Secure Network Design Techniques for Safety System Applications at Nuclear Power Plants

A Letter Report to the U.S. NRC

September 20, 2010

Prepared by:
John T. Michalski, Francis J. Wyant, David Duggan, Aura Morris, Phillip Campbell, John Clem, Raymond Parks, Luis Martinez, and Munawar Merza

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185

Prepared for:
Paul Rebstock, NRC Program Manager
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Engineering
Digital Instrumentation & Control Branch
Washington, DC 20555-0001

U.S. NRC Job Code:
JCN N6116

Abstract

This report describes a comprehensive best practice approach to the design and protection of a modern digital nuclear power plant data network (NPPDN). The important network security elements associated with the design, operation, and protection of the NPPDN are presented. This report includes an examination and discussion of newer proposed designs of modern Digital Safety Systems architectures and their potential design and operational vulnerabilities. The report explains the security issues associated with a modern NPPDN design and suggests mitigations, where appropriate, to enhance network security. Reference and discussion of the application of relevant regulatory guidance for each of the topics are also included.

Contents

Figures

Tables

This page intentionally blank

Executive Summary

This report explains the elements of network security and how they can be applied to a nuclear power plant data network (NPPDN). The report identifies and examines some practices for designing and deploying nuclear power plant (NPP) network architectures and their associated components. Throughout, the report presents observations associated with proposed design implementations. Accompanying appendices aid the understanding of security-related issues.

A digital safety system (DSS) architecture is presented along with protocols and components associated with DSS design. The report identifies three protocols including a popular Field bus protocol—Process Field Bus (PROFIBUS)—along with its important features and protections. The Ethernet protocol, its description, typical implementations, and some of its limitations and vulnerabilities are also discussed with additional information in the appendices. The fiber distributed data interface (FDDI) is presented because of its potential use in newer safety system designs.

Appendix D describes network components associated with modern best practice designs. Network component discussions include the use of virtual private networks (VPNs) and how they can be used to secure external connections originating from the plant data network. Also included is a discussion of virtual local area networks (VLANs) and how they are used to improve network security within a data plant network. Border network protection mechanisms are described, including an overview of firewall and Intrusion Detection Systems (IDSs), proper placement, and use. A section discusses host-based access control of both user and application processes, and implementation of role-based access control (RBAC).

Appendix D also discusses a wireless architecture with important elements for NPPDN protection from unauthorized access originating from the wireless medium. Finally, the report explains the importance of compliance testing to ensure that defined network security policies are being properly implemented and that policies are still relevant for NPPDN protection.

The protocols, procedures, and protections described in this report are relevant to modern networks being designed and deployed today. The primary elements necessary to provide comprehensive network security include the development of a security policy that provides a framework by which all responsible plant personal identify the important network aspects and create a plan to secure network access and operation. Protecting network access includes a discussion on important aspects of physical security implementation. The importance of maintaining security throughout the development, installation, operation, and maintenance of the network is also reviewed through proper lifecycle analysis. A large number of reported incidents involve known and addressable cyber threat vectors. Many types of security incidents could have been mitigated if better security policy, practices, and education programs were implemented instead of solely using technology-based solutions.

Modern designs of NPPDN are continuing to incorporate advances in network communications and data distribution. These advances will impact the network plant architectures associated with the operation of nuclear power electricity production. Within this push for modernization, the ability to isolate safety system processes from external, less trusted networks becomes more

difficult. Modernization of plant, safety, and control networks create the potential for secondary cyber pathways into the safety and control system networks. Proper risk mitigation starts with a comprehensive security policy management program that covers all aspects of the plant data network to include both the control and safety network systems. This policy should include both cyber and physical security to guide the proper implementation of a comprehensive, in-depth defensive strategy. This includes 1) better layering of firewall defenses, data-communications monitoring with intrusion detection and intrusion prevention systems for both wired and wireless mediums; 2) proper hardening of end-point devices and user interface configurations including authentication, patch management, and antivirus deployment; and 3) protecting internal and external data communications through the proper use of both VLAN and VPN technologies. It will also require a continual vigilance in the review and understanding of the security impacts to safety systems when newly proposed technologies are inserted.

Acronyms

ACL	access control list
AH	authentication header
AI	Acquire and Implement
AES	Advanced Encryption Standard
ANSI	American National Standard Institute
ARP	Address Resolution Protocol
BID	bridge ID
BPDU	Bridge Protocol Data Unit
CA	Certificate Authority
CAM	content addressable memory
CAT	category
CB-WFQ	class-based weighted fair queuing
CDA	critical digital asset
CDDI	Copper Distributed Data Interface
CFI	Canonical Format Indicator
CIM	customer interface management
CIP	critical infrastructure protection
CobiT	control objectives for IT and related technology
COTS	commercial-off-the-shelf
CRC	cyclic redundancy check
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access/Collision Detection
CTS	clear-to-send
DAC	dual-attached concentrator
DAS	dual-attachment station
DCS	distributed control system
DHCP	Dynamic Host Configuration Protocol
DMZ	demilitarized zone
DoS	denial-of-service
DP	Decentralized Peripheral
DS	Deliver and Support
DSL	digital subscriber line
DSS	digital safety system
DSSS	direct sequence spread spectrum
EAP	Extensible Authentication Protocol
EPROM	erasable programmable read only memory
ESP	Encapsulated Security Payload
FCC	Federal Communications Commission
FCS	Frame Check Sequence
FDDI	Fiber Distributed Data Interface
FDL	Field Bus Data Link
FDM	frequency domain multiplexing
FHHS	frequency hopping spread spectrum

FIFO	first-in-first-out
FTP	File Transfer Protocol
GUI	graphical user interface
HASH	(A well define procedure or mathematical function)
HIDS	Host-Based Intrusion Detection System
HMI	human/machine interface
HTTP	Hyper Text Transfer Protocol
HTTPS	Secure HTTP
HWFQ	hierarchical weighted fair queuing
I&C	instrumentation and control
ICCP	Inter-Control Center Communications Protocol
ICMP	Internet Control Message Protocol
ICV	Integrity Check Value
ID	identification; identifier
IDS	Intrusion Detection System
IED	intelligent electronic device
IETF	Internet Engineering Task Force
IKE	Internet Key Exchange
I/O	input/output
IP	Internet Protocol
IPS	Intrusion Prevention System
IPSec	Internet Protocol Security
ISA	Information Systems Audit
ISAKMP	Internet Security and Key Management Protocol
IS&M	Industrial, Scientific, and Medical
ISM	inventory and supply management
ISO	International Standards Organization
IT	information technology
ITGI	Information Technology Governance Institute
KVM	keyboard, video monitor, and mouse
LAN	local area network
LR-WPAN	low rate wireless personal area networks
MAC	Media Access Control
ME	Monitor and Evaluate
NAV	network allocation vector
NERC	North American Electric Reliability Corporation
NIDS	Network Intrusion Detection System
NIST	National Institute of Standards and Technology
NPP	nuclear power plant
NPPDN	nuclear power plant data network
NRC	Nuclear Regulatory Commission
NUREG	NRC report designation
OPC	Object linking and embedding for Process Control
OS	operating system
OSI	open system interconnection

PBX	private branch exchange
PDU	protocol data unit
PI	plant information
PIN	personal identification number
PKI	public key infrastructure
PLC	programmable logic controller
PMD	physical-medium-dependent
PO	Plan and Organize
PPS	physical protection system
PROFIBUS	Process Field Bus
PSTN	public switched telephone network
QDS	Qualified Display System
QoS	quality of service
RADIUS	Remote Authentication Dial-In User Service
RBAC	role-based access control
RED	Random Early Detection
RF	radio frequency
RFID	radio frequency identification
RG	Regulatory Guide
RISC	reduced instruction set computer
RTOS	real-time operating system
RTS	request-to-send
RTU	remote terminal unit
SA	security association
SAC	single-attached concentrator
SAID	Security Association Identifier
SAS	single-attachment station
SCADA	Supervisory Control and Data Acquisition
SGID	Set Group Identifier
SMIB	security management information base
SMT	station management
SNMP	Simple Network Management Protocol
SQL	Structured Query Language
SSH	Secure Shell
SSL	Secure Sockets Layer
STP	Spanning Tree Protocol
SU	switch user
SUID	Set User Identifier
TACACS	terminal access controller access control system
TCI	tag control information
TCP	Transmission Control Protocol
TCP/IP	Transmission Control Protocol/Internet Protocol
TDM	time domain multiplexing
TFTP	Trivial File Transfer Protocol
TIM	Traffic Indication Map

TLS	transport layer security
TPID	Tag Protocol Identifier
TTRT	target token rotation time
UDP	User Data Protocol
USB	Universal Serial Bus
V&V	verification and validation
VLAN	virtual local area networks
VM	virtual machine
VoIP	Voice over Internet Protocol
VPN	virtual private network
VTP	VLAN Trucking Protocol
VTY	Virtual Teletype Terminal
WAN	wide area network
WAP	wireless access point
WDM	wave division multiplexing
WEP	Wire Equivalent Privacy
WFQ	weighted fair queuing
WLAN	wireless local area network

1. Introduction

1.1 Background

The information technology (IT) communication infrastructure continues to transform the way our nation and the world communicate and conduct business. Benefits include information generation, distribution, and utilization for electric power utility companies, for whom IT is essential. However, IT interconnectivity also poses significant risks to national safety and automated infrastructures and critical operations by providing pathways—

1. To compromise utility information.
2. For unanticipated access to the safety and control systems.
3. For adversaries to gain access to critical status and control assets within utility business and safety and control networks.

Nuclear power plants (NPPs) use digital instrumentation & control (I&C) systems to monitor and control safety and non-safety–related systems. Due to obsolescence, rising maintenance costs, and the need for more efficient operations, modern digital microprocessor-based systems are replacing or partially upgrading analog safety systems. Digital systems provide a high degree of automation to enhance plant operation, reduce operator burden, and improve situational awareness during normal and accident conditions.

Digital systems are creating new challenges for the NPP industry and U.S. Nuclear Regulatory Commission (NRC) regulators, who must familiarize themselves with the new technology to ensure new systems meet all reliability, performance, and security requirements.

The main components of the safety-related I&C system are sensors and actuators that monitor and adjust plant processes, the input/output (I/O) modules that interface sensors and actuators to the control logic unit, the automatic controller for achieving and maintaining desired process states, and the operator and engineering workstations stations that allow human-in-the-loop oversight and interrogation of the safety system information. To enable these safety system components to seamlessly work together, an underlying communications network becomes the essential backbone of the design.

The communication network must be highly reliable, maintainable, and independent to provide the necessary assurance that the safety system will perform its mission. The communication network should also provide the necessary data bandwidth to convey required system-operational information to the user through a user-friendly man-machine interface.

1.2 Scope and Purpose of Report

This report explains elements of network security and how they can be applied to utility networks. This report is intended to identify *best practice* approaches for a comprehensive view of network security as it pertains to the design, operation, and maintenance of a modern nuclear power plant data network (NPPDN). In particular, the body of this report focuses on the digital safety systems (DSS) that are being incorporated into modern NPP designs. Vulnerabilities associated with the procedures, configurations, and design of an NPPDN can expose network

operations to adversaries interested in gaining a competitive advantage or wanting to compromise aspects of network operation.

Having an understanding of the best security practices for implementing a network can help plant designers plan and build more effective defenses to prevent the disruption of operations and provide for proper contingencies. Understanding network-architecture strengths and weaknesses will provide management with the criteria and guidelines to create security policies that implement comprehensive approaches for secure network operations.

1.3 Report Structure

Chapter 1 presents, background, scope, and purpose. Chapter 2 briefly describes important elements of network security and provides information about secure-network lifecycle phases. Chapter 3, the heart of this report, describes the hardware and architectural features of a modern DS-based on planned future designs for advanced nuclear plants. In chapter 3 and Appendix D, identified vulnerabilities are provided within the *security observation* discussion for each specific technology. The security observation section includes commentary on internal and external threats against the network technology and provides mitigation suggestions that can improve a network security profile. Within the discussion context the term *external threat* refers to an adversary with no authorized physical or cyber access to the NPPDN. The term *insider* can refer to different levels of privilege within an organization. The term *unprivileged insider* refers to a threat that has both physical access and cyber access to the NPP, but at some reduced or controlled level. This may include a contractor, a product vendor, or an employee with limited administrative rights. The term *privileged insider* refers to someone with primarily unlimited physical and cyber administrative access to a particular area or multiple areas of the NPP. Chapter 4 summarizes conclusions resulting from this investigation. Appendix A contains a bibliography of relevant documents compiled during the creation of the report and includes information sources for recommended protocols and procedures that comprise modern *best practices* for the network security techniques described. Appendix B contains a matrix of security findings mapped into best practices categories from 42 distinct assessment reports. These assessments were performed on various control systems and enterprise information systems connected to control systems during the past eight years by Sandia National Laboratories' Red Team and Assessment group. Appendix C lists general sources and features that contribute to the vulnerabilities associated with electric generating stations. It also identifies vulnerabilities that can be mitigated by adopting or strengthening the key elements of secure networks presented in this report. Appendix D contains a more detailed description of the overall NPPDN and provides discussion and commentary on the proper deployment of modern network-related devices, protocols, and applications. References are provided as part of the individual topical areas in this appendix. Appendix D supplements the information and insights about the particular network devices presented in chapter 3, but in relation to the NPPDN as a whole rather than focusing only on safety-related functions.

2. Elements of Secure Networks

The U.S. Code of Federal Regulations (CFR) now requires nuclear plant licensees to develop a cyber security plan that provides assurance that their digital computer and communications and networks are protected against cyber attack. Secure networks depend on management, organizational, and design elements to ensure their integrity. Important elements consist of a security policy including a framework that provides a comprehensive overview of security needs. This helps guide the identification and implementation of security products and procedures. Although this report focuses on network-based security, another important element is the physical protection of network elements and assets. Without proper physical protections in place, operations can be compromised. The design architecture can also play a part in the protection of data and processes by creating the proper cyber barriers to potential exploits. User management processes, including authentication mechanisms that validate user roles and access levels, are necessary to control access and provide audit logs of user activities. Finally, a compliance process is required to review and audit proper adherence to established security policies.

2.1 Security Policy

An overall security policy is paramount to creating an operational environment where an effective and comprehensive security program can be developed. Through security policy, management can express security requirements to be incorporated into security objectives. These objectives can provide a top level approach to assist in developing an effective security architecture. If staff have no comprehensive security policy to review, then each department of the utility company will make individual, non-comprehensive decisions on how they think best to design, operate, and maintain networked systems under their control. This can result in dissimilar approaches to establishing security features and can cause implementation of a less-than-optimal secure architecture.

A *policy framework* describes generic[1] policy items that plant managers designate as required to meet the overall security needs of the plant. In addition, a policy framework strives to describe all possibly-necessary generic policy items. The intent is not for all organizations to adopt every item, but rather to enable each organization to avoid overlooking something. Each organization should adopt just what it considers necessary to meet its security goals.

To begin a policy framework is easy—passwords should be in the list. But it is difficult to judge when the list is sufficient. To complicate matters, the list must change as threats, vulnerabilities, assets, usage, and expectations all change. The policy framework provides a current estimation of what is sufficient for the list. A good example of an application that provides a framework for policy development is the Control Objectives for IT and Related Technology (CobiT) policy framework. This application provides breadth and depth, and it is maintained and can be downloaded at the Information Technology Governance Institute (ITGI) Website, www.Itgi.org. (Appendix D, section D.1, Policy Framework Details, offers additional insights on developing a policy-based framework.)

[1] i.e., not yet specific to an organization

2.1.1 Regulatory Guidance Regarding Security Policy

In some cases, regulatory requirements will govern the scope of the security policy. For example, Regulatory Guide (RG) 5.71, Cyber Security Programs for Nuclear Facilities [1], Position C.3.5 requires a licensee to develop and maintain site-specific policies and procedures as part of the cyber security program. The RG also provides examples of security-related controls that are based on the National Institute of Standards and Technology (NIST) cyber security standards. In part, the recommended set of policy and procedures covers access controls, training, configuration management, identifying and protecting critical digital assets (CDAs) and communications, user identification and authentication, personnel security, physical and operating environment protection, incident response, and contingency planning.

In RG 1.152, Criteria for Use of Computers in Safety Systems of Nuclear Power Plants, rev. 2 [2], Position C.2.6 states that the licensee should have a digital security program in place wherein policies, standards and procedures ensure that a new installation of a computer-based safety system will not compromise the security of any system or of the plant.

2.2 Physical Security

Physical security is always the principal defense in preventing unauthorized access, corruption of informational assets, and intentional or unintentional destruction, or theft of property. Many documented cyber attacks against organizations have been initiated by having physical access to elements of the network architecture in order to execute the penetration.

Physical security is commonly referred to as "guns, gates, and guards." This simplistic label fails to capture the increasingly complex and interdependent relationship that physical security has with cyber security. While the intent of this technical report is to focus on network security issues—especially electronic security for safety and control systems—that security depends partly on physical security.

In prior decades utilities enjoyed nearly complete segregation between control networks and business/administrative networks, in terms of connectivity and communication protocols. Today that has changed; network designs support intercommunication requirements among various departments of an organization to meet real-time business needs. Furthermore, physical security at plants was designed to address assets of the physical world, i.e., people, facilities, fuel, and plant equipment.

Today's evolving utility-network environment threatens needed safety and security that isolation once helped provide. Now physical security and cyber security have significant overlap and interdependencies. Physical security contributes to the effectiveness or ineffectiveness of the cyber security and vice versa. It must be remembered that threats to computer systems and networks are not just in remote far-away places. Threats against critical cyber assets can be made by attackers (outsiders or insiders) gaining physical access to systems or network equipment. In fact, many documented cyber attacks against organizations have been initiated through physical access to elements of the network architecture, (e.g., network devices or network cabling). Having sufficient physical security for the plant's critical cyber assets is necessary to decrease the

risk of compromise by an adversary. (Appendix D, section D.2, Physical Security Details, provides more information on physical security.)

2.2.1 Regulatory Guidance Regarding Physical Security

Regulatory guidance recognizes the importance of physical security in protecting digital systems from unauthorized access. For example, part of the RG 1.152, rev. 2 discussion states that control of physical and electronic access should be implemented to prevent unauthorized changes. Additionally, Position C.2.3 states that the safety system design should address physical access to the system, based on the results of a cyber security risk analysis. Position C.2.6 of RG 1.152 requires security testing to verify and validate the physical security features in the target environment as part of the Installation, Checkout, and Acceptance Testing phase of the system lifecycle.

RG 5.71, Position C.3.4 states that 10 CFR 73.54 (b)(3) requires the licensee to incorporate their cyber security program into its physical protection program. Further, in Appendix A, section A.3.2 of RG 5.71 indicates that licensees should establish *unified* physical and cyber security controls. RG 5.71 requires physical protection of CDAs and maintenance of their operating environment so they function properly. One important aspect of this protection area is the need to control physical access to the communication pathways employed by the CDAs, as well.

2.3 System Architecture

Historically, the plant network architectures associated with nuclear-power–electricity production have used analog implementations that provided the proper isolation and physical segregation of the energy production components. These components included process control, data acquisition, performance, and safety. The isolation from external, un-trusted networks or practices allowed the plant to manage the data and communication security by physical access restrictions to the facility or through operational elements. Because of this design approach the data and the command instructions required a minimal level of additional security or oversight. Since the processes of energy production, safety, and process I&C occurred in an isolated analog environment, all communications were considered reliable and secure.

Today as organizations are facing the realities of introducing digital systems to replace antiquated analog elements, the need to modernize the data and communication infrastructure has created pressures to break the traditional barriers that isolated, and physically protected, the components of energy production. This report will address the best practices for securing the modern data and communications networks. Figure 2-1 displays a modern and integrated data and communications architecture.

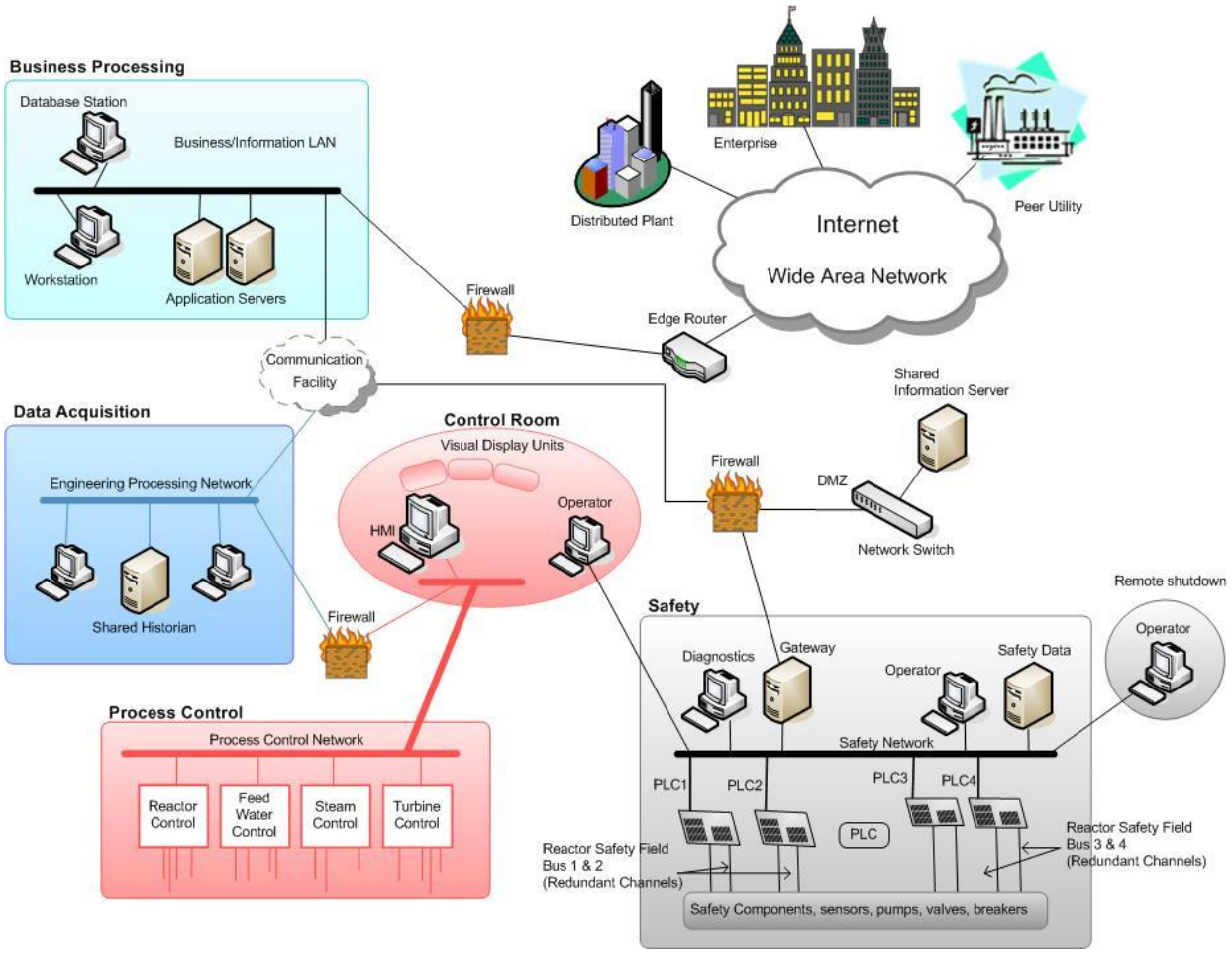

Figure 2-1. Digital Plant System Network Architecture

The architecture can provide an adversary (both internal and external) with various paths to reach data and I&C devices within the energy production components. The design of these networks requires the exchange of data from different areas of operation, which can benefit an adversary. On the other hand, as seen in Figure 2-1, networks for modern industrial process and control environments are segmented into logical (network addressing) and physical locations. They include the following:

- **Business Performance and Information Network**—IT system computers, such as operations planning and maintenance (OPM), inventory and supply management (ISM), and business performance, and customer interface management (CIM) computers are normally associated with this network segment including any general enterprise desktop with server/client applications.

- **Data Acquisition/Engineering Processing Network**—Energy processing activities include fuel efficiency analysis, energy output management, and fuel burn rate calculations; also energy balancing computers are connected to this network segment. This network segment is connected

to the business performance and information network through a firewall and also to the control room through a separate firewall.

- **Process Control Network**—This network segment includes process control devices used to monitor status and command field interface devices along with human/machine interface (HMI) consoles. Information exchanged between the field devices—and which includes measured and manipulated variable data and controller configuration information—can reside on this segment. The process control network is connected to the control room and also can be connected to the data acquisition and information network through gateway firewalls.

- **Safety Network Segment**—Safety status and protection elements reside on this network segment. It is accompanied by Class 1E safety-related alarm and display systems. Field device components and general purpose computers exchange information. Such information would include measured variable data for safety components, such as temperature and pressure sensors, pumps, valves, and electric circuit breakers. This safety segment can be connected directly to the control room and indirectly through a firewall to populate safety status information on shared information servers. Chapter 3, Digital Safety System Instrumentation & Control, provides a more detailed description.

- **Field Network Segment**—The process control and safety networks are both attached to field devices, such as sensors and actuators; Class 1E separation requirements for both safety and process control segments determine their connections. Field network devices typically use Field bus protocols that are proprietary in nature. The controllers that interface with the Field bus have less computing capability. Also, controller configuration and maintenance information can also reside on this segment.

Although networked plant safety I&C systems are being built on IT protocols and design practices, there are still operational differences between IT and safety and control systems that can impact how security measures are applied. Some differences are as follows:

- **Critical Asset Security**—In commercial IT systems, the primary asset to secure is the information, normally stored on servers. In an energy production system, the devices that control and protect the system (the safety I&C system) are just as important as the information stored on a server.

- **Availability and Reliability**—Energy production processes are 24/7. Any disruptions that create energy outages to the system are critical in nature. This implies a high level of preliminary acceptance testing prior to changing or upgrading system components.

- **Risk Management**—In energy production environments, of great importance are human safety to prevent loss of life, and public safety to prevent endangerment and to prevent loss of confidence.

- **Software and Resource Constraints**—Some systems have customized operating systems (OS) or real-time operating system (RTOS) and have embedded systems that cannot handle

typical IT software applications and practices. Safety and control networks can be more complex. Control engineers with differing levels and types of expertise (than IT staff have) manage these more complex systems.

- **Time Responses**—The response time for when an IT infrastructure server fails can be vastly different from that for a system component failure in an energy control system environment. For some energy systems, human interaction or designed automated response times are critical. Some security applications may impede or hamper the system response time to an event.

(For a more detailed discussion on implementing proper security components associated with the digital plant system network architecture [shown in Figure 2-1], see Appendix D, section D.2, Physical Security Details.)

2.3.1 Regulatory Guidance Regarding System Architecture

As part of the process for identifying critical systems and associated critical assets, RG 5.71 recommends first identifying the organization of all plant systems, equipment, communications, and networks that directly perform or support safety, security, and emergency preparedness (SSEP) functions. This is then followed by a site-specific consequence analysis to determine those critical system CDAs that if failed, compromised or exploited could impact the SSEP functions of the plant. In order to create an effective layered defense, it is essential to understand how the CDAs and critical systems fit together and their relationships to non-SSEP systems. Equally important is knowing the interconnectivity between the SSEP systems and components. Dividing the NPPDN into architecture zones helps establish clear boundaries around functional areas that define the protection level for each network segment. This approach provides a layered security design around the CDAs—much like the layers of an onion.

Position C.2 of the RG, in part, states that the cyber security plan must employ defense-in-depth protective strategies. Position C.3.2 explains that defense-in-depth is intended to provide a security architecture, which recognizes that any one point of protection may be defeated. Consequently, defense-in-depth provides layers of security and detection that provide multiple barriers that attackers must break through or bypass without being detected in order to reach critical assets. In other words, the defense-in-depth goal is to forces the attacker to perform flawlessly while being kept as blind and ignorant of the system as possible. Defense in depth strategies would include the use of Intrusion Detection Systems (IDS), firewalls, and demilitarized zones (DMZ)[2].

RG 1.152, Position C.2.5 requires that the system hardware architecture be checked for integrity and unauthorized pathways as part of validation testing. The intent is that verification of the system configuration and designed security features are correctly enabled and demonstrated by the developer prior to installation in the plant.

---

[2] Demilitarized zones are network segments located as perimeter defense points used to isolate internal (protected) network elements from external, untrusted sources.

2.4 User Management

The user and operational management element of secure networks requires management of interactions between personnel and network assets. This includes defining 1) personnel roles within the network, 2) the amount of access required to perform each role, and 3) the ability to enforce each user's role(s). These are all essential aspects of user management.

The process of defining roles becomes paramount in determining the level of access to the system it provides. Plant operation should be well defined and understood prior to setting up any role-based access control (RBAC) interface. When setting up an RBAC account, it is important the user is given no more privilege than necessary for job performance. This *least-privilege* concept requires more precision when defining a specific user's job function.

User authentication mechanisms are necessary to control access and provide audit logs of user activities on the network. The simplest user authentication is a single, personal factor like a password. This may be sufficient if there are additional physical security measures limiting access. Passwords should be strong enough to prevent password guessing within a timeline that must be calculated from the lesser of password expiration deadline or user audit log verification cycle. If the password can be determined through brute-force, dictionary attack, or HASH look-up (rainbow tables) before the password has expired or the audit log is verified, then adversaries could gain access to the host. (Appendix D, section D.13, User/Operational Management, provides a detailed discussion on user management.)

2.4.1 Regulatory Guidance Regarding User Management

RG 5.71 requires licensees to implement access controls to protect CDAs from unauthorized persons and/or process interactions. The means by which to protect the CDAs include defining access control rights and privileges, system hardening, annual auditing and management of the CDAs, and maintaining separation of duties. The item B.1.6 in Appendix B to RG 5.71 advocates employing a least-privilege philosophy as part of the access control policy. The intent is to limit the extent of damage any single insider threat may be able to cause.

Similarly, RG 1.152 states that electronic access to safety systems should be controlled through network connections as well as though maintenance equipment to prevent unauthorized changes to plant safety systems and to prevent changes to operator displays. Position C.2.1. states that remote access to the safety systems should not be implemented. Position C.2.3 states that the design phase should address logical access and communication with other systems. Position C.2.4 recommends testing the system, with scanning if warranted, to locate undocumented or malicious codes or functions that might allow unauthorized access.

2.4.2 Compliance

To provide proper adherence to established policy, each policy objective should also have a means to ensure policy implementation. Compliance assurance should include validation testing to determine if the goal of establishing a security baseline has been accomplished and is being maintained. Each appropriate area of the plant should undergo an independent audit to establish

whether the security is being maintained when measured against acceptable compliance criteria. If deficiencies are discovered, a corrective action plan should be created to resolve the security infractions. The following summarizes steps to initiate and maintain a compliance plan:

1. **Establish Compliance Requirements**—From the established security policy, once a system has been finalized for operations, *identify* the necessary security controls required to maintain proper operational security.

2. **Confirm the Baseline—***Validate* that the applicable security controls for each operations area within the plant have been installed.

3. **Audit Metrics**—Identify the type of metrics for each periodic audit to determine if the identified security control is *performing* its intended function. This will determine compliance and identify operational trends.

4. **Reporting Procedure**—Establish a *timetable* for audit reviews and send out resulting performance metrics for appropriate stakeholder review.

5. **Corrective Action**—Each stakeholder responsible for security compliance should specify what action or actions are required to remedy any non-compliance situations.

2.4.2.1 The Technical Side of Compliance Checking

It is important to regularly test the security profile of the network environment to determine if the network is operating in a secure fashion. This *security check* does not only have to be performed during the required security audit cycle; there are many applications and tools that can help the network administrator automate many of these security checks.

There are host-based scanners that can provide a report of the applications that are resident, provide account profiles to determine who is allowed on the machines, and provide a list of processes or services running on the host. This information can be reviewed to determine if the security profile is consistent with the security policy.

There are network-based scanners that use the communication protocols, such as Transmission Control Protocol (TCP)/Internet Protocol (IP) or User Datagram Protocol (UDP)/IP and Internet Control Message Protocol (ICMP)/IP to map a network to identify all active devices. These network scanners can identify the operation system and the active ports on the network devices. They can also simulate intrusions to test the installed IDS to determine if they are functioning properly. These network scanners may contain a vulnerability database that has vulnerability information from active vulnerability repositories, such as the United States Computer Emergency Readiness Team (US-CERT) or vendor advisories, such as “bugtraq” to report discovered vulnerabilities for each device. Network scanners along with host-based scanners can provide the network administrator an active view of the security posture of the system. Note that active scans of an operating network may have a detrimental effect on device operations and must always be reviewed to determine risks prior to implementation.

2.4.2.2 Regulatory Guidance Regarding Compliance

RG 1.152, rev 2, Position 2.7 states that the licensee should perform periodic testing and monitoring of the system, review system activity logs, and perform real-time monitoring during the operational phase of the system lifecycle to ensure that system security is intact.

RG 5.71 requires periodic auditing of the elements of a licensee's cyber security program. These audits ensure the program is effective and identify and correct any non-compliance findings. RG 5.71, Position C.4.3 states that 10 CFR 73.54 (g) requires the cyber security program to be reviewed as a component of the physical security program, following the periodicity requirements of 10 CFR 73.55 (m). These requirements include performing a program review at least every 24 months and the following:

- Within 12 months of initial implementation.
- When a change is made that could have an adverse impact on security.
- As deemed necessary based on site-specific criteria.

These reviews are to be documented and kept available for inspection by the NRC.

2.5 Safety System Lifecycle

DSS security addresses potential security vulnerabilities as part of the system development process, and maintaining security of the system through its lifecycle. Security vulnerabilities in DSSs may be introduced inadvertently by a design flaw, misconfiguration, improper operation, or maliciously introduced into the system.

RG 1.152 endorses IEEE Std 7-4.3.2-2003, Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating Stations, as a method that the NRC staff has deemed acceptable for satisfying NRC regulations with respect to high functional reliability and design requirements for computers used in safety systems of nuclear power plants. RG 1.152 also provides additional staff guidance concerning computer-based safety system security.

The guide introduces *phases* of DSS development and operation, from concept-to-retirement. The DSS development process should address potential security vulnerabilities in each phase of the DSS lifecycle. The following discussion highlights lifecycle phases taken from the RG:

- In the concept phase, the licensee and developer should identify safety system security capabilities that should be implemented. For example, remote access to the safety system should not be implemented; computer-based safety systems may transfer data to other systems through one-way communication pathways.

- In the development phase, the licensees and developers should define the security functional performance requirements and system configuration; interfaces external to the system; and the requirements for qualification, human factors engineering, data definitions, documentation for the software and hardware, installation and acceptance, operation and execution, and maintenance. Security requirements should be part of the overall system requirements.

- In the design phase, the safety system security requirements identified in the system requirements specification should be translated into specific design configuration items in the system design description. The safety-system security-design configuration items should address control over 1) physical and logical access to the system functions, 2) use of safety system services, and 3) data communication with other systems.

- Physical and logical access control should be based on the results of cyber-security qualitative risk analyses.

- The developer should delineate the standards and procedures that will conform with the applicable security policies to ensure the system design products (hardware and software) do not contain undocumented code (e.g., back-door coding), malicious code (e.g., intrusions, viruses, worms, Trojan horses, or bomb codes), and other unwanted or undocumented functions or applications.

- In the system (integrated hardware and software) implementation phase, the system design is transformed into code, database structures, and related machine executable representations. The implementation activity addresses hardware configuration and setup; software coding and testing; and communication configuration and set-up (including the incorporation of reused software and commercial off-the-shelf [COTS] products).

- The developer should ensure that the security design-configuration item transformations from the system design specification (system features) are correct, accurate, and complete.

- The developer should implement security procedures and standards to minimize and mitigate tampering with the developed system. The developer should account for hidden functions and vulnerable features embedded in the code and their purpose and impact on the safety system. If possible, these functions should be disabled, removed, or (as a minimum) addressed (e.g., as part of the failure modes and affects analysis of the application code) to prevent any unauthorized access.

- The objective of testing security functions (test phase) is to ensure that the system security requirements are validated by execution of integration, system, and acceptance tests where practical and necessary. Testing includes system hardware configuration (including all external connectivity), software integration testing, software qualification testing, system integration testing, system qualification testing, and system factory acceptance testing.

- The security requirements and configuration items are part of validation of the overall system requirements and design configuration items. Each system security feature should be validated to verify that the implemented system does not increase the risk of security vulnerabilities and does not reduce the reliability of safety functions.

- During installation and checkout, the safety system is installed and tested in the target environment. The system licensee should perform an acceptance review and test the safety system security features.

- The security policies, standards, and procedures should ensure that installation of the digital system will not compromise the security of the digital system, other systems, or the plant.

- The operation phase of the lifecycle process involves the licensee using the safety system in its intended operational environment. During the operations phase to ensure that the system security is intact, the licensee should apply techniques such as periodic testing and monitoring, review of system logs, and real-time monitoring where possible.

The licensee should evaluate the impact of safety system changes in the operating environment on safety system security, assess the effect on safety system security of any proposed changes, evaluate operating procedures for compliance with the intended use, and analyze security risks affecting the licensee and the system. The licensee should evaluate new security constraints in the system, assess proposed system changes and their impact on system security, and evaluate operating procedures for correctness and usability.

- The maintenance phase is activated when the licensee changes the system or associated documentation. These changes may be categorized as follows:

    - Modification (i.e., corrective, adaptive, or perfective changes)
    - Migration (i.e., the movement of system to a new operational environment)
    - Replacement (i.e., the withdrawal of active support by the operation and maintenance organization, partial or total replacement by a new system, or installation of an upgraded system).

- The licensee should develop an incident response and recovery plan for responding to digital system security incidents (e.g., intrusions, viruses, worms, Trojan horses, or bomb codes). The plan should be developed to address various loss scenarios and undesirable operations of plant digital systems, including possible interruptions in service due to the loss of system resources, data, facility, staff, and/or infrastructure. The plan should define contingencies to ensure minimal disruption to critical services in these instances.

- The licensee should perform periodic computer system security self-assessments and audits, which are key components of a good security program. The licensee should assess proposed safety system changes and their impact on safety system security, evaluate anomalies that are discovered during operation, assess migration requirements, and assess modifications made including verification and validation (V&V) tasks to ensure that vulnerabilities have not been introduced into the plant environment from modifications.

