `restricted` option once again uses the default range (1000000000 - 1000009999), so the request fails because it not within the range. (2)
- The deployment manifest requests `SYS_TIME` (it manipulates the system clock). This request fails because the SCC does not specify `SYS_TIME` either in `allowedCapabilities` or `defaultAddCapabilities` (4). The only request that passes is (3). The SC requests `runAsGroup 5678`, and which is allowed by the `runAsAny` field of the `restricted` SCC.
- As a final remark, (5) is a note to highlight that the container is assigned to the project default context value because `seLinuxContext` is set as `MustRunAs`, but lacks the specific context.



Deployment Manifest SC

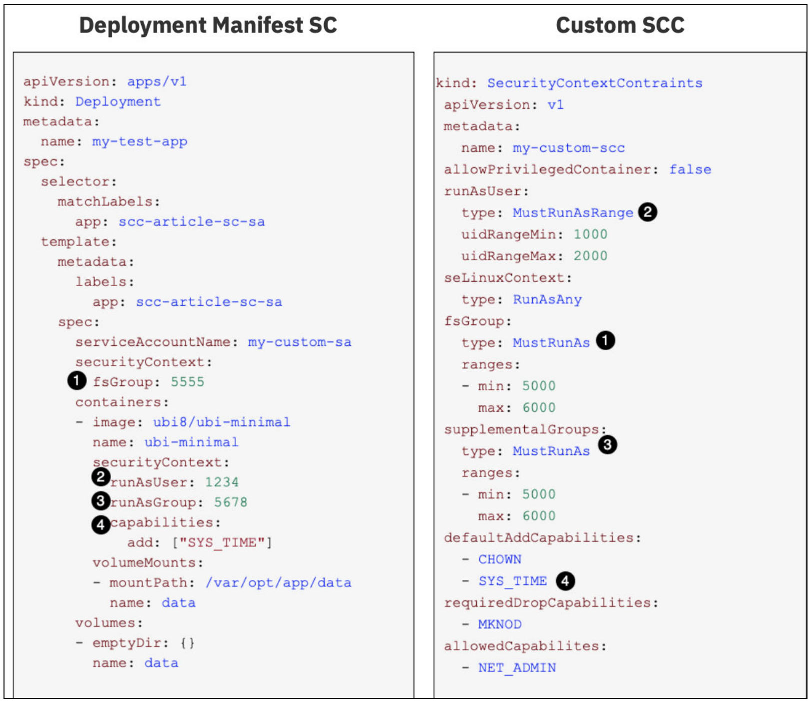
```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-test-app
spec:
  selector:
    matchLabels:
      app: scc-article-sc-sa
  template:
    metadata:
      labels:
        app: scc-article-sc-sa
    spec:
      serviceAccountName: my-custom-sa
      securityContext:
    ❶ fsGroup: 5555
      containers:
      - image: ubi8/ubi-minimal
        name: ubi-minimal
        securityContext:
       ❷runAsUser: 1234
       ❸runAsGroup: 5678
       ❹capabilities:
          add: ["SYS_TIME"]
        volumeMounts:
        - mountPath: /var/opt/app/data
          name: data
      volumes:
      - emptyDir: {}
        name: data
```

Custom SCC

```
kind: SecurityContextContraints
 apiVersion: v1
 metadata:
   name: my-custom-scc
 allowPrivilegedContainer: false
 runAsUser:
   type: MustRunAsRange ❷
   uidRangeMin: 1000
   uidRangeMax: 2000
 seLinuxContext:
   type: RunAsAny
 fsGroup:
   type: MustRunAs ❶
   ranges:
   - min: 5000
     max: 6000
 supplementalGroups:
   type: MustRunAs ❸
   ranges:
   - min: 5000
     max: 6000
 defaultAddCapabilities:
   - CHOWN
   - SYS_TIME ❹
 requiredDropCapabilities:
   - MKNOD
 allowedCapabilites:
   - NET_ADMIN
```

Figure 7-11 Validating an SC against the “restricted” SCC

Figure 7-12 shows how the SC request can be satisfied against a customized SCC (`my-custom-scc`).



Deployment Manifest SC

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-test-app
spec:
  selector:
    matchLabels:
      app: scc-article-sc-sa
  template:
    metadata:
      labels:
        app: scc-article-sc-sa
    spec:
      serviceAccountName: my-custom-sa
      securityContext:
    ❶ fsGroup: 5555
      containers:
      - image: ubi8/ubi-minimal
        name: ubi-minimal
        securityContext:
       ❷runAsUser: 1234
       ❸runAsGroup: 5678
       ❹capabilities:
          add: ["SYS_TIME"]
        volumeMounts:
        - mountPath: /var/opt/app/data
          name: data
      volumes:
      - emptyDir: {}
        name: data
```

Custom SCC