- In the retirement phase of the lifecycle, the licensee should assess the effect of replacing or removing the existing safety system security functions from the operating environment.

2.6 Federal Information Security Management Act

In 2002, the Federal Information Security Management Act (FISMA) was signed into law and requires federal agencies, or the facilities they operate, to follow and maintain compliance with certain standards and guidelines. Enforcement is accomplished through reporting to the Office of Management and Budget (OMB) and through standard or regulatory requirements that NIST created and maintains. Throughout this report are references, as background, to several NIST Special Publications (SP), which are relevant to various best practices for security cyber assets. However, there is also an underlying regulatory foundation to FISMA and there are related NIST SPs. In cases where a federal agency manages and/or operates the nuclear power plant, then FISMA applies, as do all the related requirements (largely NIST SPs).

The Federal Energy Regulatory Commission (FERC) issued a finding requiring bulk energy providers to adopt the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) standards regarding cyber security for control systems. In the order that resulted from that finding, FERC left open the option of replacing the NERC CIP standards with another standards process; referring to the NIST SP 800-82 Industrial Control Systems Security (still a draft version). An advantage to the NIST SP standard process is the inclusion of relevant stakeholders, both from vendors and operators.

However, some have argued the NIST standards process is not all encompassing—meaning NIST guidance offers little value when addressing a specific technology, such as control systems. The process itself can take years to establish new standards. When a NIST standard does not exist for a given component within a PDN, delays or later integration challenges can result in the analog-to-digital transition. Furthermore, others have argued that FISMA compliance does not, in any way, describe a secure system. U.S. Congress and OMB have realized this too and intend to revamp FISMA to add this assurance by adopting more international standards with stronger controls.

NRC and NERC have agreed to divide their statutory responsibilities for regulating and enforcing cyber security requirements at commercial NPPs operating within the United States. Through a memorandum of understanding between the two agencies[3], NRC cyber security regulations govern those digital systems and networks that can affect commercial nuclear power reactor safety, security, and emergency preparedness functions. But NRC does not govern systems within nuclear facilities—such as those related to continuity of power—that cannot have an adverse impact on safety, security, or emergency preparedness.

[3] Federal Register: January 11, 2010 (Volume 75, Number 6) Notice, Page 1416-1418, Final Memorandum of Understanding Between the U.S. Nuclear Regulatory Commission and the North American electric Reliability Corporation.

3. Digital Safety System Instrumentation and Control

To actively demonstrate the importance of each element of secure networks, an example network has been developed for the purpose of discussion. This example network design was generated based on documented information gathered from the design certification documents of a new reactor design vendors. However, no one particular design is singled out as representative for this example network; rather, design elements from each of the vendor designs reviewed are represented to one degree or another. All the primary assessment information is contained in this section.

3.1 Architecture Description

The following observations are based on the review of several DSS designs proposed for implementation in new reactor plants. This section describes general aspects of the proposed designs and characterizes the approaches based on network security best practices. Figure 3-1 captures the important features of the communication network being deployed for proposed new digital designs. This architecture is not associated with one particular NPP design, but has been generated to create observation points and discussion to help characterize the communication elements of a DSS design.

The overall safety I&C architecture can be categorized into four layers:

- **Layer 1.** The *Process Instrumentation* Communication Layer, this lowest layer provides an interface between the *instrumented* plant sensors and actuators that are used to gather near real-time information on important plant processes and that are used to protect system elements.

- **Layer 2.** The *Automated Safety Layer*, Layer 2 receives the data from the process instrumentation layer and interacts with the safety system logic block that makes decisions to perform automatic system protection functions based on *set point* levels. Layer 2 also sends process data to the supervisory operator display and sends data to any non-safety–related processes.

- **Layer 3.** The *HMI Supervisory Layer*, Layer 3 displays the data provided by the automated safety layer to operators, located in a control room, who monitor plant processes and if necessary take manual control of plant components.

- **Layer 4.** The *Non-Safety Information Layer*, Layer 4 can be segregated into two additional layers: the non-safety process control layer and the business information layer. The business information layer contains the plant information management systems that are used to provide OPM, business performance, and CIM. The non-safety process control layer provides operator console interfaces and engineering stations to monitor and review all important plant information data.

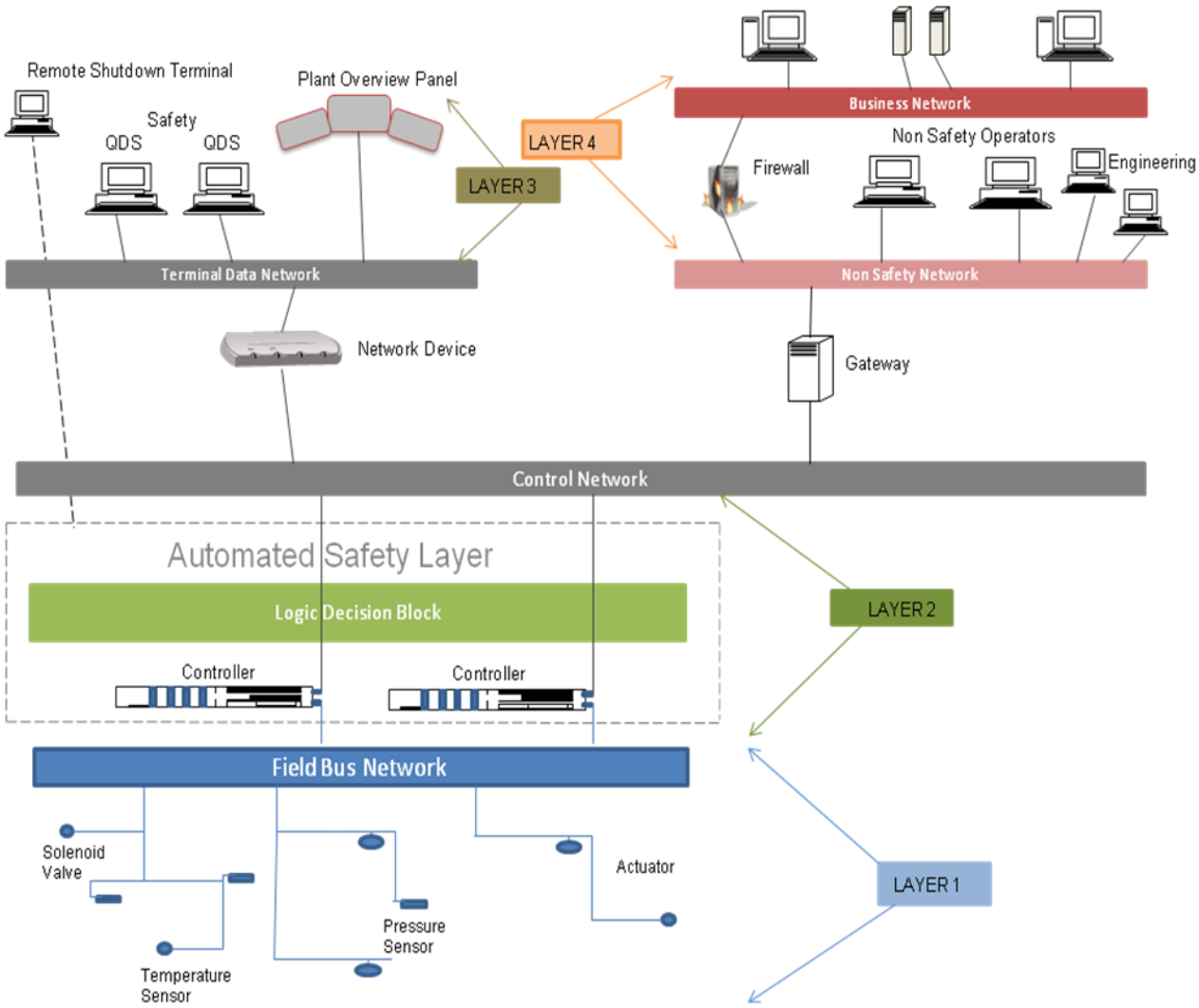

Figure 3-1. Digital Safety System Architecture

3.2 Process Instrumentation Communication Layer

3.2.1 Field Bus

As seen in Figure 3-1, the Layer 1 process instrumentation is implemented as a point-to-point or multidrop Field bus network. This network provides the interface to safety function elements, such as the protection system and safety automation system. End devices—such as temperature and pressure sensors, valves and actuators, which are used to monitor and control the operation of the plant—interact with the Field bus controllers to pass the data and commands to/from the safety logic elements.

Many Field bus protocols use the underlying RS-485 physical protocol. Formally known as TIA/EIA-485, this standard only specifies the electrical characteristics of the driver and the receiver. It does not specify or recommend any data link protocol. It can be implemented in twisted pair copper cabling or fiber optic cabling. It is used to create local and multidrop

networks for field device communications. It is inherently insecure and does not offer any security attributes.

There are a number of Field bus protocols that vary in terms of architecture deployment, data speed, I/O node densities, and operating environment. As the levels of system complexity, platform inoperability, and performance increase there is pressure within the Field bus manufacturers to move toward Ethernet-based communications. Field bus standards are being integrated into Ethernet for high performance, near real-time communications while maintaining their underlying capability to interface with equipment at the lowest level of control. The following is a list of Field bus protocols currently deployed in Field bus networks and also that have an integrated Ethernet-based equivalent:

- MODBUS Field Bus protocol
- Distributed Network Protocol Working Group 3 (DNP3)
- Lon Works
- Field Bus Foundation
- Controller Area Network (CAN)
- Device Net/Controller Net (CIP)
- Process Field Bus (PROFIBUS)

PROFIBUS protocol is commonly used for sensor operation and actuators through a centralized controller in discrete manufacturing and process control. Because of these design attributes, PROFIBUS is a popular choice for use in NPP status and control network processes and is discussed in the following section.

3.2.2 PROFIBUS

PROFIBUS is a defined protocol for Field bus communication in automation, safety, and process control technology. PROFIBUS has two protocol variations: the more popular Decentralized Peripheral (DP) and the less common process automation. PROFIBUS protocol specifies two layers within the open systems interconnect (OSI) communication stack: the application layer (also referred to as the user interface) and the data link layer. The International Standards Organization (ISO) has created a layered model called the OSI model to describe defined layers (seven in total) in a network OS. The layers provide clearly defined functions to improve Internetwork connectivity between networking nodes. Each layer has a standard-defined input and a standard-defined output. Figure 3-2 shows the OSI layer that the PROFIBUS protocol uses.[4]

[4] PROFIBUS communication protocol has been standardized under IEC 61158/IEC 61784-1 [Ref. 3 and 4].

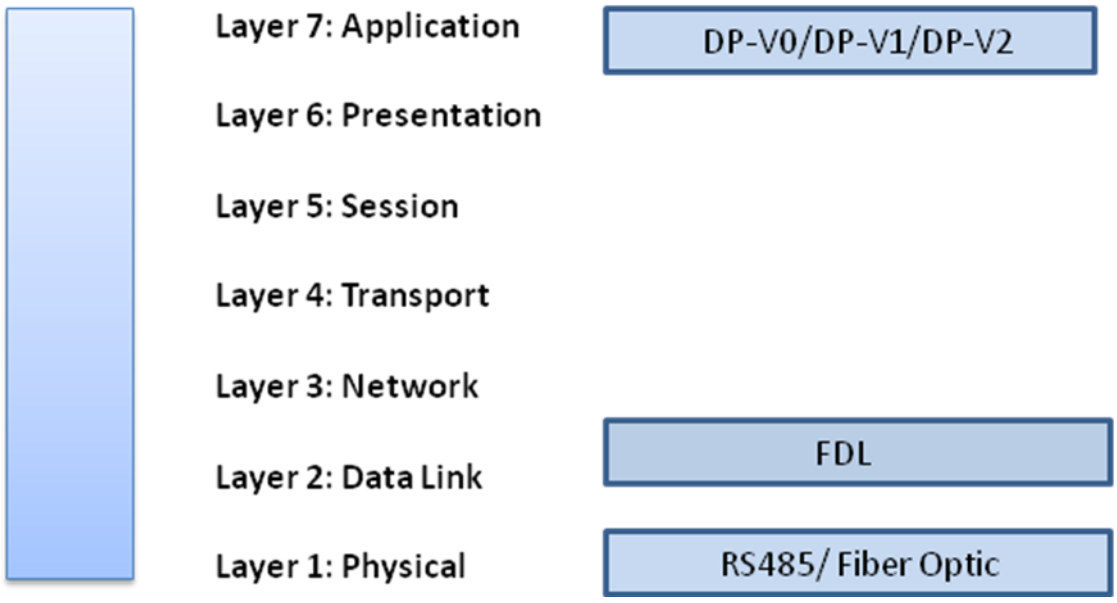

Figure 3-2. PROFIBUS and the OSI Model

Physical Layer. The physical layer for PROFIBUS can be represented by twisted pair wire using RS-485 or a fiber optic medium, which translates electrical signals to optical signals for transport.

Data Link Layer. The Field Bus Data Link Layer (FDL) arbitrates node communication among the participating devices on the PROFIBUS network. It uses an access protocol that employs a token-passing procedure with a master/slave process. The token-passing procedure guarantees that each specific master node on the PROFIBUS has a defined time frame for communicating. In this time frame the master node can access the bus and request data from any master or slave device connected to the bus. After the time frame expires, the current master node will pass the token to the next master node on the bus. The data link uses a Manchester code[5] to send digital data and clock pulse synchronization, which is embedded within the data.

Application Layer. Data are exchanged at this layer, both command/query and status. It provides the basic PROFIBUS functionality within a master/slave device construct. Masters read input and write output to/from slaves in a predictable, deterministic, cyclically fashion. The read and write commands are in the form of protocol data unit (PDU) messages and the application layer provides the proper protocol interpretation for the PDU messages. Depending on the service level required for device interaction the following application functions have been defined:

- DP-V0 for cyclic exchange of data and diagnosis.
- DP-V1 for acyclic and cyclic data exchange and alarm handling.
- DP-V2 for isochronous mode and data exchange broadcast.

[5] Defined within the IEEE 802.4 Standard [Ref. 5]

3.2.3 Field Bus Controllers

The Field bus controller provides the query and collection point for the process data being monitored. The I/O attached devices are configured and monitored by the controller. The actual design of the Field bus controller can help in providing a more robust data collection process. Independent central processing units (CPUs) for each I/O model, memory management unit, and external communication module can simplify software configuration and provide independence during system crashes. It can also facilitate independent troubleshooting and replacement without affecting other services.

Many controllers are considered *embedded* or reduced instruction set computers (RISCs). These are systems that use programs stored in flash memory that are used to *boot* the controller. This can ensure system integrity upon recovery from power failures. Controllers can be accessed remotely by plant personnel for status, configuration changes, and patch management. They are normally attached to some sort of centralized maintenance or engineering network to facilitate these interactions. A security overview for the Field bus and its controller is provided in the following security observation section.

3.2.4 Security Observations

It is important to understand that at the data exchange interface of a communication network, including the Field bus network, a communication protocol must be properly designed to detect and protect against potential security violations

The following information identifies some security vulnerabilities or limitations that exist within the PROFIBUS Field bus protocol; these are not necessarily limited to PROFIBUS. It was not the intent of this report to review all potentially used Field bus protocols and provide an individual vulnerability assessment. Rather, the limitations found in one can be used as a guide to determine the merits of others based on these findings.

Data or Message Replay. It is not only important to detect errors in receiving data; it is also important to detect repeated data strings or messages. Field bus communications should provide a sequence number scheme to detect when data are being repeated. It should also consider an acknowledgement to the originating sending node about the data reception. This can be as simple as returning the next consecutive number in the data sequence stream.

Message Time-Out Protection. Any communication protocol should be able to prevent a single node from dominating the communications channel or never completing a communication initiation. Some sort of *time-out* function should be used to override and disconnect an offending node that exceeds its communication time frame. A master/slave polling protocol should have some maximum time or maximum message size restriction. These features will help prevent a potential denial-of-service (DoS) attack against the communication protocol.

End Node Authentication. Addressing is normally done through the software setup at the controller. A graphical user interface (GUI) helps guide the operator in configuring the address for each end node. Some nodes do have hardware selected options. An adversary might use

hardware address selection if s/he can attach a field node to the Field bus at the address of an existing node and possibly create address collisions. This condition may prevent the necessary data extraction from the original monitoring device. This attack is made easier with the capability of field devices being "hot-swapped" in the field without having to shut down the process. A unique name programmed into each field device node and master controller can make masquerading adversarial devices much more difficult to impersonate active devices. Also, if the physical medium is properly protected from outside access, this can provide additional security.

Data Integrity. Data that the Field bus controller processes must be free of errors to provide the most accurate assessment of the environment being monitored. Error detection for most modern communication processes uses the cyclic redundancy check (CRC), which is an algorithm applied to the originating data (sender) and that is recalculated at the receiver to determine if it has been corrupted. The CRC process is designed to detect errors in data streams that are produced by the environment. However, note that a CRC can be recalculated by a man-in-the-middle attack simply by using software freely available (on the Internet to capture data messages, change the message, recalculate the CRC, and forward the manipulated message to the intended receiver without the knowledge of either the sender or receiver. To prevent this type of man-in-the-middle attack, a more sophisticated level of data integrity protection is required, such as a message digest algorithm[6].

Note: Any measures performed on a communication protocol to prevent its manipulation by an adversary may have an adverse impact on the operational performance of the protocol. The exact extent of such an impact depends on each vendor-specific implementation.

3.2.4.1 Maintenance and Remote Configuration Connections

Not shown in any of the digital I&C architecture diagrams being proposed for new reactor designs, but implied in some design discussions, is the existence of a maintenance network. This network is normally centralized and used by plant personnel to interact with the safety system controllers for configuration management activities. It is important to ensure that proper authentication mechanisms are in place to prevent adversarial compromise of the safety controllers.

Exploits against embedded system devices, such as programmable logic controllers ([PLCs], i.e. safety controllers) or remote terminal units (RTU) have been found in the public domain. These exploits take advantage of how the embedded system processor interacts with its firmware memory storage area, which is represented by flash (refers to the quick erase and reprogrammable memory function) also referred to as erasable programmable read-only memory (EPROM).

In many cases the exploit cannot be carried out if there is in place an authenticated process that provides the necessary interrogation of any memory update procedure. For an adversary to exploit some vulnerable aspect of a processor and its memory interaction, an unauthenticated protocol for firmware updates would have to be used. Unfortunately, a common firmware update

[6] MD5 HASH algorithm, RFC 1321 [Ref. 6].

protocol used for remote updates is Trivial File Transfer Protocol (TFTP), which does not authenticate the source or the target machine during firmware updates. Therefore, installing an update can compromise or disable the target system.

Thus, the firmware upgrade process, within the any safety or control system, should be evaluated for proper security implementation. The capability to perform remote firmware upgrades (across the maintenance network) should always be reviewed to ensure that proper authentication protection exists and that repositories for all firmware images are protected from unauthorized access and modifications. Direct access firmware upgrades are the most secure form of installing firmware. In summary, the network administrator should do the following:

- Ensure that flash update schemes require an authentication mechanism.
- Provide proper access control to protect firmware images during storage.
- Not allow remote updates to occur from hosts that reside outside of the protected safety network.
- Do not use removable media, such as thumb drives from sources that are for use on non-safety system networks, to copy upgrade images or patches.

3.2.4.2 Additional Observations

As seen in Figure 3-1, Digital Safety System Architecture, the *external threat* can be quite isolated from reaching the Field bus network. The isolation occurs because multiple layers of cyber protection can be inserted from the most external point of the public network. These *cyber layers* of protection—which can include multiple firewalls, Intrusion Prevention Systems (IPS) and a uni-directional data gateway along with proper physical protections—can help *minimize the external threat* against the Field bus network. When properly implemented these cyber protections now require the adversary to be physically located within the facility to obtain access to the Field bus network in order to touch and manipulate elements of the network. One caveat to this observation is the implementation technique used to update firmware or software on the Field bus controllers. If this technique is not properly authenticated, an external threat can take advantage of the implementation process used for software and firmware updates. Section 3.2.4.1, Maintenance and Remote Configuration Connections, previously described potential vulnerabilities and some mitigation techniques.

The *unprivileged insider threat* against the safety system Field bus can be substantially decreased if the "missing" security features (described in section 3.2.4.1, Vulnerabilities and Mitigations) are implemented as part of a Field bus protocol. Other techniques to limit an insider's privilege would include the implementation of an RBAC. This can be implemented on each controller where users can be assigned different user IDs that provide varying levels of controller capability. (For details on a centralized way of implementing RBAC, see Appendix D, section D.14, Role-Based Access Control.)

The *privileged insider* would have a larger administrative role within the facility but can possibly be limited in the number of systems that can be accessed or the location within the plant. Combining both physical protection mechanisms for personnel access control along with

restricting the number of systems that can be accessed can provide some level of protection against this type of threat.

Threats from *developer*- or vendor-based sources are primarily associated within Field bus controllers. These include default passwords and user accounts that have been configured as part of the vendor's pre-installation configuration. These potential vulnerabilities can be mitigated within a properly established security policy. Removing user accounts and default passwords should be part of a security policy established at each NPP facility. (See Appendix D, section D.1, Policy Framework Details, of this report for additional implementation information.)

3.2.5 Regulatory Guidance Regarding Field Bus Communications and Access Control

RG 1.152, Criteria for Use of Computers in Safety Systems of Nuclear Power Plants provides guidance on the inclusion of maintenance network connectivity by stating, "For digital computer-based systems, controls of both physical and electronic access to safety system and data should be provided to prevent unauthorized changes. Controls should address access via network connections and via maintenance equipment. Additionally, the design of the plant data communication systems should ensure that the systems do not present an electronic path by which a person can make unauthorized changes to plant safety systems or display erroneous plant status information to the operators."

RG 1.152 goes on to state, "Remote access to the safety system should not be implemented. Computer-based safety systems may transfer data to other systems through one-way communication pathways" (emphasis added). But in the context of the discussion in section 3.2.4.1, Maintenance and Remote Configuration Connections, of this report, the remote connection is not necessarily associated with a connection that exits in the safety system network, but refers to a communication interface that is not directly connected to the terminal port of a safety device. There can be multiple points of remote access implemented within the safety system network and should be governed by an access restriction policy. This is pointed out in clause 5.9 of IEEE Std 7-4.3.2-2003 [7], which states, "The design shall permit the administrative control of access to safety system equipment," and "controls should address access via network connections and via maintenance equipment."

As for the Field bus communication observations in section 3.2.4.1 Vulnerability and Mitigations, of this report, the following, which is contained within RG 1.152, states, "Computer-based systems (hardware and software) must be secure from electronic vulnerabilities. The consideration of hardware should include physical access control, modems, connectivity to external networks, data links, open ports, etc. Security of computer-based system software relates to the ability to prevent unauthorized, undesirable, and unsafe intrusions throughout the lifecycle of the safety system. Computer-based systems are secure from electronic vulnerabilities if unauthorized and inappropriate access and use of those systems is prevented." This statement does provide guidance, albeit broad coverage of necessary security, to address Field bus communication compromise.

NUREG/CR-6812 Emerging Technologies in Instrumentation and Controls, section 2.2.3, Safety-Related Field Bus, states, "The discussion will also address another approach to safety-

related applications of fieldbus [8]. This involves installation of a standard fieldbus system with enough redundancy, fault isolation, and diversity to achieve the required reliability and availability. In this case, a fieldbus system must be chosen that allows such configurations." This statement does mention the importance of reliability and availability, but does not include the tenants of communication integrity and authenticity.

NUREG/CR-6888, Emerging Technologies in Instrumentation and Controls: An Update, section 2.2 states, "Fieldbus standards are maturing, and networked field devices are likely to see increasing application as part of I&C upgrades for plant life extension" [9]; but this statement does not provide any observations about communication protections. Further in section 2.6, Control and Decision, a discussion centers around data acquisition to include, "The development of systems for data acquisition and control today is accomplished through the use of much higher-level computer software and programming languages. Data transfer rates continue to increase, as do communication protocols from various major commercial companies that provide hardware and software for data acquisition and control system." This section goes on to provide a modern example of data acquisition architecture, but provides no suggestions for securing the system.

Federal Regulation 10 CFR 73.54, Protection of Digital Computer and Communication Systems and Networks, identifies the need for a cyber security program and states, "The licensee shall protect digital computer and communication systems and networks associated with (i) safety-related and important-to-safety functions" [10]. This statement does not explicitly call out Field bus communication protections or how to protect them from cyber attacks, but provides broad language that can be inclusive of cyber protection mechanisms.

RG 5.71 provides acceptable methods for meeting the requirements of 10 CFR 73.54 in establishing the protection of digital computer and communication systems and networks from cyber attacks. This RG does not discuss or specify the use of specific technologies, but does provide a framework of policies and controls that licensees are expected to address in their cyber security programs. For example, the defensive architecture model (see Figure 5 in Reference 1) anticipates all safety-related equipment and functions are located within the most secure level where they are protected from access (electronic and physical) by all lower levels and where only out-bound data flow is allowed.

3.3 Automated Safety Layer Communications

As mentioned earlier, this layer receives the data from the process instrumentation layer and performs automatic system functions based on *set point* levels. This layer is also responsible for sending process data to the HMI Supervisory Layer for operator display in the control room and to other non-safety–related processes.

Process monitoring data are sent to the logic decision blocks within the Automated Safety Layer. The data can be sent directly to the decision block by an underlying physical protocol, such as RS-485; can be part of a multidrop field network, such as PROFIBUS; or may be aggregated and multiplexed through a multiplexer node. These multiple solutions can be considered *non-aggregated* feed and *aggregated* feed, respectively. Since the *non-aggregated* feed description

was part of the Field bus discussion in the previous section, we will focus on the *aggregated* feed description. *Aggregated* feed refers to the multiplexing of process sensor data, at the access point of the process sensors.

3.3.1 Multiplexers

The term *multiplexing* refers to the idea that many distinct data streams of information can share a common transmission medium. The multiplexing technique can take multiple forms, such as frequency domain multiplexing (FDM), time domain multiplexing (TDM), and wave division multiplexing (WDM). In FDM, multiple data channels are combined onto a single aggregate signal for transmission; the channels are separated in the aggregate signal by their *frequency*. In TDM, multiple data channels are combined onto a single aggregate signal for transmission; the channels are separated by sampling *time*. And in WDM, individual data channels are combined onto a single aggregate signal for transmission; the channels are separated in the aggregate light signal by the *wave*length of each propagated light signal.

The multiplexers being used on new design implementations for DSS are not described in detail. But regardless of the multiplexing technique, each multiplexing node must be properly configured to access its transmission medium of choice. All multiplexers consist of some basic components. They contain I/O channels that are used to sample input logic from the process being monitored. These lines could be bidirectional if a control function is also required. They can represent multiple serial physical protocols, such as RS-232, RS-485, RS-422, etc. The input lines are then *multiplexed* together to form an aggregate data stream. This stream of date is synchronous in nature and represents the combination of all data presented at the input channels. Each I/O channel can be independently configured for the appropriate interface specification required for the process being monitored. In most cases, channel setup and configuration can be done remotely over a maintenance network that allows an operator to have remote configuration control at a central location. Figure 3-3 shows some basic components of a multiplexer.

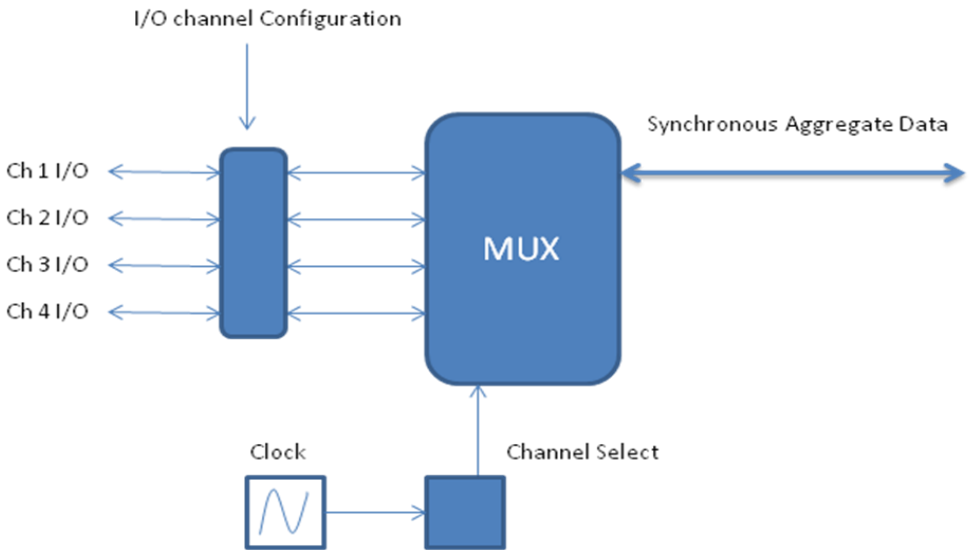

Figure 3-3. Simple Multiplexer Block Diagram

The aggregated data are put onto a transport medium and delivered to the Automated Safety-Layer logic-decision block for processing. Based on the designs that were reviewed, the Fiber Distributed Data Interface (FDDI) has been identified as a potential protocol to transport the aggregated data to the safety system logic decision block.

3.3.2 Fiber Distributed Data Interface

FDDI is an American National Standards Institute (ANSI)[7] Committee X3-T9 protocol [11] that conforms to the OSI model of functional layering. It provides 100 Mbps, token passing data transmission across dual counter rotating fiber optic rings. A Timed Token Protocol (TTP) controls the transmission of data onto the network. A station cannot transmit data onto the network until it has received the token. Once receiving the token, the FDDI node or station can transmit data as long as the TTP timer has not expired. The dual rings consist of a primary and a secondary ring. The purpose of the dual rings is to provide reliable and robust data transmission. During normal operation, the primary ring is used for data transmission, while the secondary ring remains idle until the primary ring fails. FDDI specifies the physical and media-access portions of the OSI reference model. FDDI is not actually a single specification, but is a collection of four separate specifications, each with a specific function. Combined, these specifications have the capability to provide high speed connectivity between upper-layer protocols (such as TCP/IP) and media (such as fiber-optic cabling or twisted copper pair wire). Figure 3-4 shows the FDDI OSI model relationship.

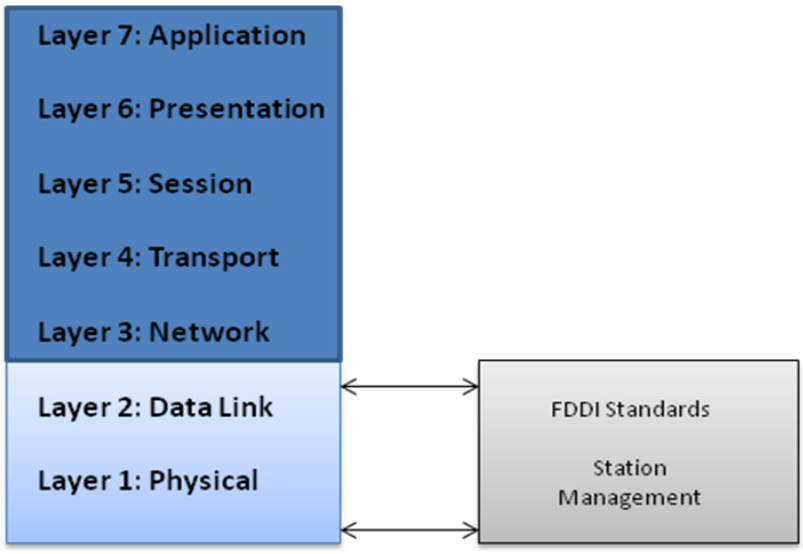

Figure 3-4. FDDI OSI Model

FDDI's four specifications are the Media Access Control (MAC), physical layer protocol (PHY), physical-medium-dependent (PMD), and Station Management (SMT) specifications. The MAC specification defines how the medium is accessed, including frame format, token handling,

[7] www.ansi.org

addressing, algorithms for calculating the CRC value, and error-recovery mechanisms. The PHY specification defines data encoding/decoding procedures, clocking requirements, and framing, among other functions. The PMD specification defines the characteristics of the transmission medium, including fiber-optic links, power levels, bit-error rates, optical components, and connectors. The SMT specification defines FDDI station configuration, ring configuration, and ring control features, including station insertion and removal, initialization, fault isolation and recovery, scheduling, and statistics collection.[8] An FDDI frame, with the number of bytes associated with each field, can be seen in Figure 3-5.

2 bytes	6 bytes	6 bytes	0-30 bytes	0-4478	4 bytes
Frame Control	Destination Address	Source Address	Route Information	Data	FCS

Figure 3-5. FDDI Protocol Frame

Frame Control. Provides information on the size of the address fields and whether the frame contains asynchronous or synchronous data.

Destination Address. Contains a unicast (singular), multicast (group), or broadcast (every station) address. FDDI destination addresses are 6 bytes long.

Source Address. Identifies the node of the originating data. FDDI source addresses are 6 bytes long.

Route Information. Provides source-route information in the form of a series of addresses associated with route bridges that forward the frame to its final destination.

Data. Contains either information destined for an upper-layer protocol or control information.

Frame Check Sequence (FCS). Calculated and inserted by the source station with a calculated CRC value dependent on frame contents. The destination node will recalculate the value to determine whether the frame was damaged in transit and will discard the frame if errors are detected.

3.3.2.1 Latency Control

One important aspects of safety system monitoring is the real-time nature of process status information. It is important to mange any latency that may be introduced by the use of a data communication protocol. FDDI addresses latency by using *timed token access*. There is a parameter that is configurable on each access node in the FDDI transmission ring that defines a Target Token Rotation Time (TTRT). This time ensures the token circulates at least once every TTRT and provides a more deterministic approach to data transport. The TTRT is the fastest time that is agreed upon by each participating node for a token to rotate all around the ring. There are two classes of TTRT frames: synchronous and asynchronous. Depending on the requirement of

[8] Cisco Internetwork Technologies Handbook, 1-58705-001-3

each access node, a s*ynchronous allocation* can be configured for nodes transporting latency sensitive traffic. Each node can be configured to send both synchronous and asynchronous traffic if needed as long as the time since the previous token arrival is less than TTRT. If TTRT was already exceeded after the synchronous frames were sent, no asynchronous frames may be sent. This mechanism allows for latency-sensitive (synchronous) traffic to take precedence or priority over other traffic.

3.3.2.2 Node Attachment

The FDDI standard defines a dual counter-routing ring for redundant data transport; it also defines four different ways a node can access the transport medium:

- A single-attachment station (SAS).
- A dual-attachment station (DAS).
- A single-attached concentrator (SAC).
- A dual-attached concentrator (DAC).

The SAS node attaches to the primary ring through a device called a concentrator (either an SAC or DAC). One advantage of connecting a node with an SAS attachment is that, when the device is powered off or removed, it will not affect the FDDI ring. However because the SAS is not connected to multiple rings, it is not afforded redundancy protection.

The DAS node attaches to both the primary and secondary FDDI rings through its two ports: port A and port B. A station node connected to both rings provides redundancy when the primary transport ring fails. But it will adversely affect the rings if they are disconnected or fail.

The SAC node attaches to only the primary transport ring of the FDDI network. Stations that are attached to it have access to the FDDI backbone, but the SAC node does not provide redundancy during primary ring failures. It does prevent stations that are attached to it from affecting the ring when they fail or are powered down. An SAC has only two different ports: port S and port M. Ports M connect to SASs, and port S connects to the FDDI network or to another SAC port M.

The DAC node attaches directly to both the primary and secondary rings and ensures that the failure or disconnection or power-down of any SAS does not bring down the ring. The DAC also provides redundancy during single ring failures for any of its attached SAS nodes. A DAC has three different types of ports: ports A, B, and M. Port A connects the input of the primary ring and the output of the secondary ring. Port B connects the output of the primary ring and the input of the secondary ring. Port M connects to the S (slave) port of an SAS. Figure 3-6 shows the device connections.

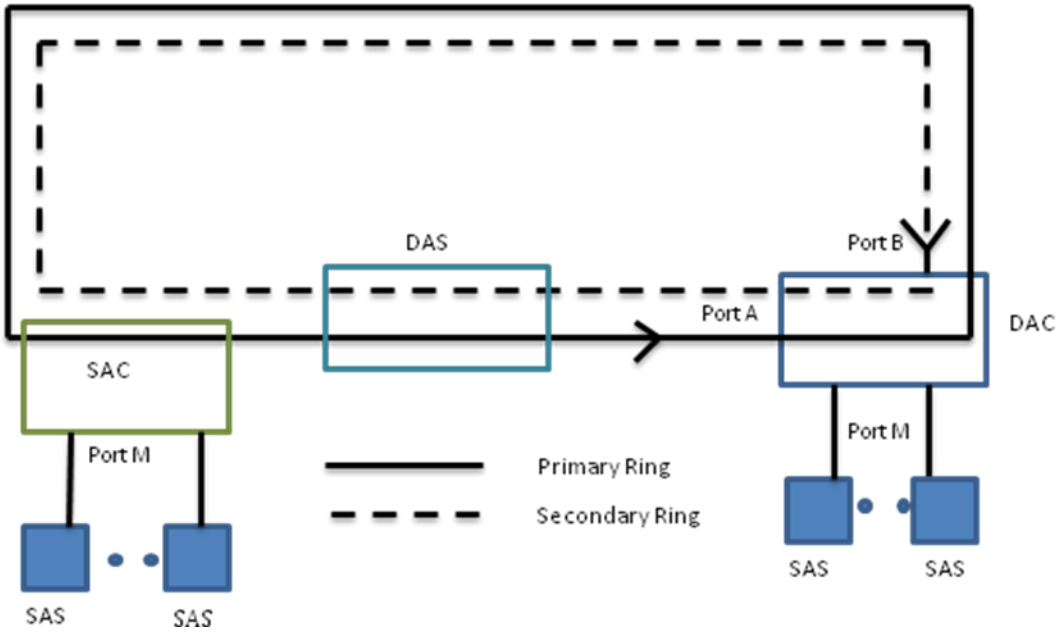

Figure 3-6. FDDI Node Attachments

3.3.2.3 Fault Tolerance

FDDI architecture allows for some fault-tolerant features to enable the network to continue to operate in lieu of some node and cable failure. More specifically, the dual ring design, the implementation of the optical bypass switch, and the dual-homing configuration support make the technology resilient to node and media failures.

3.3.2.4 Dual Ring

FDDI's primary fault-tolerant feature is the dual ring. If a station on the dual ring fails or if the cable is damaged, the dual ring is automatically *wrapped* back onto itself into a single ring. When the ring is wrapped, the dual-ring topology becomes a single-ring topology. Data continue to be transmitted on the FDDI ring without impacting the data transport. Figure 3-7 shows how the ring can restore communications if an FDDI node fails, and Figure 3-8 shows how the ring can retain communications if a portion of the media fails.

When a single node fails, as shown in Figure 3-7, devices on either side of the failed station wrap, forming a single ring. Network operation continues for the remaining stations on the ring. When a cable failure occurs, as shown in Figure 3-8, devices on either side of the cable fault wrap. Network operation continues for all stations. It should be noted that FDDI provides fault-tolerance against a single failure. The FDDI architecture is designed to accommodate any single fault, either node or cabling, but when two or more failures occur, the FDDI ring segments into two or more independent rings that are unable to communicate with each other.

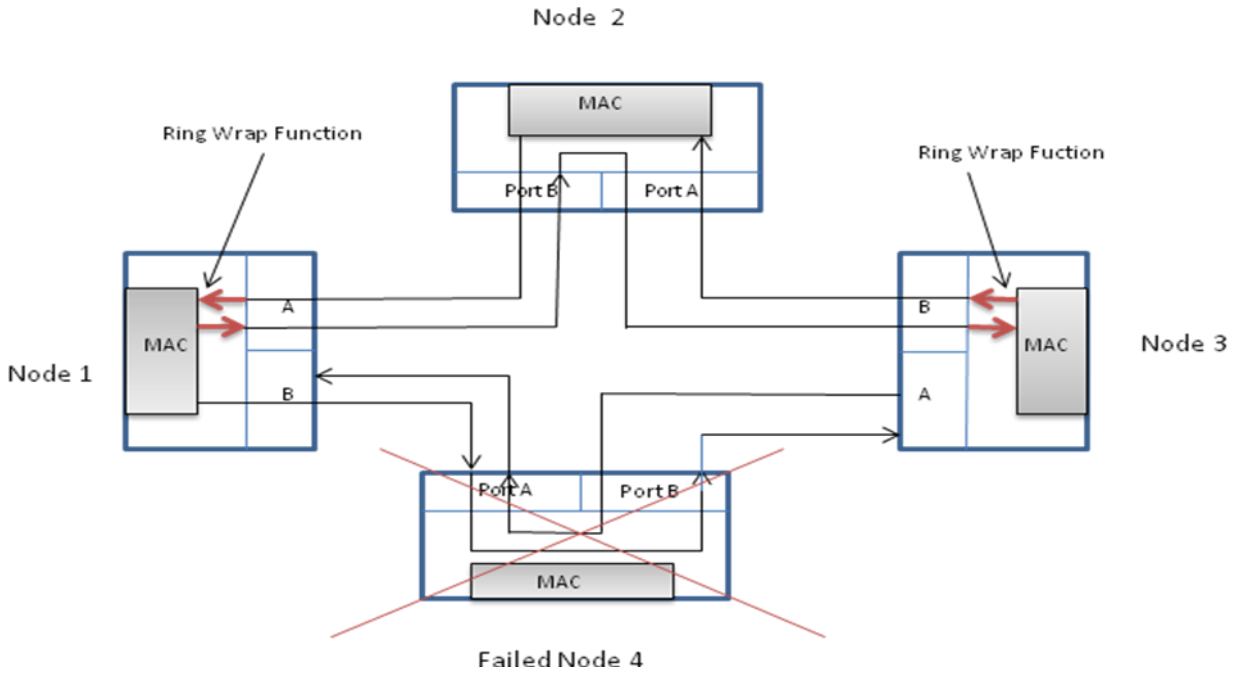

Figure 3-7. FDDI Node Failure Fault Tolerance

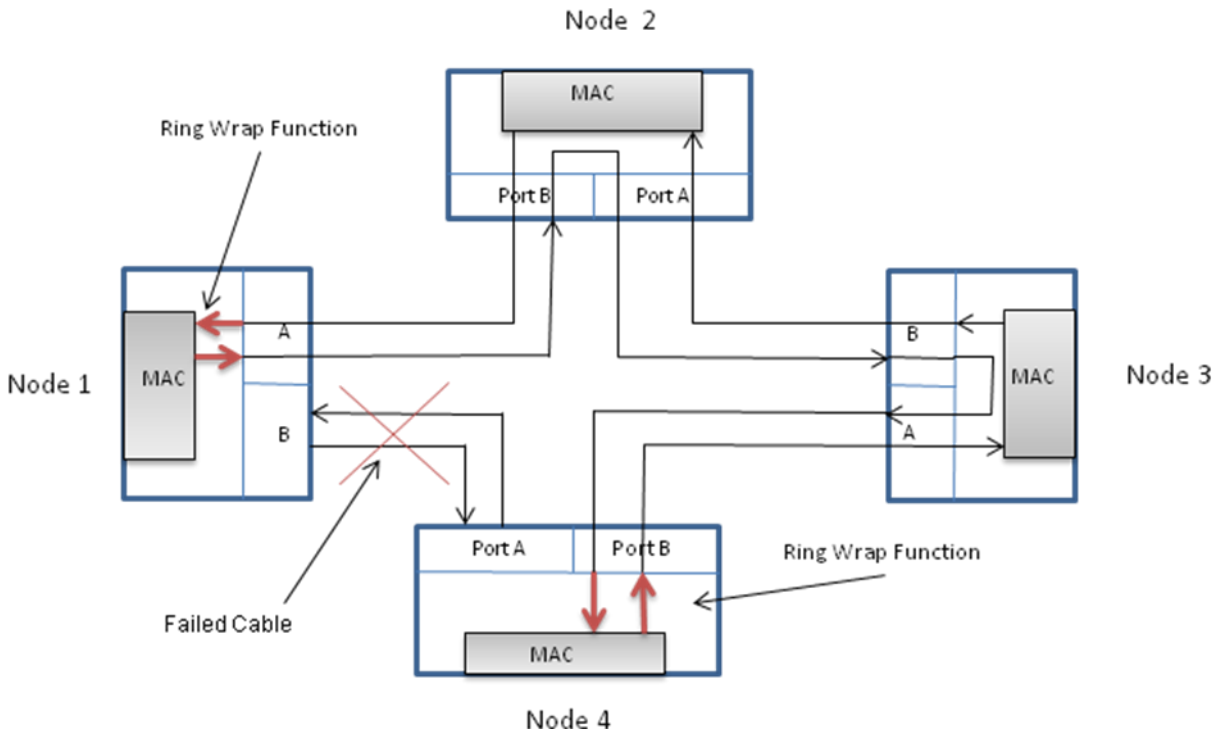

Figure 3-8. FDDI Medium Fault Tolerance

3.3.2.5 Optical Bypass Switch

Another protection against node operational faults is to use an optical bypass switch. Unlike the ring wrap function described above, the bypass switch allows for continuous dual-ring operation if a device on the dual ring fails. The optical bypass switch is a passive optical relay that is attached to both rings. The benefit of this capability is that the ring will not enter a wrapped condition in the event of a device failure and will continue to provide full ring redundancy. The disadvantage of this configuration is the loss of optical power. Each optical switch will induce an optical budget loss between 1.5 to 2.5 db. Standard FDDI fiber optic connectors will induce additional losses that, at some point—normally around 11 dB of attenuation—will prevent proper operation of the link

3.3.2.6 Dual Homing

FDDI architecture supports a fault-tolerant technique called *dual homing***.** The dual-homing technique provides a means for alternate access routes for attaching FDDI stations to two independent FDDI concentrators. One of the concentrator links is configured to be the *active link*; the other is configured to be in *passive mode.* Stations use the primary link to access the FDDI network. The passive link provides the back-up until the primary link fails. When a failure is detected, the passive link automatically becomes the active link. Figure 3-9 shows a dual-homed configuration for FDDI attached stations.

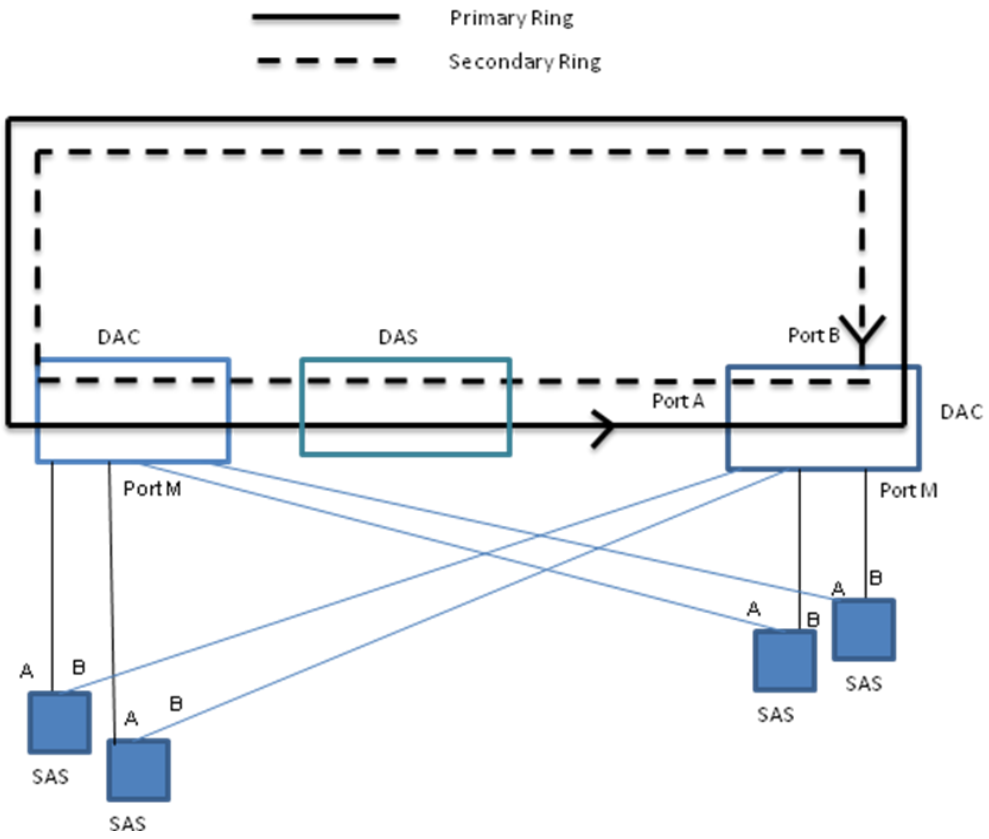

Figure 3-9. FDDI Dual Home Host Configuration

3.3.2.7 Physical Medium

The FDDI network can be built upon either twisted pair copper wire or fiber optic cabling. The Copper Distributed Data Interface (CDDI) is the standard—the FDDI backbone construction. It can use unshielded twisted pair or shielded twisted pair copper wire. CDDI can support a dual ring capacity of 100 Mbps with a maximum distance of 200 meters.

The fiber optic cabling can be comprised of multimode or single-mode fiber. Multimode allows multiple modes of light to propagate through the fiber. This propagation occurs because light enters the fiber at different angles resulting in an effect called modal dispersion. The light does not arrive at the end of the fiber at the same time. This characteristic limits the bandwidth and distances that can be accomplished using multimode fibers.

Single-mode fiber allows only one mode of light to propagate through the fiber. It uses a coherent light source, normally in the form of a laser. This prevents the modal dispersion characteristic found in multimode cable and allows single-mode fiber to deliver higher bandwidth performance over much longer distances.

3.3.3 Security Observations

3.3.3.1 Medium

FDDI uses optical fiber as the primary transmission medium, but it also can run over copper cabling. As mentioned earlier, FDDI over copper is referred to as *Copper-Distributed Data Interface*. Optical fiber has advantages over copper cabling. In particular, security, reliability, and performance all are enhanced with optical fiber media because fiber does not emit electrical signals. A physical medium that does emit electrical signals (copper) is easier to tap and could permit unauthorized access to the data that are being transmitted through the medium. In addition, fiber is immune to electrical interference from radio frequency interference (RFI) and electromagnetic interference (EMI).

3.3.3.2 Integrity and Authentication

The standard FDDI frame implementation provides an FCS, which is calculated and inserted by the source station with a calculated CRC value dependent on frame contents. The destination node will recalculate the value to determine whether the frame was damaged in transit and will discard the frame if errors are detected. This technique does provide an error detection capability, but does not protect against man-in-the-middle attacks that can capture and modify the contents of a frame, recalculate the CRC, and forward the frame out to its intended destination. For this attack to happen on the FDDI network, the adversary node would have to have access to the network and be able to masquerade their identity as the intended destination node. This can be accomplished with a technique called Address Resolution Protocol (ARP) “spoofing” attack. The ARP spoofing attack is associated with how addresses are resolved on the network. If a token is being passed along the FDDI network, the adversary would associate the address with the destination address of the token and remove it from circulation. The data within the frame could then be modified with a new CRC calculated and sent back onto the network to the originally intended destination. One means of protecting FDDI frame manipulation is with the use of FDDI Virtual Local Area Networks (VLANs).

3.3.3.3 FDDI Virtual Local Area Networks

A VLAN is a logical local area network (LAN) that extends to a group of participating access layer devices, such as Ethernet switches that create independent, isolated domains. The independence is formed by the unique VLAN identifier (ID) in which participating end nodes are assigned. The VLAN is a group of nodes with the same VLAN ID and that can share information among themselves without the information being distributed to other non-participating nodes. This prevents the data "snooping" that could occur when data are presented to other nodes associated with a larger universal domain. To be able to configure VLANs within FDDI, the interoperable LAN/metropolitan area network (MAN) security standard was designed [12]. This standard has elements that can secure the data for transport across an FDDI network by using a secure data exchange PDU. This PDU is a MAC layer frame with an 802.10 header inserted between the MAC header and the data field. The *MAC header* would be the FDDI header described earlier. Figure 3-10 shows the 802.10 header construct frame.

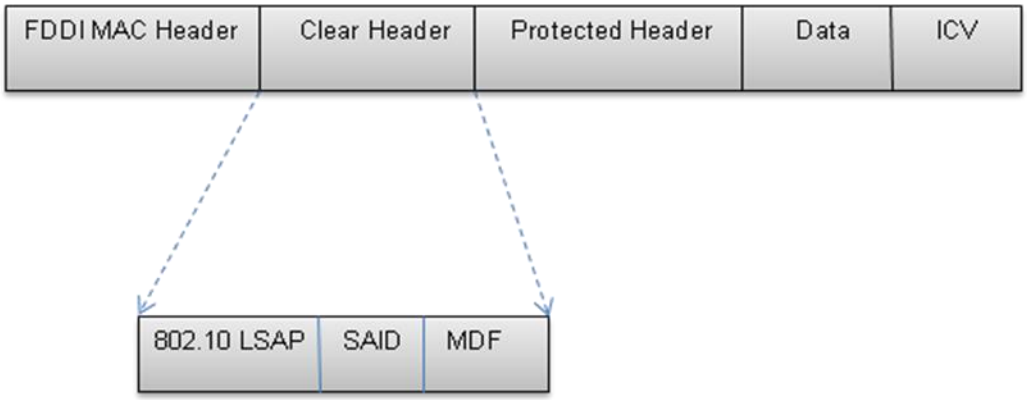

Figure 3-10. 802.10 Header Frame

As shown in Figure 3-10, the Integrity Check Value ICV) field within the 802.10 frame uses a secure cryptographic algorithm to detect modification of the data field. The *protected header* field provides a copy of the source address for validation against the original MAC header. The security management information base (SMIB) provides the Security Association Identifier (SAID), which is used as the VLAN ID. The Management Defined Field (MDF) is optional and carries information to assist with PDU processing. If secure data and authentication integrity are not required, then only the *clear header* with the Link Service Access Point (LSAP) designator is necessary. The LSAP designator within the 802.10 clear header can identify multiple data streams that may be originated from the same end device. Implementing the ICV can prevent man-in-the-middle attacks against the FDDI data frame.

3.3.3.4 FDDI Fault Tolerance

FDDI provides some fault-tolerant features. Its dual-ring architecture, the implementation of the optical bypass switch, and dual-homing support make FDDI a resilient media technology.

3.3.3.5 Maintenance and Remote Configuration Connections

During the review of the proposed new DSS designs, an indirect reference to a *maintenance* or *remote connection* capability resident on the safety and control network for network accessible

devices was alluded to. For any device—such as FDDI nodes, multiplexer nodes, Field bus controllers, or other logic controllers that have the capability to be addressed remotely—some procedures can be followed that will help provide proper protection of these assets:

- Review the requirements for each user of the system and assign and enforce (using owner or group permissions) a level of system access that is commensurate with each user's role. A role-based review would determine the level of access—and, thus, the level of system access—required to perform one's job.

- Prove authentication through some means, such as unique passwords for each user of the system, and change these passwords on a regular basis.

- Create a formalized change management process that documents all changes to the system.

- Do not allow remote updates to occur from hosts that reside outside of the protected safety network.

- Do not use removable media, such as thumb drives, from sources that are used on non-safety system networks to copy upgrade images or patches.

- Develop policies for the external equipment if the control network or field network has external access points.

3.3.3.6 FDDI Maintainability

The popularity of using FDDI for high speed backbone connectivity has been greatly reduced due to the advance of Ethernet switching technologies. Ethernet switching can provide the high speed backbone type applications that FDDI touts. It is not certain whether the availability of FDDI technologies can be maintained to support long term FDDI deployments.

As mentioned previously in the Process Instrumentation Communication Layer (section 3.2), the process monitoring can be sent directly to the decision block within the Automated Safety Layer by an underlying physical protocol, such as RS-485; or it may be aggregated and multiplexed through a multiplexer node. These multiple solutions can be considered *non-aggregated* feed and *aggregated* feed respectively. The aggregated feed refers to the multiplexing of process sensor data, at the access point of the process sensors. This is not shown in Figure 3-1, but can be considered as part of the Field bus network, since this multiplexing technique is contained within the Field bus network architecture.

3.3.4 Additional Observations

The *external threat* can be quite isolated from reaching the FDDI multiplexers because of the many layers of cyber protection inserted between the most external point onto the NPPDN to the unidirectional data gateway that resides between the safety and non-safety networks. This includes the idea that proper physical protections are in place at the NPP. One caveat to this observation is the potential remote communication technique used to determine status, or to configure or update firmware or software on the multiplexer nodes. Similar to the Field bus

controllers, an external threat can take advantage of the implementation process used for remote access to the multiplexer nodes if the access is not properly authenticated. Section 3.3.3.5, Maintenance and Remote Configuration Connections, describes the potential vulnerabilities and some mitigation techniques.

The *unprivileged insider* threat against the FDDI aggregated data feed architecture can take advantage of his/her physical location to have access onto the FDDI network in an attempt to manipulate process sensor data. But if the FDDI frame includes ICV implementation within its address header, the adversary cannot use any man-in-the-middle attacks against the process data traffic. This ICV uses a secure cryptographic algorithm to protect the process data in transit making the spoofing of process data highly unlikely. (This protection was described in section 3.3.3.3, FDDI VLANs.*)* Other techniques to limit an insider's privilege would include the implementation of an RBAC policy. This technique assigns a level of permissions associated with an insider's job profile. The use of unauthenticated remote management protocols, such as Telnet and TFTP, should be restricted when communicating with the FDDI multiplexers. These communication protocols allow the user ID and password to be passed in the clear over the network. The unprivileged insider can "sniff" these credentials to escalate his privileges on the devices. Also providing additional protections to the network in the form of Ethernet port restrictions can limit the insider from attaching unauthorized devices onto the network. (See section D.4.1, Ethernet Security Observations for more detail on these protection features.)

The *privileged insider* would have a larger administrative role within the facility, but can possibly be limited in the number of systems that can be accessed or the location within the plant. A formal process for change management should be instituted. It could include procedures, such as requiring that all configurations changes be reviewed by multiple administrators or subject matter experts to help detect malicious or accidental configurations. Combining both physical protection mechanisms for personnel access control—along with restricting the number of systems that can be accessed—can provide some level of protection against this type of threat.

Threats from *developer*- or vendor-based sources associated with FDDI multiplexers, such as Field bus controllers, include default passwords and user accounts that have been configured as part of the vendor's pre-installation configuration. These potential vulnerabilities can be mitigated within a properly established security policy. Removing user accounts and default passwords should be part of a security policy established at each NPP facility. (See section 2.1, Security Policy, in this report for additional implementation details.) As devices become more software capable, the risks associated with supply chain and software lifecycle risk increase. Implementing appropriate controls (as described in section 2.5, Safety System Lifecycle) can reduce the overall risk against this product insertion vulnerability.

3.3.5 Regulatory Guidance Regarding Multiplexers and FDDI

RG 1.152, "Criteria for Use of Computers in Safety Systems of Nuclear Power Plants," provides guidance to properly protect digital computer-based equipment, which may more broadly include a processor-based multiplexer. "For digital computer-based systems, controls of both physical and electronic access to safety system and data should be provided to prevent unauthorized changes. Controls should address access via network connections and via maintenance

equipment." Additionally, "Computer-based systems (hardware and software) must be secure from electronic vulnerabilities. The consideration of hardware should include physical access control, modems, connectivity to external networks, data links, open ports, etc." This is reiterated in Clause 5.9 of IEEE Std 7-4.3.2-2003, (referring to clause 5.9 in IEEE Std 603-1998 [13]), "Control of Access," which states, "The design shall permit the administrative control of access to safety system equipment."

Also the following observation is made in RG 1.152, "Instrumentation and control (I&C) system designs that use computers in safety systems make extensive use of advanced technology (i.e., equipment and design practices). These designs are expected to be significantly and functionally different from current designs, and may include the use of microprocessors, digital systems and displays, fiber optics, multiplexing, and different isolation techniques to achieve sufficient independence and redundancy."

With respect to FDDI there is no specific mention of this technology in the RGs. But it can be included in the topic of DSS lifecycle in RG 1.152, section 2.3.1, System Features, which states, "The safety system security requirements identified in the system requirements specification should be translated into specific design configuration items in the system design description. The safety system security design configuration items should address control over (1) physical and logical access to the system functions, (2) use of safety system services, and (3) data communication with other systems."

Similarly, access control—both physical and electronic—are two of the principal operational and technical controls for cyber security assurance framework endorsed by RG 5.71. Additionally, the RG requires methods for effectively ensuring the protection of communication interfaces to CDAs are included in a plant's cyber security program.

An observation about fiber media, which is associated with many FDDI implementations, can be found in NUREG/CR-6812, Emerging Technologies in Instrumentation & Controls, section 2.2.2.1, Optical Networking: "Fiber-optic communications will probably see increased use in nuclear power applications. However, the data throughput needs envisioned for even the most highly integrated control and information systems within a nuclear plant should be well within the current scope of the technology. The biggest challenge for safety-related applications to the field-device level appears to be environmental compatibility for fiber-optic carriers. NRC has investigated fiber-optic communications in the past and should continue to monitor the state of the technology." This observation validates the insertion of the fiber optic media in nuclear power plant architectures, but provides no additional information on the types of protocols (i.e., FDDI or Ethernet) that may use the media or information on the security advantages of using the media.

3.4 Data Communications from the Automated Safety Layer

Reviewing the proposed new DSS designs has shown some data exchange requirements from the Automated Safety Layer to both the control room (HMI Supervisory Layer) and to non-safety process and control systems. Two primary data protocols identified during the review include Field bus and Ethernet. Since the Field bus protocol has been discussed in section 3.2, with

specifics associated with a popular Field bus called PROFIBUS, this section will concentrate on the Ethernet protocol used to transport data.

3.4.1 Ethernet

The Ethernet protocol refers to a family of the LAN protocols [14]. There are two distinct modes of operation defined: half-duplex and full-duplex modes. In the half duplex mode, data are transmitted using the Carrier-Sense Multiple Access/Collision Detection (CSMA/CD) protocol on a medium that can be shared by other competing nodes. The main disadvantages of this technique are the efficiency of accessing the medium and the maximum distance limitation. Because it is a shared medium, only one node can be transmitting at any given time, and all must be able to detect that the medium is in use prior to transmitting data. In the worst case it takes twice the maximum propagation delay across the network before a station can ensure that a transmission has been successful. If a station sends a short frame, it may actually finish sending and release the medium without realizing that a collision has occurred.

The IEEE Std 802.3 design rules specify an upper limit on the maximum propagation delay in any Ethernet installation, and the minimum frame size is set to be more than twice this value. For example, the Ethernet protocol defines a minimum frame size as 64 bytes. It would take approximately 51 microseconds to send the minimum frame size at a rate of 10 Mbps. This limits the implementation efficiency for higher-rate transmission by having to wait and listen to determine if the frame was sent successfully without collision.

To overcome this data rate restriction, other Ethernet implementations have been developed that essentially create an independent LAN segment per node (Ethernet switching), thus, removing the data transmission contention of the shared medium. The following shows other Ethernet implementations and the current supported data rates:

- 10 Mbps—10Base-T Ethernet (IEEE Std 802.3).
- 100 Mbps—Fast Ethernet (IEEE Std 802.3u).
- 1000 Mbps—Gigabit Ethernet (IEEE Std 802.3z).
- 10-Gigabit—10 Gbps Ethernet (IEEE Std 802.3ae).

3.4.1.1 Frame Structure

In an Ethernet network, data are transmitted between nodes in the form of an Ethernet frame. The frame structure is comprised of the following fields:

- **Preamble**, which consists of a start frame delimiter, synchronizes the receiving stations clock with the sender node and consists of seven (7) bytes of alternating logic (10101010) and 1 byte set to (10101011).

- **Destination Address** is the address of the intended recipient of the frame. The addresses in 802.3 use globally unique hardwired 48 bit addresses.

- **Source Address** is the address of the source, in the same form as the destination address above.

- **Frame Length** is the length of the data in the Ethernet frame, which can be anything from 0 to 1500 bytes, with a minimum frame size of 64 bytes.

- **Data** file contains the information being sent by the frame.

- **Pad** compensates for small data frames that may be less than the required minimum standard. In the IEEE Std 802.3, the frame size must be at least 64 bytes long.

- **Checksum** detects errors that may have occurred during transmission.

Figure 3-11 shows the Ethernet frame format.

Figure 3-11. Ethernet Frame Structure

3.4.1.2 OSI Model

The ISO has created a layered model called the OSI model to describe defined layers (seven in total) in a network OS. The layers provide clearly defined functions to improve Internetwork connectivity between networking nodes. Each layer has a standard defined input and a standard defined output. The Ethernet protocol has layer representation in the physical layer (Layer 1) and the data link layer (Layer 2).

Physical Layer. The physical layer is concerned with the low level electronic way in which the signals are transmitted. In Ethernet, signals are transmitted using Manchester Encoding. This encoding is used to ensure that clocking data are sent along with the data, so that the sending and receiving device clocks are in sync.

Data Link Layer. Ethernet uses the CSMA/CD protocol as part of its MAC mechanism. To send out a data frame onto the LAN segment, the Ethernet node checks to determine if the medium is busy. If the LAN segment is busy, the node backs-off by a short fixed-delay time period; after the medium becomes idle, then the frame will be sent out. However, if the device detects a collision, the frame transmission stops and the station sends a jamming signal to alert other stations of the *segment-in-use* situation.

The node waits before attempting to transmit by using the *truncated binary exponential back-off algorithm*. This algorithm is based upon multiple 51.2 microsecond time slots (minimum detect time for smallest allowed Ethernet frame). The station first waits for either 0 or 1 time slots, then transmits. If there is another collision, then the station can wait for 0, 1, 2 or 3 slots before transmitting. This continues with the station choosing to wait a random number of slots from 0 to 2^k - 1 if there have been k collisions in the current transmission—until k=10 where the number of slots to choose from stops growing. After 16 continuous collisions, the MAC layer gives up and reports a failure to the layer above [15].

3.4.1.3 Ethernet Switching

The development of Ethernet switching has offered a means of using the Ethernet standard that greatly increases performance without having to replace the existing infrastructure. The Ethernet switch has been designed to divide the network into many small segmented collision domains with each port being considered a segment. This means nodes in different domains can talk simultaneously, which increases the transmission efficiency. Instead of sharing a 10 Mbps connection with many nodes, each node (a workstation or server) can have a dedicated 10 Mbps segment connected to an Ethernet port.

One of the more modern and efficient Ethernet switch configurations is hierarchical in nature; this allows the network to be designed in layers. Using the layer approach simplifies the task for network designs. Each layer can focus on specific functions, allowing the designer to choose the right features for each layer.

The hierarchical layered approach can also accommodate design changes and provide a modularity to the network design, which allows for node replication as the network grows. When a network node requires a design change, the cost of the change—and the amount of effort to induce the change—can be constrained to a small subset of the overall network. Changes on other network architectures, such as flat or meshed network architectures tend to create a large impact on the overall system. Other attributes of a hierarchical layered architecture include improved fault isolation because the interface points within the hierarchy create an easier way to identify failure points. A large Ethernet installation, such as a corporate network, would include three layers: the backbone (or core) layer, the distribution layer, and the access layer. Smaller, or segmented, installations do not require a core layer. In a nuclear plant, where segregation of tasks is necessary, a redundant two-tier architecture would be appropriate. Figure 3-12 shows a typical two-tier Ethernet Architecture.

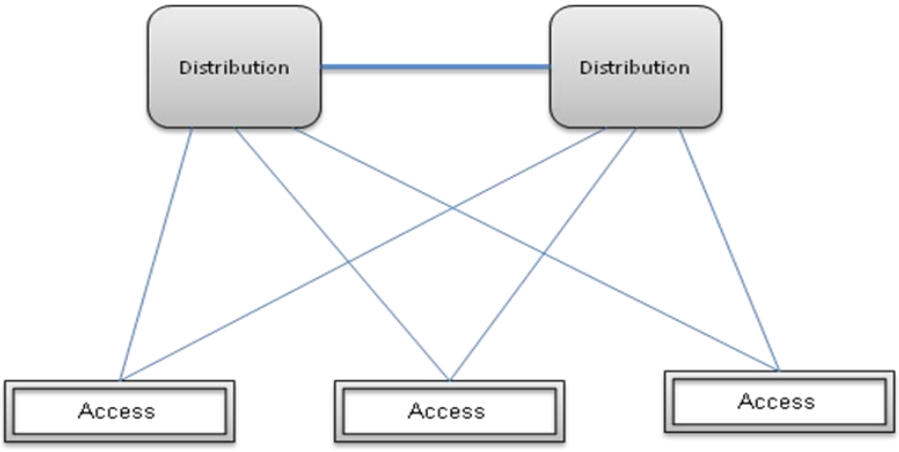

Figure 3-12. Two-Tier Ethernet Architecture

3.4.1.4 Access Layer

The access layer of the two-tier switch design allows device access to the network. This is where the originating traffic can be examined and conditioned based on network access policy. This allows the network administrator to configure port switch features for each attached device.

Access switches can provide the following:

- Link redundancy
- IEEE Std 802.1x port security
- Traffic conditioning and marking, including VLAN tagging
- Traffic aggregation for endpoint devices

3.4.1.5 Distribution Layer

The distribution layer within the two-tier switch architecture is responsible for aggregating the access layer data flows. Distribution layer switches can also function in the Layer 3 routing domain. This allows it to provide network level address distribution and network layer access control. The distribution layer can be used to filter device communications based on a network level policy or a role-base policy. Distribution layer switches can provide the following:

- Aggregation of the data traffic from access layer switches.
- Perform security filtering by the use of access control lists (ACLs).
- Link redundancy by the use of an additional distribution switch.
- Conduct route look-up service for addresses outside a common domain.

3.4.1.6 Media Types

Fast Ethernet, which is representative of many switched Ethernet applications, supports three media wiring schemes; all of these systems use hubs or switches to connect the network:

- **100Base-T4.** This uses category 3 (CAT 3) copper twisted wire pairs, that can support data up to 25 Mbps. To achieve 100 Mbps, four twisted pairs are required.

- **100Base-TX.** This uses CAT 5 copper twisted pairs; this is the most popular scheme because of the ease of distribution and its support of 100 Mbps data rate. Only two twisted pairs per station are used, which allows for full duplex transmission and is normally terminated in a RJ45 connector.

- **100Base-FX.** This uses two strands of multimode optical fiber, one for each direction, which also supports full duplex, 100 Mbps and is immune to EMI interferences. Because it uses light for data propagation, it does not create any RFI fields. It also has a greater transmission distance than its copper wire equivalent.

3.4.1.7 Ethernet VLAN

Along with network-based access control, which is administered at the application level, there is another form of need-to-know separation that can be implemented at the network device level. Network devices, primarily Ethernet switches, can be configured to separate user traffic by the administration of VLANs.

A VLAN provides a means to segregate traffic over a network, such as an Ethernet switching network, by software controls using VLAN ID tags. A switch in an internal database defines these VLAN IDs. After a VLAN has been created within the database, then end ports are

assigned. These end ports map to end user devices or to a server. A VLAN is assigned a unique number or name, which the VLAN Trucking Protocol (VTP) distributes. VTP provides the means to distribute and update the VLAN database. If a switch does not know a VLAN, then the switch (normally an Ethernet device) cannot transfer data across any of its ports. This enables the network administrator to segment users or services on a common LAN. This also provides a virtual separation of devices that may be processing sensitive information, keeping them isolated from the rest of the general devices on the LAN, regardless of their physical location.

3.4.2 Gateway Interface

A gateway is normally used to provide a point of entry to/from distinct networks. The gateway can do this by providing a physical media conversion (copper to fiber optic), a protocol conversion (serial line to Ethernet), and message translation (format A to format B). A gateway can also be associated with a route select function, which provides a route "look-up" to determine how to forward a packet out of the gateway for each addressed packet. As seen in Figure 3-1, the gateway device within the DSS design is located between the automated safety layer and the non-safety layer. This provides an alternate path (using the Ethernet protocol) for data exchange instead of the point-to-point connection using a Field bus protocol, such as PROFIBUS. The gateway provides a demarcation point between safety and non-safety–related processes.

The gateway can provide the necessary communication independence between safety and non-safety communication nodes. Isolation needs to be implemented to prevent the propagation of faults between safety channels and from safety-related processors and non-safety processors. When the gateway is properly configured, it can interrogate and restrict data communications between safety and important non-safety–related activities.

Gateways can also be configured to provide a proxy function. This is an intermediate process that provides independent connections across distinct networks. For example, if a communication device from the safety control network needs to make an outbound connection to a non-safety communication element to pass data, a proxy could be configured to provide the needed intercession. This method provides communication independence between the two communication nodes. The proxy server appears to be the content server of any requesting client on the non-safety network. This technique can also be used to hide safety control network addressing information from external (non-safety) communication elements. This can prevent device targeting by an adversary located on the non-safety network. Figure 3-13 shows this proxy function. Note that communication flows in only one direction (out) on the safety side of the gateway.

3.4.2.1 Gateway Design Considerations

Because the gateway provides both a conduit between safety and non-safety elements of the nuclear power plant, it is important to review some of the design-based attributes that can provide a more reliable and secure gateway.

3.4.2.2 Control Plane and Data Plane Separation

One advantage of a gateway is its centralized role in security. This allows many security controls to be leveraged at a single point. But this centralize approach to security controls also enables

single-point failures to disrupt data transport.
One design goal for gateways responsible for forwarding traffic is the forwarding data plane. A good design tenant of the data plane is its ability to continue to operate, even if the control plane or management network fails. Also, moving the control onto separate signaling channels and address space makes it harder to launch data-forwarding–function attacks through packets sent on the data path. Additionally, providing independent management channels allows gateway management functions to continue in the event a data plane failure occurs. Figure 3-14 shows a management/maintenance network isolated from the data transport plane.

3.4.2.3 Gateway Communication Independence

Modern DSSs contain computer systems that enable data communications within an individual safety system channel, between safety system channels and, with respect to the gateway, between safety systems and non-safety system processes. Communication independence must be maintained based on regulatory requirements.[9] Isolation between electrical and communication processes must be maintained to prevent fault propagation between safety channels and from non-safety computers to safety computers. To implement the independence of communication in new safety system designs, the gateway must afford the proper buffering between processes. The proxy function can accomplish this buffering. The proxy function can also provide electrical isolation because of its ability to create independent communication processes between two communicating entities. Fiber optic cabling can enhance electrical isolation. Another aspect of the gateway design is to prevent a non-safety workstation from affecting the operation of any safety-related equipment when it is performing its safety function [16].

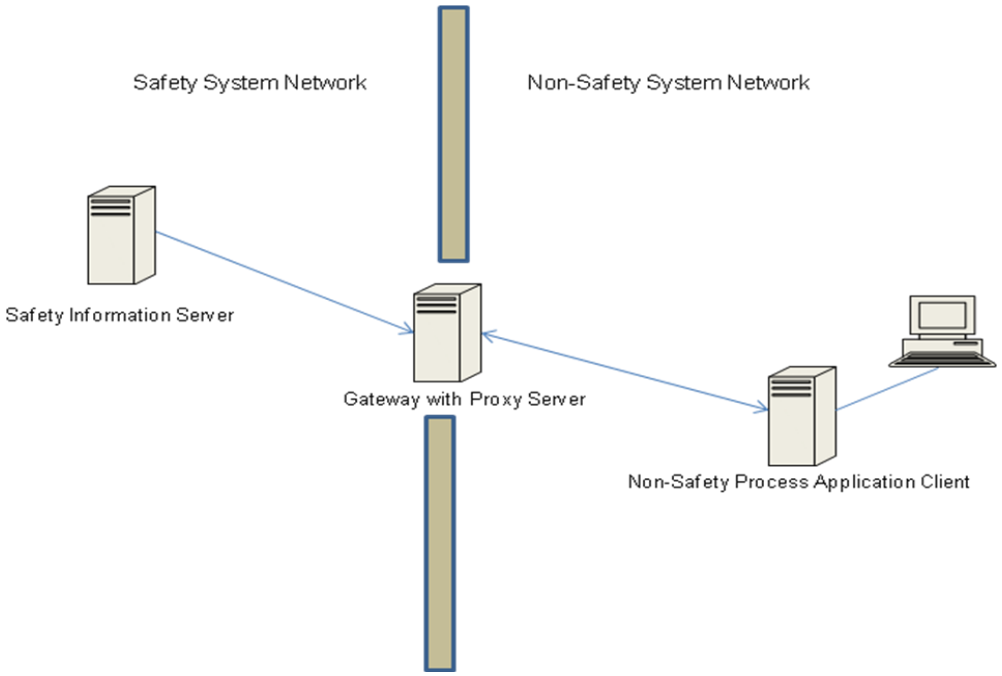

Figure 3-13. Proxy Server Function

[9] RG 1.152, Rev. 2, directs attention to Section 7.9 and Appendices 7.0-A and 7.1-C in NUREG-0800, "Standard Review Plan," for guidance on "Communication Independence."