```
kind: SecurityContextContraints
 apiVersion: v1
 metadata:
   name: my-custom-scc
 allowPrivilegedContainer: false
 runAsUser:
   type: MustRunAsRange ❷
   uidRangeMin: 1000
   uidRangeMax: 2000
 seLinuxContext:
   type: RunAsAny
 fsGroup:
   type: MustRunAs ❶
   ranges:
   - min: 5000
     max: 6000
 supplementalGroups:
   type: MustRunAs ❸
   ranges:
   - min: 5000
     max: 6000
 defaultAddCapabilities:
   - CHOWN
   - SYS_TIME ❹
 requiredDropCapabilities:
   - MKNOD
 allowedCapabilites:
   - NET_ADMIN
```

Figure 7-12 Validating an SC against a custom SCC

7.5 Monitoring and logging

Monitoring and logging enable earlier detection of vulnerabilities in Red Hat OpenShift Container Platform by providing essential context during active security incident investigations and postmortem analyses. They facilitate proactive monitoring of security-related activities and help confirm the effectiveness and integrity of the existing security configuration.

This section delves into three primary areas: monitoring containers and Red Hat OpenShift Container Storage security, audit logs, and Red Hat OpenShift File Integrity Operator monitoring.

Monitoring a containerized environment involves tracking and measuring various key performance indicators (KPIs) to help ensure the optimal performance of decoupled applications, often within a microservices architecture. Effective monitoring helps maintain application health, performance, and security, and involves several critical aspects and challenges.

Monitoring containers in a dynamic and rapidly changing environment presents several challenges:

- Rapid provisioning and termination: Containers are provisioned and terminated quickly, making it difficult to track changes in environments with continuously fluctuating numbers of containers and instances. This rapid churn requires monitoring tools to be highly adaptive and capable of real-time tracking.
- Ephemeral nature of containers: Because containers are temporary, their metrics, logs, and other data disappear immediately on termination. It is crucial to collect and store this data in a central location before the containers shut down, which requires a robust logging and monitoring infrastructure that can handle high data throughput and ensure data persistence.
- Shared resources: Containers share resources such as memory, CPU, and operating systems, complicating the measurement of individual container performance. Resource contention and performance bottlenecks can arise, making it essential to have tools that can isolate and identify issues at a granular level.
- Inadequacy of traditional monitoring tools: Many conventional monitoring tools are often insufficient for effectively monitoring containerized environments due to their inability to handle the dynamic and scalable nature of containers. Traditional tools might not provide the necessary visibility into container orchestration layers or the ephemeral nature of container lifecycles.

To address these challenges, several strategies can be employed:

- Monitoring the entire stack: Achieving full application visibility requires monitoring the entire stack, including containers, clusters, networking, and inter-container communications. This holistic approach helps ensure that all aspects of the application and infrastructure are observed, providing a complete picture of system health and performance.
- Granular visibility: Multiple levels of granularity are required to get a comprehensive picture. Drilling down by degrees of granularity helps pinpoint the exact locations of issues. This strategy involves monitoring at the node, pod, and container levels, and observing network traffic, storage I/O, and other critical metrics.

- Contextualized alerting: In a containerized environment, alerts should include relevant context because an issue in one container might be related to its interaction with another container. Contextualized alerts help you understand the root cause of issues and take corrective actions.

The benefits of monitoring containers are substantial:

- Problem resolution: Monitoring helps determine the cause of issues, facilitates their resolution, and enables the cataloging of the “lessons learned” for future reference. This continuous improvement cycle helps in building more resilient and reliable applications.
- Resource usage analysis: Monitoring enables the analysis of how containerized applications use cloud resources and help with cost apportionment. By understanding resource utilization patterns, organizations can optimize their infrastructure and reduce costs.
- Future resource planning: Historical monitoring data helps organizations plan future computing resource requirements. This data-driven approach helps ensure that adequate resources are allocated to meet future demand, avoiding both under-provisioning and over-provisioning.

The Red Hat OpenShift Container Monitoring Platform addresses many of these monitoring challenges through a preconfigured, automatically updated stack that is based on Prometheus, Grafana, and Alertmanager. Key components of this platform are as follows:

- Prometheus: Used as a back end to store time-series data, Prometheus is an open-source solution for cloud-native architecture monitoring. It offers powerful querying capabilities and a flexible data model, making it suitable for a wide range of monitoring scenarios.
- Alertmanager: Handles alarms and sends notifications. It integrates seamlessly with Prometheus, enabling sophisticated alerting rules and notification mechanisms. Alertmanager supports multiple notification channels, including email, Slack, and PagerDuty, which helps ensure that alerts reach the correct people at the correct time.
- Grafana: Provides visual data representation through graphs. Grafana's rich visualization capabilities enable users to create dynamic and interactive dashboards, making it simpler to interpret monitoring data and identify trends and anomalies.

The platform includes default alerts that notify administrators immediately about cluster issues. Default dashboards in the Red Hat OpenShift Container Platform web console offer visual representations of cluster metrics, which help with a quick understanding of cluster states. The “Observe” section of the web console enables access to metrics, alerts, monitoring dashboards, and metrics targets. Cluster administrators can optionally enable monitoring for user-defined projects, enabling customized monitoring of services and pods. This flexibility helps ensure that different teams and projects can tailor monitoring to their specific needs.

As cloud-native applications continue to grow in scale and complexity, Application Performance Monitoring (APM) observability provides constant visibility into the health of the app and its infrastructure. APM observability is crucial for highly distributed and scalable cloud-native and hybrid apps, which helps ensure optimal performance and resiliency.

IBM Instana enhances the observability and APM functions that are provided by the default Red Hat OpenShift Container monitoring tools. Instana is an automated system and APM service that visualizes performance through machine learning-generated graphs. It increases application performance and reliability through deep observability and applied intelligence. Instana excels in cloud-based microservices architectures, enabling development teams to iterate quickly and address issues before they impact customers.

Instana provides several key capabilities:

- Automatic discovery and instrumentation: Instana automatically discovers applications and their dependencies, and uses them without requiring manual intervention. This capability reduces the impact that is associated with setting up monitoring and helps ensure that all components are monitored from the outset.
- Real-time data collection: Instana collects data in real time, providing immediate insights into application performance and health. This real-time visibility is critical for identifying and resolving issues before they affect users.
- Machine learning-based analytics: Instana uses machine learning algorithms to analyze performance data and detect anomalies. This predictive capability helps identify potential issues early so that you can take preemptive action.
- Comprehensive dashboards: Instana offers comprehensive dashboards that provide a unified view of application performance, infrastructure health, and user experience. You can customize these dashboards to meet the specific needs of different stakeholders, from developers to operations teams.

By integrating IBM Instana with Red Hat OpenShift, organizations can elevate their monitoring and observability capabilities, which help ensure that their cloud-native applications remain performant, resilient, and reliable.

In addition to monitoring containers and application performance, maintaining security and integrity within the Red Hat OpenShift environment involves leveraging audit logs and the Red Hat OpenShift File Integrity Operator.

Audit logs provide a detailed record of all activities and changes within the system. They are crucial for tracking user actions, detecting unauthorized access, and investigating security incidents. Effective audit logging helps in maintaining compliance with regulatory requirements and provides an audit trail that can be used for forensic analysis.

The Red Hat OpenShift File Integrity Operator enhances security by monitoring file integrity within the cluster. It detects unauthorized changes to critical system files, which help ensure that the integrity of the operating environment is maintained. The File Integrity Operator works by periodically checking the hashes of monitored files and comparing them to known good values. Any discrepancies trigger alerts, enabling administrators to investigate and remediate potential security breaches.

7.6 Authorization and authentication

Users accessing the Red Hat OpenShift Container Platform must authenticate initially to the cluster. Authentication verifies the identity of the user that makes the requests to the platform's API. Then, the authorization layer evaluates the user's permissions to determine whether the requested actions are permitted. The configuration of authentication settings within the platform is managed by the cluster administrator.

The authentication process in Red Hat OpenShift Container Platform involves multiple layers to help ensure secure access to its resources. Users authenticate primarily through OAuth access tokens or X.509 client certificates. OAuth tokens are obtained through the platform's built-in OAuth server, which supports authentication flows such as Authorization Code Flow and Implicit Flow. The server integrates seamlessly with various identity providers, including LDAP, Keystone, GitHub, and Google, enabling organizations to leverage existing user management systems securely.

X.509 client certificates are used for HTTPS-based authentication, providing a robust mechanism for verifying the identity of clients interacting with the Red Hat OpenShift apiserver. These certificates are verified against a trusted CA bundle, which helps ensure the integrity and authenticity of client connections.

In Red Hat OpenShift, users are classified into different categories based on their roles and responsibilities within the platform. Regular users are typically individuals who interact directly with applications and services that are deployed on Red Hat OpenShift. System users are automatically generated during the platform's setup and associated with specific system-level tasks, such as managing cluster nodes or running infrastructure-related operations.

Service accounts represent a specialized type of system user that is tailored for project-specific roles and permissions. These accounts enable automated processes within projects, which help ensure that applications and services can securely access resources without compromising system integrity.

Groups play a pivotal role in managing authorization policies across Red Hat OpenShift environments. Users can be organized into groups, facilitating streamlined assignment of permissions and simplifying the enforcement of access control policies. Alongside user-defined groups, Red Hat OpenShift automatically provisions virtual groups, which include system-defined roles and default access configurations. This hierarchical group structure helps ensure efficient management of user permissions while adhering to organizational security policies and compliance requirements.

The internal OAuth server in Red Hat OpenShift acts as a central authority for managing authentication and authorization workflows. It issues and validates OAuth tokens that are used by clients to authenticate API requests, which help ensure that only authorized users and applications can access protected resources. Administrators can configure the OAuth server to integrate seamlessly with various identity providers, including htpasswd, Keystone, LDAP, and external OAuth providers like GitHub or Google. Each identity provider offers distinct authentication mechanisms, such as simple bind authentication for LDAP or OAuth 2.0 flows for external identity providers, enhancing flexibility and compatibility with diverse organizational environments.

RBAC is fundamental to enforcing granular access control policies within Red Hat OpenShift, which enable administrators to define fine-grained permissions through roles and role bindings. Roles specify a set of permissions (verbs) that dictate actions users can perform on specific API resources (objects). Role bindings associate these roles with individual users, groups, or service accounts, enabling administrators to implement the principle of least privilege effectively.

ClusterRoles extend RBAC capabilities by providing cluster-wide permissions that apply to all users within the platform. ClusterRoleBindings establish associations between ClusterRoles and subjects (users or groups), which enable administrators to manage permissions consistently across large-scale deployments.

Prometheus system metrics capture comprehensive data of authentication attempts within Red Hat OpenShift, providing administrators with actionable insights into access patterns and security incidents. Metrics include counts of successful and failed login attempts across different authentication methods, such as password-based authentication through a command-line interface (CLI) or web console logins. Monitoring these metrics enables proactive detection of anomalous login activities, supporting timely mitigation of security threats and optimization of authentication workflows.

Administrators can configure and manage RBAC roles and role bindings by using a CLI or a GUI that is provided by Red Hat OpenShift. Practical examples illustrate the steps for creating, modifying, and deleting roles and bindings, which help ensure precise control over access permissions across diverse user populations and project environments. RBAC strategies empower organizations to align access policies with business requirements, enforcing security best practices while facilitating seamless collaboration and application deployment within Red Hat OpenShift Container Platform.

7.7 Tools

There are multiple tools that are available to help you set up and monitor security in your Red Hat OpenShift environment. This section describes some of them.

7.7.1 Aqua

This section describes Aqua, which is a robust security tool that safeguards workloads that are hosted on Red Hat OpenShift running on IBM Power servers. Developed by an IBM Business Partner, Aqua addresses the intricate security challenges that are inherent in cloud-native environments, spanning the entire lifecycle of containerized applications.

Aqua is positioned as a pivotal component in the Cloud-Native Application Protection Platform (CNAPP). This platform helps secure applications from development to deployment and run time on cloud-native architectures. It supports the shift from traditional software security models that rely on vendor-provided patches to a proactive, integrated security approach that is suited for DevOps environments.

Here is a list of Aqua features and capabilities:

- Image scanning: Aqua performs comprehensive vulnerability and malware scans on container images. This scanning process is integral to the CI/CD pipeline, which helps ensure that vulnerabilities are identified and mitigated early in the development cycle. Aqua uses both proprietary scanning engines and open-source tools to detect security issues, which help ensure that only secure images are deployed into production.
- Runtime protection: Once containers are deployed, Aqua provides robust runtime protection, which includes implementing network security policies, access controls, and process-level isolation to prevent unauthorized access, privilege escalation, and network-based attacks. This proactive approach minimizes security risks during application execution.
- Compliance and governance: Aqua enables organizations to enforce compliance with regulatory standards and internal security policies. It offers detailed auditing and reporting capabilities, which are essential for demonstrating compliance and maintaining a security posture across diverse environments and regulatory frameworks.
- Centralized management: The Aqua platform provides a unified management interface through a web-based console and APIs. This centralized management facilitates seamless oversight and control across multiple Kubernetes clusters and namespaces, enhancing operational efficiency and security management at scale.
- Secrets management: Aqua helps ensure the secure management of secrets, credentials, and sensitive data within container environments. It offers features such as secure storage, encryption, and fine-grained access controls to protect critical information from unauthorized access and breaches.

Aqua integrates seamlessly with Red Hat OpenShift on IBM Power by deploying an Aqua Enforcer container on each node within the cluster. These enforcers communicate with the Aqua Security Control Plane, enabling the enforcement of security policies and providing real-time visibility into the security status of the cluster. This integration augments native Red Hat OpenShift security controls, enhancing the overall security posture without compromising the platform compatibility or performance.

Recognizing the trend toward hybrid and multi-cloud deployments, Aqua supports security management across diverse infrastructure environments. It enables organizations to maintain consistent security policies and compliance measures across on-premises data centers and public cloud platforms, which help reduce the attack surface and mitigate risks that are associated with complex deployment landscapes.

7.7.2 Red Hat Advanced Cluster Security

Red Hat Advanced Cluster Security for Kubernetes is a Kubernetes-native security platform that you can use to build, deploy, and run cloud-native applications with more security.

The solution helps protect containerized Kubernetes workloads in all major clouds and hybrid platforms, including Red Hat OpenShift, Amazon Elastic Kubernetes Service (EKS), Microsoft Azure Kubernetes Service (AKS), and Google Kubernetes Engine (GKE).

Red Hat Advanced Cluster Security for Kubernetes is included with Red Hat OpenShift Platform Plus, which is a complete set of powerful, optimized tools to secure, protect, and manage the applications. For more information, see Red Hat Advanced Cluster Security for Kubernetes.

A good feature of Red Hat ACS is that it works to prevent risky workloads from being deployed or running. Red Hat Advanced Cluster Security monitors, collects, and evaluates system-level events such as process execution, network connections and flows, and privilege escalation within each container in your Kubernetes environments. Combined with behavioral baselines and “allowlisting”, it detects anomalous activity that is indicative of malicious intent such as active malware, cryptomining, unauthorized credential access, intrusions, and lateral movement.

For more information about the full features of Red Hat Advanced Cluster Security, see the Red Hat ACS Data Sheet.

8

Certifications

Security standards are a set of guidelines and best practices that organizations can follow to protect their sensitive information and systems from cyberthreats. These standards are developed by various organizations and agencies, such as the International Organization for Standardization (ISO) and the National Institute of Standards and Technology (NIST).

Certification for security standards is a formal process that verifies an organization's adherence to specific security guidelines and best practices. It involves an independent assessment by a certified third-party auditor.

IBM continuously works to maintain certification for industry security standards to provide their clients with a product base that helps them build systems that are compliant to the relevant industry standards.

This chapter describes the following topics:

- 8.1, “Security standards and certifications” on page 230
- 8.2, “Federal Information Processing Standards” on page 231
- 8.3, “ISO standards” on page 232
- 8.4, “Security Technical Implementation Guides” on page 232
- 8.5, “Center for Internet Security” on page 233

8.1 Security standards and certifications

Every enterprise has a set of services or goods that they produce and sell. During the process of creating those goods or services, an enterprise is interacting with consumers, clients, IBM Business Partners, and suppliers, and gathering data from and about those entities. During these interactions, data is gathered and kept. It is the responsibility of each enterprise to manage the data that is collected in a responsible way, and governments worldwide have created regulations about how an enterprise should control and protect that data. These regulations differ depending on the geography or the industry that the enterprise is in.

8.1.1 Security standards

Security standards are a set of guidelines and best practices that organizations can follow to protect their sensitive information and systems from cyberthreats. These standards are developed by various organizations and agencies, such as the ISO and the NIST.

Security standards are important because they provide the following items:

- Consistency: Standards provide a consistent framework for implementing security measures, making it simpler to manage and maintain security across an organization.
- Compliance: Many industries have specific regulations that require adherence to certain security standards.
- Risk reduction: Following security standards helps organizations identify and mitigate potential risks, reducing the likelihood of cyberattacks.
- Enhanced reputation: Demonstrating commitment to security standards can improve an organization's reputation and customer trust.

By understanding and implementing security standards, organizations can improve their cybersecurity posture and protect their valuable assets.

8.1.2 Certifications

Certifications for security standards provide third-party validation that an enterprise is compliant with specific security standards. Certification demonstrates a commitment to robust security practices, reducing the risk of data breaches and cyberattacks.

Certification provides the following benefits:

- Build customer trust.

  Certified organizations gain the trust of customers and partners, especially ones in highly regulated industries.
- Show regulatory compliance.

  Many industries have specific regulations that require adherence to certain security standards. Certification can help organizations meet these requirements.
- Create a competitive advantage.

  Certification can differentiate an organization from competitors, showcasing a strong security culture.

By investing in certification, organizations can demonstrate their commitment to security, protect their valuable assets, and build trust with their stakeholders.

8.1.3 IBM certifications

IBM is dedicated to enhancing security and ensuring compliance. IBM has obtained a wide range of certifications to empower its clients in building secure solutions that adhere to applicable government regulations. The following sections provide an overview of these certifications.

8.2 Federal Information Processing Standards

The Federal Information Processing Standards (FIPS) are a set of publicly announced standards that the NIST developed for use in computer of agencies and contractors. FIPS standards establish requirements for helping ensure computer security and interoperability, and are intended for cases in which suitable industry standards do not exist.

If you are focusing on IBM Power servers security standards, you should learn more about IBM PCIe Cryptographic Coprocessors, which are a family of high-performance Hardware Security Modules (HSMs). These programmable PCI Express (PCIe) cards work with Power servers to offload computationally intensive cryptographic processes, such as secure payments, or transactions from the host server. Using these HSMs, you gain performance and architectural advantages and enable future growth by offloading cryptographic processing from the host server, in addition to delivering high-speed cryptographic functions for data encryption and digital signing, secure storage of signing keys, or custom cryptographic applications.

IBM PCIe Cryptographic Coprocessors meet FIPS PUB 140-2, Security Requirements for Cryptographic Modules, Overall Security Level 4, which is the highest level of certification that is achievable. For more information, see IBM PCIe Cryptographic Coprocessor.

Each IBM HSM device offers the highest cryptographic security that is available commercially. FIPS PUB 140-2 defines security requirements for cryptographic modules. It is issued by the NIST and widely used as a measure of the security of HSMs. The cryptographic processes of each IBM HSM are performed within an enclosure on the HSM that provides complete physical security. For more information, see IBM Cryptographic HSM Highlights.

8.2.1 Software layer support

The security standards that are described in this chapter also apply to the IBM Java SDK that works on IBM Power. For more information about this topic, see IBM SDK, Java Technology Edition 8 and Cryptographic Module Validation Program.

IBM AIX supports FIPS. For more information about FIPS, see IBM AIX 7.x Security Technical Implementation Guide and AIX FIPS Crypto Module for OpenSSL FIPS 140-2 Non-Proprietary Security Policy Document Version 1.9.

If you are using Red Hat Enterprise Linux (RHEL) CoreOS machines in your Red Hat OpenShift cluster, you can apply FIPS PUB 140-2 when the machines are deployed based on the status of certain installation options that govern the cluster options, which you can change during cluster deployment. With RHEL machines, you must enable FIPS mode when you install the operating system on the machines that you plan to use as worker machines. These configuration methods help ensure that your cluster meets the requirements of a FIPS compliance audit, that is, only FIPS-validated or Modules In Process cryptography packages are enabled before the initial system start.

For more information, see Enabling FIPS Compliance in Red Hat OpenShift Cluster Platform on IBM Power.

8.3 ISO standards

Why do you need ISO standards? From a client's perspective, they choose an ISO-certified supplier for general trust and the reliability of the services that are provided. Clients must work with reliable suppliers for different reasons, such as audited processes, documented processes, and other requirements that can be extra proof of professionalism and long-term stability. An ISO-certified supplier is often more reliable than one who does not have any ISO certifications. This statement also applies to other reputable certifications, like the System and Organization Controls 2 Type II or Payment Card Industry Data Security Standard (PCI DSS).

From a supplier's perspective, having ISO certifications in place leads to continuous improvement of the services that are offered. Another benefit is the presence of auditable processes: There is a difference between having only various internal processes, and the processes that can be audited. Even more so, there are clear guidelines in place about how to audit these processes.

Common Criteria (CC) (ISO 15408) is the only global, mutually recognized product security standard. The goal of the CC is to develop confidence and trust in the security characteristics of a system and in the processes that are used to develop and support it.

The ISO 15408 international standard is specifically for computer security certification. For more information and a full description of the ISO 15408 standard, see ISO/IEC 15408-1:2022.

8.4 Security Technical Implementation Guides

The United States Department of Defense (DoD) systems have another layer of requirements that are promulgated by the Defense Information Systems Agency (DISA). Though more of a set of guidelines than a certification, the Security Technical Implementation Guides (STIGs) and Security Requirements Guides for the DoD information technology systems describe security hardening guidelines that are mandated by DODI 8500.01.

Federal IT security professionals within the DoD must comply with the STIG technical testing and hardening frameworks. According to DISA, STIGs "are the configuration standards for DoD [information assurance, or IA] and IA-enabled devices and system. The STIGs contain technical guidance to 'lock down' information systems/software that might otherwise be vulnerable to a malicious computer attack."[1]

You can search through a publicly available document library of STIGs. Table 8-1 on page 233 lists some specific operating systems and components and their corresponding STIGs.

[1] Source: https://disa.mil/.

Table 8-1 STIGs for various IBM components

Product	Link to STIG
Hardware Management Console (HMC)	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_IBM_HMC_Y24M07_STIG.zip
AIX 7.1 and AIX 7.2	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_IBM_AIX_7-x_V2R9_STIG.zip
Red Hat Enterprise Linux 8	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_RHEL_8_V1R14_STIG.zip
Red Hat Enterprise Linux 9	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_RHEL_9_V2R1_STIG.zip
SUSE Linux Enterprise Server 12	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_SLES_12_V2R13_STIG.zip
SUSE Linux Enterprise Server 15	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_SLES_15_V2R1_STIG.zip
IBM Db2 10.5	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_IBM_DB2_V10-5_LUW_V2R1_STIG.zip
IBM WebSphere® Liberty Server	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_IBM_WebSphere_Liberty_Server_V1R2_STIG.zip
IBM WebSphere Traditional 9.x	https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_IBM_WebSphere_Traditional_V9-x_V1R1_STIG.zip

8.5 Center for Internet Security

The Center for Internet Security (CIS), which is a nonprofit that was founded in October 2000, unites the global IT community to develop, validate, and promote best practices in cybersecurity. Over the years, CIS has created and distributed numerous no-charge tools and solutions to help enhance cybersecurity readiness for organizations of all sizes.

CIS is best known for its CIS Controls, which is a comprehensive framework that has 20 essential safeguards and countermeasures to improve cyberdefense. These controls offer a prioritized checklist that organizations can use to reduce their vulnerability to cyberattacks. Also, CIS produces CIS Benchmarks, which provide best practices for secure system configurations, referencing these controls to guide organizations in building stronger security measures.

8.5.1 CIS Benchmarks

Created by a global network of cybersecurity experts, CIS Benchmarks provide best practices for securely configuring IT systems, software, networks, and cloud infrastructure. Published by the CIS, at the time of writing, there are over 140 CIS Benchmarks across seven core technology categories.

These benchmarks are developed through a consensus-based process involving cybersecurity professionals and subject matter experts worldwide. This collaborative approach helps ensure that security best practices are continuously updated and validated.

CIS Benchmarks align closely with-security and data privacy regulatory frameworks, including the NIST Cybersecurity Framework, the PCI DSS, the Health Insurance Portability and Accountability Act (HIPAA), and ISO/EIC 2700. As a result, any organization that operates in an industry that is governed by these types of regulations can progress toward compliance by adhering to CIS Benchmarks. Also, CIS Controls and CIS Hardened Images can help support an organization's compliance with the European Union's (EU) General Data Protection Regulation (GDPR).

Each CIS Benchmark offers configuration recommendations that are organized into two profile levels: Level 1 and Level 2.

- Level 1 profiles provide base-level configurations that are simpler to implement with minimal impact on business operations.
- Level 2 profiles are for high-security environments, requiring more detailed planning and coordination to implement while minimizing business disruption.

There are seven core categories of CIS Benchmarks:

1. Operating systems benchmarks cover security configurations of core operating systems, such as Microsoft Windows, Linux, and Apple OSX. These benchmarks include best practices for local and remote access restrictions, user profiles, driver installation protocols, and internet browser configurations.
2. Server software benchmarks cover security configurations of server software, including Microsoft Windows Server, SQL Server, VMware, Docker, and Kubernetes. These benchmarks include best practices for configuring Kubernetes PKI certificates, application programming interface (API) server settings, server admin controls, vNetwork policies, and storage restrictions.
3. Cloud provider benchmarks address security configurations for Amazon Web Services (AWS), Microsoft Azure, Google, IBM Cloud, and other public clouds. They include guidelines for configuring identity and access management (IAM), system logging protocols, network configurations, and regulatory compliance safeguards.
4. Mobile device benchmarks address mobile operating systems, including iOS and Android, and focus on areas such as developer options and settings, OS privacy configurations, browser settings, and app permissions.
5. Network device benchmarks offer general and vendor-specific security configuration guidelines for network devices and applicable hardware from Cisco, Palo Alto Networks, Juniper, and others.
6. Desktop software benchmarks cover security configurations for some of the most commonly used desktop software applications, including Microsoft Office and Exchange Server, Google Chrome, Mozilla Firefox, and Safari Browser. These benchmarks focus on email privacy and server settings, mobile device management (MDM), default browser settings, and third-party software blocking.
7. Multi-function print device benchmarks outline security best practices for configuring multi-function printers in office settings and cover such topics as firmware updating, TCP/IP configurations, wireless access configuration, user management, and file sharing.

At the time of writing, there are more than 100 CIS Benchmarks that are available as downloadable, no-charge PDFs for non-commercial use.

CIS Benchmarks for IBM Power

CIS has developed specific benchmarks for AIX, IBM i, and various Linux distributions that run on IBM Power.

Here are some CIS Benchmarks for that are relevant to IBM Power servers:

- CIS Benchmark for IBM AIX

  This benchmark provides security configuration guidelines for AIX, which is commonly used on IBM Power servers. It includes best practices for system configuration to enhance security and reduce vulnerabilities.

- CIS Benchmark for IBM i

  This benchmark offers best practices for securely configuring IBM i. It focuses on system settings, security policies, and configurations to improve the overall security posture.

- CIS Benchmarks for Linux

  For IBM Power servers running Linux, there is a generic Linux benchmark, and benchmarks for RHEL, SUSE Enterprise Linux, and Ubuntu Linux.

These benchmarks are regularly updated to reflect the latest security practices and vulnerabilities. You can find the most recent versions and more information on the CIS website or through their publications and resources. Table 8-2 provides a more comprehensive list.

Table 8-2 CIS compliance of IBM Power supported operating systems

Operating system	Recent versions available for CIS Benchmark
AIX	IBM AIX 7.2 (1.1.0) IBM AIX 7.1 (2.1.0)
IBM i	IBM i V7R5M0 (2.0.0) IBM i V7R4M0 (2.0.0) IBM i V7R3M0 (1.0.0)
RHEL	Red Hat Enterprise Linux 9 (2.0.0) Red Hat Enterprise Linux 8 (3.0.0) Red Hat Enterprise Linux 8 STIG (1.0.0)
SUSE Linux Enterprise Server	SUSE Linux Enterprise 15 (1.1.1) SUSE Linux Enterprise 12 (3.1.0)
Ubuntu Linux	Ubuntu Linux 22.04 LTS (2.0.0)

For more information, see the CIS official website.

9

IBM PowerSC

IBM PowerSC is a security and compliance solution that is optimized for virtualized environments on IBM Power servers running AIX, IBM i, or Linux.

PowerSC is on top of the IBM Power server stack, integrating security features that are built at different layers. You can now centrally manage security and compliance on Power servers for all IBM AIX and Linux on Power endpoints. In this way, you can get better support for compliance audits, including the General Data Protection Regulation (GDPR).

This chapter describes the following topics:

- 9.1, “Compliance automation” on page 238
- 9.2, “Real-time file integrity monitoring” on page 238
- 9.3, “Endpoint detection and response” on page 239
- 9.4, “Anti-malware integration” on page 239
- 9.5, “Multi-factor authentication” on page 240

9.1 Compliance automation

The PowerSC Security and Compliance Automation feature is an automated method to configure and audit systems in accordance with Department of Defense (DoD) Security Technical Implementation Guides (STIGs), the Payment Card Industry Data Security Standard (PCI DSS), the Sarbanes-Oxley act and COBIT compliance (SOX/COBIT), the Health Insurance Portability and Accountability Act (HIPAA), Center for Internet Security (CIS) benchmarks compliance for AIX, and IBM i best practices.

PowerSC helps to automate the configuration and monitoring of systems that must be compliant with the PCI DSS. Therefore, the PowerSC Security and Compliance Automation feature is an accurate and complete method of security configuration automation that is used to meet the IT compliance requirements of the DoD UNIX STIG, the PCI DSS, the SOX/COBIT, and HIPAA.

The PowerSC Security and Compliance Automation feature creates and updates ready XML profiles that are used by IBM Compliance Expert Express (ICEE) edition. You can use the PowerSC XML profiles with the `pscxpert` command.

The preconfigured compliance profiles that are delivered with PowerSC reduce the administrative workload of interpreting compliance documentation and implementing the standards as specific system configuration parameters. This technology reduces the cost of compliance configuration and auditing by automating the processes. IBM PowerSC is designed to help effectively manage the system requirements that are associated with external standard compliance, which can potentially reduce costs and improve compliance.

For more information, see this IBM PowerSC document.

9.2 Real-time file integrity monitoring

The PowerSC GUI includes file integrity monitoring (FIM). FIM includes Real-Time Compliance (RTC) for AIX, IBM i `audit`, and Linux `auditd` events. By using the PowerSC GUI, you can configure and view real-time events, which means that you can manage extensive profiles by editing and customizing the reporting capabilities.

FIM monitors critical files on a system that contain sensitive data, such as configuration details and user information. From a security perspective, it is important to monitor changes that are made to these sensitive files. FIM can also monitor changes to binary files and libraries.

PowerSC can generate real-time alerts whenever the contents of a monitored file change or and when a file’s characteristics are modified. By using the Autonomic Health Advisor File System (AHAFS) event monitoring technology, PowerSC RTC monitors all changes and generates alerts by using the following methods:

- Sends email alerts.
- Logs a message to a file.
- Sends an SNMP message to your monitoring server.
- Sens an alert to the PowerSC GUI server.

For more information, see the following resources:

- What's new in PowerSC
- PowerSC GUI description

9.3 Endpoint detection and response

With the recent increase in ransomware and other cybersecurity attacks, PowerSC now has endpoint detection and response (EDR) capabilities. Aspects of EDR include the following items:

- Intrusion detection and prevention
- Log inspection and analysis
- Incident response, and event triggering and filtering

For more information, see the following resources:

- IBM PowerSC 2.2
- IBM Support

One of the EDR forms in PowerSC is that you can configure intrusion detection and prevention systems (IDPS) for a specific endpoint. For AIX, you can use the PowerSC GUI to use the Internet Protocol Security (IPsec) facility of AIX to define parameters for intrusion detection. The IPsec facility of AIX must be installed on the AIX endpoint. For Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server, you must install the `psad` package on each endpoint on which you want to run `psad`, as described in Installing PowerSC on Linux systems, before you can use it with the PowerSC GUI.