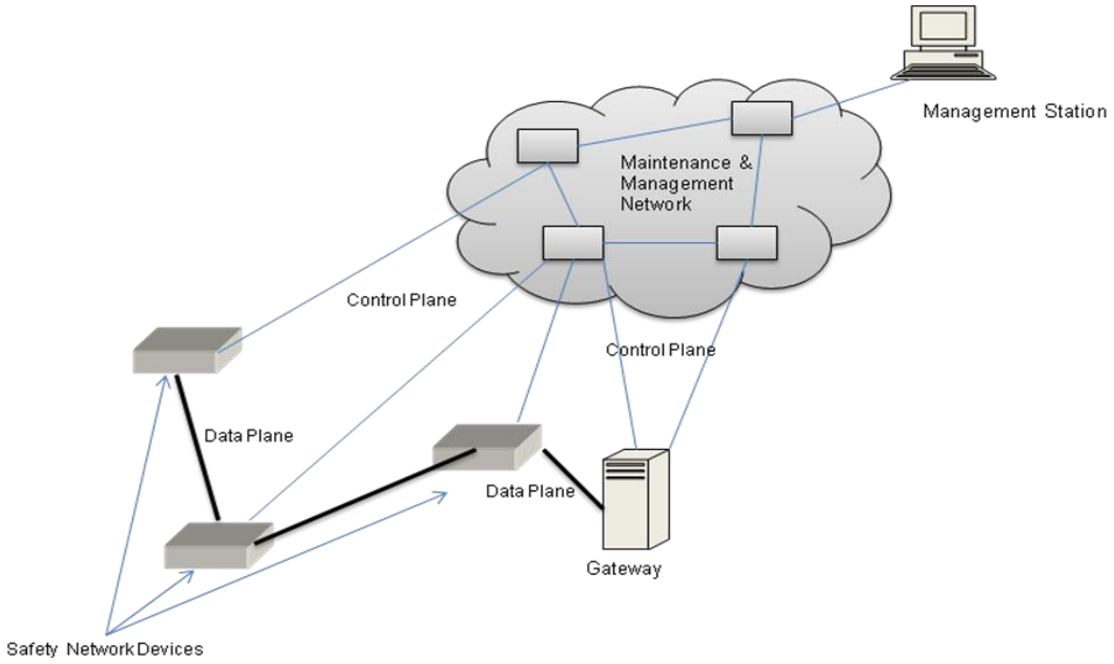

Figure 3-14. Out-of-Band Management & Maintenance Network

3.4.3 Security Observations

A secure communication network needs to meet all requirements classified as high safety, high maintainability, and user convenience requirements. Based on these requirements, several design concepts—such as a simple structure, standardized protocol, near real-time data transport, high-speed data rate, a hierarchical architecture, data surety, and enhanced network management—are identified as important elements in providing the necessary features for a network transporting safety-related information. Ethernet has been identified as a candidate protocol for providing data transport services from the Automated Safety Layer to the HMI Supervisory safety Layer and the Non-Safety Layer of the process information and control system.

3.4.3.1 Ethernet Observations

As previously described in section 3.4.1, the original Ethernet standard uses a protocol called CSMA/CD. This protocol provides contention-based access control to participating nodes on the system. This mechanism does not provide a deterministic approach to traffic management, but allows a more data responsive approach to data management; Ethernet switching creates an independent collision domain for each participating node, essentially suspending the CSMA/CD protocol. (See section 3.5.1, Deterministic Ethernet and Traffic Segregation, for a discussion of deterministic Ethernet.)

3.4.3.2 Ethernet Switching Vulnerabilities

The majority of attacks against the Layer 2 (Ethernet) protocol exploit the inability of a device to track the attacker who can, therefore, perform undetected malicious actions on the forwarding path to manipulate data and alter the data path. The following is a list of potential attacks that can be leveraged against Ethernet switched networks:

- MAC flooding attack
- ARP attack
- Spanning-tree attack

(For more detailed descriptions of these attacks and suggested mitigation measures, see Appendix D, section D.4.1, Ethernet Security Observations.)

3.4.3.3 VLAN Security Vulnerabilities

As discussed earlier in the Ethernet switching section, Ethernet VLANs can provide an additional layer of network data segregation to prevent unwanted data distribution and analysis ("snooping").

The attacks against a VLAN layered Ethernet network are associated with taking advantage of non-secure protocol interactions. The following is a list of potential attacks that can be leveraged specifically against VLAN Ethernet networks:

- Double-encapsulated IEEE Std 802.1Q/nested VLAN attacks
- VTP revision attacks

(For more detailed descriptions of these attacks and suggested ways to defend against these exploits, see Appendix D, section D.4.2, Ethernet VLANs.)

3.4.3.4 Gateway Security Observations

The gateway device is integral to segregating safety and non-safety applications. Because of this role, any interactions and configuration control should be authenticated. In respect to the gateway, two types of authentication can be identified: user authentication and network service authentication.

User authentication includes traditional computer authentication, such as logging into a computer or activating an HMI to adjust a process. Network service authentication can be regarded as the ability for networked devices to distinguish between authorized and unauthorized remote requests for data or to perform some action.

Some of the network accesses to the host OS are for system administration and should require authentication to that higher level of access. Many network accesses are through an application, frequently through a client-server model. The user access to the host should be regulated by the applications in that model. Web (HTTP)[10] access through the network should be regulated in the same manner, with the recognition that the protocol is stateless, thus, every automatic action should be considered a fresh access.

User authentication mechanisms are necessary to control access and provide audit logs of user activities on hosts. The simplest user authentication is a single, personal factor like a password.

[10] Hyper Text Transfer Protocol

This may be sufficient if there are additional physical security measures limiting access. (See Appendix D, section D.2, Physical Security Details, for information on physical access control.)

Data communications associated with the Automated Safety Layer can include using the Ethernet protocol for data transport to the HMI Supervisory Layer. This allows safety system operator review and data transport through a gateway device to send status information to non-safety network elements. The potential to compromise the safety network and/or safety data increases because of the requirement to properly configure the Ethernet node devices and the gateway.

3.4.3.5 Additional Observations

The *external threat* can potentially take advantage of the management network if it crosses between the non-safety network element and the safety network. as shown previously in Figure 3-14. The external adversary needs only to reach the management station(s) located on a non-safety network to be able to potentially influence the network elements within the safety network. This external threat can be reduced substantially if the network management workstations cannot be reached over a network connection. The gateway provides a demarcation point between safety and non-safety–related processes. The non-safety network, as shown in Figure 3-15, may contain connections to other plant data subnets including aspects of the business network. This cyber touch point can provide to the *external threat* a potential cyber access to the gateway device. The gateway access control and configuration become paramount in resisting any attempts by an external adversary. The gateway, along with properly configured firewall between the non-safety network and elements of the business network, can create a *layered* defense that reduces the external threat.

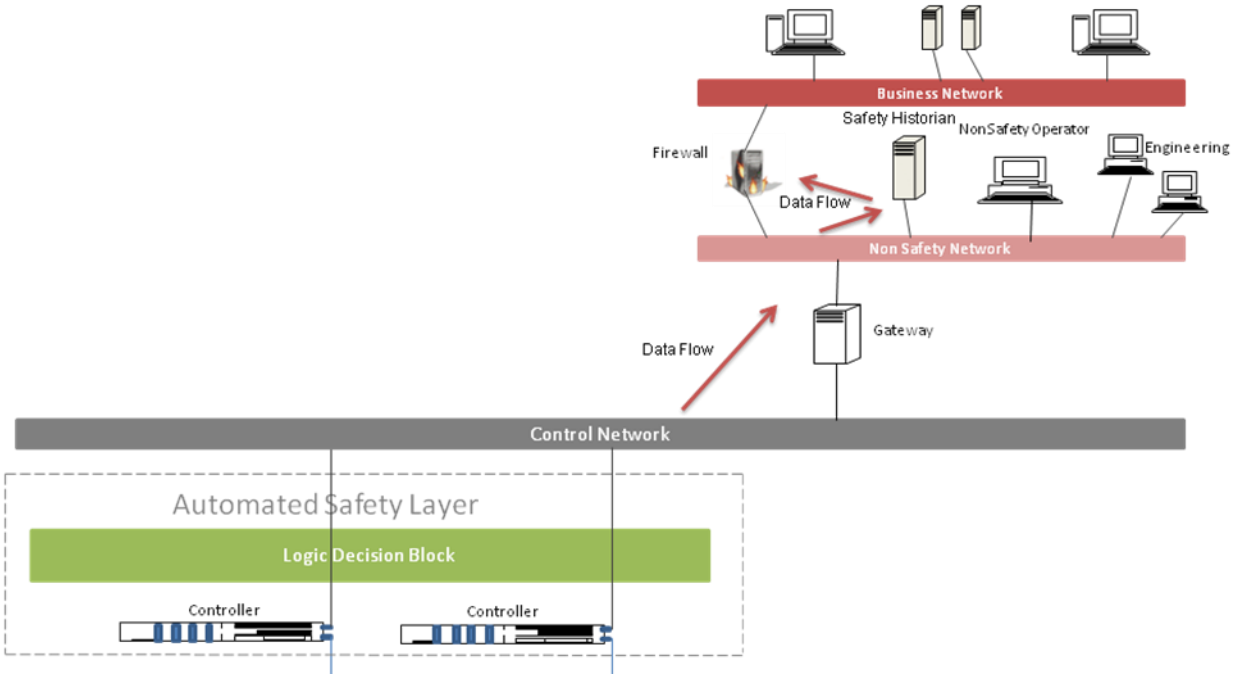

Figure 3-15. Safety to Non-Safety Data Flow

An *unprivileged insider* can take advantage of being on the Ethernet network by having the capability of physically attaching to an Ethernet port. Using unauthenticated remote management protocols, such as Telnet and TFTP, should be restricted when communicating with the Ethernet switch and the gateway. These communication protocols allow the user ID and password to be passed in the clear over the network. The unprivileged insider with physical access to the network can attach a device that can "sniff" these credentials as they travel across the network. Once these credentials have been captured, they can be used to escalate his/her privileges on the devices. The Ethernet switch and Ethernet protocol have some potential vulnerabilities that can be used to manipulate and disrupt communications. (These exploits, along with available mitigation techniques to help prevent attacks against the network, are discussed in more detail in Appendix D, section D.4.1, Ethernet Security Observations.)

The *privileged insider* would have a larger administrative role within the facility, but can possibly be limited in the number of systems that can be accessed or the location within the plant. Implementing a security policy that dictates the procedure on providing gateway access and configuration changes can provide some accountability to detect unauthorized changes. A formal process for change management should be instituted. It could include procedures, such as requiring that all configuration changes be reviewed by multiple administrators or subject matter experts to help detect malicious or accidental configurations. Implementing event logging as part of the access control process can help identify users, thus, possibly deterring malicious activity. Combining both physical protection mechanisms for personnel access control, along with restricting the number of systems that can be accessed, can provide some level of protection against this type of threat. The *privileged* insider is the most difficult threat to overcome.

Threats from *developer*- or vendor-based sources are associated with the Ethernet switches and the gateway device. Like many network capable products, switches and gateway devices may include factory default passwords and user accounts that have been configured as part of the vendor's pre-installation configuration. These potential vulnerabilities can be mitigated within a properly established security policy. Removing user accounts and default passwords should be part of a security policy established at each NPP facility. (See section 2.1, Security Policy in this report for additional implementation details). As devices become more software capable and dependent, the risks associated with supply chain and software lifecycle increase. Implementing appropriate controls (as described in section 2.5, Safety System Lifecycle) can reduce the overall risk against this product insertion vulnerability.

3.4.4 Regulatory Guidance Regarding Ethernet and Gateway Interfaces

There are no RGs that explicitly address the network communication protocol Ethernet associated with any network device. But the need to review the configuration setup is in the RG 1.152, "Criteria for Use of Computers in Safety Systems of Nuclear Power Plants," section 2.4, "Implementation Phase," which states, "The implementation activity addresses hardware configuration and setup; software coding and testing; and communication configuration and set-up [including the incorporation of reused software and commercial off-the-shelf (COTS products]." This would include any deployed Ethernet devices within the safety system network.

With respect to a gateway interface in RG 1.152, a reference is made to clause 5.6(a) of IEEE Std 7-4.3.2-2003 that "barrier requirements shall be identified to provide adequate confidence that the non-safety functions cannot interfere with the performance of the safety functions of the software of firmware." Also, regarding control of access, RG 1.152 states, "the design of the plant data communication systems should ensure that the systems do not present an electronic path by which a person can make unauthorized changes to plant safety systems or display erroneous plant status information to the operators".

Also in RG 1.152, section 2.5, Test Phase, provides guidance on configuration testing of security functions, which would encompass both the Ethernet and the gateway devices. The following statement is made: "The objective of testing security functions is to ensure that the system security requirements are validated by execution of integration, system, and acceptance tests where practical and necessary."

Thereafter, RG 1.152, section 2.5.1, System Features, refers to the validation of security features, which would include both the Ethernet and gateway devices. The following statement is made: "Each system security feature should be validated to verify that the implemented system does not increase the risk of security vulnerabilities and does not reduce the reliability of safety functions."

RG 5.71 does not address network communication and control devices specifically; however, Regulatory Position C.1 does note, "The rule [10] specifies that digital computer and communication systems and networks associated with these functions must be protected from cyber attacks that would adversely impact the integrity or confidentiality of data and software; deny access to systems, services, or data; or provide an adverse impact to the operations of systems, networks, and associated equipment." Therefore, the elements of the cyber security program framework that this RG endorses also include protection of the interconnecting networks that interface with CDAs.

In the Digital I&C Interim Staff Guidance (DI&C-ISG-04), section 3, discusses non-safety systems controlling the operation of safety-related equipment. It is mentioned that this should only occur if the following restrictions are enforced. It states, "A non-safety station should not affect the operation of safety-related equipment when the safety-related equipment is performing its safety function. This provision should be implemented within the safety-related system, and must be unaffected by any operation, malfunction, design error, software error, or communication error in the non-safety equipment."

3.5 HMI Supervisory Layer

The HMI Supervisory Layer is primarily associated with the potential architecture and devices found in a control room of a reactor facility. This is where the safety information is displayed and acted upon by the control room staff. The safety information arrives at the display consoles by multiple paths that include point-to-point (non-routed hardwired) connections using a serial protocol, such as PROFIBUS, and through a network connection using a network protocol, such as Ethernet. This is also the location for hardwired I&C connections that allow operators direct control of some actuators to be activated during emergency response situations.

Although network connections are being integrated into modern control room designs, back-up systems are still comprised of point-to-point (hardwired) connections.

The HMI Supervisory Layer contains the Qualified Display Systems (QDSs) and a plant overview panel in which operators monitor reactor status. The HMI is normally presented in the form of a console with a graphical user interface that displays status information about plant operations. HMI designs are incorporating more Web and Internet technologies because of the ease of user interaction, which makes data more accessible. The trend is to connect QDS to a terminal data network that allows for the inter-exchange of monitor information. This terminal data network can be comprised of a network protocol, such as Ethernet.

As mentioned earlier Ethernet technology has moved from a collision detection *shared* 10 Mbps LAN to a collision-free full duplex 100 Mbps or even a 1000 Mbps LAN. Although the contention has been removed and the speed increased, can it be considered a *deterministic* network connection?

Deterministic is commonly recognized as the ability of a system to respond with a consistent and predictable time delay between input and response. The cycle time between a master and slave type architecture, which can be deployed in a Field bus protocol such as PROFIBUS, can be calculated. For example, assume a polling message of 700 bytes is broadcast from a master node (operating at a transmit rate of 1 Mbps) to many multidrop slave nodes on a fiber optic medium. Assume the propagation delay on the fiber optic medium is negligible (approximately 1 micro-second) between the master and the 20 slave devices. So, the time required for the polling message to reach a slave device would be 700 x 8 (8 bits = 1 byte) x 1 microsecond = 5,600 microseconds. Also assume the processing delay for each slave is 10 microseconds and the response from the slave to the master is 700 bytes of information. The total query and response time for the polling cycle would be 5,600 microseconds + 10 microseconds + 5600 microseconds, for a total time of 11,210 microseconds, or approximately 11 milliseconds.

Ethernet switches process Ethernet frames using a *store-and-forward* technique. For example, if the master and slave process were replaced with a 100 Mbps Ethernet switch and a processing delay of 5 microseconds per switch port, the total send and response time would be as follows: 700 x 8 (8 bits = 1byte) x .01 microsecond (the clocking rate of a 100 Mbps transmission) x 5 microsecond (port processing delay) = 280 microseconds. If the processing delay for the end node to respond is 10 microseconds, the total cycle time would be 280 microseconds x 2 (round trip time) + 10 microseconds (end node processing delay) = 570 microseconds or approximately 0.6 milliseconds.

As this example shows, replacing a master/slave 1 Mbps architecture with a 100 Mbps Ethernet switch can reduce the cycle time dramatically. But the 11-millisecond cycle time is considered *deterministic* because it has no store-and-forward potential delays to encounter. Although the Ethernet switch has the advantage of reducing the overall latency in the cycle time, its store-and-forward technique cannot guarantee consistent latency. When the latency in a switched Ethernet

environment can be consistently predicted, Ethernet could be considered a *deterministic* technology.

3.5.1 Deterministic Ethernet and Traffic Segregation

To develop a more near real-time approach to Ethernet networks, designers have employed data identification and segregation schemes to filter and manipulate traffic based on priorities. Some of the key developments in providing a more deterministic approach to data traffic over Ethernet networks have been realized from the following:

1. Full duplex channels and higher link speeds.
2. Virtual LAN construction (IEEE Std 802.1Q).
3. Priority queuing (IEEE Std 802.1p).
4. Rapid Spanning Tree Protocol (IEEE Std 802.1w) for multiple switch networks.

Because items 1 and item 2 have been discussed earlier in this report (see section 3.4.1,Ethernet), they will not be included in the following discussion.

3.5.1.1 Priority Queuing

One of the most important standards that have been developed that position Ethernet networks to realize a more deterministic means of establishing data latency is the development of a quality of service (QoS) attribute that enables priority queuing and processing of critical data. QoS is a mechanism to allow better handling of data that passes over a network.

A marking scheme has been developed to discriminate from different types of data flowing across an Ethernet switched network. The priority queuing standard defines a 4-byte tag, which is inserted into the Ethernet frame header to identify different types of traffic. The IEEE Std 802.1p protocol sets a 3-bit value in the MAC header to indicate prioritization [17]. This 3-bit value provides priority levels ranging from 0 to 7, with level 7 representing the highest priority. This permits packets to cluster and form different traffic classes. Thus, when network congestion occurs, those packets with higher priorities will receive preferential treatment, while low priority packets will be kept on hold. Figure 3-16 shows an Ethernet header with the QoS field.

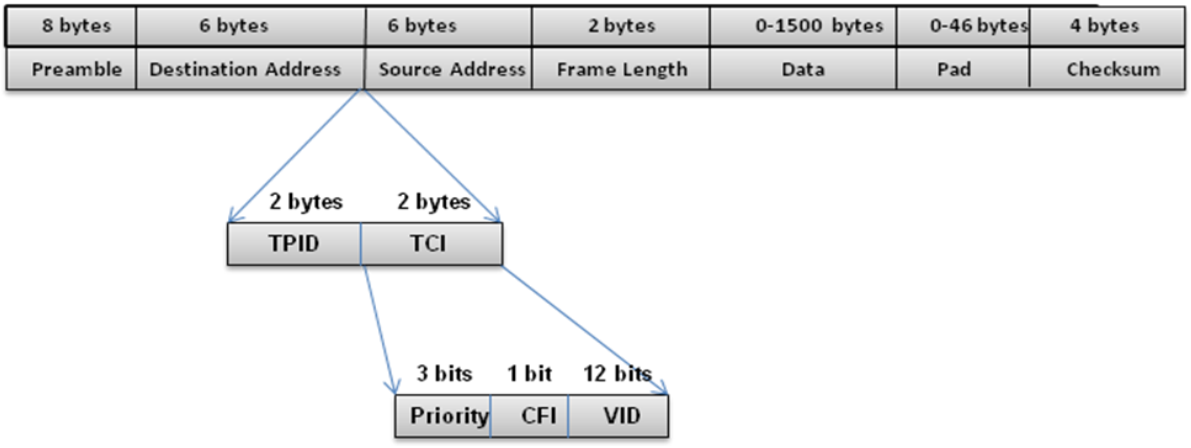

Figure 3-16. Ethernet Frame with Priority Tag

The IEEE Std 802.1p is an extension of the IEEE Std 802.1Q (VLANs tagging) standard. The 802.1p standard defines a tag that appends to an Ethernet MAC frame. The tag control information (TCI) field of the VLAN tag has three parts: the priority field (3 bits), the Canonical Format Indicator ([CFI] 1 bit), and the VLAN ID (12 bits). The rest of the Ethernet frames are the same as defined previously (in section 3.4.1,Ethernet).

1. **TPID**—Tag Protocol Identifier has a defined value of 8100 in hex. When a frame has the Ethernet type equal to 8100, this frame carries the tag IEEE 802.1Q / 802.1p.

2. **TCI**—Tag control information field including user priority, CFI and VLAN ID.

    ° **User Priority**—Defines user priority, using eight priority levels (2^3). IEEE 802.1p defines the operation for these three user priority bits.

    ° **CFI**—Canonical Format Indicator is always set to zero for Ethernet switches. CFI is used for compatibility reason between an Ethernet type network and a token ring type network. If a frame received at an Ethernet port has a CFI set to 1, then that frame should not be forwarded as it is to an untagged port.

    ° **VID**—VLAN ID, the identification of the VLAN, is basically used by the standard IEEE 802.1Q [18]. It has 12 bits and allows for the identification of 4096 (2^{12}) different VLANs.

3.5.1.2 Queuing Techniques

The implementation of priory is carried out within the queue of an Ethernet switch. Different techniques can be used when building queues for data prioritization. The simplest approach is the first-in-first-out (FIFO), which prioritizes data processing in the order it has been received. In a true FIFO system, all packets are stored in one queue. The following list identifies and defines other popular forms of queuing:

- **Weighted Fair Queuing.** WFQ is a flow-based queuing algorithm that schedules low-volume traffic first, while letting high-volume traffic share the remaining bandwidth. This is handled by assigning a weight to each flow, where lower weights are the first to be serviced.

- **Class-Based Weighted Fair Queuing.** Improvements to weighted fair queuing (WFQ) include CB-WFQ, where each type of traffic is assigned to a class, and each class is given its own queue. CB-WFQ allows easier queue management.

- **Hierarchical Weighted Fair Queuing.** HWFQ is another improvement to simple WFQ. In HWFQ, the network device monitors the worst-case packet delay for each queue and adjusts queue priorities automatically.

- **Random Early Detection.** RED attempts to alert the devices originating the stream that congestion is causing packet loss. RED simply drops packets if too many are received. This

causes the devices that are sending the packets to notice a problem and reduce their transmissions.

- **Weighted Random Early Detection.** WRED, an improvement to RED is RED that utilizes the IP headers priority value to determine which packets to drop.

3.5.1.3 Traffic Shaping and Rate Limiting

Another method of implementing QoS, traffic shaping uses pre-defined bandwidth allocations to determine which traffic flows should be delayed (queued) when they exceed their allocated bandwidth. A pro-active means of traffic shaping, rate limiting allows the changing of priority fields of data flows or allows dropping of the packets.

3.5.1.4 Rapid Spanning Tree Protocol

The Spanning Tree Protocol (STP; IEEE Std 802.1D [19]) is a loop-prevention protocol that is implemented at the data link layer. This technology allows switches to communicate with one another to discover and map physical loops in the network. The STP creates a tree structure, which has loop-free leaves and branches, and that spans the entire Layer 2 network. Two STP-network attributes are used to create its logical tree-like structure: the *bridge ID (BID)* and the *path cost*.

Rapid Spanning Tree Protocol (RSTP; IEEE Std 802.1w [20]) allows the STP tree structure to converge more quickly. This allows information about defective links to be propagated more quickly, thus, allowing the network to recover from network disruptions.

To create the STP "tree" structure, a *root bridge* needs to be designated. An election process accomplishes this. All switches/bridges send out Bridge Protocol Data Units (BPDU) advertising the following attributes:

- Root bridge identification (root BID)
- Root path cost
- Sender BID
- Port ID

The important attribute in the root bridge election process is the root BID. All switches participating in the election process choose the root bridge based on the lowest BID. To define a root bridge, the operator must change the value of the first two bytes of the BID to a lower value than factory default. Otherwise, the switch in the network that has the lowest value MAC address will be defined as the root bridge. If this root bridge has a less than desirable location within the network, a less then optimal switching path for data transport could be constructed.

STP was designed to ensure a loop-less network environment. There are three basic steps in which STP establishes its topology: 1) electing the root bridge, 2) selecting one root port on every non-root bridge, and 3) selecting one designated port per network segment. Electing the root bridge is done by exchanging Layer 2 BPDUs. When the STP is in use, every port on a switch goes through several stages. The bridge with the lowest ID becomes the root bridge.

When sending BPDUs, the switch sets the root BID to its own ID. Because every switch stops sending BPDUs when it receives a BPDU with a lower root ID than its own, eventually the only switch sending BPDUs is the root bridge.

3.5.2 Security Observations

When deploying a terminal data network for control room visualization using Ethernet, it is important to assess the need for near real-time data transport. Without the implementation of a priority queuing scheme, the un-deterministic nature of Ethernet communications may prevent information from being shared or distributed in a deterministic *timely* fashion. The vulnerabilities associated with data communications on the HMI Supervisory Layer can be primarily associated with a failure to transport data across the network when it is requested and with the possible modification of data for purposes of deception or unauthorized control.

3.5.2.1 Denial-of-Service

Another reason to implement a priority queuing scheme within an Ethernet network is to reduce the chances of a DoS situation. The potential corruption of data or loss of packets due to a large and sudden increase of traffic can prevent data from being shared.

An occurrence of a large and unexpected increase in network traffic happened at the Davis-Besse Nuclear Power Station in 2003. A computer network server was infected with the Slammer MS-SQL server worm through an unsecured contractor network connection. This allowed access to the network through a pathway that bypassed the plant firewall. The worm affected both the business network and the plant network (although not the safety system). As a consequence, large amounts of data were sent onto the plant site networks (a DoS attack). The large amounts of data caused many of the plant site computers to cease communicating with other computers on the networks [21]. This DoS situation may have been reduced for critical data streams if a prioritized queuing had been implemented within the Ethernet network.

3.5.2.2 Priority Queuing Compatibility

It is important to note that IEEE Std 802.1p is not backward compatible and can lead to instability on networks with non-IEEE Std 802.1p switches. Non-IEEE Std 802.1p compliant devices can misinterpret the header used by the IEEE Std 802.1p protocol. It is important that the Ethernet switches, Ethernet cards, and Ethernet device drivers are all IEEE Std 802.1p compatible when deploying a priority queuing scheme on a switched Ethernet network.

3.5.2.3 Potential Qualified Display Systems Exploit

When a QDS series is on a terminal data network—which may be comprised of Ethernet switches—it is important to note that on a LAN environment all local users of the environment may have unauthorized access to any QDS resident on the LAN. In some cases this unauthorized access is gained because of the un-authenticated protocols that the QDS uses, and which are not properly protected. This lack of protection can result in an adversary on the terminal network obtaining QDS access through an Ethernet switch port. Once an adversary can gain the highest privilege level on the operator interface, s/he may be able to manipulate elements of its operator display. This type of attack has been documented in a report to the U.S. NRC [22]. Proper host-based security management procedures, along with using only authenticated network access, can prevent these types of attacks. Also, if functions associated with the status and control of safety

information need to be truly independent, then hosting different functions on physically different QDSs can establish this independence [16]. This will prevent a single compromised QDS from allowing an adversary greater control over many functions.

3.5.2.4 Spanning Tree Exploit

An attack vector can disrupt the switch spanning-trees, destabilize their MAC address tables, and hold the network in a constant state of re-election of the root bridge. This can be achieved because there is no authentication mechanism built into the STP.

By crafting BPDUs of a non-existent switch with an ID of 1, the adversary can elect its non-existent switch as the root bridge. Using a minimal "maximum age" for the crafted packets, and not sending BPDUs within that time, will cause another election on the network, during which the adversary will start sending bogus BPDUs, once again winning elections and becoming the root bridge. By repeating this process, the network will be in a constant state of re-electing the root bridge. It will fail to converge, thus, reducing data traffic and saturating the network with BPDU frames.

3.5.2.5 Spanning-Tree Attack Protection Measures

Some simple methods can prevent the exploitation of the STP vulnerability in a network. For any STP attack to be feasible, the switch must accept BPDUs on a port that the attacker has access to. It is, therefore, possible to make such an attack impossible by denying access to STP-enabled ports to ordinary users. This can be done by disabling STP on access ports, having port security enabled on all user ports, and restricting physical access to network equipment.

With STP disabled on user ports, the attacker would have to access the switch physically and use a switch-to-switch port to connect his/her computer to (assuming all non-used ports are either disabled or have STP disabled). If physical access to network devices cannot be restricted, other measures must be taken to ensure network security. Port security is a feature that allows the switch to accept frames from only a given number (usually the first learned) of source MAC addresses. Enabling port security on user ports will make the attack unfeasible without prior network "sniffing" or hijacking a user workstation.

Data flow communications with the HMI Supervisory Layer are normally associated with elements of the Automated Safety Layer. Communications within this interaction can include both Field bus protocols, such as PROFIBUS, and the switched Ethernet protocol for data transport as shown in Figure 3-17.

3.5.2.6 Additional Observations

The HMI Supervisory Layer is quite isolated from the *external threat*, with the assumption that the terminal data network that supports this layer has no external connections to other points within the plant data network. It is important to note that the means of updating software or adding patches to any element of the HMI Supervisory Layer needs to be evaluated for potential compromise due to infected media. The external threat can use infected media to bypass access control features and gain admission to a network segment that seems isolated from any external network connectivity. (See Appendix D, section D.16, Malicious Software Protection, for details on software protection procedures.)

An *unprivileged insider* who has access to either of these data transport networks can potentially manipulate or deny data traffic to the HMI Supervisory Layer. This layer, as shown in Figure 3-17, is also comprised of a terminal data network. The terminal data network may use Ethernet to share display data and, thus, can be subject to potential attacks associated with the Ethernet protocol and the deployed Ethernet devices. (For a detailed description of potential attacks that can be leveraged against Ethernet network, see Appendix D, section D.4.1, Ethernet Security Observations. Appendix D also identifies mitigation techniques and procedures that can reduce the insider threat.) The *unprivileged insider* may also take advantage of any improper or weak authentication practices associated with user access controls on the HMI workstations. (See Appendix D, section D.13.1, Host Access Control, for details of proper access control implementations.)

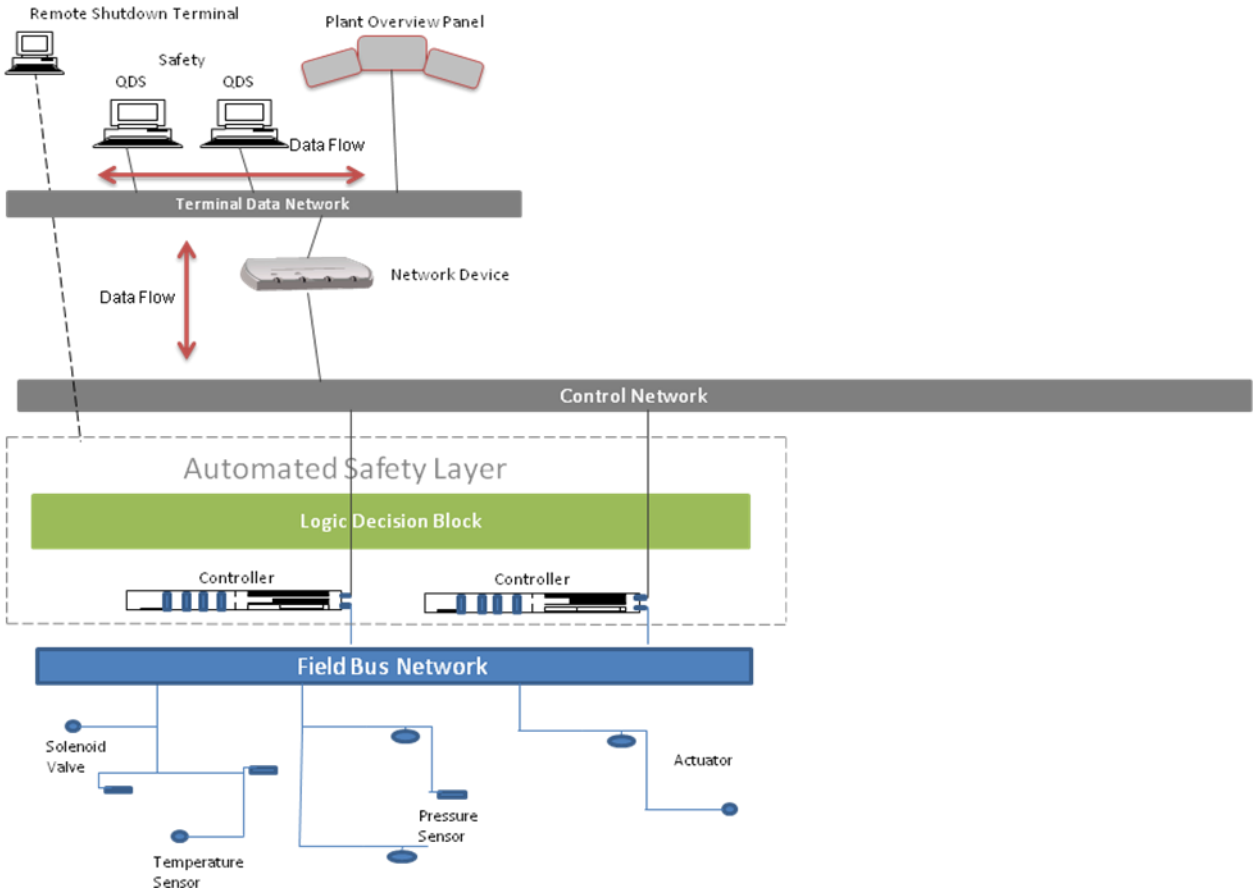

Figure 3-17. HMI Supervisory Layer Data Flows

Threats from developer- or vendor-based sources would be associated with the Ethernet switches, qualified display systems, or other similar operator workstations and applications. These devices may include services running on OSs that have been programmed with back-doors for vendor access. Using removable media, such as thumb drives, to update software or to apply patches may provide an avenue for compromise or vulnerable to virus infection. A proper security policy should include practices and procedures that help provide a more secure operating environment.

Implementing appropriate controls (as described in section 2.5, Safety System Lifecycle) can reduce the overall risk against potential product vulnerabilities.

3.5.3 Regulatory Guidance Regarding HMIs and Deterministic Communications

RG 1.152 and RG 5.71 do not specifically address HMI, but they can be captured within the category of digital computer-based systems. In that respect, some security criteria are identified to include access control. As stated in the discussion, "Computer-based systems are secure from electronic vulnerabilities if unauthorized and inappropriate access and use of those systems are prevented." And also, "Controls should address access via network connections and via maintenance equipment. Additionally, the design of the plant data communication systems should ensure that the systems do not present an electronic path by which a person can make unauthorized changes to plant safety systems or display erroneous plant status information to the operators."

No RGs were identified that discussed the specific requirements of deterministic communications, in particular Ethernet. In NUREG/CR-6812, Emerging Technologies in Instrumentation & Controls, section 2.2.6, Network Design, the following statement is made concerning network design in nuclear power plants: "Evaluation of emerging network design approaches may become an important consideration in the review of future power plants with highly interconnected, distributed computing environments for autonomous control and information systems. The trends in network design should be monitored, in particular, configuration approaches for high network availability and the robustness of high-speed switching routers and network interface processing."

And, more specifically, in section 2.2.3, Safety-Related Field Bus, the following observation is made about deterministic Ethernet: "High-speed Ethernet is being implemented, and more effort is being made to provide precise and deterministic bus timing in order to support control loops and safety actions. Given the current state of competitive pressure among the vendors, it seems inevitable that fieldbus systems will adapt networking developments occurring for general computer systems and telecommunications."

3.6 Non-Safety Information Layer

The business information layer of the Non-Safety Information Layer is comprised of IT system computers, such as OPM, ISM, and business performance CIM computers. This layer has access to external networks, such as the Internet, and must be properly protected to prevent compromise of any its operations. Modern data communication architectures for the business information layer create separate zones or domains that normally have a single entry and exit point to the backbone network (i.e., data plant network). Within these zones the communication between computers is based on trust and all devices within the same zone can communicate with one another only through need-to-know access segregation**.**

There may be zones constructed that have a trust relationship with one another, or there may be multiple zones with varying levels of trust. There may also be single devices that may not be part of any zone but may be allowed to communicate with specific zones. Each zone access point must be properly protected to prevent unwanted data flows from exiting or entering the zone

perimeter. (For a full architectural description of zone design and protection, see Appendix D, section D.12, Intrusion Monitoring and Sensor Deployment.)

3.6.1 Perimeter Defense

When securing the business information layer, it is easier to start at the edges or perimeter. This is where the business layer interfaces with other sections of plant operations and with the rest of the world. The broadest based protections generally occur when securing the perimeter of the network.

3.6.1.1 Firewalls

A firewall should protect each security zone. The firewall should be situated at the zone entry point or at the backbone connection point. This firewall should allow communication to stations in its protected zone only from trusted zones or trusted devices from the backbone or the office network. As it is relatively easy to spoof IP or MAC addresses, the access control should not be based on packet (address) filtering techniques alone, but on other techniques for ensuring proper data flow, such as *stateful* inspection. Stateful inspection can determine which end device initiated the connection request and determine if it is allowed. Also external, distant connections may need the assurance of cryptographic-based protocols, such as the virtual private network (VPN). (Examples of VPN protocols can be found in Appendix D, section D.3, Virtual Private Networks.)