The PowerSC GUI uiAgent monitors the endpoint for port scan attacks on the ports that are listed in the IPsec filter rules. By default, PowerSC creates an IPv4 rule in `/etc/idp/filter.rules` to monitor operating system network ports. PowerSC also creates the `/var/adm/ipsec.log` log file. The IPsec facility of AIX also parses IPv6 rules in `/etc/idp/filter.rules`, and the IPv6 addresses appear in the event list.

For more setup information, see the following resources:

- Configuring Intrusion Detection and Prevention (IDP) for AIX endpoints
- Configuring the Intrusion Detection System (IDS)

9.4 Anti-malware integration

This section shows how IBM PowerSC can act as an anti-malware defense.

IBM PowerSC can integrate with the Comprehensive Malware Detection and Removal (ClamAV) global antivirus software toolkit to help prevent malware attacks and detect trojans, viruses, malware, and other malicious threats by scanning all incoming data to prevent malware from being installed and infecting the server.

Through the PowerSC server UI, you can configure anti-malware settings for specific endpoints. Then, ClamAV moves or copies any detected malware to the quarantine directory on the PowerSC uiAgent, assigning a time-stamped prefix and nullifying file permissions to prevent access. ClamAV is not included in the initial PowerSC package, so you must install it on the uiAgent before you can use it with the PowerSC GUI.

For more information about installing the ClamAV toolkit, see the following resources:

- Installing anti-malware on AIX
- Installing anti-malware on Red Hat Enterprise Linux Server or SUSE Linux Enterprise Server
- Installing and configuring anti-malware on IBM i

For more information about a generic PowerSC ClamAV detailed configuration and features, see Configuring anti-malware.

9.5 Multi-factor authentication

This section shows how IBM PowerSC can be an authentication server with multi-layered levels of authentication.

IBM PowerSC can deploy multi-factor authentication (MFA) for mitigating the risk of a data breach that is caused by compromised credentials. PowerSC Multi-Factor Authentication provides numerous flexible options for implementing MFA on Power. PowerSC Multi-Factor Authentication is implemented with a Pluggable Authentication Module (PAM), and can be used on AIX, Virtual I/O Server (VIOS), RHEL, SUSE Linux Enterprise Server, IBM i, Hardware Management Console (HMC), and PowerSC Graphical User Interface server.

The National Institute of Standards and Technology (NIST) defines MFA as authentication that uses two or more factors to achieve authentication. Factors include “something that you know”, such as a password or personal identification number (PIN); “something that you have”, such as a cryptographic identification device or a token; or “something that you are”, such as a biometric.

IBM PowerSC authentication factors improve the security of user accounts. The user either provides the credentials directly in the application (in-band) or out-of-band.

For in-band authentication, users can generate a token to satisfy a policy and use that token to directly log in. Out-of-band authentication enables users to authenticate on a user-specific web page with one or more authentication methods to retrieve a cache token credential (CTC) that they then use to log in. For more information, see Out-of-band authentication type.

IBM PowerSC MFA server can be installed on AIX, IBM i, or Linux. For more information about installation procedures, see the following resources:

- Installing IBM PowerSC MFA server on AIX
- Installing and configuring IBM PowerSC MFA server on IBM PASE for i
- Installing an IBM PowerSC MFA server on Linux
- IBM PowerSC Multi-Factor Authentication Version 2.2.0 User's Guide
- IBM PowerSC Multi-Factor Authentication Version 2.2.0 Installation and Configuration

9.5.1 IBM PowerSC MFA high availability

Because IBM PowerSC depends on a PostgreSQL database, you may configure the IBM PowerSC MFA PostgreSQL database for streaming replication between the primary IBM PowerSC MFA server and a secondary IBM PowerSC MFA server. The feature improves availability, but it is not a load-balancing solution.

In the replication model, the PostgreSQL database on the secondary IBM PowerSC MFA server is a read-only copy of the database that is on the primary IBM PowerSC MFA server. If the database on the primary IBM PowerSC MFA server becomes unavailable, you can promote the database on the secondary IBM PowerSC MFA server to be the primary server. Only one database can be the primary database at any time.

Before you configure IBM PowerSC MFA for high availability (HA), meet the following prerequisites:

- The primary and secondary server must use the same operating system.
- Updates to any files in `/opt/IBM/powersc/MFA/mfadb` are not preserved if you reinstall the IBM PowerSC MFA server.
- If the secondary server uses RHEL or SUSE Linux Enterprise Server, install PostgreSQL, openCryptoki, and opencryptoki-swtok on the secondary server.

For more information, see Configuring IBM PowerSC MFA for high availability.

10

IBM Power Virtual Server security

As the migration of IT services and workloads to the cloud accelerates, cloud security becomes paramount. Given the hybrid cloud's essential role in deploying critical business applications, prioritizing security is imperative.

With the introduction of IBM Power Virtual Server and its ability to run AIX, IBM i, and Linux on Power in the cloud, understanding Power Virtual Server security is crucial for establishing a reliable and secure environment.

This chapter gives a high-level overview of security in Power Virtual Server. For more information, see 10.7, “Additional references” on page 247.

This chapter describes the following topics:

10.1 Introducing Power Virtual Server

The IBM Power Virtual Server offering enables users to deploy IBM Power servers running AIX, IBM i, or Linux into the cloud. This section provides details about the various security features and compliance that are met by Power Virtual Server.

10.2 Authentication and authorization

Power Virtual Server user access is controlled by using the IBM Cloud Identity and Access Management (IAM) service. Table 10-1 shows the IAM platform access roles and the corresponding type of control that is allowed by Power Virtual Server.

Table 10-1 IAM platform access roles

Platform access role	Type of access allowed
Viewer	View instances and list instances.
Operator	View instances and manage aliases, bindings (IBM Power Virtual Server Private Cloud only), and credentials.
Editor	View instances, list instances, create instances, and delete instances.
Administrator	View instances, list instances, create instances, delete instances, and assign policies to other users.

You can use the service access roles to define the actions that the users can perform on Power Virtual Server resources. Table 10-2 shows the IAM service access roles and the corresponding actions that a user can complete by using the Power Virtual Server.

Table 10-2 IAM service access roles

Service access role	Description of actions
Reader	View all resources (such as Secure Shell (SSH) keys, storage volumes, and network settings). You cannot change the resources.
Manager	Configure all resources. You can perform the following actions: ► Create instances. ► Increase storage volume sizes. ► Create SSH keys. ► Modify network settings. ► Create boot images. ► Delete storage volumes.

When you assign access to the Power Virtual Server service, you can set the access scope to:

- All resources
- Specific resources, which support the following selections:
  - Resource group
  - Service instance

Power Virtual Server requires extra access for features such as Direct Link, Transit Gateway service, and Virtual Private Cloud. You might require these extra access capabilities based on your resource requirements. Table 10-3 shows the additional access roles that are required for the corresponding type of services that are allowed by Power Virtual Server.

Table 10-3 Additional access roles

Additional access role	Resources Attribute
Editor, Manager, Operator, Reader, and Viewer	Power Virtual Server service
Editor, Manager, Operator, Reader, Viewer, and Virtual Private Network (VPN) Client	IBM Virtual Private Cloud Infrastructure Services service
Editor, Operator, and Viewer	Transit Gateway service
Reader and Viewer	All resources in the account (Including future IAM enabled services)
Editor, Operator, and Viewer	Direct Link service
Viewer	All resource group
Viewer	Red Hat Satellite service

10.3 IBM Cloud Key Management Services

IBM Cloud provides two Cloud Key Management Services that integrate with Power Virtual Server workloads:

- IBM Cloud Hyper Protect Crypto Services (HPCS) is a dedicated key management service and Hardware Security Module (HSM) that is based on IBM Cloud. You can integrate HPCS with Power Virtual Server to securely store and protect encryption key information for AIX and Linux.
- IBM Key Protect is a full-service, multi-tenant encryption solution that enables data to be secured and stored in IBM Cloud with envelope encryption techniques. You can integrate Key Protect with Power Virtual Server to securely store and protect encryption key information for AIX and Linux.

10.4 Network interfaces

Power Virtual Server is logically isolated from all other public cloud tenants and infrastructure components, creating a private, secure place on the public cloud. This isolation includes all logical partitions (LPARs), networks, and storage. When you create a Power Virtual Server, you can choose a connectivity type from the various available options. When you create an LPAR on Power Virtual Server, you can bind a public interface or multiple private network interfaces to it. A public network interface is implemented by using an IBM Cloud Virtual Router Appliance (VRA) and a Direct Link Connect connection. The public network is protected by a firewall and supports SSH (port 22), HTTPS (port 443), Ping, IBM i, and 5250 terminal emulation with Secure Sockets Layer (SSL) (port 992) network protocols and ports. Private network interfaces use a Direct Link Connect connection to connect to your IBM Cloud account network and resources.

10.5 Network security

Power Virtual Server internal networks are isolated. To meet different network requirements, IBM Cloud offers many connectivity options to multiple environments on the internal private network, public networks, or on-premises networks. Cloud Connections (also called Direct Link Connect) facilitates connectivity between Power Virtual Server and other IBM Cloud environments (Classic, VPC, and others). Cloud Connections can connect directly to specific VPCs, the classic environment, or to an IBM Cloud Transit Gateway. The IBM Cloud Transit Gateway is a network service that interconnects Power Virtual Server, IBM Cloud VPCs, and the classic infrastructure, which enables users to build a global network. Transit Gateway is deployed in a hub and spoke model, where Transit Gateway is the hub, and IBM Cloud VPC, Power Virtual Server, and the classic infrastructure are the spokes. Transit Gateways can be scoped locally (Local Transit Gateway) or across regions (Global Transit Gateways). Using Transit Gateway, environments can be configured across geographies for high availability and disaster recovery (HADR). A Generic Routing Encapsulation (GRE) tunnel connects two endpoints (a firewall or a router and another network appliance) in a point-to-point logical link. Finally, Power Edge Router (PER) is a high-performance networking component that provides direct access to the IBM Cloud services from the Power Virtual Server workspace. It also provides a direct access to the Power Virtual Server from a client-managed environment by using a Direct Link connect or Direct Link dedicated.

10.6 Security and compliance

Power Virtual Server meets the following industry compliance requirements:

- GDPR, as a data protection law framework across the European Union (EU).
- Financial Services Validated, per the IBM Cloud framework for financial services control requirements.
- System and Organization Controls audits of internal controls at a service organization, which are implemented to protect client-owned data in client financial reporting. Also includes audits that are based on the Statement on Standards for Attestation Engagements (SSAE 18) and International Standards for Assurance Engagements No. 3402 (ISAE 3402). The following System and Organization Controls reports are available for Power Virtual Server:
  - System and Organization Controls 1 Type II report
  - System and Organization Controls 2 Type II report
- International Organization for Standardization (ISO): The Power Virtual Server provides services that are delivered from global data centers that are a component of the IBM Cloud IaaS ISO certification. The ISO certification covers a family of four standards:
  - ISO/IEC 27001:2013
  - ISO/IEC 27017:2015
  - ISO/IEC 27018:2019
  - ISO/IEC 27701:2019
- The Payment Card Industry Data Security Standard (PCI DSS). A Service Responsibility Matrix (SRM) guide for Power Virtual Server is available on request.
- US Health Insurance Portability and Accountability Act (HIPAA) to build HIPAA-ready environments and applications by using Power Virtual Server.

10.7 Additional references

The following links provide additional information on security options in Power Virtual Server.

- *IBM Power Systems Cloud Security Guide: Protect IT Infrastructure In All Layers*, REDP-5659
- Getting started with IBM Power Virtual Server
- Network security
- What is IBM Power Virtual Server?
- Power Virtual Server with VPC landing zone
- IBM Cloud PowerVS networking concepts
- Managing identity and access management (IAM) for IBM Power Virtual Servers

11

Lessons learned and future directions in IBM Power security

This chapter provides a summary of lessons that were learned during the writing of this publication. Also, it presents the findings of a recent IBM study on threats in the cybersecurity environment at the time of writing.

This chapter describes the following topics:

11.1 Lessons that were learned from real-world breaches

A strong security culture is the backbone of any effective security program. People are the final line of defense, and their awareness and actions directly impact an organization's vulnerability. Although security policies and procedures are essential, their success hinges on employees understanding the risks and consistently practicing safe behaviors.

Although learning from real-world incidents is valuable, proactive measures are crucial to prevent costly breaches. Security experts emphasize the importance of cultivating a security-conscious workforce through targeted training and awareness campaigns. By fostering a culture where security is a shared responsibility, organizations can reduce their risk exposure.

11.1.1 Recommendations to reduce data breach costs

IBM published the Cost of a Data Breach Report 2024 that listed the findings of research from IBM and the Ponemon Institute. The report provides insights from the experiences of 604 organizations and 3,556 cybersecurity and business leaders that experienced a breach. Out of the research came the following best practices about on how to mitigate the risks of a breach:

- Comprehensive data visibility: Gain a complete understanding of data locations (on-premises, cloud, and others), and implement robust data security measures across all environments.
- Artificial intelligence (AI) and automation: Leverage AI and automation for enhanced threat detection, response, and vulnerability management.
- Generative AI security: Prioritize security in AI initiatives by protecting data, models, and infrastructure.
- Incident response preparedness: Conduct regular cyberrange simulations and train employees in incident response procedures.

By following these best practices, organizations can reduce the financial and reputational impact of a data breach.

11.1.2 Summary of IBM X-Force Threat Intelligence Index 2024

IBM X-Force® is a team of elite cybersecurity professionals, including hackers, incident responders, researchers, and analysts. With a deep understanding of the threat landscape, they offer a comprehensive approach to defending against cyberattacks. The Red Team thinks like attackers to uncover vulnerabilities, while the Incident Response team focuses on preventing, detecting, and responding to threats. The researchers stay ahead of emerging threats, and the analysts transform complex data into actionable insights.

IBM X-Force published the IBM X-Force Threat Intelligence Index 2024. Here is a summary of the findings:

- Identity-centric attacks: Cybercriminals increasingly target identities as the easiest point of entry, with a rise in credential theft and abuse.
- Ransomware decline, data theft surge: While ransomware attacks decreased, data theft and leaks became the primary motivation for cyberattacks.
- Infostealer malware growth: The use of infostealer malware to steal credentials rose tremendously, fueling the dark web's stolen credential market.

- Misconfiguration and legitimate tool abuse: Security misconfiguration and the misuse of legitimate tools contributed to breaches.
- Emergence of AI as a target: The rapid adoption of AI creates an attack surface, and cybercriminals are likely to focus on AI platforms once they achieve market dominance.
- Manufacturing remains a prime target: The manufacturing industry remains the most targeted sector, with malware and ransomware as the primary threats.

Overall, the report highlights a shift in cybercrime tactics toward identity-based attacks and data theft while also warning of the growing threat that is posed by AI. Organizations must prioritize identity protection, implement strong security measures, and stay vigilant against evolving threats.

11.1.3 Best practices for data breach prevention

This IBM blog on data breach prevention presents the following best practices. It states that "To enhance your cyber resilience, it is vital to build security in every stage of software and hardware development". You can strengthen your data breach prevention strategy as follows:

- Proactive risk management: Employ a zero-trust security framework and conduct rigorous testing to identify and eliminate vulnerabilities before breaches occur.
- Data protection: Safeguard sensitive information with multi-factor authentication (MFA), strong passwords, and employee training to prevent data loss and identity theft.
- Business continuity: Implement robust data backup strategies and well-rehearsed incident response plans (IRPs) to minimize downtime and financial losses if there are emergencies.

11.1.4 Summary

The importance of fixing the basics is key. Security is built from steps such as asset inventory, patching, and training. Here are some important points to consider:

- Develop an automated methodology for secure assessments and detection.
- Establish a risk management framework that includes cyberinsurance.
- Maintain a dedicated environment for testing security patches.
- Ensure that rollback options are available in all scenarios.

11.2 Basic AIX security strategies and best practices

Security is an important topic for a good reason. So much personally identifiable information is stored online that security break-ins most likely have affected everyone reading this publication. The penalties for breaches of the various standards, such as the Health Insurance Portability and Accountability Act (HIPAA), are significant. Good security requires a multi-layered approach that starts with people, then physical security, and then the various layers. Look at the whole environment and see how security can be applied at each level.

This section describes some of the basics of locking down your logical partitions (LPARs). This lockdown is not done by default, but it is fairly simple to do. It includes default permissions and umasks; good usernames and passwords; logging, patching, and removing insecure daemons; and integrating Lightweight Directory Access Protocol (LDAP) or Active Directory (AD).

11.2.1 Usernames and passwords

A username and password combination is one of the most basic protections. To use longer usernames and passwords, make a system change. Changing the username length is required if you want to integrate with LDAP or AD, and it requires a restart. To increase the maximum username length to 36, run the following command:

```
chdev -l sys0 -a max_logname=36
```

To check the setting, run the following command:

```
# lsattr -El sys0 | grep max_log
max_logname     36           Maximum login name length at boot time             True
```

This change requires a restart of the LPAR. To have longer passwords, use the `chsec` command. The following version configures the system to use ssha256 (up to 255 characters) for passwords. The next time local users change their password, they will get a longer and more secure password.

```
chsec -f /etc/security/login.cfg -s usw "pwd_algorithm=ssha256"
```

To check the setting, run the following command:

```
getconf PASS_MAX
255¬†
```

As a best practice, set the system to automatically create home directories, which are important in an LDAP or AD environment. Example 11-1 shows how to accomplish this task.

Example 11-1 Automatically creating home directories

```
chsec -f /etc/security/login.cfg -s usw -a mkhomeatlogin=true
tail /etc/security/login.cfg
    pwd_algorithm = ssha256
        mkhomeatlogin = true
```

11.2.2 Logging

Logging is a critical part of any system-protection strategy. Without logs, it is impossible to know what has been happening on the system. The `syslog` daemon (`syslogd`) starts by default on AIX, but the log configuration file is not set up to log everything; you must correctly set up `/etc/syslog.conf`. It is a best practice to set up a separate file system for logs (such as `/usr/local/logs`) rather than use the default of `/var/spool` because if `/var` fills up, the system crashes; if your separate file system fills up, it stops logging. Although file systems should be monitored, it is a best practice to store logs in their own file system to protect against large logs bringing down the system.

Logs can be written to a file, sent to the console, logged to a central host across the network (but the traffic can be substantial), emailed to an administrator, sent to all logged-in users, or any combination of these methods. The most commonly used method is writing to a file in a file system. Once the file system is set up, create a `/etc/syslog.conf` file.

Example 11-2 on page 253 shows an example file that writes to a local file system. It keeps the logs to no more than 2 MB, and then rotates and compresses them. It keeps the last 10 logs. Do this task on all LPARs and Virtual I/O Servers (VIOSs).

Example 11-2 Sample log configuration

```
mail.debug      /usr/local/logs/mailog rotate size 2m files 10 compress
*.emerg         /usr/local/logs/syslog rotate size 2m files 10 compress
*.alert         /usr/local/logs/syslog rotate size 2m files 10 compress
*.crit          /usr/local/logs/syslog rotate size 2m files 10 compress
*.err           /usr/local/logs/syslog rotate size 2m files 10 compress
auth.notice     /usr/local/logs/infolog rotate size 2m files 10 compress
*.info          /usr/local/logs/messages rotate size 2m files 10 compress
```

Go into `/usr/local/logs` and create each of the files that are shown in Example 11-2 by running `touch`. Now, you can stop (`stopsrc -s syslogd`) and start (`startsrc -s syslogd`) the logging daemon.

11.2.3 Insecure daemons

The `/etc/inetd.conf` file is over 120 lines long. Some items are commented out, but most of them are not. Many of the protocols that are in there have known security holes. As a best practice, after setting up an LPAR, secure those protocols. Save a copy of `/etc/inetd.conf` as `/etc/inetd.conf-orig`, and then delete everything in `/etc/inetd.conf` except the items that you want to keep. Then, do a refresh by running `-s inetd`. Do this task on all your LPARs and VIOSs. Typically, the edited `inetd.conf` looks like Example 11-3.

Example 11-3 The inetd.conf file