3.6.1.2 Intrusion Detection

An IDS is a type of security monitoring system for both network and host-based traffic. A Network IDS (NIDS) analyzes information from various areas of the network for any security concerns. It can be configured to identify intrusions or attacks originating from outside an organization network or to identify attacks or misuse from within the organizational protected boundaries. (See Appendix D, section D.10, Intrusion Detection, for additional IDS discussions.)

3.6.1.3 Host Access Control

The term *host* or *workstation* normally refers to a device that contains an OS. The OS allows users of different roles to interact with applications and utility services on the host or workstation. The OS can be seen as one layer of defense to protect applications and sensitive information. It can also contain a Host IDS (HIDS). If configured properly, it can ensure that only designated persons can make changes to system configurations and security policies. (See Appendix D, section D.13.1, Host Access Control, for more details.)

3.6.2 Security Observations

The following listed observations provide procedures for providing and maintaining host-based security on computer-based systems used for operations. These recommended procedures apply within the safety system network as well as the non-safety system assets.

3.6.2.1 Host Access and Protection

Prior to granting a user access to any computer system, a review of user requirements should be validated. A role-based review should identify the level of access and, thus, the level of system resources, required to perform one’s job.

A password maintenance process for the system should be created that includes a means to track password changes, the strength of the password, and the frequency of change.

A log file directory should be created on each host system that documents user log-ons and tracks important events, such as file additions and/or changes. This log directory should be protected with appropriate directory and file permissions.

A formalized change management process should be created that documents all changes to the system. This will help identify the current OS software for potential upgrade protections.

IT personnel responsible for cyber-asset configuration and control should participate in regular and documented training sessions that include vendor-supported network devices. Training should provide information on the latest approaches to host-based and network security.

(For additional information on host access and protection, see Appendix D, section D.13.1, Host Access Control.)

3.6.2.2 Application Protections

Identify all services and applications on all hosts to determine if they are needed for operations. If some services are not needed for normal operating conditions, the system security could be improved by disabling them.

Review all directories and file permissions on the system. Determine which files must be highly protected and only allow administrative access to these files and/or directories.

Provide a timely means of scheduling patch audit reviews to determine if any network devices (e.g., firewalls and network switches) need updated patches. This also applies to all NPPDN devices and protocols. (For additional application security discussions, see Appendix D, section D.15, Application Access and Control.)

3.6.2.3 Physical Protection Measures

The previous discussions did not mention any physical protection mechanisms that should be in place to protect the network from unauthorized access. A physical security review should be completed for *physical access* to both the safety and non-safety network and other sensitive access points. (See Appendix D, section D.2, Physical Security Details, for a detailed description on proper physical security protections.)

The Non-Safety Information Layer is comprised of an Ethernet switched network that allows communication interactions between the devices resident on this network. This network can contain computer servers and workstations that are running on general OSs supporting a variety of applications. This network also provides an interface to the gateway that is the demarcation point between safety and non-safety processes. Since the variety of computer systems and applications are much more abundant on this network, the potential for vulnerabilities residing on these computers increases along with their potential exploit. Operational management and software lifecycle procedures guided by a safety policy becomes paramount to reduce the ability to exploit vulnerabilities by external and internal threats.

Because the Non-Safety Information Layer has potential interfaces to other networks within the data plant network, it is much more open to attacks from an *external threat*. As shown in Figure 3-18, the interaction between the business network and the non-safety network allows for a cyber path from the open Internet. The external threat is separated from this network segment by a firewall and potential protections, such as intrusion detection and intrusion protection systems. These protection systems can provide proper isolation from the adversary if configured and maintained appropriately. Since the non-safety network resides on a different network segment than the business network, an additional firewall would have to be penetrated to allow the external adversary direct access to elements of this network.

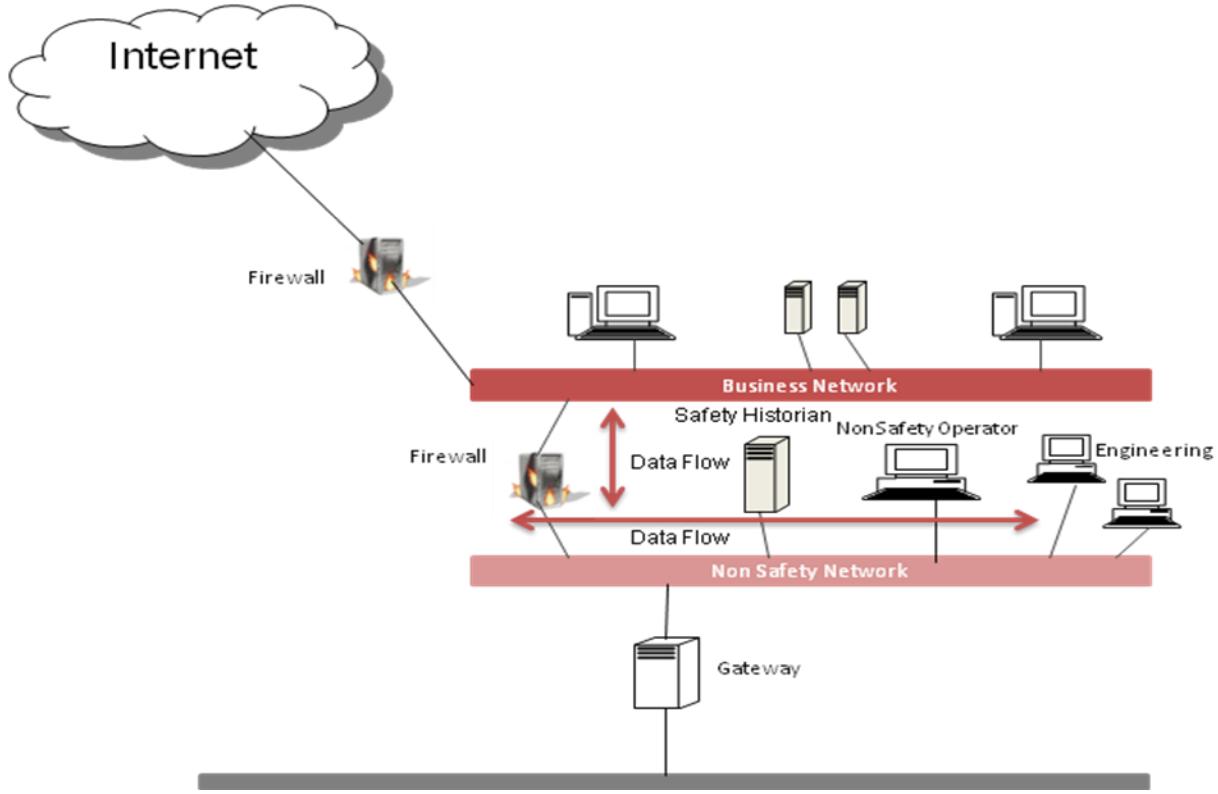

Figure 3-18. Non-Safety Network Data Flows

3.6.2.4 Additional Observations

An *unprivileged insider* who has access to the network can potentially manipulate or deny legitimate traffic from reaching its intended destination. This can be accomplished by taking advantage of some of the potential vulnerabilities associated with a deployed Ethernet network. (For more details associated with Ethernet vulnerabilities, see Appendix D, section D.4.1, Ethernet Security Observations.) The unprivileged insider may gain access also to the gateway device if proper authentication mechanisms to protect it from unauthorized access are not in place. (See Appendix D, section D.13, User/Operational Management, for details of proper user access controls.)

The *privileged insider* would have a larger administrative role within the facility, but can possibly be limited in the number of systems that can be accessed or the location within the plant. Having a security policy in place that dictates the procedure on providing firewall or gateway access and configuration changes can provide some means to detect unauthorized changes. A formal process for change management should be instituted. It could include procedures, such as requiring that all configurations changes be reviewed by multiple administrators or subject matter experts to help detect malicious or accidental configurations. Providing logging as part of the access control process can help identify users and possibly deter malicious activity. Combining both physical protection mechanisms for personnel access control— along with restricting the number of systems that can be accessed—can provide some level of protection against this type of threat. The privileged insider is the most difficult threat to overcome.

Threats from *developer*- or vendor-based sources associated with the non-safety network increase dramatically with the number of computer systems and applications resident on the non-safety network. These systems may have Trojan programs running to allow back-door access to applications. They will also be more likely to be virus-infected due to the OS commonality and applications resident on these systems. Proper host access controls, application controls, and malicious software protection become paramount in providing the proper line of defense against compromise. Using removable media, such as thumb drives, to update software or to apply patches may also provide an avenue for compromise. A proper security policy should include practices and procedures that promote a more secure operating environment. Implementing appropriate controls (as described in Appendix D, section D.13.1, Host Access Control) can reduce the overall risk against potential product vulnerabilities.

3.6.3 Regulatory Guidance Regarding Modern Network Communications Security Practices

10 CFR 73.54, Protection of Digital Computer and Communication Systems and Networks, contains some important aspects related to network security, but not specifically related to safety systems. It requires each licensee to submit a cyber security plan. This plan is not only associated with the safety system but other aspects of network communications, as stated in paragraph (a), "Each licensee subject to the requirements of this section shall provide high assurance that digital computer and communication systems and networks are adequately protected against cyber attacks." This is described in more detail in (1), "The licensee shall protect digital computer and communication systems and networks associated with: (ii) Security functions; (iv) Support system and equipment which, if compromised, would adversely impact safety, security, or emergency preparedness functions." And further, in (2) the following statement is made: "The licensee shall protect the systems and networks identified in paragraph (a)(1) of this section from cyber attacks that would: (i) Adversely impact the integrity or confidentiality of data and/or software; (ii) Deny access to systems, services, and/or data; and (iii) Adversely impact the operation of systems, networks, and associated equipment."

There are also references about the need to identity import assets of the nuclear power plant communication network that would be import to protect against cyber attacks and for the cyber security program to be designed to carry out the following: (c)(2) "Apply and maintain defense-in-depth protective strategies to ensure the capability to detect, respond to, and recover from

cyber attacks." This statement can encompass the installation and operation of a cyber IDS and firewall.

The aspects of host-based access control for non-safety system computers are not specifically called out in an RG. Access control can be found in Clause 5.9 of IEEE Std 7-4.3.2-2003 (referring to Clause 5.9 in IEEE Std 603-1998), Control of Access, which states, "The design shall permit the administrative control of access to safety system equipment," but also applies to the overall design of the generating station with the following statement: "These administrative controls shall be supported by provisions within the safety systems, by provision in the generation station design, or by a combination thereof." The non-safety system is called out specifically to prevent unauthorized access by its proper design in the following statement also found in the section B discussion of Control of Access in RG 1.152, rev. 2, which states, "The design of the plant data communication systems should ensure the systems do not present an electronic path by which a person can make unauthorized changes to plant safety systems or display erroneous plant status information to the operators."

In a similar fashion, RG 5.71 bases its regulatory positions concerning an effective cyber security program on the standards provided in NIST SP 800-53, Recommended Security Controls for Federal Information Systems [23]; and NIST SP 800-82, Guide to Industrial Control Systems Security [24], as offering a comprehensive approach for complying with the cyber security requirements specified by 10 CFR 73.54.

This page intentionally blank

4. Summary and Conclusions

This report provides the reader with an understanding of the elements of network security and how they can be applied to a nuclear power plant data network. Some modern practices of designing and deploying NPP network architectures and their associated components were identified and examined. Observations associated with proposed design implementations were presented throughout the report and accompanying appendices to provide an understanding of security-related issues.

A DSS architecture was presented along with identified protocols and components associated with its design. Three protocols were identified to include a popular Field bus protocol called PROFIBUS, which was described along with its important features and protections. The Ethernet protocol, its description, typical implementations, and some of its limitations and vulnerabilities were also discussed with additional information provided in Appendix D.4. The FDDI was presented because of its potential use in newer safety system designs.

Network components associated with modern best practice designs are described in Appendix D. Network component discussions included the use of VPNs and how they can be used to secure external connections originating from the plant data network. A discussion of VLANs and how they are used to improve network security within a data plant network was also provided. Border network protection mechanisms were described including an overview of firewall and IDS, proper placement, and use were described. A section on host-based access control identified the modern means of providing both user and application protections, including a discussion on the RBAC implementation.

A wireless architecture is also discussed in Appendix D to show some of the important elements for protection of the NPPDN from unauthorized access originating from the wireless medium. And finally, the importance of implementing compliance testing to ensure that defined network security policies were being implemented and still relevant for NPPDN protection was provided.

The protocols, procedures, and protections described in this report are relevant to modern networks being designed and deployed today. The primary elements necessary to provide comprehensive network security include the development of a security policy that provides a framework for all responsible plant personal to identify the important aspects of the network and to create a plan for securing its access and operation. Protecting access to the network includes a discussion on important aspects of physical security implementation. The importance of maintaining security throughout the development, installation, operation and maintenance of the network is also reviewed through proper lifecycle analysis. For example, a perimeter firewall protecting the external access to the plant data network offers no protection against internally released viruses originating from mobile laptop computers or removable media connected to the control or safety network. A large number of reported incidents involve known and addressable threat vectors. Many of these types of security incidence could have been mitigated if better security policy, practices, and education programs were implemented rather than through solely technology-based solutions.

Modern NPPDN designs are continuing to incorporate advances in network communications and data distribution. These advances will impact the network plant architectures associated with the operation of nuclear power electricity production. The previous review has provided observations into the direction and the implementation of more modern DSSs and the modern NPPDN in which they reside. Within this push for modernization, the ability to isolate safety system processes from external, less-trusted networks becomes more difficult. Modernization of plant, safety, and control networks creates the potential for secondary cyber pathways into the safety and control system networks. Proper risk mitigation starts with a comprehensive security policy management program that covers all aspects of the plant data network to include both the control and safety network systems. This policy should include both cyber and physical security to guide the proper implementation of a comprehensive, in-depth defensive strategy. This includes 1) better layering of firewall defenses, data-communication monitoring with IDS and Intrusion Prevention Systems for both the wired and the wireless media; 2) the proper hardening of end-point devices and user interface configurations, which include authentication, patch management, and antivirus deployment; and 3) the protection of both internal and external data communications through the proper use of both VLAN and VPN technologies. It will also require a continual vigilance in the review and understanding of the security impacts to safety systems when newly proposed technologies are inserted.

5. References

1. Regulatory Guide 5.71. *Cyber Security Programs for Nuclear Facilities*. U.S. Nuclear Regulatory Commission.

2. Regulatory Guide 1.152. *Criteria for Use of Computers in Safety Systems of Nuclear Power Plants*. U.S. Nuclear Regulatory Commission.

3. IEC 61158. *Industrial Communication Networks—Fieldbus Specifications*.

4. IEC 61784-1. *Industrial Communication Networks, Profiles—Part 1: Fieldbus Profiles.*

5. IEEE Std 802.4. *IEEE Standard for Local Area and Metropolitan Area Networks.*

6. Rivest, R. *The MD5 Message-Digest Algorithm*. IETF RFC 1321, April 1992.

7. IEEE Std 7-4.3.2. *Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating Stations*. 2003.

8. NUREG/CR-6812. *Emerging Technologies in Instrumentation and Controls.* March 2003.

9. NUREG/CR-6888. *Emerging Technologies in Instrumentation and Controls: An Update.* January 2006.

10. 10 CFR Part 73.54. *Protection of Digital Computer and Communication Systems and Networks*. U.S. Nuclear Regulatory Commission, Washington, DC.

11. ANSI/ISO 9314. *Information Processing Systems, Fiber Distributed Data Interface*—Part 1: *Token Ring Physical Layer Protocol (PHY);* Part 2: *Token Ring Media Access Control (MAC);* Part 3*: Physical Layer Medium Dependent (PMD);* Part 6: *Station Management (SMT).*

12. IEEE Std 802.10. *Standards for Local and Metropolitan Area Networks: Standards for Interoperable LAN/MAN Security (SILS).* 1998.

13. IEEE Std 603. *IEEE Standard Criteria for Safety Systems for Nuclear Power Generating Stations,* sections 5.4, 7.1-C. 1998.

14. IEEE Std 802.3. *IEEE Standard for Information Technology; Telecommunication and Information Exchange between Systems: Local and Metropolitan Area Networks—Specific Requirements.*

15. IEEE Std 802.3. Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications.

16. Digital I&C Interim Staff Guidance (DI&C-ISG-04) Task Working Group #4: *Highly-Integrated Control Rooms—Communications Issues, Interim Staff Guidance*, revision 0.

17. IEEE Std 802.1p. *Traffic Class Expediting and Dynamic Multicast Filtering*, (merged into IEEE Std 802.1D, Ref. 18).

18. IEEE Std 802.1Q. IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks.

19. IEEE Std 802.1D. IEEE Standard for Local and Metropolitan Area Networks—Media Access Control (MAC) Bridges. 2004.

20. IEEE Std 802.1w. IEEE Standard for Local and Metropolitan Area Networks—Common Specification. Part 3: Media Access Control (MAC) Bridges—Amendment 2: Rapid Reconfiguration. 2001.

21. Nuclear Regulatory Commission (U.S. NRC). Information Notice (IN) 2003-14. *Potential Vulnerability of Plant Computer Network to Worm Infection.* U.S. Nuclear Regulatory Commission, Washington, D.C. August 29, 2003.

22. Michalski, J. T., et al. *Vulnerability Assessment of the Common Q Digital Safety System to Cyber Threats*. Sandia National Laboratories, prepared for the Division of Engineering, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission. June 2009. (Not publicly available.)

23. NIST SP 800-53, rev. 3. *Recommended Security Controls for Federal Information Systems.* National Institute of Standards and Technology, Gaithersburg, MD. August 2009.

24. NIST SP 800-82. *Guide to Industrial Control Systems Security*. National Institute of Standards and Technology, Gaithersburg, MD. September 29, 2008.

Appendix A: Bibliography

American Society of Mechanical Engineers (ASME). *Nuclear Quality Assurance (NQA) Standards*, (i.e., ASME NQA-1, 1989 ed), NQA-2a-1990 addenda (Part 2.7) to ASME NQA-2, 1989 ed.

Anderson, Robert H., Thomas Bozek, Tom Longstaff, Wayne Meitzler, Michael Skroch, and Ken Van Wyk. Conference Proceedings: *Research on Mitigating the Insider Threat to Information Systems—#2*. CF-163-DARPA, RAND National Defense Research Institute, The RAND Corporation, Santa Monica CA. August 2000.

ANSI/ISA-TR99.00.01-2004. *Security Technologies for Manufacturing and Control Systems*.

ANSI/ISA-TR99.00.02-2004. *Integrating Electronic Security into the Manufacturing and Control Systems Environment*.

Clauset, A., M. Young, and K.S. Gleditsch. "On the Frequency of Severe Terrorist Events." In *J Conflict Resolution*. 51(1):58-88. 2007.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.3, Configuration Management. October 2005.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.4, Defense in Depth. October 2005.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.7, Host and Device Security. October 2005.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.11, Intrusion Detection Systems, October 2005.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.12, Logging. October 2005.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.13, Software Updates. October 2005.

Department of Homeland Security. *Control Systems Cyber Security: Defense in Depth Strategies*, external report. INL/EXT-06-11478. May 2006.

Dierks, T., et. al, Internet Engineering Task Force (IETF), Network Working Group. *TLS Protocol Version 1.0, Request for Comments 2246*. The Internet Society. 1999.

Faria, Daniel B. and David R. Cheriton. "DoS and Authentication in Wireless Public Access

Networks." In *Proceedings of the First ACM Workshop on Wireless Security.* (WiSe'02*)*. September 2002.

Freier, Alan O., et. al. *Internet Engineering Task Force (IETF), Transport Layer Security Working Group, the SSL Protocol*, version 3.0. Internet draft. 1996.

Garcia, Mary Lynn. *The Design and Evaluation of Physical Protection Systems*, 2nd. ed. Butterworth-Heinemann, Burlington, MA, 2007.

Garcia, Mary Lynn. *Vulnerability Assessment of Physical Protection Systems*. Butterworth-Heinemann, Burlington, MA. 2006.

Huber, Lt. Col. Arthur F. II, and Jennifer M. Scott. "The Role and Nature of Anti-tamper Techniques in U.S. Defense Acquisition." In *Acquisition Review Quarterly,* fall 1999.

IEC 60960. *Functional Design Criteria for Safety Parameter Display System for Nuclear Power Stations*, ed. 1.0. 1988.

IEC 60964. *Nuclear Power Plants: Control Rooms—Design*, ed. 2.0. 2009.

IEC 60965. *Nuclear Power Plants: Control Rooms—Supplementary Control Points for Reactor Shutdown without Access to the Main Control Room*, ed. 2.0. 2009.

IEC 61772. *Nuclear Power Plants: Control Rooms—Application of Visual Display Units*, ed. 2.0. 2009.

IEC 61850. *Communication Networks and Systems in Substations*. 14 parts issued as International Standard between 2002 and 2004.

IEEE Std 7-4.3.2. S*tandard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating Stations*. 2003.

IEEE Std 7-4.3.2., Annex E. *Diversity Requirements Determination, Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating.* Institute of Electrical and Electronics Engineers, Stations. 1993.

IEEE Std 7-4.3.2., Annex G. *Bibliography, Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating.* Institute of Electrical and Electronics Engineers, Stations. 1993.

IEEE Std 603-1998. *IEEE Standard Criteria for Safety Systems for Nuclear Power Generating Stations*, sections 5.4, 7.1-C. July 1998.

IEEE Std 802.1D. *IEEE Standard for Local and Metropolitan Area Networks—Media Access Control (MAC) Bridges*. 2004

IEEE Std 802.1Q. *Standards for Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks*. 2003.

IEEE Std 802.1w. *IEEE Standard for Local and Metropolitan Area Networks—Common Specification.* Part 3: *Media Access Control (MAC) Bridges.* Amendment 2: *Rapid Reconfiguration.* 2001.

IEEE Std 802.1X. *Standard for Local and Metropolitan Area Networks, Port Based Network Access Control.* 2001.

IEEE Std 802.3x, 802.3y. Supplements to ISO/IEC 8802-3. *1996 Specifications for 802.3 Full Duplex Operation and Physical Layer Specification 100Mb/s*. 1997.

IEEE Std 802.10. *Standards for Interoperable LAN/MAN Security.* 1989.

IEEE Std 802.11. *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. 2007.

IEEE Std 802.15.4. *Wireless Medium access Control (MAC) and Physical Layer (PHY) Specifications for Low-rate Wireless Personal Area Networks (LR-WPANs).* 2003.

IEEE Std 1012-2004. Annex C. *Definition of Independent V&V*, revision of IEEE std 1012-1998.

IEEE Std 1012-2004. Annex F. *Example of V&V Organizational Relationship to Other Project Responsibilities*, revision of IEEE std 1012-1998.

IEEE 1074.1. *Guide for Developing Software Life Cycle Processes*. 1995.

Information Systems Audit and Control Association (ISACA). *Control Objectives for Information Technology (COBIT).* 1998.

ISA-SP100. *Wireless Systems for Automation Standards Committee*. <http://www.isa.org> May 2006.

ISO/IEC 12207. *Software Lifecycle Process*. 1995.

ISO/IEC 15408-2. *Information Technology Security Functional Requirements*.

ISO/IEC 2001:2005. *Information Technology Security Techniques—Information Security Management Systems Requirements.*

ISO/IEC 27000. *Series on Information Security Management*.

Kent, S., R. Atkinson. *Security Architecture for the Internet Protoco*l. IETF RFC 2401. November 1998.

Kent, S., R. Atkinson. *IP Authentication Header*. IETF RFC 2402. November 1998.

Michalski, J. T., et al. *Vulnerability Assessment of the Common Q Digital Safety System to Cyber Threats*. A Letter Report to the U.S. NRC, Sandia National Laboratories. June 2009 (limited release).

National Infrastructure Security Co-ordination Centre (NISCC). *Good Practice Guide on Firewall Deployment for SCADA and Process Control Networks*, February 2005.

National Institute of Standards & Technology (NIST), Melton, Ron, et al. *System Protection Profile: Industrial Control Systems.* <http://www.isd.mel.nist.gov/projects/processcontrol/SPP-ICSv1.0.pdf.>

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1, STOE Security Functional Requirements, version 1.0. April 2004.

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1.11, Intrusion Detection and Response, version 1.0. April 2004.

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1.18, *Secure Communications Channels*, version 1.0. April 2004.

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1.9, Firewall Access Control, version 1.0. April 2004.

National Institute of Standards and Technology (NIST). A*n Introduction to Role Based Access Control*. CSL Bulletin on RBAC. December 1995.

National Institute of Standards and Technology (NIST). *Engineering Principles for Information Technology Security***,** Special Publication 800-2.7, section 2.3. 2001.

National Institute of Standards and Technology (NIST). NIST 800-12. *Introduction to Computer Security*.

National Institute of Standards and Technology (NIST). NIST 800-14, *Generally Accepted System Security Principles.*

National Institute of Standards and Technology (NIST), NIST 800-18. *Guide for Developing Security*.

National Institute of Standards and Technology (NIST). *Guidelines on Firewalls and Firewall Policy*, NIST Special Publication 800-41 5, rev. 1. September 2009.

National Institute of Standards and Technology (NIST). NIST 800-53. *Recommended Security Controls*.

National Institute of Standards and Technology (NIST). *Guide to Intrusion Detection and Prevention Systems*. NIST publication 800-94. February 2007.

National Institute of Standards and Technology (NIST). *Establishing Wireless Robust Security Networks: A Guide to IEEE 802.11i*. NIST Special Publication 800-97.

NUREG/CR-6263. *High Integrity Software for Nuclear Power Plants*. ISO/IEC 12207, Software Lifecycle Process. 1995.

NUREG/CR-6882. *Assessment of Wireless Technologies and Their Application at Nuclear Facilities.* July 2006.

NUREG/CR-6939. *Coexistence Assessment of Industrial Wireless Protocols in the Nuclear Facility Environment.* July 2007.

Rigney, C., et. al. IETF standard 2865, obsoletes 2138. *Remote Authentication Dial-In User Service (RADIUS).* The Internet Society. June 2000.

U.S. Nuclear Regulatory Commission. Regulatory Guide (RG) 1.152. *Criteria for Use of Computers in Safety Systems of Nuclear Power Plants,* rev. 2. January 2006.

Whitehead, D. W., Potter, C.S., O'Connor S.L. *Nuclear Power Plant Security Assessment Technical Manual,* SAND2007-5591. Sandia National Laboratories, Albuquerque, NM. September 2007.

Wyss, G. D., D. Pless, R. Rhea, C. Silva, P. Kaplan, R. Aguilar, and S. Conrad. *Total Risk Assessment Methodology,* SAND2009-0178, Official Use Only document. Sandia National Laboratories, Albuquerque, NM. January 2009.

Wyss, G., P. Sholander, J. Darby, and J. Phelan. "Identifying and Defeating Blended Cyber-Physical Security Threats," *In The Guardian: InfraGard National Members Alliance Quarterly Newsletter.* <http://www.infragardmembers.org/modules/content/index.php?id=38. <http://www.infragardmembers.org/modules/content/index.php?id=38> Issue 5. Spring 2007.

Ylonen, T., et. al. Internet Engineering Task Force (IETF), Network Working Group. *The Secure Shell (SSH) Transport Layer Protocol.* The Internet Society. 2006.

Ylonen, T., et.al. Internet Engineering Task Force (IETF), Network Working Group. *The Secure Shell (SSH) Protocol Architecture*. The Internet Society. 2006.

This page intentionally blank

Appendix B: Site Inspections, Lessons Learned

This appendix contains an accounting of security findings mapped into the best practices element categories from distinct assessment reports. These assessments were performed on various control systems and enterprise information systems connected to control systems over the course of the past eight years. In these reports, there were a total of 103 security findings. Table B-1 below contains the mapping of those findings into the categories presented in section 2, Elements of Secure Networks, of the main report. An individual finding might be mapped into more than one category; therefore, the sum of the number of findings will not equal 103. This table provides the reader with information on the areas within utility companies that are the most prevalent for security infractions.

Table B-1. Best Practice Findings

Best Practice Category	Number
Security Policy—Security policy did not exist or was inadequate.	33
Physical Security—Physical security in general did not exist or was inadequate.	3
Physical Perimeter—Physical security of specific critical components did not exist or was inadequate.	4
System Architecture—System architecture was incomplete, incorrect, or nonexistent for important security-related components.	37
System Architecture Components—Critical components to the security architecture were used incorrectly, lacked appropriate features, or were placed in the wrong location within the architecture.	15
User/Operational Management—Lack of management of, or tools for, configuration management within the system.	42
Host Access Control—Access control to host systems did not exist or was inadequate.	9
Application Configuration & Control—Configuration of applications with security implications was incorrect.	14
Compliance—Compliance with existing security policy was not adequate.	32

The greatest numbers of findings were associated with weaknesses in User/Operational Management applications. For the most part, these involved a lack of procedures to implement configuration controls and methods to determine compliance with policies. System Architecture-related issues were the next largest group of findings, primarily involving the misapplication of failure to implement important protective devices in the network. On the other hand, Physical Security-related matters were not found to be prevalent areas of concern. System lifecycle issues were not addressed as part of these system assessments.

This page intentionally blank

Appendix C: Electricity Generation Vulnerability Observations

The following list of vulnerabilities presented in this section were the output of a brainstorming session among electric utility subject matter experts. This information can be a means to determine some of the general sources of vulnerabilities that can be leveraged against an electricity generation provider and a nuclear power plant in particular. The vulnerabilities can be broken down into three general sources that can be used to the advantage of an adversary:

Type 1: The systems operational environment threatens proper system function.
Type 2: The system is inherently weak against known threats.
Type 3: Security practices are not present or are inadequate.

Associated vulnerabilities that can be reduced by adoption of, or improvement in best practices, are shown in ***bold type***. These are featured in the report and are associated with the following topics: security policy; physical security; system architecture; user management including host access control, application access control, malicious software and compliance checking; and overall system lifecycle. Other aspects of electricity generation that are not addressed by the best practices report, but can expose the business side of electricity generation to exploit, are listed as factual statements.

1. Global economy and market (Type 1)
- 1.1 Lack of control over product source inhibits thorough security.
  - 1.1.1 Rapid pace of technology inhibits *lockdown* of configurations to stabilize feature sets.
  - 1.1.2 High rate of new technology adoption due to economic pressure.
- 1.2 Economic pressure to reduce operating costs can have adverse effect on security.
  - 1.2.1 May lead to *monoculture* of cookie-cutter security solutions.
  - 1.2.2 Non-maintainable product logistics (obsolescence of critical components).
- 1.3 Operator expertise is low.
  - 1.3.1 High turnover due to lack of compensation.
  - ***1.3.2 Inadequate security-based training.***
- 1.4 Deregulation of energy generation.
  - 1.4.1 Centralized generation control with independent system operator.
    - ***1.4.1.1 Additional cyber-secure architecture is required.***

2. The physical structure (Type 2)
- 2.1 Transmission lines and substations can be remote and hard to protect, even if monitored.
- 2.2 Access protection of cyber system physical components is weak or absent.
- 2.3 Physical structure is exposed to environmental stressors.
- 2.4 Transmission lines use metal core conductors and are susceptible to electro-magnetic pulse discharges.
- 2.5 Generation requires large area of protected enclosures.

3. Grid Architecture characteristics (Type 2)
- 3.1 System organization

3.1.1 Power grid is highly interconnected; every part affects every other part.
3.1.2 Power grid is very complex; its behavior is difficult to predict.
3.1.3 Centralized control designs may amplify the effects of an attack.

3.2 System geography
3.2.1 Long distances between sources and loads.
3.2.2 Critical functions are geographically concentrated.

3.3 System relationship to other systems
3.3.1 Interdependence of infrastructures increases number of attack paths.
3.3.1.1 Banking and finance needed to provide payroll and support costs.
3.3.1.2 Transportation required for fuel acquisition and waste management.
3.3.1.3 Maintainable water source for steam generation and cooling.
3.3.1.4 Telecommunications, external network connectivity (the Internet and the long haul communications structure it relies upon).

4. Engineering characteristics (Type 2)

4.1 Operating procedures provide inadequate protection.
4.2 Operating systems (OSs) and energy management software have exploitable bugs.
4.3 Commercial-off-the-shelf (COTS) hardware and OSs
4.3.1 Widespread use of inherently insecure legacy components.
4.3.2 Information about flaws and exploitable elements of information technology (IT) COTS products is openly available.
4.3.3 Information networks generally use IT COTS components and protocols.
4.3.4 Control/safety systems use IT COTS components.
4.3.5 Metal core communication cables radiate emissions that can be intercepted.

5. Operating characteristics (Types 2 & 3)

5.1 High-level design decisions
5.1.1 Grid operation depends on frequency; source, load, and grid frequencies must match.
5.1.2 Status and control communications use IT networks.
5.1.3 Control systems interface with administration systems.
5.1.4 PLCs, RTUs, front end processors, protection relays, etc., utilize firmware/software.
5.1.5 Autonomous components can act without human oversight to execute improper behavior.

5.2 Operational environment
5.2.1 Active defense during an event lags offense (i.e., reactive-based defense).
5.2.2 Pace of cyber operations effectively removes humans from decision loop.
5.2.3 Near real-time information sharing between multiple utility companies.
5.2.4 Reliance on external timing sources for accurate event logging (e.g., network time protocol and global positioning system).
5.2.5 Grid is often operated at near-maximum capacity.
5.2.6 Demand can be large relative to capacity.

5.2.7 Short demand cycles of power purchasing and transmission.

5.3 Inappropriate use of system resources.

5.3.1 Control system hosts used for non-control applications or devices (e.g., peer-to-peer messaging applications, e-mail, Voice-over Internet Protocol, Web browsers and servers).

5.3.2 Control system network used inappropriately (i.e., increased traffic load due to non-control system applications).

6. Control/Safety system cyber defense (Types 2 & 3)

6.1 Information about operations and grid status is readily available to adversary.

6.2 Awareness of cyber situation is poor.

6.2.1 Anomalous and inappropriate activity is not uniformly detected and reported.

6.2.2 Cyber system logs are not used to detect intrusions and operational anomalies.

6.3 Exploitable network cyber paths

6.3.1 Trust relationships between control system hosts and administration hosts.

6.3.2 System administration mechanisms inadequately scrutinized or maintained.

6.3.3 Non-dedicated channels used for command and control (sharing of different data types on a single local area network, no quality-of-service enforcement).

6.3.4 Wireless communication not adequately secured.

6.3.5 Modem communication not adequately secured.

6.3.6 Wireless communication not adequately secured.

6.3.7 Internal network communication not adequately secured.

6.3.8 Firewalls/Intrusion Detection System/Intrusion Prevention System not properly configured.

6.4 Access control issues

6.4.1 Weak passwords

6.4.2 Improper change management controls

6.4.3 No role-based access restrictions

6.4.4 Areas exposed by access control issues.

6.4.4.1 Maintenance and upgrade access

6.4.4.2 Host software/firmware images

6.4.4.3 Host operational settings

6.4.4.4 Remote access to control system functions

6.4.4.5 Sharing of removable media (compact discs, digital video discs, thumb drives, etc.).

6.4.5 Authentication is improper, inadequate, or absent.

6.4.5.1 Unauthenticated protocols for control system firmware/software upgrades.

6.4.5.2 Unauthenticated protocols (e.g., SNMP, Telnet, Trivial File Transfer Protocol and File Transfer Protocol) allow unauthorized access to control system components.

6.4.5.3 Unauthenticated control system command and control data.

This page intentionally blank

Appendix D: Additional Network Security Discussions

The topics presented in Appendix D are intended to provide stand-alone, detailed discussions on the various elements that comprise a secure network. Included are matters concerning policy, physical protection, system architecture (including the variety of equipment and software that make up a network), and user/operational management. Each topical section presents basic background and introductory information and then discusses the security-related observations that attempt to point out the specific vulnerabilities and threats associated with that particular network device. Threats from external, insiders, and vendor-based adversaries are discussed. Mitigation techniques are also recommended. Additional sources of information and detail are provided where applicable.

D.1 Policy Framework Details

This section defines *policy* and the concept of a *policy framework*. This section then introduces how one particular policy framework, namely control objectives for IT and related technology (CobiT) can be used to developed a detailed security policy for an organization [D.1-1]. There is multiple policy framework software available for use and the discussion of CobiT is for the purpose of providing an example.

A *policy* describes what is to be done. It can also describe the following:

- How it is to be done?
- When it is to be done?
- Who is to do it?
- What could be done instead?
- What would happen if it were not done?
- Why it is necessary to be done at all?

In general, written policy is preferred over common understanding, but the former cannot spell out everything and, at some point, must rely upon the latter. There are two fundamental problems with policy as it relates to cyber security:

- How much detail should each policy item cover?
- How much ground should the policy set cover?

The solution to the first problem has to be specific to each organization. The solution to the second problems is the goal of a policy framework. Each organization should adopt just what it considers that it needs. The value of a policy framework is that it provides a current estimation of what is sufficient for the list. The policy framework should provide breadth and depth of detail to accurately capture all the salient elements of the policy. As an example of detail, the remainder of this section briefly describes the CobiT methodology.

CobiT is hierarchically organized, using four levels, as shown in Table D-1 below.

Table D-1. CobiT Structure

Level	Name	Number of Items at this Level
1	Domain	4
2	Process	34
3	Control Objective	~250
4	Control Practice Statement	~1600

The hierarchical structure supports the proposition of sufficient breadth and depth of policy detail. With this structure in hand, policy developers can consider if there is a fifth domain, for example. Policy developers can consider if there is an additional process for a given domain, etc.

The names of the four domains are listed in bold, followed by their two-letter abbreviations. Process names (for some of the four to 13 processes for each domain) are listed below each domain name.

Domain—Plan and Organize (PO)
- PO1 Define a Strategic Plan
- PO2 Define the Information Architecture
- …
- PO10 Manage Projects

Domain—Acquire and Implement (AI)
- AI1 Identify Automated Solutions
- AI2 Acquire and Maintain Application Software
- …
- AI7 Install and Accredit Solutions and Changes

Domain—Deliver and Support (DS)
- DS1 Define and Manage Service Levels
- DS2 Manage Third-Party Services
- …
- DS13 Manage Operations

Domain—Monitor and Evaluate (ME)
- ME1 Monitor and Evaluate information technology (IT) Performance
- ME2 Monitor and Evaluate Internal Control
- …
- ME4 Provide IT Governance

A policy framework seeks to enable an organization to consider what it needs. The intention is not that everything must be implemented today or ever; the intention is that a policy framework provides the extent of what should be considered. Another aspect of a security policy is the need to identify and, at some level, quantify risk. Realistically, a security policy is not going to

eliminate all risks to the operation of the network and its critical assets, but it should help identify residual risk. Therefore, security policy allows a level of risk acceptance, which becomes part of the overall cohesive security policy, as described in ISO 27001:2005 [D.1-2].

CobiT has been used as an example. There is extensive guidance elsewhere, as well. The Information Technology Governance Institute (ITGI) provides *mapping* documents that describe how CobiT relates to other security policy guidance, such as ISO/IEC 27001 [D.1-2], NIST 800-14 [D.1-3], and ISO/IEC 15408-2 [D.1-4]. RG 5.71 Cyber Security Programs for Nuclear Facilities, section C.2, Elements of a Cyber Security Plan, provides some additional insights of the elements required to creating a cyber security plan to implement an organizations security policy.

Some standards and guides pertaining to security policy development are listed below:

ISO/IEC 27001:2005. *Information Technology Security Techniques Information Security Management Systems Requirements* [D.1-2] specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining, and improving a documented Information Security Management System within the context of the organization's overall business risks.

ISO/IEC 15408-2:2008 provides information on the functional requirements [D.1-4] needed to be assessed in an organization security review, as when constructing a security policy. It catalogues security functional components that will meet most common security needs of an organization. It also provides guidance on the specifics of customizing security requirements to meet unique needs that the organization identifies.

NIST 800-14. *Generally Accepted Principles and Practices for Securing Information Technology Systems* [D.1-3] section 3.1 Policy, describes some needs for a policy and policy ability to define program goals including those directed at facilities, hardware, software, information, and personnel. It also mentions the need for policy to set an organization’s strategic direction, assess responsibilities ,and address compliance issues.

NIST SP 800-53, *Recommended Security Controls for Federal Information Systems*, is a guide for assessing the security controls in federal information systems and a guide for organizations in building effective security assessment plans.

References:

D.1-1. Information Systems Audit and Control Association (ISACA). *Control Objectives for Information Technology (CobiT).* 1998.

D.1-2. ISO 27001:2005. Information technology: Security Techniques—Information Security Management Systems, Requirements.

D.1-3. NIST 800-14. Generally Accepted System Security Principles.

D.1-4. ISO/IEC 15408-2. Information Technology Security Functional Requirements.

D.2 Physical Security Details

This technical report focuses on network security issues, especially electronic security, for Safety and Control Systems. However, physical security must not be ignored. Physical security is always the principal defense in preventing unauthorized access, corruption of informational assets, and intentional or unintentional destruction of property. Many documented cyber attacks against organizations have been initiated by having physical access to elements of the network architecture in order to execute the penetration.

Physical security is commonly referred to as "guns, gates, and guards." This simplistic label fails to capture the increasingly complex and interdependent relationship that physical security has with cyber security. While the focus of this technical report is network security issues—especially electronic security for Safety and Control Systems—that security is partly dependent on physical security.

In prior decades utilities enjoyed nearly complete segregation between control networks and business/administrative networks, in terms of connectivity and communication protocols. Furthermore, physical security at utilities was designed to address assets of the physical world (i.e., people, facilities, fuel, and plant equipment).

Today's evolving utility network environment threatens system safety and security that isolation once helped provide. Now physical security and cyber security have significant overlap and interdependencies. Physical security contributes to the effectiveness or ineffectiveness of cyber security, and vice versa. It must be remembered that threats to computer systems and networks are not just originated at remote, far-away places. Threats against critical cyber assets can be made by attackers (outsiders or insiders) gaining physical access to systems or network equipment. In fact, many cyber attacks have been realized because physical access to elements of the network architecture was available. Having sufficient physical security for plant critical cyber assets is a needed element in decreasing the risk of adversary compromise.[11]

D.2.1 Essential Strategy for Effective Physical Security at the Plant

Most adversary objectives can be distilled into one of two categories: theft or sabotage (regardless of whether the target is information, electronic data, or physical assets). The goal of protection systems is to prevent either from occurring. For any critical security environment, an integrated physical protection system is required to prevent unauthorized access to critical assets, including those cyber systems that contribute to successful operations, regardless of which network they may reside in (if any at all).

---

[11] This section addresses security and not safety. Some facilities might integrate the security function with safety functions including emergency response for fire, medical, and environmental situations. Such integration would include additional sensors, systems, personnel, policies, procedures, etc., not discussed below. Readers requiring a comprehensive understanding of physical security strategy and techniques for nuclear power plants should read *Nuclear Power Plant Security Assessment Technical Manual*, SAND2007-5591, Sandia National Laboratories, Albuquerque, NM, September 2007, a revision to NUREG/CR-1345.

A balanced physical protection system (PPS) is one that has an adequate level of effectiveness against defined physical threats along all possible physical pathways and one that maintains balance with other considerations including cost, safety, and structural integrity. In addition to adequate technology, a balanced PPS includes policies, personnel, procedures, training, testing, and maintenance. In general, a PPS with well designed, multiple layers and redundancies increases the PPS performance capabilities. An additional consideration is the need to eliminate single points of failure. Finally, note that, like cyber security, physical security is not a product but a process; the lack of sufficient attention to the various elements of a PPS—whether technology, procedures, or training—will result in sub-optimal effectiveness.

The basic elements required for an effective PPS are detection, delay, and response. Deterrence (having good lighting, warning signs, security guards in visible locations, etc.) is difficult to measure; thus, it is not considered a key element of PPS design and operation for high-consequence security applications as there is a lack of sufficient statistical evidence confirming its effectiveness. Deterrence is a commonly used approach for reducing the insider threat. A final piece to the security puzzle that is frequently overlooked is systems engineering and integration.

D.2.1.1 Detection

Detection is the first step required for successful execution of the security function. The purpose of detection is to alert security personnel to the presence of unauthorized personnel and violations of established security perimeters. Positive detection must result in response by an official guard-force, possibly including local law enforcement. Detection involves two key elements: 1) sensing a boundary violation, and 2) correct human assessment of the sensed violation. Another component of the detection function is the requirement that a facility must permit authorized access (to employees, contractors, visitors, etc.). Reliability is a key measure of performance for the detection function; measurement should be made across people, processes, and technologies. Figure D-1 depicts the sequence of detection and reporting events.

Figure D-1. Summary of the Detection Function for Physical Security

Sensors (Exterior and Interior)

Employment of multiple sensor technologies utilizing differing phenomenologies (e.g., balanced magnetic door switches along with monostatic microwave motion detectors) is recommended. These multiple sensor technologies minimize the consequences of failure in one set of sensors due to adversary capability or adverse environmental/weather conditions that degrade sensor performance. Multiple sensors can also be deployed in an architecture that increases the size of the overall sensing field, making it more difficult for an adversary to evade. Different sensor types have advantages and disadvantages; it is essential to understand these when choosing what sensors to deploy or when a site is already dependent on them. When designing and placing sensors, the recommended strategy is to move them as far as possible from the critical assets and

as close as possible to the defined perimeter. Performance objectives for sensors include the probability of detection, nuisance alarm rate, and vulnerability to defeat. Sensors must be tested per established procedures at a documented frequency. This is the only way that non-functional sensors can be identified and repaired or replaced in a timely manner.

Assessment
Detection also includes successful and timely communication of alarms to assessment personnel and response forces. Once a sensor has tripped due to a possible unauthorized access attempt, an electronic alarm must be transmitted to assessment personnel. Assessment is the process of verifying the veracity of the received alarm. This process can be done in a fixed location by local or central alarm monitoring personnel, by mobile guard-force personnel, or by a combination of both. It is common for a PPS to integrate security cameras, lighting, and digital video recorders to support personnel responsible for this function. Each technology contributing to this function must be properly evaluated for capability, reliability, vulnerability to defeat, and ease of use.

Entry/Exit (Access) Control
Facility access control addresses the operational requirements of a facility. Access control consists of policies, procedures, and systems used to verify entry authorization and support contraband detection (for both entry and exit control). Access control systems must be integrated into the detection function of the PPS. Methods and technologies supporting personnel entry authorization include encoded badges, smartcards, personal identification numbers (PIN), biometric identification, and voice recognition. Methods and technologies supporting contraband detection include manual search and detectors capable of sensing for metal, explosives, and nuclear material. The technical performance of associated vendor products varies widely; thus, it is essential that the chosen equipment is thoroughly evaluated before selection and placement. In general, access control systems using multiple, complimentary techniques and technologies should reduce the probability of false alarms while increasing the probability of detection. Key measures of performance for access control components include throughput, false accept/positive rate (allowing facility access to unauthorized entities or material), and false rejection/negative (denying access to authorized entities/permitted material).

D.2.1.2 Delay
Physical security requires the placement of engineered delay features, particularly for facilities containing critical assets (people, materials, systems, etc.). The need is to increase adversary task time, thereby enabling the guard-force to respond in time to prevent a loss of assets through theft or sabotage. Fences, gates, controlled entry access points, activated delays, locks, reinforced doors and walls, anti-tampers and other barriers are examples of delays. A guiding principle in the placement of delays is to maximize delay as close as possible to critical assets. Delays should be chosen that are appropriate to the loss they are trying to prevent and according to the threat model/security scenarios adopted by the site.

Engineered delays have differing performance characteristics, so it is crucial for the security engineer to understand the impact of choosing one vendor product over another. The key measure of performance for any delay is the time required to defeat the obstacle by the adversary (i.e., the

increase in adversary task time) after detection and the corresponding time increase the guard force has to effectively respond. Figure D-2 depicts the delay function.

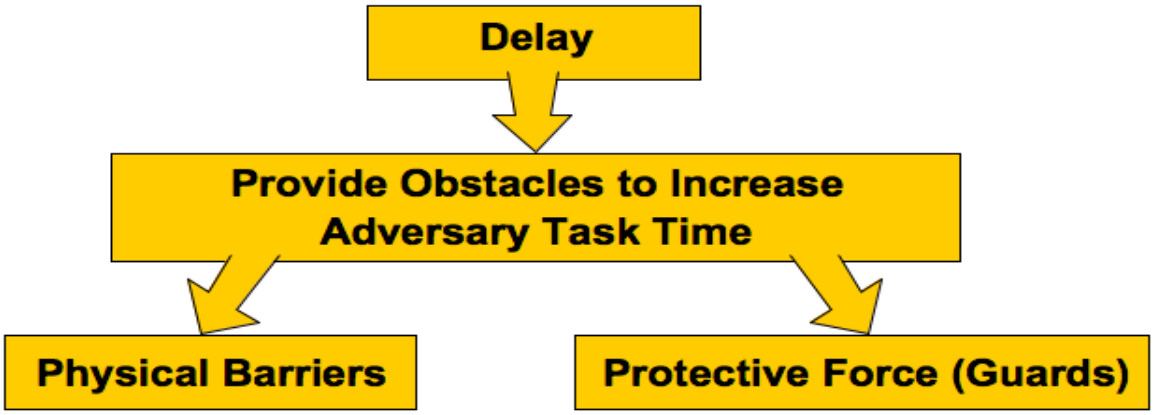

Figure D-2. Summary of the Delay Function for Physical Security

D.2.1.3 Response

Actions taken by a response force (e.g., guards, armed guards, and police) to prevent adversary success comprise the response function. Example strategies for response include interruption, containment, and neutralization of the adversary to prevent loss or sabotage of critical assets, or to recover critical assets. Key considerations include robust communications capabilities (including redundant methods and systems of communication), dissemination of accurate information, time required by the response force to deploy, numbers of responders, and capabilities/professionalism of the responders (this includes training, tactics, and procedures as well as equipment). Other key considerations are how and how well the response function is integrated with local law enforcement for adversary intelligence sharing and for supplemental, on-site response. Figure D-3 shows a response function.

Figure D-3. Summary of the Response Function for Physical Security

Performance measures for the response function include probability of communication from detection personnel to responders, the time needed for effective communication, the probability that forces are deployed to the adversary's position, the time needed by responders to deploy, and the effectiveness of the response force.

D.2.1.4 Systems Engineering and Integration

Many PPS have been equipped and outfitted with modern technology, have grown drastically over time, and lack end-to-end lifecycle-based engineering. Key engineering issues in PPS include system-of-systems design and analysis, PPS security (discussed below), performance-

based measurement and balanced security (discussed above in Appendix D, section D.2.1, Essential Strategy for Effective Physical Security at the Plant).

Vendors of security products for the PPS may specialize in one or more components and applications. But it is critical for facility stakeholders to identify experienced engineers able to provide an independent, objective design for a new PPS or existing PPS upgrades. The integration function must ensure, through careful configuration, implementation, and testing that selected technologies achieve the required performance levels and contribute effectively to a robust PPS.

Security of the PPS Network

The PPS network is a control systems network; thus, much of the network security guidance for control systems contained in this document is applicable to the PPS network. The security of the PPS in part depends upon the security of the computers, devices, and networks that host the PPS. Market demand for security components that are network-addressable have resulted in PPS components—such as cameras, and sensors with TCP/IP-based communication capabilities. Nearly every new device can have an IP address. Many devices, such as cameras, support remote configuration. Because of the critical security functions that the PPS network enables, it requires a secure configuration.

PPS network infrastructure, as with control systems networks, should be configured to protect devices and computers from malicious network traffic, while the PPS network must be protected from rogue devices. Default configurations should be eliminated where possible and replaced with secure configurations. Any wireless access points (WAPs) used in the PPS should be secured to prevent unauthorized devices and computers from secretly attaching to the network. All platforms used in the PPS should be configured with the minimum required functionality; unused services and ports should be disabled. Strong passwords should be applied at the host, operating system (OS), and application levels where possible. Remote management should only be used where essential and with secure, encrypted protocols. Use of encrypted channels and/or authentication to support data collection (e.g., device logs, video streams, etc.) and control commands are recommended. Systems in the PPS network should be properly segregated from the administrative network to further safeguard critical PPS functions from inadvertent or poorly designed connectivity. This is not to say that systems in the PPS network and systems in the administrative network cannot exchange information; rather the connectivity must be securely devised. Finally, new systems and applications should be evaluated in test networks before integration into the operational PPS network.

D.2.2 Physical Security for Critical Cyber Assets

Cyber systems personnel are not generally accustomed to discussing threats to their computers and information in terms of theft or sabotage. Common terms that cyber systems personnel use to evaluate threat focus on attributes of data, chiefly, the following: *availability*, *integrity*, and *authenticity* (the physical security term *theft* relates to *availability*, and the term *sabotage* applies to all three). A successful attack on any one of these attributes could mean loss of control in a control systems environment.

Cyber assets (computers, switches, WAPs, modems, Voice over Internet Protocol (VoIP) systems, email, Websites, Universal Serial Bus [USB]/thumb drives, etc.) are ubiquitous and represent a reliable, if not constant, opportunity for adversary compromise, and their compromise could have significant adverse consequences for the stakeholders of distributed control systems (DCSs). Therefore, rigorous efforts to secure cyber assets should be pursued. Physical security controls (primarily supporting the delay function) can be incorporated into the cyber asset protection scheme. Security design and upgrades should also be measured against other operational requirements to identify potential conflicts and ensure the optimal solution is reached.

Locks

Locks are a primary means of providing physical security for cyber assets. Locks for any application offer varying degrees of performance depending on design, materials, and workmanship. Different kinds of locks have distinct advantages and disadvantages and what is suitable for one application may not be for another. Locks do not guarantee security; they may deter certain adversaries and only add delay for other adversaries.

Room Locks—Typical room locks include cipher, key, keyless, and card or token-based mechanisms. Cipher locks and some keyless programmable locks require knowledge of a combination. Depending on how many personnel have a need-to-know who has the combination can become a difficult problem; also organizational requirements must include changing the combination upon personnel loss of need-to-know (due to a change in role or termination of employment, etc.). Keyless, programmable locks offer the advantage of having no physical keys that can be lost or copied; cores do not need to be changed or re-keyed when a key is lost. Some programmable, digital access-control locks can even be programmed with a duress alert code. Proximity-based, token locks have unique identifiers that make it easy to customize access to multiple rooms for any authorized individual. If the tokens are lost they can be removed from the system and replaced to prevent unauthorized access. Card-based access is similar to that of token-based access. For high-security applications, it is recommended that access to rooms containing critical cyber assets follow a multifactor authentication scheme, such as a card-based scheme with a pass code (or PIN). Electromagnetic locks are preferred for high-security applications as they rely on the bond strength of powerful magnets. Simple magnetic locks have fail-safe performance (meaning the locks unlock) when power is lost. Therefore, it is essential to determine the need for fail-secure units and identify contingencies for emergency exits for staff in power-lost conditions.

Logging/auditing should be enabled for all access attempts/visits to secure rooms whether with a paper log or with an electronic log associated with electronic access control.

Equipment Locks—Mechanical locks are available for most cyber assets including IT equipment racks, server bays, and slot locks for individual computers. Slot locks typically are cable locks that are easily defeated. Secure physical enclosures can be used to house individual workstations.

Other Locks—Padlocks can be used to secure equipment containers. Certain critical applications could be designed to support the use of a physical key to enable or lock-out control of a particular workstation, or require multiple keys to enable two-person control.

Doors, Walls, Ceilings, and Floors
To further secure critical assets, hardening features can be incorporated into doors, walls, ceilings, and floors. Doors should be self-closing and have no hold-open feature. Doors with windows (and locks) can be used for rooms containing cyber assets, allowing people outside of the room some view. Walls can be constructed with windows to allow people outside of the room a view of operations and equipment within the room. Hardened glass can be used to add delay.

Monitoring Systems
Recording systems (video surveillance) can be used to monitor the actions of personnel in a critical environment. While this will not provide protection against imminent theft and sabotage scenarios during an attack, it could provide some level of deterrence for less serious scenarios (employee misuse of equipment and resources, etc.).

Sensors and Alarms
Use of detection-oriented sensors and alarms close to critical assets provides little to no benefit to the effectiveness of the PPS (recall the principle that sensors for detection should be placed as close to the perimeter as possible). However, use of sensors and alarms (especially audible or visual alarms), even on sensitive equipment housings, can deter certain adversaries (insiders) and/or cause an attacker to alter their plan, lose their ability to think clearly, or otherwise interfere with the attacker's execution. However, this is not measurable and less reliable than adding more delay. Another benefit of audible and visual alarms is to assist responders to establish the attacker pathway. Unless these sensors and alarms are verified with a positive assessment (e.g., confirmation using video surveillance), they cannot be relied upon.

Anti-Tampers
Anti-tamper mechanisms can be used to support deterrence, detection, and delay. Anti-tamper tape with unique identification numbers can be applied to computer chassis, racks, and other equipment that, when opened, make it difficult for an adversary to replace or repair. For detection, this only works if the anti-tamper tapes are regularly checked for compromise. Anti-tamper tape does not add delay, but could further deter insiders.

Anti-tamper screws, bolts, and other hardware can be used to secure mechanical housings. Use of such mechanisms can add delay, but can also add to the time required for maintenance of the system being protected.

D.2.3 Security Observations

Any site security can be weakened (accidentally or purposefully) or hardened by employees, contractors, and visitors. Many times the maintenance and upgrades associated with security controls are the responsibility of the product vendor, who can provide another avenue of exploit. Effective security requires teamwork between different stakeholders, from executive management and human resources to IT personnel and control systems operators. Effective

security also requires the integration of multiple program elements: physical, technical, and administrative.

Administrative elements that contribute to effective security include the following:

- Strong policies for information protection and system use.
- Technical standards establishing performance criteria for security controls.
- Documented procedures to ensure configurations and implementations meet applicable standards and policy requirements.
- Regularly scheduled security awareness training and briefings (cyber, physical, insider threat, etc.) to foster a strong security culture.
- Technical training to ensure proper execution of duties and resource use.
- Personnel screening (background checks, drug testing for certain occupations).
- User and administrator account registration (assists with deactivation of computing privileges upon termination or suspense).
- Separation of duty and/or two-person control for critical functions.
- Non-retaliatory reporting environment (encourages employee cooperation).
- Performance of risk, vulnerability, and other security assessments.

Some standards guides for physical security are listed below:

- NIST SP 800-14. *Generally Accepted Principles and Practices for Securing Information Technology Systems*, section 3.10, Physical and Environmental Security, contains information about the importance of physical access control of personnel, equipment, and media from buildings and the need to physically protect elements of the network from compromise including cyber assets, such as computer servers. This protection should include all elements required for the systems operation. This can encompass many of the approaches previously discussed in Appendix D, section D.2, Physical Security Details.

- CIP-006-1**.** *Cyber Security-Physical Security of Cyber Assets*, describes the required characteristics of a physical security implementation approach to protect cyber assets of power utility companies. Its requirements are defined in sections R1 through R6, which include physical access controls, such as card key, special locks, biometrics, access monitoring (e.g., alarm systems), and human observation, logging (e.g., computerized logging, video recording and manual logging), and testing of the physical security system. These elements are required to meet the North American Electric Reliability Corporation (NERC) critical infrastructure protection (CIP) requirements for proper physical security protection of electric power companies.

- ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 10 Physical Security Controls, describes the classes of physical security devices and typical deployments.

Additional information about physical security methods and techniques is available from the following resources:

Garcia, Mary Lynn. *The Design and Evaluation of Physical Protection Systems,* 2d ed. Butterworth-Heinemann, Burlington, MA. 2007.

Garcia, Mary Lynn. *Vulnerability Assessment of Physical Protection Systems*. Butterworth-Heinemann, Burlington, MA. 2006.

Wyss, G.D., D. Pless, R. Rhea, C. Silva, P. Kaplan, R. Aguilar, and S. Conrad. *Total Risk Assessment Methodology,* SAND2009-0178. Official Use Only document. Sandia National Laboratories, Albuquerque, NM. January 2009.

Whitehead, D. W., Potter, C.S., O'Connor S.L. *Nuclear Power Plant Security Assessment Technical Manual,* SAND2007-5591. Sandia National Laboratories, Albuquerque, NM. September 2007.

Wyss, G.D., P. Sholander, J. Darby, and J. Phelan. "Identifying and Defeating Blended Cyber-Physical Security Threats." In *The Guardian: InfraGard National Members Alliance Quarterly Newsletter*. <http://www.infragardmembers.org/modules/content/index.php?id=38> <http://www.infragardmembers.org/modules/content/index.php?id=38> Issue 5. Spring 2007.

Clauset, A., M. Young, and K.S. Gleditsch. "On the Frequency of Severe Terrorist Events." In *J Conflict Resolution* 51(1):58-88 (2007).

Anderson, Robert H., Thomas Bozek, Tom Longstaff, Wayne Meitzler, Michael Skroch, and Ken Van Wyk. Conference Proceedings: *Research on Mitigating the Insider Threat to Information Systems–#2*. CF-163-DARPA, RAND National Defense Research Institute, The RAND Corporation, Santa Monica, CA. August 2000.

Huber, Lt. Col. Arthur F. II, and Jennifer M. Scott. *"The Role and Nature of Anti-tamper Techniques in U.S. Defense Acquisition."* In Acquisition Review Quarterly. Fall 1999.

Defense-in-Depth, a Principal of Design

The ideal of defense-in-depth when associated with protecting a network asset simply means having a defensive strategy that includes multiple layers of different security methods. If one layer of the defense is breached, then another layer can be used to protect the asset. This is a modern approach to network security architectures.

References D.2-14 through D.2-8, below, provide some examples of defense-in-depth strategies and guidance for identifying appropriate security controls that can be implemented to protect network assets.

References:

D.2-1 NIST Special Publication 800-14. *Generally Accepted Principles and Practices for Securing Information Technology Systems,* section 3.10, Physical and Environmental Security. September 1996.

D.2-2 CIP-006-1. Cyber Security-Physical Security of Cyber Assets. May 2008.

D.2-3 ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems,* section 10, Physical Security Controls. October 2007.

D.2-4 Department of Homeland Security. *Control Systems Cyber Security: Defense in Depth Strategies*, external report INL/EXT-06-11478. May 2006.

D.2-5 *Engineering Principles for Information Technology Security*, Special Publication 800-2.7, section 2.3, National Institute of Standards and Technology. 2001.

D.2-6 ISO/IEC 27000 Series on Information Security Management System. 2005.

D.2-7 NIST 800-18. *Guide for Developing Security*, February 2006.

D.2-8 Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.4, Defense in Depth. October 2005.

Sections Description

The following sections, D.3 through D.12, describe and discuss system components that, when properly configured and utilized can provide a *defense-in-depth* deployment strategy in a nuclear power plant data network (NPPDN) system architecture. Following the description of each network component a *security observation* discussion is provided. This discussion is intended to point out the weaknesses and vulnerabilities associated with the particular network component or protocol that an adversary may exploit. Mitigation techniques that can be used to improve the security of the component are identified and discussed. These sections provide a framework for ensuring the security of the overall nuclear power plant (NPP) network.

D.3 Virtual Private Networks

A virtual private network (VPN) is a private network that operates between two participating nodes and can use a public network infrastructure. It maintains privacy by encrypting the data between the nodes. Prior to VPNs, a utility company that wanted to keep its data transfers private had to build and maintain a private or *leased* line network to exchange data between external nodes.

Virtual Private Networks in the Plant Data Network Environment

Referring to Figure 2-1, Digital Plant System Network Architecture, a VPN could be constructed from the firewall of the business information network to an external enterprise location across an *un-trusted* network, such as the Internet. It would also be appropriate to construct VPNs in different locations of the plant, or even at a distant regulatory partner location, to allow accesses to a shared information server located at the local plant.

Also referring to Figure 2-1, it would be appropriate to create a VPN from one utility company to another to share electricity production and capacity information in order to coordinate the production of electricity on the electric grid.

Other uses could include VPN implementation between any network device or management station to improve the security of internal *trusted* networks and to restrict access to control system hosts and to the controllers in order to improve data security.

D.3.1 Network Layer Virtual Private Network

One popular VPN is constructed by the use of the Internet Protocol Security (IPSec). IPSec is a standard developed by the Internet Engineering Task Force (IETF) to provide secure communications over public Internet Protocol (IP) networks (Internet). At the network level, IPSec supports peer authentication, data origin authentication, data confidentiality, data integrity, and replay protection. IPSec is normally used with Internet Key Exchange (IKE) for key management. IPSec supports most modern encryption algorithms, such as Advanced Encryption Standard (AES), Data Encryption Standard, its more secure Triple Data Encryption Standard (3DES) version, and Rivest cipher. It also supports integrity mechanisms that use popular integrity HASH algorithms, such as message digest, and secure HASH algorithm, and authentication using X.509 certificates. IPSec can be implemented in a host-to-host fashion or a gateway-to-gateway implementation.

IPSec can be used in one of two different modes: Authentication Header (AH) or Encapsulating Security Payload (ESP). These modes are called transport and tunnel mode, respectively. In tunnel mode, the IP datagram is fully encapsulated by a new IP datagram using the IPSec. Tunnel mode provides for *authentication, confidentiality*, and *integrity* of the data stream. In transport mode, only the payload of the IP datagram is handled by the IPSec, inserting the IPSec header between the IP header and the upper-layer protocol header. Transport mode provides only authentication and integrity of the data, not confidentiality.

The IKE protocol has two negotiation phases: phase1 and phase 2. Phase 1 initiates negotiation between two participating gateways. The gateways set up a two-way Internet Security and Key Management Protocol (ISAKMP) security association (SA), which they can then use to handle phase 2 negotiations. One such SA between a pair of gateways can handle negotiations for multiple tunnels. Using the ISAKMP, the gateways negotiate IPSec (ESP and/or AH) SAs as required. IPSec SAs are unidirectional (a different key is used in each direction) and are always negotiated in pairs to handle two-way traffic. There may be more than one pair defined between two gateways.

Both phases use the User Datagram Protocol (UDP) and port 500 for their negotiations. After both IKE phases are complete, IPSec SAs are then instructed to carry the encrypted data which use the ESP or AH protocol.

For a more detailed description of IPSec and IKE, refer to the following documents:

S.Kent, R.Atkinson. *Security Architecture for the Internet Protocol.* IETF RFC 2401. November 1998.

S.Kent, R.Atkinson. *IP Authentication Header.* IETF RFC 2402. November 1998.

Modes of Operation

IPSec tunnel mode gateway configurations are required to support tunnel mode connections. In this mode the gateways provide tunnels for use by client machines behind the gateways. The client machines have no need to provide IPSec processing; all they have to do is route data to gateways. IPSec tunnel mode is popular in site-to-site VPN implementations because it can be realized in a network device, such as a gateway router, without modifying any client or server applications. Figure D-4 shows a typical tunnel mode deployment, which is being used to provide private data exchange between participating utility control centers. The application protocol being used is called Inter-Control Center Communications Protocol (ICCP). This protocol is commonly used in utility communications to share inter-utility data between connected systems of the utility industry.

IPSec transport mode can also be implemented between two chosen hosts—for example, between an ICCP client host and an ICCP server. Each end host must support IPSec and be able to negotiate an authenticated link between host machines (as opposed to security gateways). IPSec is implemented at Layer 3 of the open system interconnect (OSI) network stack to encapsulate IP packets. After a VPN tunnel has been established per tunnel mode, application data, such as ICCP, can be encapsulated and sent through the tunnel. The above example used communications between utility companies to coordinate the delivery of power, but the endpoints could have been a regulatory node with a distant interface to a utility organization to query and review required plant data information.

The primary purpose of an IPSec gateway is to decide which flows are to be protected between two distant end points. Profiles are created to provide the ability to isolate communication between hosts, such as trusted servers, and any pre-determined end devices. Thus, regardless of the means of communicating—private wide area network (WAN) or public Internet—the remote egress gateway must use IPSec to negotiate trust and to secure IP traffic end-to-end with the destination computer located behind the corresponding ingress gateway.

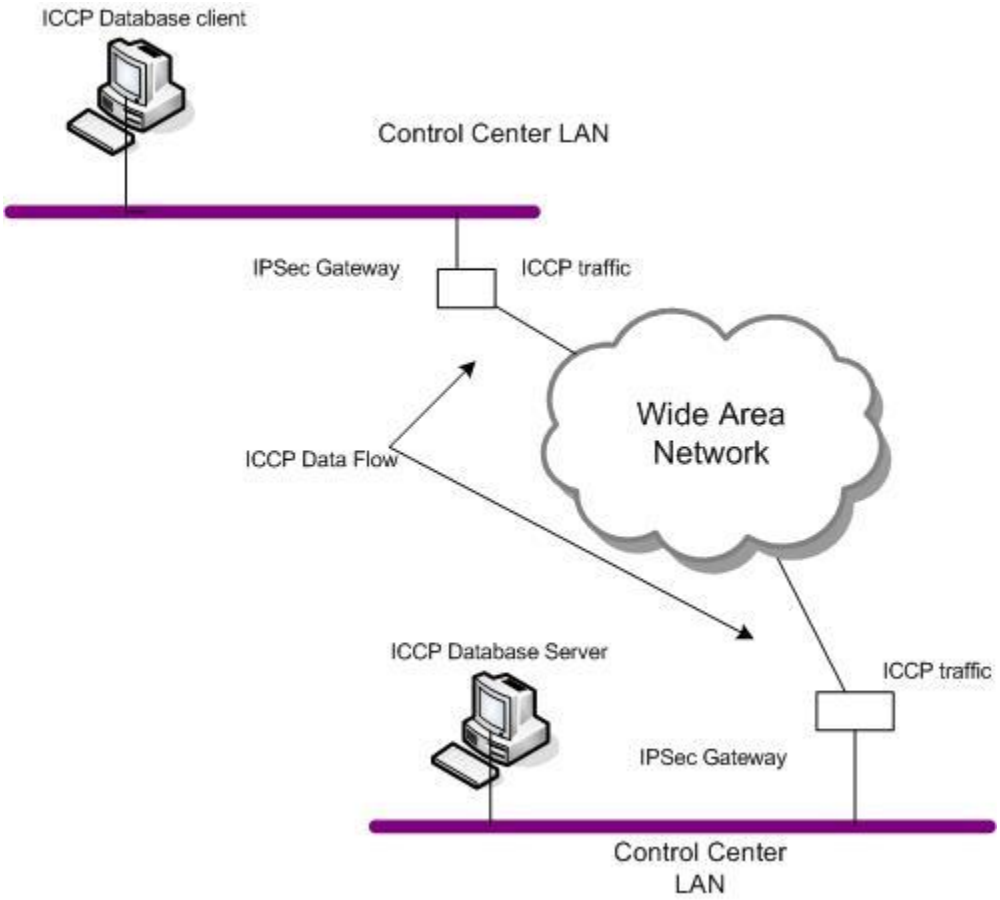

Figure D-4. IP Sec VPN Tunnel Mode Example

With respect to ICCP, there are two ways to approach this profile configuration. The first is to use a less granular configuration that provides IPSec encryption *for all* communications between identified end hosts and that does not require port filtering. The second way is to use fine granularity in the form of a port filter, which can identify a specific application like ICCP and provide IPSec encryption *for only* that application.

Single Host Isolation, No Port Filtering

In the case of an ICCP server and a distant host or hosts, each connection will be identified and authenticated by its IP address. This provides a bulk approach to data confidentiality by encrypting all communications between end points regardless of whether a higher layer of encryption is being applied, such as with the use of an application layer VPN using Secure Sockets Layer (SSL). This double encryption can provide an additional layer of protection by obfuscating the original IP addresses of the end host participating in the communications; but this may cause additional processing burdens and delays associated with data transmission.

Single Host Isolation with Port Filtering

Another approach that can be pursued to isolate data flows—in the scenario of transporting ICCP data streams originating and terminating at the same server—is by using port filtering for flow identification. As part of the IPSec configuration profile, an access control list is created that

identifies each host allowed into the protected domain to communicate with a particular host. In the case of ICCP, this could be the ICCP server. To identify the type of communications taking place between the two endpoints, an additional filter can be enabled that allows the gateway to peer into the transport layer header and identify the port being addressed by the client/server session. If the port address is determined to be ICCP, it is then pushed through the tunnel for data encryption. This allows for a granular approach to data encryption. Figure D-5 shows this inspection process.

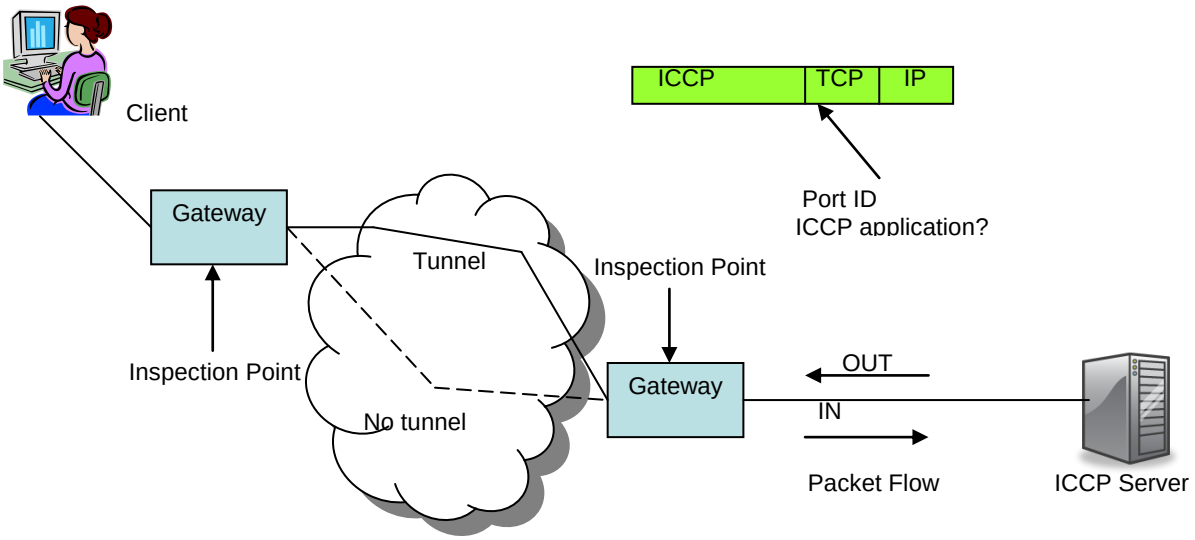

Figure D-5. IPSec Port Filtering Implementation

The purpose of IPSec domain isolation is to mitigate the risk posed to trusted resources. Implementation technologies, such as gateways for filtering and authentication, can help protect utility data assets. The gateway solution allows only those end-nodes that can meet some specific security profile to interact with trusted resources. End-nodes that are untrusted are denied access. By creating this trusted environment and restricting the permitted communications inside and outside of this environment, the utility company can reduce the overall risk to its data assets.

As part of a layered approach to network security, one additional security asset that needs to be mentioned as part of the inspection architecture shown in Figure D-5 is the addition of an Intrusion Detection System (IDS). Because the encrypted IPSec data stream is not encrypted until it reaches the demarcation point, represented by the gateway, an IDS has complete knowledge of all activity on its protected domain. This is the advantage of using IPSec in tunnel mode as opposed to transport mode. (For a description of an IDS, see Appendix D, section D.10, Intrusion Detection.)

The following steps provide some configuration guidelines when building an IPSec VPN:

1. Determine network design details to include the encryption policy, identified host, and networks that will be protected, and the IPSec features that will be used. Allow any preconfigured firewalls to pass IPSec negotiation ports (UDP ports 50 and 51).

2. Configure the mode for creating security associations, static or dynamic. The process of securing data between multiple users using IPSec starts with defining an SA. An SA, uniquely identified by a multiple-bit number called a security parameter index, is constructed by identifying the following parameters in a transform set:
   - Source and destination IP address of the peers participating in the creating and termination of the IPSec tunnel.
   - IPSec encapsulation protocol (AH or ESP).
   - The encryption algorithm and secret key used by the IPSec protocol.
   - The authentication algorithm used to authenticate IPSec packets.
   - IPSec mode (tunnel or transport).
   - Lifetime of the security association.