```
#ftp stream      tcp6  nowait   root /usr/sbin/ftpd ftpd
#telnet stream   tcp6  nowait   root /usr/sbin/telnetd telnetd -a
#xmquery dgram   udp6  wait     root /usr/bin/xmtopas
#dtspcd stream   tcp   nowait   root /usr/dt/bin/dtspcd /usr/dt/bin/dtspcd
```

The file is only four lines and everything is commented out. On a NIM server, you see `tftp` and `bootp` uncommented. Occasionally, when you do maintenance. it uncomments or adds services. When the file is only 4 lines, you can see immediately what was uncommented or added.

As a best practice, do not use `ftp` and `telnet` because they are insecure; use `ssh` and `sftp` instead. If you must use `telnet` or `ftp`, then you can uncomment them, but they send passwords and other items in clear text.

As a best practice, look at `/etc/rc.tcpip` to see whether `snmp`, `sendmail`, and other daemons are starting. If you need `snmp` or `sendmail`, then configure them to keep hackers from exploiting them.

11.2.4 Time synchronization

To help ensure that your timestamps match across your enterprise, implement the Network Time Protocol (NTP). Most companies have an NTP server set up on their AD servers. Set up NTP on your LPARs and VIOSs.

11.2.5 Patching

At a minimum, make sure that you are running a fully supported version of the OS (VIOS, AIX, IBM i, or Linux). To check, use the Fix Level Recommendation Tool (FLRT). As a best practice, keep your patching up to date to proactively solve problems.

In AIX and VIOS environments, there are two different kinds of patching:

- Fix packs (Technology Levels and Service Packs (SPs))
- Emergency fixes or interim fixes

Fix packs are installed by using `install`. Emergency fixes and interim fixes are installed by using `emgr`.

Technology Levels and SPs are found at IBM Fix Central. Check there regularly for updates to your LPARs, VIOSs, server and I/O firmware, and Hardware Management Consoles (HMCs).

Also, there are products that are installed (even at the latest SP) that need updating, such as Java, OpenSSH, and OpenSSL. Java patches are downloaded at IBM Fix Central. OpenSSH and OpenSSL are downloaded at the Web Applications website. As a best practice, perform a full patching window every 6 months unless there is an emergency. You can use the FLRT and Fix Level Recommendation Tool Vulnerability Checker (FLRTVC) tools to determine what patching must occur.

As a best practice, first update the HMC, then the server firmware, then the I/O firmware and VIOS servers, and then the LPARs. However, you should look at the readme file and description files for every update to make sure that IBM does not have prerequisites that must be followed. There are also some requirements for IBM Power9 and adapter firmware because of the new Trusted Boot settings.

FLRTVC (emergency fixes and interim fixes)

FLRTVC comes in two versions:

- A script that you can run on the system that uses data from a file (`apar.csv`) to compare installed file sets and interim fixes against known security problems.
- Use the web-based online tool (FLRTVC Online). With this option, you can upload the output from two commands to the web page, which produces the output that you need to identify security holes that must be closed. As a best practice, run it directly on the LPAR or VIOS. For more information, see 11.3, “Fix Level Recommendation Tool for IBM Power” on page 256.

To run `flrtvc`, complete the following steps:

1. Download the compressed file and then decompress it. You might need to download the `apar.csv` file. If your LPAR or VIOS does not have access to IBM, get the file from your IBM representative and upload it to your LPAR or VIOS.
2. Edit the script and change `SKIPDOWNLOAD` from `0` to `1`. Now, the script looks for the `apar.csv` file in the same directory that the script is in. After the file is found, run it in compact mode and produce an output file as follows:

   ```
   cd  /directory where flrtvc is
   ksh93  ./flrtvc.ksh  >systemname-flrtvc-output.csv
   ```

3. Run `sftp` or `scp` (as ASCII) to download the `systemname-flrtvc-output.csv` file to your computer and open it with Excel as a `.csv` file. The delimiter is `|`.

You can write scripts that `grep` certain things in the output and email them to yourself.

The compact output from the FLRTVC script is best viewed in a spreadsheet. It is broken down into the following columns:

- Fileset: Shows the name of the file set, for example, `bos.net.tcp.client`.
- Current Version: Shows the currently installed level, for example, 7.1.3.45.
- Type: Either `sec` (for security) or `hiper`.
- EFix: You see a value here only if the actual emergency fix for the problem is installed.
- Abstract: A description of the problem, for example, Vulnerability in BIND.
- Unsafe Versions: A list of the file set levels that are impacted, for example, 7.1.3.0 - 7.1.3.45.
- APAR: Provides the APAR number, for example, CVE-2015-5477 or IV75031.
- Bulletin URL: Provides the URL where you can go to read about the vulnerability to get more information.
- Download URL: Provides the links to get the fix.

You can run `flrtvc` ahead of time and then download and prestage the updates. `flrtvc` typically identifies emergency fixes and interim fixes that you need, and the Java, OpenSSH, OpenSSL, and other updates that must go on the system.

11.2.6 Server firmware and I/O firmware

Many security and other fixes are done in firmware. Server firmware and I/O firmware updates can resolve many issues, and should be done at least once a year, preferably every 6 months. These updates can be downloaded ahead of time and prestaged so that you do not have to depend on IBM websites during a maintenance window.

As a best practice, wait until firmware, Technology Levels or SPs have been out for at least 1 -2 months before you update them. Then, update your NIM server and migrate the updates to test, dev, QA, and production.

11.2.7 Active Directory and LDAP integration

When users are created on AIX systems (or any UNIX like system), they are assigned a default user ID (UID) and group ID (GID) number. The system permissions for files use those numbers. For that reason, you want to be sure that a user has the same UID and GID on every system. Using the same IDs and setting permissions correctly should help ensure that the users cannot access files for which they do not have privileges. You can either do this task manually (huge spreadsheets) or integrate your normal users into an AD or LDAP environment. If you decide to do the integration, work closely with your AD or LDAP admin, and the security team to do so.

Do not put root or other system accounts under the control of AD or LDAP. Those accounts must be local. Restrict those accounts to console access only.

11.2.8 Enhanced access

If you have admins or other users who need enhanced access, provide it by using `sudo` or another tool. If multiple users are logging in as root, then there is no accountability. Using `sudo` causes everything to be logged.

To do this task, you can go to the IBM Linux toolbox, download `yum.sh`, and run it. This process installs `rpm` and `yum` (requires `ftp` access to IBM repositories). When `yum` is installed use `yum` to install `sudo` or other tools. Then, you can use `visudo` to put together the rules. You can set a user as root with or without a password, and you can also restrict them to using only certain commands as root. This access is useful for level 1 support and DBAs who need privileges to perform certain tasks.

11.2.9 Backups

As a best practice, take regular `mksysb` (OS bootable) backups. Make sure that these bare metal `mksysb` backups are part of any backup and disaster recovery (DR) plan. An mksysb should be taken at least monthly, and before and after any system maintenance. Also, as a best practice, have two disks (even on the Storage Area Network (SAN)) on the system that is reserved for rootvg. One is active, and the other is one is used to take an `alt_disk_copy` backup of rootvg before you make changes.

11.2.10 A multi-silo approach to security

Security is a multi-silo approach where everyone works together to provide the multiple layers that help ensure that your systems are as secure as possible. The security team should be able to provide you with policies regarding usernames and settings. If you re integrating with AD or LDAP, then most of those policies are implemented there. Although it takes time to implement basic security measures, it is worth it to make your systems harder to break into.

11.2.11 References

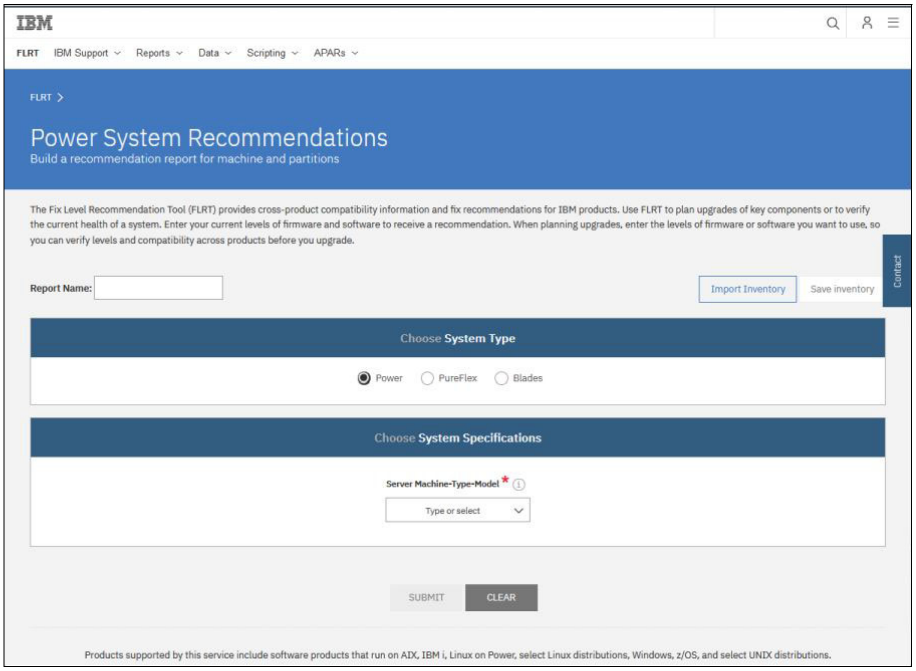
For more information about setting up your AIX security, see the following resources:

- FLRT home page
- FLRTVC home page
- The `apar.csv` file
- FLRTVC Online Tool
- IBM Fix Central (fixes and updates)
- FLRT Lite (check firmware and supported software levels)
- Web Applications (OpenSSH, LDAP, OpenSSL, and Kerberos)
- AIX Linux Toolbox

11.3 Fix Level Recommendation Tool for IBM Power

The FLRT provides cross-product compatibility information and fix recommendations for IBM products. Use FLRT to plan upgrades of key components or to verify the health of a system. Enter your levels of firmware and software to receive a recommendation. When planning upgrades, enter the levels of firmware or software that you want to use so that you can verify levels and compatibility across products before you upgrade.

Figure 11-1 shows an image of the Fix Level Recommendation Tool for IBM Power.



Power System Recommendations

Build a recommendation report for machine and partitions

The Fix Level Recommendation Tool (FLRT) provides cross-product compatibility information and fix recommendations for IBM products. Use FLRT to plan upgrades of key components or to verify the current health of a system. Enter your current levels of firmware and software to receive a recommendation. When planning upgrades, enter the levels of firmware or software you want to use, so you can verify levels and compatibility across products before you upgrade.

Report Name:

Choose System Type

● Power ○ PureFlex ○ Blades

Choose System Specifications

Server Machine-Type-Model *

Type or select

Products supported by this service include software products that run on AIX, IBM i, Linux on Power, select Linux distributions, Windows, z/OS, and select UNIX distributions.

Figure 11-1 The IBM Fix Level Recommendation Tool for IBM Power

Note: For more information, see the Fix Level Recommendation Tool for IBM Power.

11.4 Physical security

Physical security is a critical component of safeguarding people, property, and assets. Although locks and alarms are foundational, they represent only part of the solution.

Protecting hardware, data, and backup systems from damage or theft is paramount. A robust physical security framework is essential for any organization, serving as the bedrock on which other security measures are built. Without it, securing information, software, user access, and networks becomes more challenging.

Beyond internal systems, physical security encompasses protecting facilities and equipment from external threats. Building structures, such as fences, gates, and doors, form the initial defense against unauthorized access. A comprehensive approach considers both internal and external factors to create a secure environment.

11.4.1 Key physical security measures: a layered approach

Access control encompasses the measures that are taken to limit exposure of certain assets to authorized personnel only. Examples of these barriers often include ID badges, keypads, and security guards. However, these obstacles can vary in terms of method, approach, and cost.

Effective physical security is essential for protecting facilities, assets, and personnel. A comprehensive strategy involves a layered approach that combines various security measures to deter, detect, delay, and respond to potential threats.

Deterrence

Discourage unauthorized access through visible security measures:

- Clear signage indicating surveillance
- Robust physical barriers like fences and gates
- High-quality security cameras
- Controlled access systems (card readers, and keypads)

Detection

Identify potential threats early by using the following items:

- Motion sensors and alarms
- Advanced video analytics
- Environmental sensors (temperature and humidity)
- Real-time monitoring systems

Delay

Hinder intruders and buy time for response by using the following items:

- Multiple points of entry and exit
- Sturdy doors, locks, and window reinforcements
- Access control measures (biometrics and mobile credentials)
- Security personnel or guards

Response

Swiftly address security incidents by using the following items:

- Emergency response plans and procedures
- Integration of security systems with communication tools
- Trained personnel for incident management
- Collaboration with law enforcement

By strategically combining these elements, organizations can create a robust physical security framework that mitigates risks and protects critical assets.

11.4.2 Perimeter security and beyond

Having physical protection through solid building construction and perimeter protection and control is part of the equation. Maximizing physical security measures to limit and control who has access to sites, facilities, and materials is paramount. Also, a good physical security process includes monitoring, emergency preparedness, reliable power supplies, adequate climate control, and effective system documentations.

Perimeter security forms the initial line of defense for any facility. Physical barriers like fences, gates, and surveillance systems create a deterrent against unauthorized access. Strategic landscaping and lighting can further enhance perimeter protection by improving visibility and restricting movement.

Access control

Granting authorized access while preventing unauthorized entry is crucial. Modern access control systems, such as key cards, biometric readers, and mobile credentials, offer convenience and security. Restricted areas demand stricter controls, often incorporating MFA and surveillance.

Monitoring and detection

Surveillance systems, including cameras and sensors, play a vital role in detecting and deterring threats. Video analytics and sensor technology provide real-time monitoring and can trigger alerts for suspicious activity. Detailed logs of access attempts and system events are essential for incident investigation and security audits.

The human factor

Although advanced technology is a cornerstone of modern security, human involvement is equally critical. Trained security personnel and informed employees form a powerful defense against threats. Regular security drills and comprehensive training empower staff to recognize and respond to potential dangers.

Even the most sophisticated security systems are only as effective as the people who use them. Employees who understand their role in security can enhance a facility's protection. By equipping your staff with the knowledge and skills to handle emergencies, you create a safer environment for everyone.

A comprehensive physical security strategy integrates these elements to create a layered defense. By combining physical barriers, access control, monitoring, and human involvement, organizations can effectively protect their assets and personnel.

A

IBM Technology Expert Labs offerings

IBM Technology Expert Labs is a professional services organization of highly experienced product specialists. This team offers deep technical expertise across various areas, including IBM Data and Artificial Intelligence (AI), IBM Automation®, IBM Sustainability, IBM Security, IBM Software-Defined Networking, IBM Power, IBM Storage, IBM Z and LinuxONE, IBM GDPS®, and IBM Cloud.

They use methodologies, practices, and patterns to help partners develop complex solutions, achieve better business outcomes, and drive client adoption of IBM software, servers, and storage.

This appendix describes security offerings from IBM Technology Expert Labs. For more information about Technology Expert Labs broader offerings, see the Technology Expert Labs website.

This appendix describes the following topics:

- “Security assessment for IBM Power from IBM Technology Expert Labs” on page 262
- “Security and Compliance Tools for IBM i” on page 264

Security assessment for IBM Power from IBM Technology Expert Labs

If you want to avoid the pressure and confusion of ensuring that your IBM Power environment is safe and secure, there might be no better way than to have IBM secure your environment for you by employing the services of IBM Technology Expert Labs.

By engaging IBM Technology Expert Labs, they can help secure your IBM Power environment by assessing your setup. The purpose of this activity is to help you assess system security on IBM Power. It provides a comprehensive security analysis of either a single AIX, IBM i, or Linux instance, or a single Red Hat OpenShift cluster.

This service can help you address issues that affect IT compliance and governance standards.

Assessing IBM Power Security for AIX, Linux, or Red Hat OpenShift

The goal of this service is to help a client assess system security on IBM Power by providing a thorough security analysis of an AIX instance, Linux instance, or Red Hat OpenShift cluster. This service is aimed at helping the client address issues that are related to IT compliance and governance standards.

IBM Technology Expert Labs perform the following tasks:

1. Conduct a comprehensive security analysis of the following environments:
   - AIX
   - Red Hat Enterprise Linux (RHEL)
   - SUSE Linux Enterprise Server
   - Ubuntu
   - A Red Hat OpenShift 4 cluster
2. Evaluate the security configuration details of the AIX, Linux, or Red Hat OpenShift server.
3. For Red Hat OpenShift, they analyze security recommendations for master node configuration files, the application programming interface (API) server, the controller manager, the scheduler, etcd, the control plane configuration, worker nodes, and the kubelet configuration.
4. For AIX or Linux, they review administrative privileges, logging, monitoring, vulnerability management, malware defenses, and the limitation and control of network ports, protocols, and services.
5. Provide guidance about security best practices based on the CIS Critical Controls and CIS Benchmarks.
6. Offer detailed recommendations for potential adjustments and remediation to enhance overall security.

Assessing IBM Power Security for IBM i

The purpose of this activity is to help a client assess system security on IBM Power. It provides a comprehensive security analysis of an IBM i environment. The service is designed to help the client address issues that affect IT compliance and governance standards.

IBM Technology Expert Labs perform the following tasks:

1. Perform a comprehensive analysis of the security of the IBM i environment.
2. Identify a comprehensive analysis of user profiles, special authorities, and System Service Tool (SST) Profiles.
3. Assess the authentication of the password policy, default and dictionary passwords, user groups, multi-factor authentication (MFA), and single sign-on (SSO).
4. Perform a comprehensive analysis of confidentiality and data access to library authorities, the public authority, and the Integrated File System (IFS) root.
5. Assess the system integrity and configuration for system values, Product Temporary Fixes (PTFs), and work management.
6. Assess logging and auditing with audit journaling, and syslog.
7. Assess IP network settings, for example, Transport Layer Security (TLS), NetServer shares, distributed data management and distributed relational database architecture (DDM and DRDA), guest access, exit points, and ransomware.
8. Provide in-depth recommendations for potential adjustments and remediation, if necessary, to improve overall security.

Other offerings

These security assessment offerings, and a wide range of offerings that cover areas such as performance and availability, are provided by IBM Technology Expert Labs and are generally available worldwide.

Here is the list of IBM Power offerings that are available at the time of writing:

- Assess IBM Power System Health
- Assess IBM Power Availability
- Assess IBM Power Performance
- Assess IBM Power Database Performance
- Assess IBM Power Security for AIX, Linux, or Red Hat OpenShift
- Assess IBM Power Security for IBM i
- Assess IBM PowerVM Health
- Assess IBM Power Capacity
- Assess Oracle Licensing
- Assess IBM i Performance
- Assess Db2 Mirror for IBM i
- Plan Migration to IBM Power10
- Plan Oracle Exadata Migration to IBM Power
- Install and Configure Linux on IBM Power
- Install and Configure IBM License Management Tool for Software Asset Management
- Install and Configure Security and Compliance Tools for IBM i
- Build IBM PowerVM Recovery Manager
- Build IBM PowerHA
- Build IBM PowerSC
- Build IBM PowerHA SystemMirror® for AIX
- Build IBM PowerHA SystemMirror for IBM i
- Build HA/DR Solution with PowerHA Tools for IBM i IASP Manager
- Build Safeguarded Copy with IBM i
- Build Cyber Vault with IBM i
- Build Full System FlashCopy and Replication for IBM i
- Migrate to IBM i Infrastructure
- Perform IBM i Security Services

- Perform AIX and Linux Security Optimization
- Perform AIX Upgrade
- IBM Expertise Connect for AIX on IBM Power
- IBM Expertise Connect for IBM i on IBM Power

The complete list of standard services that are offered by IBM Technology Expert Labs for IBM Power can be found at IBM Technology Expert Labs Power Offerings.

The offerings might differ in each geographical region. For more information about details that are specific to your region, contact an IBM Technology Expert Labs representative.

Security and Compliance Tools for IBM i

The IBM Technology Expert Labs team for IBM i Security is an IBM team that helps clients make the most of their IBM i purchase by offering services such as security assessments and system hardening, and developing IBM i utilities. This family of utilities is known as Security and Compliance Tools for IBM i.

The utilities range from simple to complex, and they complement the tools that are provided natively in IBM i. Each tool has its own purchase price and is available directly from IBM Technology Expert Labs.

Here is a summary of the tools:

- Compliance Automation Reporting Tool (CART)

  After a security assessment and subsequent remediation, systems must be monitored to maintain compliance. Without monitoring, the state of the system is unknown. Your system might have been secure at one point, but without ongoing monitoring, you cannot be sure of your current status. Although there are many security tools that are available, most of them do not focus on IBM i. In fact, several do not even run on IBM i or analyze IBM i security attributes. For this reason, the IBM Technology Expert Labs security and database teams collaborated to create a tool that is specifically for IBM i, which leverages the unique features of the system. This tool provides built-in reports and dashboards for monitoring security attributes that highlight where vulnerabilities or configuration mistakes might exist.

- Advanced Authentication for IBM i

  The primary purpose of this tool is to provide a second factor that users must enter when attempting to gain access to a system. In addition to the standard user password (which should expire regularly), users must provide a unique 6-digit code that changes every 30 seconds on a hardware token or software pap. This code is known as a time-based one-time password (TOTP), and it is based on RFC 6238. This approach forces users to provide something they know (their standard password) and also something they have (the hardware token or software pap). Without both items, access to the system is denied.

- Syslog Reporting Manager for IBM i

  The primary purpose of this tool is to provide a simple way to extract native IBM i logs and send them to a centralized Security Information and Event Management (SIEM) solution. It does this task by extracting entries from various native IBM i logging facilities, transforming them into properly formatted syslog messages (per RFC 3164 and 5424), and sending them to a central collection system. In addition to the native logs, the tool can also monitor and report on changes to IFS files.

- Network Interface Firewall for IBM i

  The primary purpose of this tool is to restrict access to various remote services on IBM i to only an approved list of users. If a user is not authorized to use that particular service, then they are blocked. Blocking is done per service, per user, and even by IP address, if wanted. For example, you can allow user JSMITH access over FTP only if the request is coming from another server.

- Privileged Elevation Tool for IBM i

  Without careful control, privileged users can pose a risk to your system security. This tool enables the security administrator to reduce privileged accounts, with a mechanism to temporarily elevate privileges to users when needed. The Privileged Elevation Tool for IBM i is fully auditable and provides notifications when invoked. With this tool, you can enforce compliance to industry guidelines for privileged users.

- Single Sign-On Suite for IBM i

  IBM Technology Expert Labs Power Delivery Practice provides a suite of tools to set up SSO with IBM i. SSO involves setting up Network Authentication Services and then mapping Windows user profiles to IBM i user profiles by using Enterprise Identity Mapping (EIM).

- Password Validation Tool for IBM i

  The primary purpose of this tool is to provide a more complex method of password validation than the operating system alone provides. Despite warnings, one in five users choose a non-compliant password to protect their identity. This tool validates that passwords meet company and industry recommended rules and guidelines.

- Certificate Expiration Manager

  The primary purpose of this tool is to provide a simple way to be notified about upcoming certificate expirations. In a modern network, TLS encryption is crucial to providing encrypted communications. But this encryption works only when a certificate has a valid date range. Expired certificates can lead to outages in an otherwise healthy network. Therefore, keeping your certificates valid is a key item. The Certificate Expiration Manager notifies you before certificate expiration so that you can ensure uninterrupted service.

For more information about these offerings, see IBM i Security.

B

Ecosystem and products

One path that you can choose to properly secure your environment is to employ solutions and services that can do the job for you. Although it is not intended to be a comprehensive list, this appendix lists some of these security solutions and services that are available on IBM Power through IBM or third-party providers.

This appendix describes the following topics:

- “BigFix (HCL Technologies)” on page 268
- “IBM QRadar Suite (Palo Alto Networks)” on page 268
- “Trend Vision One (Trend Micro)” on page 270
- “Anypoint Flex Gateway (Salesforce/Mulesoft)” on page 271
- “Active IBM i security ecosystem companies” on page 273

BigFix (HCL Technologies)

If you choose to automate the assessment and application of security patches and fixes in your IBM Power environment, one option is to use an automation management solution, such as BigFix. Formerly from IBM, BigFix (now owned by HCL) is a solution for endpoint and security management. If you want to stay ahead of cyberattacks, improve workstation and server patching, proactively resolve tickets, or enhance your digital employee experience, BigFix enables you to do it from a central platform.

With BigFix, you can accomplish the following tasks:

- Revolutionize workspace and enterprise management.

  Leverage artificial intelligence (AI) technologies to elevate your digital experience and automate infrastructures with seamless, secure, and AI-enabled intelligent management.

- Automate IT operations and lower costs.

  By leveraging the world's largest library of automations and a comprehensive unified endpoint management solution, you can automate OS and software patch management and streamline management processes. BigFix can reduce IT complexity and cost by consolidating patch management, software asset management, and endpoint security with a single, comprehensive, BigFix offering that supports multiple operating systems.

- Achieve and maintain continuous compliance.

  Automatically bring non-compliant endpoints back to a compliant state by using industry checklists that contain over 44,000 compliance checks. Help ensure continuous compliance with constant, low-impact monitoring and automatic remediation that protects your endpoints against cybersecurity threats and provides near real-time compliance reporting.

- Discover, prioritize, and remediate vulnerabilities fast.

  Leverage the world's fastest vulnerability remediation solution to discover, prioritize, and mitigate critical security vulnerabilities by using threat resources from MITRE and the Cybersecurity and Infrastructure Security Agency (CISA) and vulnerability scan data from Tenable, Rapid7, Qualys, and others.

BigFix provides an AI Digital+ endpoint management platform that leverages AI to improve employee experience and intelligently automate infrastructure management. It aims to secure and manage endpoints across nearly 100 different operating systems, help ensure continuous compliance to industry benchmarks, and revolutionize vulnerability management with cybersecurity analytics.

Note: For more information, see BigFix.

IBM QRadar Suite (Palo Alto Networks)

IBM QRadar Suite is a modernized threat detection and response solution that is designed to unify the security analyst experience and accelerate their speed across the full incident lifecycle. The portfolio is embedded with enterprise-grade AI and automation to dramatically increase analyst productivity, helping resource-strained security teams work more effectively across core technologies.

Palo Alto Networks acquired the QRadar Suite SaaS offerings on Sept 4, 2024. QRadar Suite SaaS offerings integrated into Cortex XSIAM.

Note: QRadar on-premises remains with IBM.

With a common user interface, shared insights, and connected workflows, this solution offers integrated products for the following areas:

- Endpoint security (endpoint detection and response (EDR) and Managed Detection and Response (MDR))

  EDR solutions are important because endpoints are the most exposed and exploited part of any network. The rise of malicious and automated cyberactivity targeting endpoints leaves organizations struggling against attackers who exploit zero-day vulnerabilities with a barrage of ransomware attacks.

  IBM QRadar EDR provides a more holistic EDR approach:

  - Remediates known and unknown endpoint threats in near real time with intelligent automation.
  - Enables informed decision-making with attack visualization storyboards.
  - Automates alert management to reduce analyst fatigue and focus on threats that matter.
  - Empowers staff and helps safeguard business continuity with advanced continuous learning AI capabilities and a simple interface.

- Security Information and Event Management (SIEM)

  As the cost of a data breach rises and cyberattacks become more sophisticated, the role of security operations center (SOC) analysts is more critical than ever. IBM QRadar SIEM has advanced AI, powerful threat intelligence, and access to the latest detection content.

  IBM QRadar SIEM uses multiple layers of AI and automation to enhance alert enrichment, threat prioritization, and incident correlation. It presents related alerts cohesively in a unified dashboard, reducing noise and saving time. IBM QRadar SIEM helps maximize your security team's productivity by providing a unified experience across all SOC tools, with integrated, advanced AI and automation capabilities.

- SOAR

  The IBM QRadar SOAR platform can optimize your security team's decision-making processes, improve your SOC efficiency, and help ensure that your incident response processes are met with an intelligent automation and orchestration solution.

  Winner of a Red Dot User Interface Design Award, IBM QRadar SOAR helps your organization accomplish the following tasks:

  - Cut response time with dynamic playbooks, customizable and automated workflows, and recommended responses.
  - Streamline incident response processes by time-stamping key actions and helping with threat intelligence and response.
  - Manage incident responses to over 200 international privacy and data breach regulations with Breach Response.

For more information, see the QRadar web page.

Trend Vision One (Trend Micro)

Trend Vision One is a solution for IBM Power customers who want protection across clouds, networks, devices, and endpoints through an AI-powered cybersecurity platform. With full support to run all components of the Trend Vision One platform on IBM Power, Trend Vision One aims to provide administration and DevOps teams greater control over their environment with central visibility and management. Using Trend Vision One on IBM Power can help your organization modernize, simplify, and converge your security operations, enabling better protection against cyberthreats across diverse hybrid IT environments.

In today's complex threat environment, the ability to stay ahead of adversaries, design for resilience, and create secure work environments is paramount. Trend Micro XDR services are engineered to provide advanced threat defense through technologies and human intelligence that proactively monitor, detect, investigate, and respond to attacks. The IBM Power partnership helps ensure that data is protected with comprehensive end-to-end security at every layer of the stack. These integrated security features are designed to help ensure compliance with security regulatory requirements.

Trend Vision One delivers real-time insights to your executive dashboard. No more manual tasks; just efficient, informed decision-making. While IBM Power frees client resources so that they can focus on strategic business outcomes, Trend Vision One automates cybersecurity reporting and playbooks for more efficient and productive security operations. Security teams can stay ahead of compliance regulations, with real-time updates helping to ensure that their enterprise security posture remains robust.

Server and workload protection features

Trend Vision One provides the following server and workload protection features:

- Intrusion and vulnerability prevention:
  - Protect your environment from attacks of known and zero-day vulnerabilities, SQL injections, cross-site scripting, and other web application vulnerabilities.
  - Use intrusion prevention rules when patches are unavailable for known vulnerabilities in applications or operating systems.
  - Intercept traffic that attempts to exploit unpatched vulnerabilities, keeping your assets protected until patches are released, tested, and deployed.
- File integrity monitoring (FIM):
  - Scan for unexpected changes to registry values and keys, services, processes, installed software, ports, and files.
  - Using a baseline secure state as a reference, the integrity monitoring module scans these items and logs an event if unexpected changes are detected.
- Log inspection:
  - Identify events that might be buried in your operating system and application logs.
  - Send these events to a SIEM system or centralized logging server for correlation, reporting, and archiving.
- SAP Scanner:
  - Scan files on demand to protect critical information within SAP environments.
  - Experience seamless certified integrations with both SAP NetWeaver and the SAP HANA platform.

- Analyze uploaded data and identify possible malicious script content that might be embedded or disguised within documents.
- Auto-tag and report malicious content to SAP systems through the NetWeaver Virus Scan Interface (VSI), where administrators can set or enforce policies and actions.

Other features

Trend Vision One includes other features:

- Anti-malware
- Web reputation service
- Activity monitoring
- Activity firewall
- Application control
- Behavioral analysis
- Machine learning
- EDR and XDR
- Device control
- Virtualization protection

> **Note:** For more information about Trend Vision One, see Trend Vision One. For more information about the Trend Vision One on IBM Power solution, see Endpoint Security Solution - Trend Vision One.

Anypoint Flex Gateway (Salesforce/Mulesoft)

To fully realize the value of enterprise data, businesses often use application programming interfaces (APIs). APIs improve existing products, operations, and systems; open new streams of revenue; and provide richer insights that result in enhanced business strategies and provide richer customer experiences.

When you transport data through APIs, you must have a protection layer to help ensure the security of data, and limit accessibility only to known actors. Mulesoft and IBM collaborated to produce such a layer: Anypoint Flex Gateway on IBM Power.

In today's digital landscape, seamless connectivity and rapid data exchange are crucial for business success. Organizations constantly seek innovative solutions to streamline operations, and MuleSoft Anypoint Flex Gateway provides that capability.

Many companies leverage Salesforce's MuleSoft to manage and secure APIs across cloud-native, containerized environments. Now, IBM Power users can tap into the power of Anypoint Flex Gateway's advanced API protection layer to modernize applications and accelerate API-driven initiatives.

Empowering integration on IBM Power

IBM Power is renowned for its robust performance, reliability, and scalability. With the native integration of Anypoint Flex Gateway, businesses that use IBM Power servers can leverage one of the industry-leading API management platforms from MuleSoft to seamlessly connect diverse systems, applications, and data sources.

MuleSoft Anypoint Flex Gateway is an Envoy-based, ultra-fast, and lightweight API gateway. Designed for seamless integration with DevOps and continuous integration and continuous deployment (CI/CD) workflows, Anypoint Flex Gateway delivers the performance that is needed for demanding applications and microservices while helping ensure enterprise-grade security and manageability across any environment.

This synergy unlocks new levels of agility, innovation, and efficiency for your digital transformation journey. This native integration enables a smooth installation and operation of the API gateway, effectively safeguarding your IBM Power applications.

Key use cases and benefits

This approach empowers you to accelerate modernization with API-led integration. You can enable a hybrid retail model with a container-based solution and an API integration layer or simplify SAP S/4HANA integration with other systems.

Deploying Anypoint Flex Gateway close to your IBM Power hosted applications, APIs, and data enhances the customer experience, enforces security policies, reduces data latency, and boosts application performance. You can deploy the gateway on Red Hat OpenShift, Red Hat Enterprise Linux (RHEL), and SUSE Linux Enterprise Server.

Here are the key benefits:

- Seamless connectivity: Connect seamlessly across on-premises, cloud, and hybrid environments, facilitating real-time data exchange and decision-making.
- Unified integration platform: Access a unified platform for integration and API management, which streamlines development, deployment, and management of integration solutions; reduces complexity; and accelerates time-to-market.
- Scalability and flexibility: Handle a few transactions or millions of events with unmatched scalability and flexibility, adapting to evolving business needs and helping ensure integration solutions.
- Integration: Connect applications, data sources, and devices across your IBM Power server environment.
- Dynamic scaling: Scale your integration infrastructure dynamically to meet evolving business demands without compromising performance.
- Unwavering reliability: Help ensure continuous operation and data integrity with resilient integration solutions for IBM Power servers.
- Enhanced security: Safeguard your critical assets and data with enterprise-grade security features that are embedded within Anypoint Flex Gateway.

By combining the strengths of MuleSoft's Anypoint Platform with the performance and reliability of IBM Power servers, businesses can confidently embark on their digital transformation journeys equipped with the tools and capabilities to drive innovation, agility, and growth.

Note: IBM and Mulesoft's partnership announcement for Anypoint Flex Gateway on IBM Power can be found at IBM and MuleSoft expand global relationship to accelerate modernization on IBM Power. The solution brief can be found at MuleSoft + IBM Power: Modernize and Manage Application Connectivity.

Active IBM i security ecosystem companies

The IBM i ecosystem includes several companies that are dedicated to enhancing security for IBM i environments. These companies provide a range of solutions to address various security needs, which help ensure robust protection for IBM i servers.

This section describes some notable companies that are active in IBM i security.

Fortra (formerly HelpSystems)

Fortra offers a comprehensive suite of security solutions to protect IBM i environments. Their products cover areas such as data encryption, compliance management, and threat detection. Fortra's IBM i security solutions are renowned for their robust features and comprehensive reporting capabilities. They cater to many industries, including finance, healthcare, and retail, and help ensure that their clients meet stringent regulatory requirements and safeguard critical data.

Raz-Lee Security

Raz-Lee Security specializes in providing advanced security solutions for IBM i. Their offerings include tools for real-time threat detection, audit and compliance management, and vulnerability assessment. The Raz-Lee iSecurity suite is highly regarded for its powerful and customizable security modules, which help organizations proactively manage and mitigate security risks. Their customer base spans various sectors such as banking, insurance, manufacturing, and government, reflecting their ability to address diverse security challenges across different industries.

Precisely

Precisely provides a range of IBM i solutions that help ensure data integrity, availability, security, and compliance. Their IBM i security solutions include tools for access control, monitoring, privacy, and malware defense. Precisely is known for its robust, scalable solutions that can integrate seamlessly into existing IT infrastructures. These solutions deliver market-leading IBM i security capabilities that help organizations successfully comply with cybersecurity regulations and reduce security vulnerabilities. Also, these security offerings seamlessly integrate with Precisely IBM i high availability (HA) solutions to deliver a greater level of business resilience. Precisely customers range from large enterprises to small and medium businesses (SMBs) in sectors like telecommunications, financial services, and logistics.

Precisely also offers a no-charge assessment tool for IBM i. Assure Security Risk Assessment checks over a dozen categories of security values, compares them to best practices, reports on findings, and makes recommendations. You can find this security risk assessment at Assure Security.

Fresche Solutions

Fresche Solutions offers a comprehensive IBM i Security Suite to protect IBM i servers from modern security threats. Their solutions include tools for real-time monitoring, vulnerability assessment, and compliance management. The Fresche security suite is noted for its innovative approach to security management, combining ease of deployment with powerful analytical capabilities. Their customer base includes businesses of all sizes, from SMBs to large enterprises, in industries such as retail, manufacturing, and services, demonstrating their versatile and scalable security offerings.

These companies, among others, play a vital role in the IBM i ecosystem by continuously innovating and providing security solutions that are tailored to the unique needs of IBM i users. Their diverse customer bases and strong industry reputations underscore their effectiveness in delivering reliable, high-quality security solutions.

Abbreviations and acronyms

ABAC	attribute-based access control
ACE	Access Control Entry
ACF	Advance Crypto Facility
ACL	access control list
AD	Microsoft Active Directory
AES	Advanced Encryption Standard
AHAFS	Autonomic Health Advisor File System
AI	artificial intelligence
AKS	Azure Kubernetes Service
API	application programming interface
APM	Application Performance Monitoring
ASMI	Advanced System Management Interface
ASP	auxiliary storage pool
AWS	Amazon Web Services
BMC	Baseboard Management Controller
BRMS	Backup Recovery and Media Services
CA	certificate authority
CART	Compliance Automation Reporting Tool
CC	Common Criteria
CCA	Common Cryptographic Architecture
CCIN	Customer Card Identification Number
CCPA	California Consumer Privacy Act
CCTK	Cryptographic Coprocessor Toolkit
CHIM	Cryptographic Hardware Initialization and Maintenance
CI/CD	continuous integration and continuous deployment
CIMOM	Common Information Model Object Manager
CIS	Center for Internet Security
CISA	Cybersecurity and Infrastructure Security Agency
CL	control language
ClamAV	Comprehensive Malware Detection and Removal
CLI	command-line interface
CLiC	CryptoLight for C library
CMC	Cloud Management Console
CMS	Cryptographic Message Syntax
CMS	Certificate Management Services
CNAPP	Cloud-Native Application Protection Platform
CNO	Cluster Network Operator
CoD	Capacity on Demand
CSI	Container Storage Interface
CSO	Chief Security Officer
CSR	certificate signing request
CTC	cache token credential
DAC	discretionary access control
DAS	direct attached storage
DAST	Dynamic Application Security Testing
DCM	Digital Certificate Manager
DISA	Defense Information Systems Agency
DN	distinguished name
DoD	Department of Defense
DoS	denial of service
DR	disaster recovery
DST	Dedicated Service Tool
EA	extended attribute
EAR	Export Administration Regulations
eBMC	Enterprise Baseboard Management Controller
eBPF	Extended Berkeley Packet Filter
ECC	Error Correcting Code
EDR	endpoint detection and response
EFF	Electronic Frontier Foundation
EFS	Encrypted File System
EIM	Enterprise Identity Mapping
EKS	Elastic Kubernetes Service
ELK	Elasticsearch, Logstash, and Kibana
EPEL	Extra Packages for Enterprise Linux
EU	European Union
FHE	Fully Homomorphic Encryption
FIM	File Integrity Monitoring

FIPS	Federal Information Processing Standards
FISMA	Federal Information Security Management Act
FLRT	Fix Level Recommendation Tool
FLRTVC	Fix Level Recommendation Tool Vulnerability Checker
FSP	Flexible Service Processor
GDPR	General Data Protection Regulation
Gen3	Generation 3
GID	group ID
GKE	Google Kubernetes Engine
GLVM	geographical logical volume manager
GPL	GNU Public License
GRE	Generic Routing Encapsulation
HA	high availability
HADR	high availability and disaster recovery
HIDS	host-based intrusion detection system
HIPAA	Health Insurance Portability and Accountability Act
HMC	Hardware Management Console
HNDL	Harvest Now, Decrypt Later
HPCS	Hyper Protect Crypto Services
HSM	Hardware Security Module
IAM	identity and access management
IASP	independent auxiliary storage pool
IBM	International Business Machines Corporation
ICAT	Interactive Code Analysis Tool
ICEE	IBM Compliance Expert Express
ICS	internet connection server
IDPS	intrusion detection and prevention systems
IDS	intrusion detection systems
IETF	Internet Engineering Task Force
IFS	Integrated File System
IKE	Internet Key Exchange
IMM	Integrated Management Module
IPA	Identity, Policy, and Audit
IPL	initial program load
IPS	intrusion prevention systems
IPsec	Internet Protocol Security
IRP	Incident Response Plan
ISA	instruction set architecture
ISO	International Organization for Standardization
ITAR	International Traffic in Arms Regulations
JFS	Journaled File System
JFS2	Enhanced Journaled File System
JSSE	Java Secure Socket Extension
KDB	kernel debugger
KMIP	Key Management Interoperability Protocol
KPI	key performance indicator
KST	Kernel Security Table
KVM	Kernel-based Virtual Machine
LDAP	Lightweight Directory Access Protocol
LIC	Licensed Internal Code
LPAR	logical partition
LUKS	Linux Unified Key Setup
LV	logical volume
LWE	Learning With Errors
MAC	mandatory access control
MDM	mobile device management
MDR	Managed Detection and Response
MFA	multi-factor authentication
MSS	Merkle Signature Scheme
MTMS	machine serial number
NAS	network-attached storage
NAT	network address translation
NFS	Network File System
NIDS	network intrusion detection system
NIST	National Institute of Standards and Technology
NSM	Network Security Monitoring
NTP	Network Time Protocol
NX	Nest Accelerator
ODM	Object Data Manager
OVAL	Open Vulnerability and Assessment Language
PAM	Pluggable Authentication Module
PASE for i	IBM Portable Application Solutions Environment for i
PCI DSS	Payment Card Industry Data Security Standard
PCIe	PCI Express
PEP2	Power Enterprise Pools 2

PER	Power Edge Router
PFS	physical file system
PFW	partition firmware
PIN	personal identification number
PKCS	Public Key Cryptography Standards
PKS	platform keystore
PQC	Post-Quantum Cryptography
PTF	Product Temporary Fix
PV	physical volume
PVC	Persistent Volume Claim
QSE	Quantum-Safe Encryption
RAS	reliability, availability, and serviceability
RBAC	role-based access control
RCAC	row and column access control
RCS	Right Computer Systems
RHEL	Red Hat Enterprise Linux
ROP	return-oriented programming
RTC	Real Time Compliance
S2I	Source-to-Image
SAN	Storage Area Network
SAST	Static Application Security Testing
SAVSYS	Save System
SC	security context
SCC	Security Context Constraints
SCUP	Smart Card Utility Program
SDN	software-defined networking
SIEM	Security Information and Event Management
SLA	service-level agreement
SMBs	small and medium businesses
SOC	System and Organization Controls
SOC	security operations center
SP	Service Pack
SRM	Service Responsibility Matrix
SSG	SCAP Security Guide
SSH	Secure Shell
SSL	Secure Sockets Layer
SSLv2	Secure Sockets Layer 2.0
SSLv3	Secure Sockets Layer 3.0
SSO	single sign-on
SST	System Service Tools
STIG	Security Technical Implementation Guide
STRSST	Start SST
SVIPC	System V Interprocess Communication
TCB	Trusted Computing Base
TE	Trusted Execution
TLS	Transport Layer Security
TME	Transparent Memory Encryption
TOTP	time-based one-time password
TPM	Trusted Platform Module
TSD	Trusted Signature Database
UFW	Uncomplicated Firewall
UID	user ID
UOV	Unbalanced Oil and Vinegar
vHMC	virtual Hardware Management Console
VIOS	Virtual I/O Server
VM	virtual machine
VMM	virtual machine monitor
VPC	IBM Virtual Private Cloud
VPN	virtual private network
VRA	Virtual Router Appliance
VSI	Virus Scan Interface
vTPM	virtual Trusted Platform Module
WORM	write-once, read-many
XSRF	cross-site request forgery

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *Data Resiliency Designs: A Deep Dive into IBM Storage Safeguarded Snapshots*, REDP-5737
- *IBM Power Systems Cloud Security Guide: Protect IT Infrastructure In All Layers*, REDP-5659
- *IBM Storage DS8000 Safeguarded Copy: Updated for DS8000 Release 9.3.2*, REDP-5506
- *Implementing, Tuning, and Optimizing Workloads with Red Hat OpenShift on IBM Power*, SG24-8537
- *Introduction to IBM PowerVM*, SG24-8535
- *Security Implementation with Red Hat OpenShift on IBM Power Systems*, REDP-5690

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- Cloud Management Console Cloud Connector Security white paper:

  https://www.ibm.com/downloads/cas/OGGYD9OY
- IBM AIX Documentation on Security:

  https://www.ibm.com/docs/en/aix/7.3?topic=security
- Modernizing Business for Hybrid Cloud on Red Hat OpenShift Video Series:

  https://community.ibm.com/community/user/power/blogs/jenna-murillo/2024/01/29/modernizing-business-for-hybrid-cloud-on-openshift
- Red Hat OpenShift Documentation on Configuring your Firewall:

  https://docs.openshift.com/container-platform/4.12/installing/install_config/configuring-firewall.html

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

IBM Power Security Catalog

SG24-8568-00

ISBN 0738462012

(0.5” spine)
0.475”<->0.873”
250 <-> 459 pages

SG24-8568-00

ISBN 0738462012

Printed in U.S.A.