D.3.2 Application Layer Virtual Private Network

Two popular forms of VPN are SSL and its more modern versions, Transport Layer Security (TLS) and Secure Shell (SSH). The SSLv3.0 protocol can provide an encrypted tunnel between two participating nodes. TLS has a few more features. Both SSLv3.0 and TLSv1.0 are used to protect application traffic that can be exchanged between end nodes. The protocols are also application independent, which means higher level protocols can be layered on top of SSL or TLS. The protocol allows client/server applications to be protected from tampering, forgery, and viewing. IEC TC57 Working Group 15 (Data and Communication Security) is addressing cyber security of control center and substation communications, which includes introducing an SSL implementation for data communication authentication and confidentiality.

One popular form of the SSL application is used to protect Web traffic or secure HTTP (HTTPS). This implementation is built upon a public key infrastructure (PKI) that can provide application authentication through public and private key pairs. With respect to HTTPS, the client public key is embedded into the user Web browser and becomes transparent to the user. However, SSL is not limited to securing just HTTP traffic; SSL can secure many different application layer programs, including any of the control center protocols, such as ICCP, as seen in Figure D-6.

The primary security need, prior to utility-data transmission from one node to another, is the verification that each participating end node can be trusted. In other words, each node has a predefined mechanism that can validate identity. As part of the application VPN process, the authenticity of communicating end nodes is through digitally signed certificates. This process relies on PKI installation. This technique of proving each other's identity will take place prior to any data exchange between nodes. When an application on a client utility node calls a secure application on another node, the secure application layer will initiate the request for a certificate exchange. How long it takes prior to the resolution of trust will depend on 1) the speeds of the processors on each node performing the certificate exchange, 2) the transmission delay caused by all the intermediate communication infrastructure nodes, and 3) any additional layers of security that must be engaged to process the transmitted data. A properly configured certificate exchange provides the application transaction user with *end-node authentication and negotiated data confidentiality*.

For additional application layer VPN integration and design details, see the following standards documents:

Alan O. Freier et al. Internet Engineering Task Force (IETF). Transport Layer Security Working Group. *SSL Protocol*, version 3.0. Internet draft. 1996.

T. Dierks et al. Internet Engineering Task Force (IETF). Network Working Group. *TLS Protocol*, version 1.0, Request for Comments 2246. The Internet Society. 1999.

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1.18, Secure Communications Channels, version 1.0. April 2004.

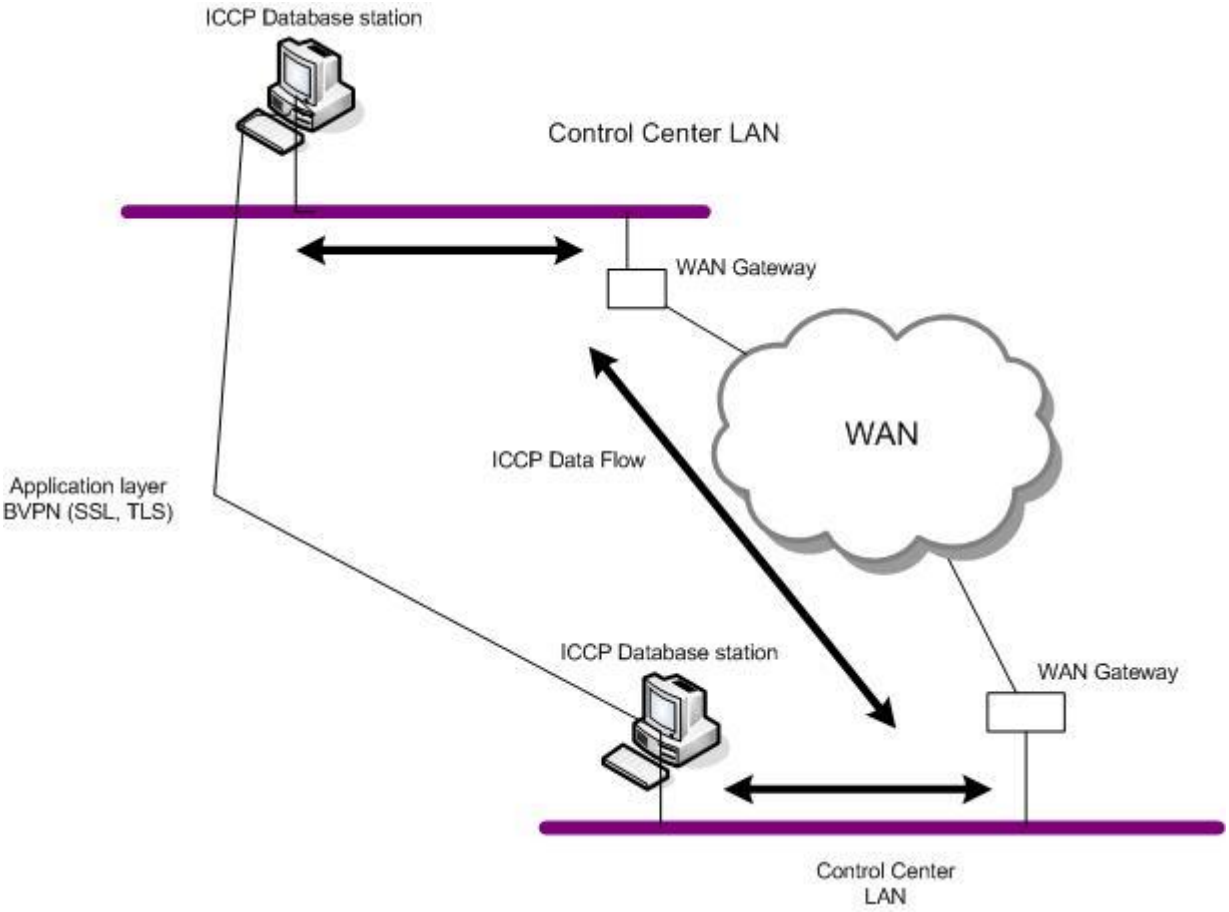

Figure D-6. Application Layer VPN

Another application layer VPN is SSH version 2 (SSHv2). This VPN protocol is used to secure communications between network devices. It has been primarily used on UNIX-based systems to access "shell accounts" remotely in a secure fashion. It is a secure replacement for the Telnet protocol, which also is used to provide access to network machines. Telnet allows user passwords and data to be transmitted in the clear. The SSH encrypts remote sessions, which

provides confidentiality and data integrity over non-secured networks. It is recommended for use by network and system administrators when remotely configuring network devices.

Embedding Virtual Private Networks

It is also possible to tunnel one VPN, such as SSL, through another, such as IPSec. This results in overlaying the IPSec and SSL VPN technologies on each other in order to provide secure access to and through security perimeters. For example, as seen previously in Figure D-4, a company may set up an IPSec VPN to provide secure access to the edge perimeters of each participating company node. The IPSec VPN would terminate at the perimeter edge. While each computer host within the security perimeter may allow an SSL VPN to protect an application, this would be tunneled through the IPSec VPN and would terminate at the distant computer host. The following documents provide standard descriptions and applications of the SSH protocol:

T. Ylonen, et.al. Internet Engineering Task Force (IETF), Network Working Group. *The Secure Shell (SSH) Protocol Architecture*. The Internet Society, 2006.

T. Ylonen, et al. Internet Engineering Task Force (IETF), Network Working Group. *The Secure Shell (SSH) Transport Layer Protocol*. The Internet Society, 2006.

D.3.3 Security Observations

Because commercial VPNs were built around commercial OS and protocols, there is a lack of VPN products available for controller products, such as programmable logic controllers (PLCs), remote terminal units (RTUs), and DCSs. VPNs also add processing delays due to the protocol overhead and the encryption algorithms they support.

A VPN should not be considered a complete network security solution, but as one layer in multiple security layers. A VPN does not protect the network or host against malicious software, such as viruses or Trojans. Proper host access controls, application controls, and malicious software protection are important protection mechanisms to prevent VPN compromise.

A VPN does not provide protection against *any insider* who has access to the host or network device that supports a VPN. AVPN does not provide user-based access control. There is a need to have a process in place to identify a user who is logging into a particular application service. This will prevent unauthorized access to the application server or workstation. This process, when guided by policy, can be implemented in the form of user authentication to the workstation and/or to the server where the application resides. This can provide the proper restrictions to application access on a per-user basis.

User authentication can be implemented locally for each machine or globally by the use of user role-base authentication services providing a role-based access control (RBAC). This essentially translates a *user's role* to application permission. (See Appendix D, section D.13.1, Host Access Control in this report for more access control details.)

Both the network layer and application layer VPNs use cryptographic algorithms to provide end node authentication and data integrity. Therefore, these layers provide good protection against

external threats attempting to compromise or manipulate data.

Threats from developer- or vendor-based sources are primarily associated with software lifecycle issues. Software update processes that use remote access connectivity or removable media, such as thumb drives, can be vulnerable to malicious code insertion. These potential vulnerabilities can be mitigated within a properly established security policy to provide secure guidance for software update and patch procedures.

Internet Protocol Security

As previously described, inserting IPSec encryption services can protect data streams from a demarcation point between the local area network (LAN) and the WAN. Although the Layer 3 protection mechanism could be deployed at the workstation or server, it is recommended to insert this protection at the WAN entry point. This allows for other security monitoring technologies, such as IDSs and intrusion prevention systems, to continually monitor the data transmission and reception streams for abnormal content or behavior. Layer 3 encryption services provide data integrity, data confidentiality, application port confidentially, and end node authentication.

Note that when referring to *authentication* of IPSec VPN nodes, this term is associated with network node end points, not users. Another form of user-level authentication is needed if using a VPN to provide access for remote users. This could be in the form of an application requiring a user ID and password.

Although IPSec is a defined standard, there are still interoperability problems; not all products interoperate in all modes. IPSec authentication typically works using a static IP address as a distinguishing name for identification purposes. Therefore, it does not work well in a Dynamic Host Configuration Protocol (DHCP) environment (where IP addresses are dynamically assigned, typical of dial-up connections to Internet service providers).

Also, if digital certificates are being used for end point authentication, establishing a PKI can add to the complexity of the overall implementation because of the need to include and manage a root certificate server.

An important observation about implementing an IPSec VPN is its path through a router. If a router in the path of an IPSec VPN is running a Network Address Translation function, it will break the IPSec tunnel connection. (A Network Address Translation essentially swaps an internal *private* IP address, originally within the internal network, with an external *public* IP address, used to route through the public [Internet] network.) The IPSec tunnel connection breaks because the internal IP address is used as part of the information needed to create the cyclic redundancy check (CRC) value that IPSec uses to verify that the header information has not been changed on its route through the public network.

Secure Sockets Layer/Transport Layer Security

The SSL/TLS protocol is implemented using the Transmission Control Protocol (TCP) (a transport protocol) and does not provide support for UDP data application traffic. Another potential security issue with SSL/TLS is that many applications that use SSL/TLS require the

server to provide only a certificate for authentication without implementing any client authentication. This aspect should be considered if both the server and the client in a connection must be authenticated. The SSL/TLS VPN does not provide the following protections:

- Application software validation (software version security).
- Application identification (does not hide port numbers).
- User identification (an actual person).
- Address information (does not hide originating host address).

Secure Shell
A security flaw in the SSH version 1 leaves it vulnerable to man-in-the-middle attacks. Redesigned to prevent this vulnerability, SSHv2 is recommended for use. SSH is not designed to be incorporated into network gateways, such as routers or firewalls, as a complete VPN solution for data traffic. The primary use for SSH is for scripting applications, such as RTU connections; these allow remote user account access that can protect user ID and password authentications from being intercepted by an adversary. Typically, this application is used to protect network management login accounts during remote access to network devices.

Note: Key length is an important security parameter when used to protect data confidentiality. Key length and the cryptographic algorithm selected can have impact on the operational performance of the key management protocol along with an adversary's ability to compromise confidentiality. A publication produced by NIST [D.3-1] helps in understanding the choices associated with key management.

Some standards and guides for VPN security are listed below:

NIST SP 800-113. *Guide to SSL VPNs*, provides guidelines on implementing a secure sockets Layer (SSL) virtual private network (VPN). It mentions that SSL can also be referred to as TLS and that IPsec is a complementary VPN.

NIST SP 800-77. *Guide to IPsec VPNs*, provides guidelines for using security controls, in particular IPsec VPN implementations that can provide data protection for TCP/IP networks. It includes some typical architecture implementations, IPsec fundamental discussions, planning, and implementation.

ISO/IEC 18028-5:2006, defines techniques for securing inter-network connections that are established using VPNs.

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 7.3, Virtual Private Networks, discusses typical deployment and security vulnerabilities addressed by VPNs.

Reference:
D.3-1 *Recommendation for Key Management*, Special Publication 800-57 Part 1, NIST, March 2007.

D.4 Ethernet

The Ethernet protocol is associated with Layer 2 of the OSI[12] model used to describe a communication process that sends data between two distinct nodes across a communication network. The Ethernet protocol was originally designed as a shared medium protocol. The data transmission from one node to another was propagated on the entire network segment. This shared medium meant that all nodes on the network shared a common communication path or *domain*; only one node could be transmitting data at any given time, or a *collision* would result. If one node were transmitting data, then all other nodes would wait until they sensed the transmission media were not in use and then would attempt to communicate. This method created a contention for media access and created communication bottlenecks as more nodes required access to the media.

The development of Ethernet switching offers a means of using the Ethernet standard, greatly increasing performance without having to replace the existing infrastructure. The Ethernet switch has been designed to divide the network into many small segments. Each segment can then separate and isolate each transmitting node into separate collision domains. This means nodes in different collision domains can talk simultaneously, which increases the transmission efficiency. Instead of sharing a 10-Mbps connection with many nodes, each node (a workstation or server) can have a dedicated 10-Mbps segment connected to an Ethernet switch. This allows simultaneous communications, which can occur at 10 Mbps speeds or greater. Thus, transmission bandwidth substantially increases.

One popular and efficient Ethernet switch configuration is hierarchical in nature. This allows the network to be designed in layers. Using the layer approach simplifies the task for network designs. Each layer can focus on specific functions, allowing the designer to choose the right features for each layer.

The hierarchical layered approach can also accommodate design changes. The hierarchical layered approach also provides modularity to the network design, which allows for node replication as the network grows. When a network node requires a design change, the cost of the change, and the amount of effort to induce the change, can be constrained to a small subset of the overall network. Changes on other network architectures, such as flat or meshed network architectures, tend to create a large impact on the overall system. Other attributes of a hierarchical layered architecture include improved fault isolation because the interface points within the hierarchy make identifying failure points easier. A hierarchical switch network normally includes three layers: the backbone (or core), a distribution layer, and an access layer. Figure D-7 shows a typical hierarchal switch network with a description of the functionality of each layer.

The upper, or core, layer is the high-speed switching backbone and is designed to switch packets at aggregate rates for all incoming and outgoing data flows. This network layer should not perform any packet inspection or manipulation, such as access lists and filtering. Packet

[12] http://www.networkdictionary.com/protocols/osimodel.php

inspection or manipulation will adversely impact the performance of the high-speed backbone. Packet inspection should take place at the boundary between the core and the distribution layer.

The distribution layer of the network, which is the demarcation point between the access and core layers, helps define and differentiate the core from the rest of the network. This layer provides boundary definition and is where access control lists can be applied to enforce packet policies.

The access layer is the point where local end devices are allowed access into the network. This layer may use filters, such as Media Access Control (MAC) addresses or virtual LAN (VLAN) markings, to optimize the needs of a particular set of devices.

Additional information is available in the following standard:

IEEE Std 802.3x, 802.3y, Supplements to ISO/IEC 8802-3, 1996 Specifications for 802.3. *Full Duplex Operation and Physical Layer Specification 100Mb/s*, 1997.

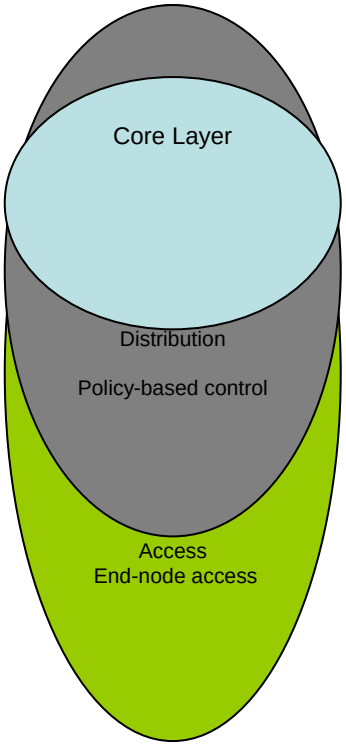

Figure D-7. Hierarchical Ethernet Architecture

D.4.1 Ethernet Security Observations

The Ethernet switched architecture comprises most of the networks associated with NPPDNs. Both the Non-Safety Information Layer and some proposed safety networks utilize Ethernet switches. The overall security profile of a network should always include attention to every participating component and its communication layer. The OSI communication stack is only as

robust as its weakest link; therefore, equal attention should be paid to any of its layers to ensure its entire structure is sound.

The majority of attacks against the Layer 2 (Ethernet) protocol exploit the inability of a device to track the attacker. Therefore, the attacker can perform undetected malicious actions on the forwarding path to alter it and then implement the change.

The following is a list of potential attacks that can be leveraged against Ethernet *switched* networks:

- MAC flooding attack
- Address Resolution Protocol (ARP) attack
- Spanning-tree attack

D.4.1.1 MAC Flooding Attack

A form of denial-of-service (DoS) attack, located at the MAC layer of the OSI model and implemented in a brute force way, is referred to as a *MAC attack*. (The MAC attack takes its name from the acronym for Media Access Control.) This attack takes advantage of the memory needed to store the MAC address to port mappings within Ethernet switches.

When Ethernet switching is used to provide network communications, the switch builds a content addressable memory (CAM) table. The CAM table maps the source MAC address of an Ethernet frame and its associated port. This allows the switch to determine the destination port of a transmitted Ethernet frame. All workstations and servers on a local segment have a unique MAC address, which is associated by the Ethernet switch with its interconnected port. This mapping allows the switch to direct the Ethernet frames to their proper destination. By using ARP, the switch retrieves the information to build the CAM tables .

An adversary who wants to deplete the available memory space creates a script that sends out a large volume of gratuitous ARPs. These are stored in the CAM tables of the Ethernet switch; this creates a DoS situation against the CAM table. The volume of ARPs is greater than the designed capacity of the Ethernet switch. Therefore, the switch stops forwarding Ethernet frames from the source device to the destination device. At this point many switches default to a broadcast mode for all incoming frames, sending out each frame to all switch ports. This allows the viewing of traffic from hosts. Figure D-8 shows the MAC flooding attack. Several programs are available to perform this task, such as *macof*, part of the *dsniff* suite of tools, easily acquired over the Internet.

MAC Flooding Attack Protection

Some Ethernet switches, which support *port security, can be* used to constrain the connectivity of a device based on that device's MAC layer address. Limiting the number of MAC addresses that can be associated with a single port can prevent a MAC flooding attack. Thus, the identification of device traffic can be mapped directly to its port of origin.

D.4.1.2 ARP Attack

Another attack that can be launched against Ethernet networks is the man-in-the-middle attack. It occurs at the data link layer of the network and is called an ARP spoof or ARP attack. A feature provided when using the gratuitous ARP protocol allows this attack to be implemented and carried out. This feature allows a host, upon boot up, to deliver to all listening hosts and Ethernet switch ports certain information. This information allows these hosts and network devices to update their mapping between an existing IP address and a new MAC address. When a gratuitous ARP is received by hosts and Ethernet switch ports, all previous ARP cache and CAM table entries—containing the previous IP-to-MAC address mapping of the host—are overwritten by the newly requested association. This allows an adversary with a cleverly written script to issue gratuitous ARPs so new IP-to-MAC mappings will be created. The intent is to redirect traffic flow through the adversary's connected port.

As seen in Figure D-9, the adversary starts by sending to Host A a forged gratuitous ARP packet with Host B's IP address and the attacker's MAC address. The adversary also sends to Host B a forged gratuitous ARP packet with Host A's IP address and the adversary's MAC address. Now, all of hosts A and B's traffic will go to the adversary, who can review and/or modify the traffic prior to sending it on to hosts B or A.

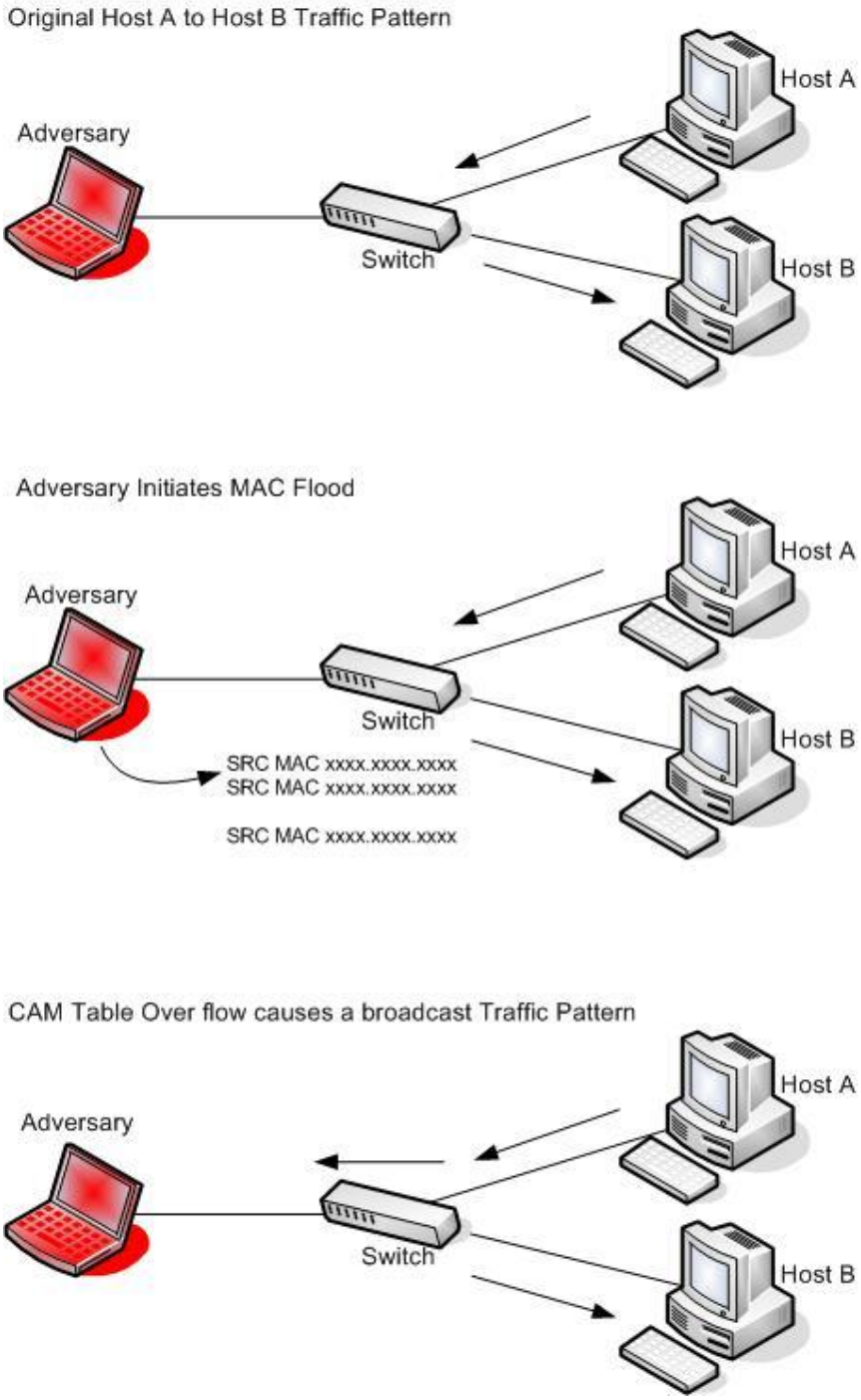

Figure D-8. MAC Flooding Attack

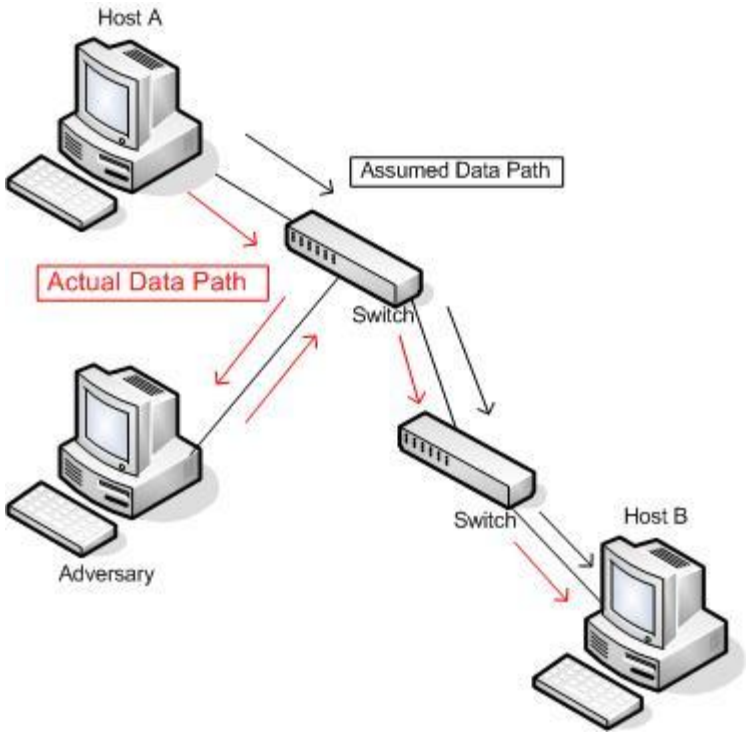

Figure D-9. ARP Attack

ARP Attack Protection

This type of attack can be prevented either by 1) blocking the direct communication between the attacker and the attacked device at the Layer 2 port, or 2) embedding more intelligence into the network so that the network can check the forwarded ARP packets for identity correctness. Available on some Ethernet switches, a feature called ARP inspection prevents ARP spoofing by ensuring an attacker cannot hijack the user's default gateway address. This feature prevents malicious users from impersonating other hosts or routers. It does this by inspecting all ARP packets. It enables the network administrator to configure a set of order-dependent rules within the security access control list (ACL) framework, to prevent the attack described above.[13]

D.4.1.3 Spanning-Tree Attack

The Spanning Tree Protocol (STP) is a loop-prevention protocol that is implemented at the data link layer. This technology allows switches to communicate with one another to discover and map physical loops in the network. The STP creates a tree structure that has loop-free leaves and branches and spans the entire Layer 2 network. There are two attributes within a spanning tree network that are used to create its logical tree like structure: they are the *bridge ID* (BID) and the *path cost*.

A BID is a single 8-byte field that is composed of two subfields: a low-order subfield and a high-order subfield. The low-order subfield is comprised of the 6-byte MAC address of the device and the high-order field contains what is referred to as the bridge priority; this is a two-

[13] Cisco Catalyst OS Software Product Bulletin, http://www.cisco.com

byte field. The default bridge priority field is normally set to the value 32,768 (base ten), which is half the maximum setting of 65,535. Bridges (switches) use the concept of cost to evaluate how close they are to other switches. The cost of a path is based on the speed of the media that the information will transverse. In essence, the faster the media, the lower the path cost. To create the "tree" structure in the STP, a *root* bridge needs to be designated. An election process accomplishes this. All switches/bridges send out Bridge Protocol Data Units (BPDU) advertising the following attributes:

- Root bridge identification (root BID)
- Root path cost
- Sender BID
- Port ID

The important attribute in the root bridge election process is the root BID. All switches participating in the election process choose the root bridge based on the lowest BID. To define a root bridge the operator must change the value of the first two bytes of the BID to a lower value than factory default. If this is not done then the switch in the network that has the lowest value MAC address will be defined as the root bridge. If this root bridge has less than a desirable location within the network, a less then optimal switching path for data transport could be constructed.

STP was designed to ensure a loop-less network environment. There are three basic steps in which STP establishes its topology: (1) electing the root bridge, (2) selecting one root port on every non-root bridge, and (3) selecting one designated port per network segment. Electing the root bridge is done by exchanging Layer 2 BPDUs. When the STP is in use, every port on a switch goes through several stages. The bridge with the lowest ID becomes the root bridge. When sending BPDUs the switch sets the root BID to its own ID. Because every switch stops sending BPDUs when it receives a BPDU with a lower root ID than its own, eventually the only switch sending BPDUs is the root bridge.

An attack vector can disrupt the switch spanning-trees, destabilize their MAC address-tables and hold the network in a constant state of reelecting the root bridge. This can be achieved because there is no authentication mechanism built into the STP.

By crafting BPDUs of a non-existent switch with an ID of 1, the adversary can elect its non-existent switch as the root bridge. By using a minimal max-age for the crafted packets, and not sending BPDUs within that time, will cause another election on the network, during which the adversary will start sending bogus BPDUs, once again winning elections and becoming the root bridge.

By repeating this process, the network will be in a constant state of re-electing the root bridge, and will fail to converge, thus, reducing data traffic and saturating the network with BPDU frames. Figure D-10 shows the adversary's STP attack.

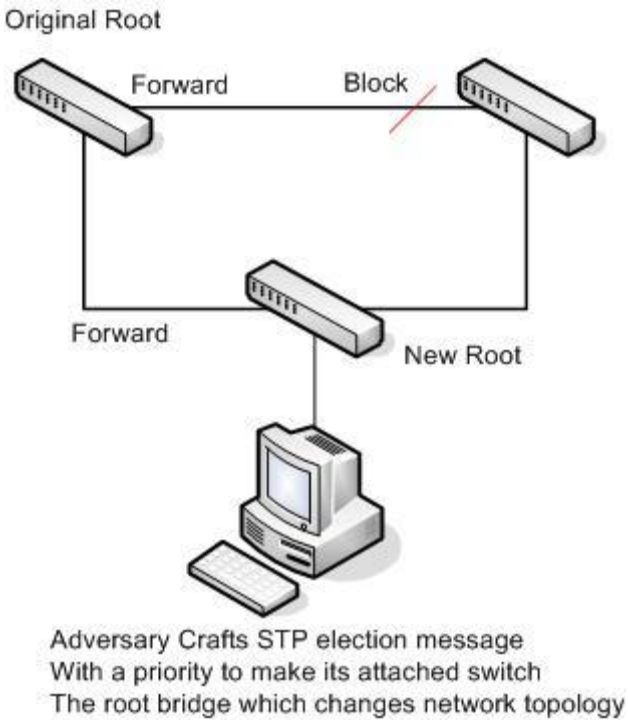

Figure D-10. Spanning-Tree Attack

Spanning-Tree Attack Protection

There are a couple of simple methods to prevent the exploitation of the STP vulnerability in a network. For any STP attack to be feasible, the switch must accept BPDUs on a port that the attacker has access to. It is therefore possible to make such an attack impossible by denying access to STP enabled ports to non-privileged users. This can be done by disabling STP on access ports, having port security enabled on all user ports, and restricting physical access to network equipment.

With STP disabled on user ports, the attacker would have to access the switch physically and use a switch-to-switch port to connect his computer (assuming all non-used ports are either disabled or have STP disabled). If physical access to network devices cannot be restricted, other measures must be taken to ensure network security. Port security is a feature that allows the switch to accept frames from only a given number (usually the first learned) of source MAC addresses. Enabling port security on user ports will make the attack unfeasible without prior network "sniffing" or hijacking a user workstation.

D.4.1.4 Additional Observations

To implement the attacks previously described against the switched Ethernet protocol, an *external threat* would need physical access to the network. Any Layer 3 device (router) that sits between the external threat and the Ethernet network would prevent the implementation of these attacks. This is because these attacks are all associated with the Layer 2 communication process. If proper physical protections at the NPP are in place the *external threat* can be prevented from implementing these attacks.

The *unprivileged insider threat* can be a concern because of the physical access required to attach a device to the Ethernet network. An unmonitored network would be unable to detect or prevent an unprivileged insider from initiating some of these exploits. However, there are features that can be included in the Ethernet switch network that can help prevent these attacks. Some Ethernet switches can run a Simple Network Management Protocol (SNMP) and be configured to send out a status message to a network management system alerting the network manager whenever a new device is plugged into an Ethernet port. Also there are security features available that allow an Ethernet switch port to learn the MAC address of the Ethernet device plugged into the port and allow only that particular address to send and receive information from the originating port. Another good security procedure to prevent open ports from being improperly used is to simply disable the port. Proper device access controls, which include proper user authentication and role-based capability are important to provide layers of defense to protect the Ethernet network from unprivileged insider manipulation.

The *privileged insider* would have a larger administrative role within the facility. This could include the management of the Ethernet switch network. It may be possible to limit the number of systems that a single administrator can access or limit the locations where administrators are allowed access within the plant. A formal process for change management should be instituted. It could include procedures, such as requiring that multiple administrators or subject matter experts review all configuration changes to help detect malicious or accidental configurations. Combining both physical protection mechanisms for personnel access control along with restricting the number of systems that can be accessed can provide some level of protection against this type of threat. The privileged insider is the most difficult threat to overcome.

Threats from *developer*- or vendor-based sources associated with the Ethernet switch depend on the sophistication level of the Ethernet device. Low end function devices are quite simplistic in their operation and would have a minimal attack surface for exploitation. The upper end of Ethernet functionally would include many of the features seen in OSs associated with computer systems. These features could include Web servers for remote HTTP access and configuration, SNMP Management Information Base utilities that allow information to be queried from the switch, and Trivial File Transfer Protocol (TFTP) to allow external configuration files to be downloaded to the switch. These are all possible avenues to allow misappropriation of the Ethernet product. Designing a network that includes proper monitoring techniques to detect unusual or inappropriate actions from network devices can help detect any malicious vendor content. (See Appendix D, section D.11, Intrusion Prevention Systems for more details on detection techniques.)

D.4.2 Ethernet Virtual Local Area Networks

Along with network-based access control, which is administered at the application level, there is another form of need-to-know separation that can be implemented at the network device level. Network devices, primarily Ethernet switches, can be configured to separate user traffic by the administration of VLANs.

A switch in an internal database defines the VLANs. After a VLAN has been created within the database, then end ports are assigned. These end ports map to end user devices such as a

workstation or a server. A VLAN is assigned a unique number or name, which is distributed by the VLAN Trucking Protocol (VTP). VTP provides the means of distributing and updating the VLAN database. If a switch does not know a VLAN, then the switch (normally an Ethernet device) cannot transfer data across any of its ports. This provides the network administrator the ability to segment users or services on a common LAN and provides a virtual separation of users that need access to sensitive information from the rest of the general users on the LAN, regardless of their physical location.

The VLAN Trunking Protocol (VTP) is a Layer 2 messaging protocol that maintains VLAN configuration consistency by managing the addition, deletion, and renaming of VLANs on a network-wide basis. When a new VLAN is configured on a VLAN supported switch, the VLAN configuration information is distributed through the VTP protocol through all switches in the domain. This reduces the need to configure the same VLAN everywhere.

Each switch can be configured in one of three possible VTP modes: server mode, client mode, or transparent mode. Each switch can be part of only a single VTP management domain at any given time; each mode is defined below;

Server Mode: Once the VTP is configured on an Ethernet switch, the default mode used is Server Mode. In any given VTP management domain, at least one switch must be in Server Mode. When in Server Mode, a switch can be used to add, delete, and modify VLANs; this information will be passed to all other switches in the VTP management domain.

Client Mode: When a switch is configured to use VTP Client Mode, it is simply the recipient of any VLANs added, deleted, or modified by a switch in Server Mode within the same management domain. A switch in VTP Client Mode cannot make any changes to VLAN information.

Transparent Mode: A switch in VTP Transparent Mode will pass VTP updates received by switches in Server Mode to other switches in the VTP management domain; however, it will not actually process the contents of these messages. When individual VLANs are added, deleted, or modified on a switch running in Transparent Mode, the changes are local to that particular switch only, and are not passed on to other switches in the VTP management domain.

Additional information on VLANs is available from the following standards:

IEEE STD 802.1Q. Standards for Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks. IEEE Computer Society, 2003.

IEEE STD 802.1X. Standard for Local and Metropolitan Area Networks, Port Based Network Access Control. IEEE Computer Society, 2001.

D.4.3 Ethernet VLAN Security Observations

Modern NPP data networks are comprised of Ethernet network architectures. Note that Ethernet switches were not designed as security devices, but this usage has been incorporated over time

and is supplementary to their main use as devices that improve network performance. If a switch is used for security reasons, then security relies on the correct configuration of the switch; this includes user understanding of the standards that the switch software is based upon, and the correct implementation of those standards.

The attacks against a VLAN layered Ethernet network are associated with taking advantage of non-secure protocol interactions. This section describes some of the attacks and suggests ways to defend against these exploits.

The following is a list of potential attacks that can be leveraged specifically against VLAN Ethernet networks:

- Double-encapsulated 802.1Q/nested VLAN attacks
- VTP revision attacks

D.4.3.1 Double-Encapsulated 802.1Q/Nested VLAN Attack

While internal to a switch, VLAN numbers and identification are carried in a special extended tag format that allows the forwarding path to maintain VLAN isolation from end to end without any loss of information. The tagging rules are dictated by standards, such as Inter-Switch Link or IEEE Std 802.1Q.

Since every packet always gets a tag, there is no risk of identity loss and, therefore, of security weaknesses. But the IEEE committee that defined 802.1Q decided to allow backward compatibility and, thus, to support the so-called native VLAN. A native VLAN is not associated explicitly to any tag on an 802.1Q link. This VLAN is implicitly used for all the untagged traffic received on an 802.1Q capable port. This capability is desirable because it allows 802.1Q capable ports to talk to old 802.3 ports directly by sending and receiving untagged traffic. However, it may be very detrimental because packets associated with the native VLAN lose their tags—for example, their VLAN identity enforcement, as well as their Class of Service (802.1p bits)—when transmitted over an 802.1Q link. For this reason the use of the native VLAN should be avoided.

A double-encapsulated 802.1Q packets may be injected into the network from a device whose VLAN happens to be the native VLAN of a trunk. Then, the VLAN identification of those packets cannot be preserved from end to end. This is because the 802.1Q trunk would always modify the packets by stripping their outer tag. After the external tag is removed, the internal tag permanently becomes the packet's only VLAN identifier. Therefore, by double-encapsulating packets with two different tags, traffic can be made to hop across VLANs. See Figure D-11 for a depiction of this scenario.

Double Encapsulation Attack Protection

This scenario can be considered a misconfiguration, since the 802.1Q standard does not necessarily force the users in these cases to employ the native VLAN. The proper configuration is to clear the native VLAN from all 802.1Q trunks. In cases where the native VLAN cannot be cleared, always choose an unused VLAN as the native VLAN for all the trunks and do not use the VLAN for any other purpose. Protocols like STP, DTP, and UDLD should be the only

rightful users of the native VLAN and their traffic should be completely isolated from any data packets. [14]

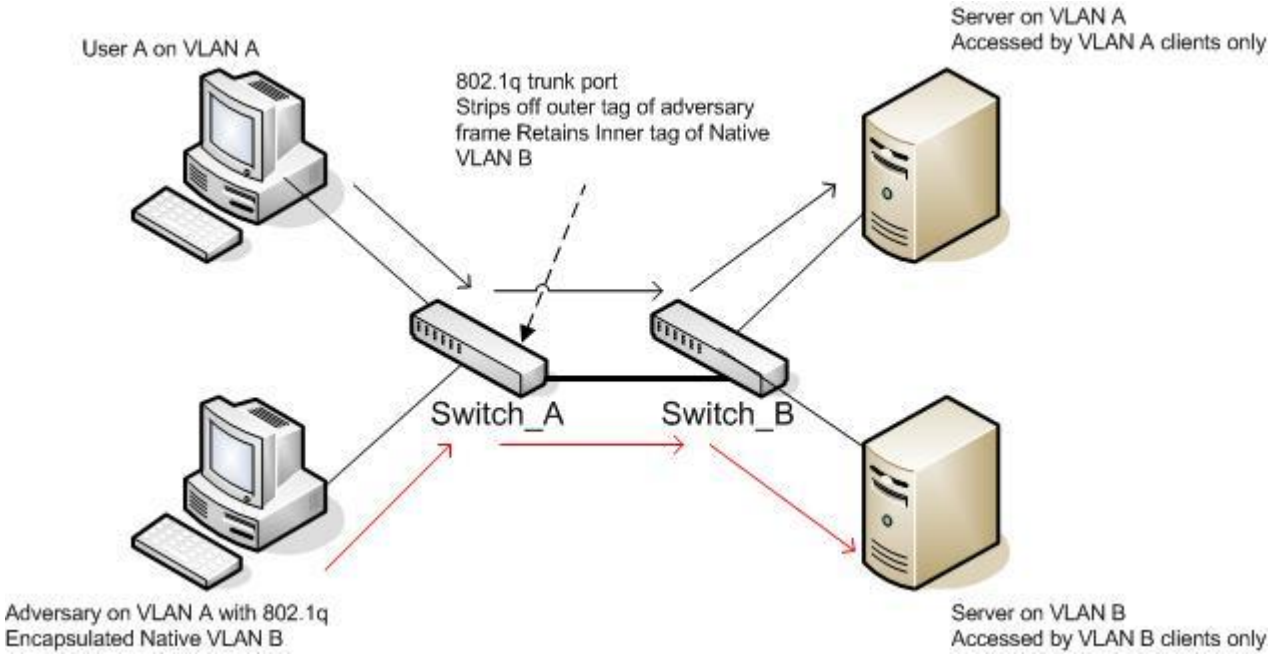

Figure D-11. Double Encapsulated 802.1Q VLAN Attack

D.4.3.2 VTP Revision Attack

As previously described in this section, Ethernet VTP is used to distribute up-to-date information on all VLANs associated with interconnected switches within the same VTP domain. The VTP domain is a logical group of switches that will share VLAN information. Each switch can belong only to a single VTP domain; the domain must be the same between interconnected trunk ports for communication to take place. By default, switches are configured to be VTP servers without a VTP domain.

To join a VTP domain, the switch must be manually configured, or by connecting the switch to a VTP domain through a trunk port (to receive domain information through a VTP advertisement). When changes are made to a VLAN configuration on a VTP server, VTP advertisements are sent out over all trunk ports to propagate the changes to the rest of the domain. To track changes to the VLAN configuration, VTP relies on a revision number. When a VTP domain is initially configured, the revision number is 0. Each time the VLAN database is edited, the revision number is incremented by 1. If a switch receives a VTP advertisement with a higher revision number, the information in the advertisement overwrites the information stored in non-volatile random access memory.

The adversary can use VTP to their advantage to remove all VLANs (except the default VLANs) on a network. The adversary exploits VTP by connecting into a switch and establishing a trunk

[14] WLAN Security White Paper [Cisco catalyst 6500 Series Switches] Cisco Systems Inc.

between his/her computer and the switch. The adversary then sends a VTP message to the switch with a higher configuration revision number than the current VTP server advertising no VLANs configured. This causes all switches to update their VTP configuration database with the adversary's update, which removes all the *non-defaults* from their VLAN database and allowing the adversary to be on the same default VLAN. Figure D-12 displays this attack.

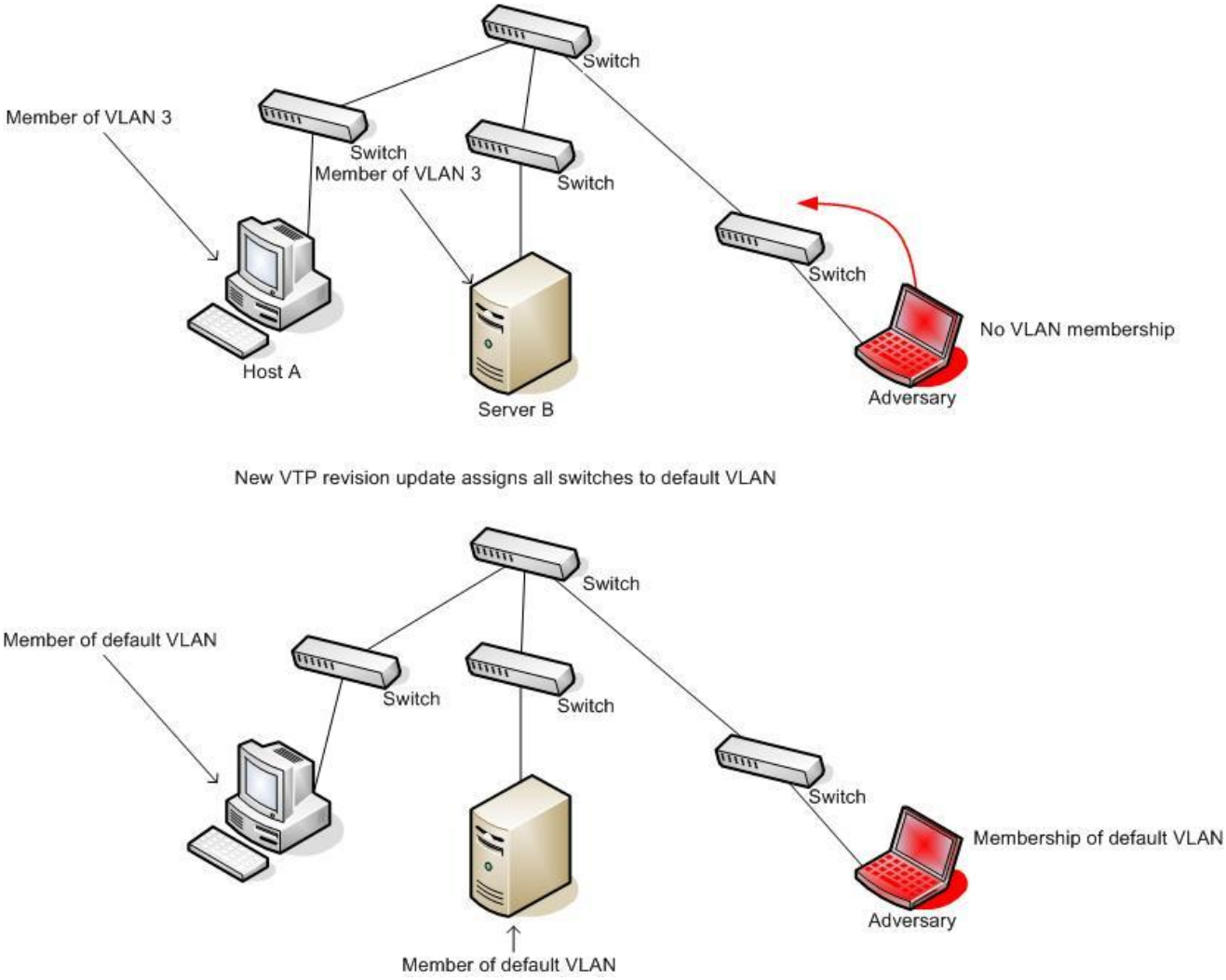

Figure D-12. VTP Revision Attack

VTP Revision Attack Protection

Requiring a password for VTP revision advertisements will prevent a rogue switch from connecting to the network and participating in its advertised configuration. If a password for VTP is set, the password must be configured on all switches in the VTP domain. It is also important to provide MAC address mapping to active ports, thus, preventing the introduction of rogue MAC addresses and disabling ports not in use. Turning the auto-trunking feature *off* on all the switches that do not require trunking will also tighten security.

D.4.3.3 Additional Observations

The *external threat* would have the same limitations attacking Ethernet VLANs as attacking the Ethernet Switch. The external threat would need physical access to the network. Any Layer 3 device (router) that sits between the external threat and the Ethernet network would prevent implementation of attacks. If the NPP has proper physical protections at in place, the external threat can be mitigated.

The *unprivileged insider* threat can be a concern because of the physical access required to attach a device to the Ethernet network and, thus, to the Ethernet VLANs. All previous recommendations on the protection of the Ethernet switch mentioned in the threat discussion (section D.4.1) apply. In addition requiring a password for VTP revision, advertisements will prevent a rogue switch from connecting to the network and from participating in advertised revision announcements.

The *privileged insider* would have a larger administrative role within the facility. This could include the management of the Ethernet switch network. It may be possible to limit the number of systems that can be accessed by a single administrator or the locations that are allowed access within the plant. Combining both physical protection mechanisms for personnel access control along with restricting the number of systems that can be accessed can provide some level of protection against this type of threat. The privileged insider is the most difficult threat to overcome.

Threats from developer or vendor-based sources associated with the deployment of Ethernet switch that support VLANs are the same as the Ethernet switch protections. Designing a network that includes proper monitoring techniques to detect unusual or inappropriate actions from network devices can help detect any malicious vendor content. (See Appendix D, section D.11, Intrusion Prevention Systems for more details on detection techniques.)

Some standards and guides for Ethernet deployment security are listed below:

IEEE.1AE-2006. *IEEE Standard for Local and Metropolitan Area Networks Media Access Control (MAC) Security*, provides an overview of MAC layer associated with Ethernet deployment. It describes how to provide secure provisioning of the MAC service to include secure VLAN tagging.

National Security Agency Report 133-010R-2004. *Cisco IOS Ethernet Switch Security Configuration Guide*, provides specific information on configuring and deploying Cisco series Ethernet switches in a network environment. Although the actual configuration details are specific to the Cisco product, the overall configuration approach can be used by all Ethernet switch deployments.

ISO/IEC 18028-3 Part 3. *Securing Communications between Networks Using Security Gateway,* section 7.1, Switches, describes Ethernet switches and their relationship to security gateways (firewalls).

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 6.3, Virtual Networks, describes the Ethernet VLAN technology, some known issues, and the vulnerabilities addressed by this technology.

D.5 Programmable Logic Controller

PLCs are used in both Supervisory Control and Data Acquisition (SCADA) and DCSs as the control and status component that manages processes through feedback control and status. PLCs are connected by a communication bus that allows an operator to review both configuration and status of electricity generation components under the PLC control. In Figure 2-1, Digital Plant System Network Architecture, four PLCs are seen in the safety network. Each PLC can be queried by the safety operator console that is located in the control room or an auxiliary room used for *remote shutdown* of electricity generation components. These *electricity generation* components can be comprised of sensors, pumps, valves, or breakers and are attached to the PLC through a Field bus.

PLCs have a user-programmable memory for storing instructions for the purpose of implementing specific functions, such as I/O control, logic, timing, counting, communication, and data and file processing. The PLC is accessible through a programming interface, located on an engineering workstation. Data can be stored in a data Historian, labeled in Figure 2-1 as a *safety data* server, which is connected to the same network bus as the PLCs. For the Manufacturing and Control System environment, *edge* devices like RTUs, PLCs, and controllers are arguably as important, if not more important, than the host computers. They perform measurement functions, make logic and control calculations, and issue commands that modify the operation of the process.

D.5.1 Security Observations

PLC devices are embedded computers that contain a real-time operating system (RTOS) or reduced instruction set computer (RISC) for their basic operation. Furthermore, the nature of industrial control requires that these devices accept parameters, commands, and even downloads of new programs through a network connection. The combination of limited internal security features—plus the requirement that devices accept commands sent over the network—make these systems vulnerable to cyber attacks unless they are on a truly isolated network. The problem is further aggravated by the trend to *Internet enable* these devices by adding convenience features like Web servers for remote administration.

Most RTOSs have no mechanism for denying access to system resources unless there is a timing conflict. Embedded systems typically use a memory space that is available to all processes. As a result, malicious programs that are introduced into an embedded device (e.g., through its network connection) are free to read and modify any data and circumvent the normal operation of the device. In some deployments of PLCs, vendors are given access to the PLC remotely for maintenance and upgrades. This practice should be captured in the security policy and proper authentication mechanisms, such as user ID profiles and passwords implemented to prevent compromise of the device. (See Appendix D, section D.8, Landline Modem Access, for additional remote connectivity details.)

Exploits against embedded system devices such as programmable logic controllers have been found in the public domain. These exploits take advantage of how the embedded system processor interacts with its firmware memory storage area, which is represented by flash (refers to the quick erase and reprogrammable memory function) also referred to as erasable programmable read only memory (EPROM).

To reduce the potential for malicious programs to be propagated into RTOSs or embedded devices, such as PLCs, the following can provide a secure approach to preventing these types of attacks:

- Ensure that flash update schemes require an authentication mechanism.
- Provide proper access control to protect firmware images during storage.
- Do not allow remote updates to occur that reside off the protected control system LAN, and add firewall rules to enforce this policy.
- Do not use removable media, such as thumb drives, from sources that are used on non-safety system networks to copy upgrade images or patches.

D.5.1.1 Additional Observations

For the *unprivileged insider threat* to be able to exploit some vulnerable aspect of a processor and its memory interaction, an unauthenticated protocol for firmware updates would have to be used. Unfortunately, a common firmware update protocol used for remote updates is TFTP, which does not authenticate the source or the target machine during firmware updates. Because of this, the possibility exists that installing an update can compromise or disable the target system. This threat can decreased substantially if the procedures described above are implemented as part of the PLC software update process. Other techniques to limit insider privilege would include the implementation of an RBAC. This can be implemented on each PLC where users can be assigned different user IDs that provide varying levels of controller capability, such as uploading firmware updates.

The *external threat* can be quite isolated from reaching the PLC network. This isolation occurs because the safety network is not directly connected to the most external point of the public network. There are multiple firewalls, Intrusion Prevention Systems (IPSs), and a uni-directional data gateway in the cyber pathway. The proper implementation of these cyber protections along with proper physical protections help isolate the external threat. One caveat to this observation is the implementation technique used to update firmware or software on the PLC controllers. An external threat can take advantage of the implementation process used for software and firmware updates if this technique is not properly authenticated.

The *privileged insider* would have a larger administrative role within the facility, but can possibly be limited in the number of systems that can be accessed or in the location of the PLCs within the plant. Combining both physical protection mechanisms for personnel access control along with restricting the number of systems that can be accessed provides some level of protection against this type of threat.

Threats from *developer*- or vendor-based sources are primarily associated with default passwords and user accounts that have been implemented as part of the vendor pre-installation configuration. These potential vulnerabilities can be mitigated within a properly established security policy. Removing user accounts and default passwords should be part of a security policy established at each NPP facility. (See section 2.1, Security Policy, in this report for additional implementation details.)

A standards guide for embedded and real-time systems is listed below:

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems,* section 9.2, Real-time and Embedded Operation Systems, describes typical deployment of these types of systems, some known issues and weaknesses and protections that can be afforded by these systems

D.6 Shared Server

As electric utility plant processes move toward more digital integration and, thus, more advanced capability, the data that can be extracted will increase in volume and importance. The need to share this information is the prime reason for the data Historian. The Historian sever, sometimes referred to as the plant information (PI) server, has the capability to record and store large amounts of data that can be used for both *real-time* and historical data analysis. Because of this capability, the Historian has become a vital source of process information and provides the means to analyze processes and system performance. The important aspect of any Historian is the ability to allow for the data it contains to be distributed to the appropriate staff, making it useful for information gathering. This need for information storage, retrieval, and sharing also makes this server susceptible to compromise.

The Historian can support many types of interfaces, flat files, Web servers, direct database connections, such as Structured Query Language (SQL) and more modern interfaces, such as message oriented middleware, where the communication is based on a publish/subscribe method. This method is popular because it is easy to add additional subscribers of the data without the need to reconfigure the server. All of these *ease-of-use* implementations are very attractive for administrators who are responsible for installing and maintaining these data repositories.

D.6.1 Remote Access Servers

Another important aspect of secure communications is the need to enforce access level or need-to-know authority. Access control can be implemented on individual workstations and servers or as a network level implementation, (such as an RBAC service, which provides a system-level means of translating a *user role* to application permission). If there is a need for remote access to the shared Information Server, such as an ICCP server or possibly an Historian, then there are common applications available to provide a means of enforcing a remote access policy.

Two popular applications are the terminal access controller access-control system (TACACS+) (a Cisco base product), and Remote Authentication Dial In User Service (RADIUS). Both of these applications supply authentication, authorization, and accounting protocols to protect access to services on the hosted network.

TACACS+, a Cisco, Inc. proprietary implementation, is a client/server protocol. The client takes the form of a network access server, which sends requests to and receives responses from the server. The server or servers supply the authentication, authorization, and accounting services.

RADIUS, which is described in IETF RFC 2865[D.6-1] is another form of access control and can be enforced for remote access security and provide authentication and authorization of those allowed to gain access to the LAN. Simple authorization methods use a database of username and passwords on the terminal server or access server. More advanced authorization systems use methods, such as a centralized token card systems and Kerberos.

Reference:

D.6-1 C. Rigney et al. IETF standard 2865, obsoletes 2138. *Remote Authentication Dial-In User Service (RADIUS).* The Internet Society, June 2000.

D.6.2 Security Observations

An important aspect of securing shared server transactions lies outside any direct association with the application or connection setup processes, which are the disabling of unnecessary services or ports on the server. This process is sometimes referred to as *hardening* a server. Along with file permissions and password policies, this process provides a more secure server and constitutes a good security practice.

Every service that is running on a server adds to the size of the attack surface for an adversary. Reducing the number of unnecessary services increases the protection of the server. The first step in hardening the server is to determine all the essential services. Services not considered essential can often be disabled without any negative effect on the operation of the server. There may also be services on a system that support many media capable protocols and participate in remote access services that are not needed in a utility environment. The services that can be disabled will depend on what applications and functions the server must support.

Prior to turning off any ports or services, it is important to note there may be some dependencies that at first seem isolated from a primary service, but which will prevent the primary service from running without the supporting service. Some OS companies, such as Microsoft, have posted guidelines on determining which services are considered vital for the operating environment and those that can be disabled without impacting operations. This may also help the administrator identify related service dependencies. When disabling services, it is important to proceed in a linear fashion—disabling a single service at a time, reviewing the action, and recording any unexpected events. (See Appendix D, section D.13.1, Host Access Control, for additional security observations that pertain to shared servers.)

Unfortunately, many vendors who use some of the more popular protocols for database interaction—such as Wonderware Industrial SQLserver, OSIsoft PI, Westinghouse eDB process Historian, or Microsoft .NET, Active server pages—are all subject to database software vulnerabilities. Some vulnerabilities, such as SQL injection, take advantage of the protocol interexchange between *data retrieval* and database command sequences. Many types of attacks

are enabled because of the lack of authentication of end users that are allowed access to the Historian or PI server.

If appropriate authentication is implemented for users of a shared server using RADIUS or TACACS+ remote access server, the *external threat* can be thwarted from accessing and manipulated data on the shared data server. If the external threat is able to review the external data streams between a valid remote user and the data server, they may be able to identify the IP address and application port of the front end firewall that is facilitating the authentication of the external user and the internal authentication server (TACACS or RADIUS) exchanges. Although the user ID and password should be encrypted, the address information to interact with the firewall can be used to inject many server interaction requests in an attempt to create a DoS against the remote server. This can prevent legitimate users from reaching the server for valid interactions. One way to combat this attack is to provide the firewall with additional user information of valid remote clients in order to "weed out" connection request to the data server originating from invalid address ranges.

D.6.2.1 Additional Observations

The best way to reduce the threat from an *unprivileged insider* would be to implement an appropriate access control to the shared data server, the remote access server, and the firewall. These need-to-know access restrictions can restrict the insider to limited privileges on any of these devices. It is also important to implement proper secure network management protocols when interacting with the servers or the firewall. Applications, such as TFTP and Telnet, should not be used for network management interaction. These protocols pass the user ID and password over the network in clear text. An insider that has physical access to the network may be able to "sniff" these credentials over the network and obtain upper level privileges.

The *privileged insider* would have a larger administrative role within the facility, but can possibly be limited in the number of systems that can be accessed or to specific locations within the plant. Implementing a user logging function to monitor user access, which is sent to a central repository not accessible to the administrator, can help track users signing onto the system. Combining both physical protection mechanisms for personnel access control along with restricting the number of systems that can be accessed can provide some level of protection against the privileged insider threat.

Threats from *developer* or vendor-based sources are primarily associated with default passwords and user accounts that have been implemented as part of the vendor pre-installation configuration. Also the available services that are allowed to run on the initial system may have vulnerabilities associated with their operations. These potential vulnerabilities can be mitigated within a properly established security policy. Removing user accounts, default passwords and *hardening* the sever to run only essential services should be part of a security policy established at each NPP facility.

Some standards and guides for remote access server security are listed below:

RFC 2865. *Remote Authentication Dial In User Service (RADIUS)*, section 8, Security Considerations, describes some standard server configurations that support the topic discussed

previously in this section. The discussion includes the secure means of identifying a user by a single authentication method, the proper storage and protection of passwords, and the importance of proper access control to prevent unprivileged server access.

NIST SP 800-46 revision 1. *Guide to Enterprise Telework and Remote Access Security*, section 3, Remote Access Server Security, reiterates the importance of proper remote access server security by ensuring they are kept properly patched, are operated using an organization defined security configuration, and only managed from trusted hosts by authorized administrators. It also discusses the remote access server placement to include server performance, authentication, etc.

ISO/IEC 18028-4:205, provides guidance for securely using remote access. It discusses the authentication issues related to remote access and provides support when setting up remote access securely for servers.

NIST SP 800-123. *Guide to General Server Security*, provides information on planning, implementing, and maintain security of a server

National Security Agency (NSA). *Guidance on Operation System Configurations*, provides configuration guidance for a variety of OSs for the purpose of creating a secure baseline configuration. This information is available on the Web at <http://www.nsa.gov/ia/guidance/security_configuration_guides/operating_systems.shtml>

D.7 Wireless

The use of wireless communications in electric utility companies has traditionally been associated with the connection of distant substations—through radio, microwave, or sometimes satellite—to provide distant reach back. With the introduction of substation automation, primarily the adoption of IEC 61850 [D.7-1], the use of wireless applications is expanding.

Some applications that are arising from the implementation of IEC 61850 are the introduction of distributed sensors using wireless integrated networks to instrument substations. These sensors, referred to as intelligent electronic devices (IEDs), are changing the direction of traditional monitor and control techniques along with the approaches to security. Substation automation is going beyond standard supervisory control to provide added capability and information that further improves operations and maintenance. Applications can include remote access IEDs, relay configuration control, waveform and voltage monitoring, event acknowledgement, diagnostic and troubleshooting information, metering, switching, and video for remote security applications.

The boundary protection mechanisms that are associated with *wired* technologies do not apply to the *wireless* application implementations. Therefore, it is important to understand the wireless protocols being used within the electric utility environment and, more importantly, how to properly secure them. Two primary protocols being deployed for utility use today are the IEEE Standard 802.11 [D.7-2] suite, which includes 802.11a [D.7-3], 802.11b [D.7-4], 802.11g [D.7-5], and 802.15.4 [D.7-6], referred to as ZigBee.

The 802.11 suite of protocols is found in the Industrial, Scientific, and Medical (IS&M) frequency spectrum. The three primary IS&M frequencies bands are 868 MHz, 915 MHz, and 2.4 GHz. the Federal Communications Commission (FCC) has reserved these frequency bands for unlicensed, low-power radio frequency (RF) operation as defined by FCC Part 15. Many of the 802.11 suite of devices can operate in an *ad-hoc* mode, where each device can independently communicate with other peers in its transmission domain. But in many utility implementations, which require "reach back" into a utility network, they are configured to operate in *infrastructure* mode with the introduction of a WAP.

The ZigBee wireless technology is associated with IEEE Std802.15.4, the low rate wireless personal area networks (LR-WPAN) standard. ZigBee is popular because of low-power and low-cost and its ability to create peer-to-peer wireless multi-hop networks. It is self-organizing and supports multi-hop routed network topologies. It also can support the AES. Like the 802.11 protocol suite, ZigBee also is assigned to the IS&M frequency spectrum and can use the 802.11 protocol scheme to aggregate ZigBee network traffic and provide a *bridge* between its ZigBee created network. The protocol can be used to aggregate ZigBee network traffic and reach-back to the wired network by the means of a WAP.

Wireless Access Point

Wireless access points are dedicated hardware devices with built-in network adapters, which are designed to *bridge* the wireless and wired networks. Wireless access points by default send out *beacon* frames to announce themselves so clients can find them and initiate a connection. The access point service set identifier is sent out in the clear; this makes it easy for unauthorized clients to attempt access to the network. To prevent unauthorized access to the network, an authorization process is needed, such as the one described in the IEEE Std 802.1X [D.7-7] wireless authentication standard.

802.1X Network Authentication

802.1X is an authentication method that requires a wireless client to authenticate itself to the wireless local area network (WLAN) access point prior to gaining network access. The authentication protocol requires the client to send its identity to the access point, which is then forwarded to an authentication server, such as a RADIUS server. Using an algorithm, the authentication server checks the identity of the client and responds to the access point with either an *accept* or *reject* message. The 802.1X network authentication protocol is part of the 802.11g wireless standard. Along with 802.1X there are other security layers that can be added to protect the network from compromise. For example as seen in Figure D-13, an IDS can be added to the wired side of the access point to monitor data streams for malicious activity originating from the wireless domain. Also a firewall can be configured to filter incoming and outgoing network connections originating and terminating into the wireless domain. Along with a firewall, a wireless IDS can be deployed to monitor the wireless interactions between the wireless nodes. The wireless IDS can be incorporated within the access point or can be an independent device. This can help identify unauthorized rogue wireless devices or attacks being conducted by an adversary with malicious intent. This *defense-in-depth* approach allows the network administrators to implement the needed layers of security that will best fit each design circumstance.

Figure D-13 shows an implementation of some of the previously mentioned security elements, on an extension to the plant data network. It shows several ZigBee networks and an 802.11 implementation that have access points attached to a firewall. Each access point would provide MAC source address filtering while the firewall would implement Layer 3 IP address and application filtering. Alternatively, each participating end node may require an authentication look-up, which would be directed toward to the RADIUS server. The RADIUS server may have each node's public authentication token, which the certificate authority server has signed. An IDS is also located on the on the wired network, along with IDS sensors to monitor and report suspicious activities or attacks.

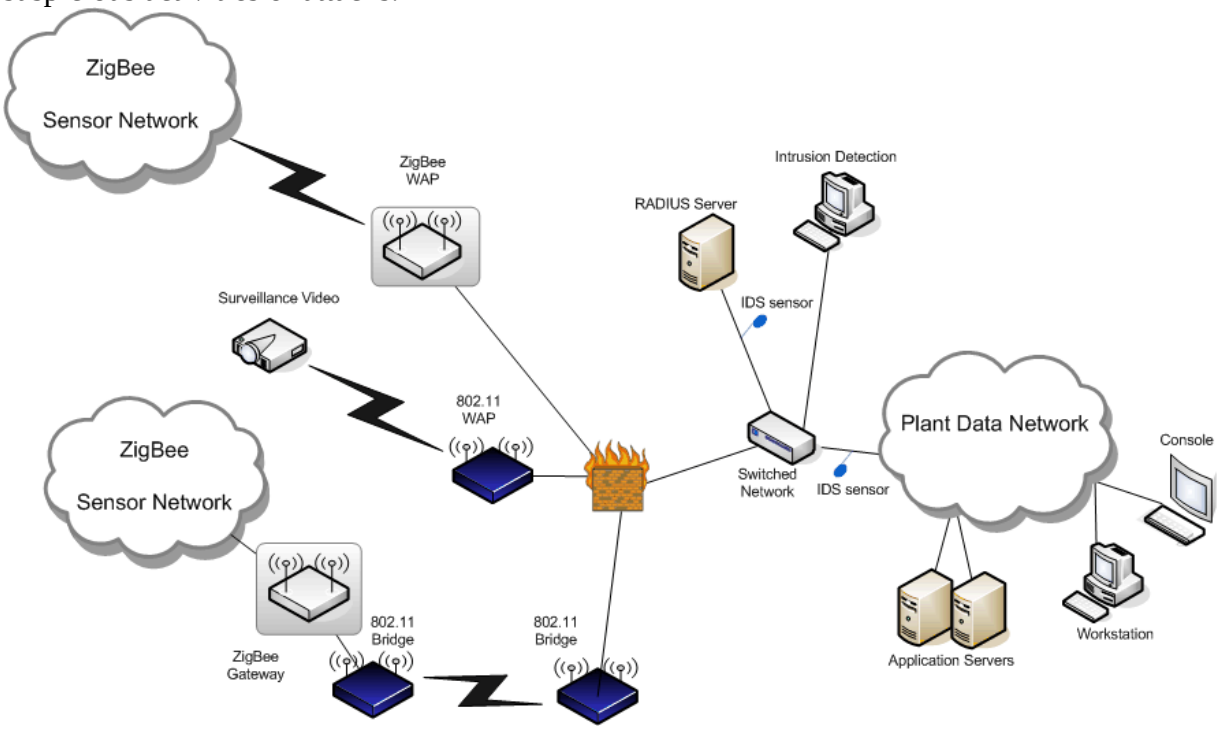

Figure D-13. Wireless Network Architecture

Wireless Installation and Performance

Prior to the integration of a wireless network at a facility, some necessary preplanning is required. The first step requires a site survey. The wireless propagation pattern should be identified to verify the required coverage area. Reducing the wireless field strength and direction to only what is necessary will also reduce the wireless footprint that an adversary can identify. The site survey should also include potential wireless interference areas that need to be compensated to provide the desired coverage.

The ISA working group SP100 [D.7-8] was formed to address the performance and cost needs of a wide range of industrial applications, such as monitoring, logging, and alerting. Since the industrial environment may include high power interference sources, the standard will also address network robustness. The working group will address coexistence with other wireless devices anticipated in the industrial work space, such as IEEE Standards 802.11X, 802.15X,

802.16X: cell phones, radio frequency ID (RFID), ISA-SP100.14, and others. The information this working group provides can help in utility system design and help assess the reliability and performance requirements of an application prior to its deployment in the field. Included in the performance of the wireless medium is the sense of quality-of-service (QoS) to provide a guarantee of service. In near real-time monitoring systems, the predicted response times are important parameters to include because of the contention-based approach to the wireless transmission medium. Priority queuing schemes at WAPs can help provide a more deterministic probability of receiving messages within a predetermined time frame.

Additional wireless network-related information may be found in the following documents:

NUREG/CR-6882. Assessment of Wireless Technologies and Their Application at Nuclear Facilities. July 2006.

NUREG/CR-6939. Coexistence Assessment of Industrial Wireless Protocols in the Nuclear Facility Environment. July 2007.

Establishing Wireless Robust Security Networks, A Guide to IEEE 802.11i, NIST Special Publication 800-97.

D.7.1 Security Observations

Another external path into the data plant network is associated with wireless connectivity. More utilities are installing wireless networks to facilitate transmission of sensor and instrumentation information associated with elements of electricity generation located in substations. This status information is fed back to the utility monitoring networks to assist operators in assessing operational parameters of their process assets.

With the introduction of the wireless element into the NPP generation facility networks, it is important to protect this new environment from adversary node "spoofing," data manipulation, data insertion, and communication disruption. The *wired* data network must be protected from intrusions that are attempted from the wireless network elements. To defend against these types of insertions the wireless network will need to have a secure layered protection as proposed through the previous section of this report.

The most popular widespread wireless protocol today is 802.11. The primary reasons are its license-free use of spectrum, efficient, and high speed data channel, and its inexpensive hardware interface. Unfortunately, this same widespread deployment makes 802.11-based networks an attractive target for potential adversaries. Earlier 802.11 security protocols, such as Wired Equivalent Privacy (WEP), had vulnerabilities in its encryption mechanisms. But today, these vulnerabilities have been primarily overcome with the replacement of WEP with stronger security protocols, such as Wireless Access Point, 802.11i [D.7-9] and 802.1X. The primary vulnerabilities that exist today are associated with the physical spectrum used for wireless wave propagation and the 802.11 MAC layer.

Spectral Interference

The wireless medium is subject to electromagnetic interference. This interference can be the result of industrial factors, such generators, motors, and other radio spectrum interference. If enough electromagnetic energy is coupled into a wireless receiver, it can affect the availability and/or the reliability of the wireless node. A site survey should take into account the placement location of access points to reduce the impact of environmental interference. The interference can also be caused by an adversary purposely directing spectral transmissions at the access point devices in the attempt to disrupt "jam" communications.

Spread Spectrum

Because the 802.11 devices were designed to operate in the unlicensed inventory-management-and-supply band in a local area. They were designed to reduce co-located transmissions from interfering with other communicating nodes. This design provides a means to spread the transmitted signal over a wide *spectrum* of radio frequencies, minimizing the impact of narrowband interference. There are currently two different spread spectrum techniques, both using a coded pattern to send and receive information. They are f*requency hopping spread spectrum (FHSS)* and d*irect sequence spread spectrum (DSSS).*

Frequency Hopping Spread Spectrum

FHSS works by separating a narrow band communication stream into a wide spectrum of radio frequencies. A defined, but random-appearing, pattern of non-sequential bands is used with successive parts being transmitted over the next frequency band in the pattern. The distant receiver is configured to receive the signals in the same pattern. The receiver then reassembles the pieces into the original signal.

Frequency hopping reduces the electrical noise that is present to reduce co-channel interference caused by other co-located radio communications operating in narrow bands of the spectrum. This also increases its immunity to purposed "jamming" attempts by an adversary. For an adversary to purposely disrupt communications, s/he must either direct a large transmission energy source over the entire spectrum of frequencies used by the frequency hopper, or monitor and decipher the frequency hopping code in order to match the "jamming" signal to that of the transmitting FHSS node.

Direct Sequence Spread Spectrum

DSSS also spreads information over a larger spectrum, but uses a different technique than FHSS. Instead of splitting a data communication stream signal into different frequency segments, DSSS encodes each data bit into a longer bit string, called a *chip*. A chip can vary in size; usually between 11 to 20 bits are used for the chip, depending on the application. The chip is then used to modulate the signal that the radio transmitter generates, spreading the signal out over a wide band of frequencies. The receiver uses the same chipping code to receive the unique signature across the frequency spectrum. It then decodes the signal back to the original data.

DSSS has the same advantages of FHSS, except DSSS can be much more difficult for an adversary to determine the sequencing scheme as the size of the chipping bit increases. This prevents an adversary from identifying the spectrum pattern to sequence an attack against a transmission source.

802.11 MAC Layer
The 802.11 MAC layer incorporates functionality that allows wireless nodes to discover networks, request-to-join-and-leave networks, coordinate access to the wireless medium, and send data. These protocol features are encompassed in the three types of MAC frames:

- Management frame
- Control frame
- Data frame

The vulnerabilities with the MAC layer are most often associated with node identity and media access control. Node identity vulnerabilities occur because of the implicit trust 802.11 has in the source address of a wireless node. As with standard Ethernet protocol, the wireless nodes are assigned a unique MAC address. This MAC address can be used to validate an end node and, thus, an adversary may attempt to "spoof" the address. The following are some attacks that can be leverage against the 802.11 system based on elements of the protocol design:

De-Authentication Attack
The de-authentication attack takes advantage of how an 802.11 client communicates with the access point node. If proper authentication security is in place, a node authenticates itself to the access point prior to being allowed to communicate further. Part of the authentication process is the ability to send out a *de-authentication* message that disconnects the active authenticated session. This message is not part of the protected authentication process. This can allow an adversary to "spoof" a message on the part of either the client or the access point. In response, the access point or client will exit the authenticated state and will reject all further "unauthenticated" packets until authentication is reestablished. Figure D-14 shows this interaction.

Disassociation Attack
Another vulnerability that can be exploited against 802.11 is the disassociation attack. This is similar to the de-authentication attack. The 802.11 standard allows wireless nodes to associate themselves to multiple access points. This normally happens when a client node is roaming and detects a stronger access point signal and sends out an association request to the new access point, while currently attached to another access point. This association message is needed to allow the client and access point to agree on which access point shall have responsibility for forwarding packets to and from the wired network for the client. As with de-authentication frame previously discussed, the 802.11 provides a disassociation message that is not authenticated and, thus, an adversary can exploit this interaction in the same manner as the de-authentication attack.

Power Save Mode Attack
To conserve power, the 802.11 protocol allows clients to enter a sleep mode during which they are unable to communicate. Prior to entering the sleep mode the client sends an announcement to the access point so that the access point can buffer any inbound traffic destined for the sleep mode node. The sleep mode is configured to allow the client to awake and poll the access point for any buffered traffic. If there are buffered data, the data are sent to the client for processing

and the contents of the access point buffer are deleted. By "spoofing" the client's polling message, an adversary can cause the access point to discard client packets while it is in sleep mode.

Another exploit can potentially fool the client node into thinking there are no buffered packets waiting at the access point even though they may exist. The access point sends out periodic broadcast packets called Traffic Indication Maps (TIMs) to indicate which nodes have buffered packets. If the TIM message is "spoofed," an adversary can advertise that there are no buffered packets awaiting the client, and the client will return to its sleep mode. Another exploit associated with the sleep mode is the reliance of the clients on synchronization information that is also periodically sent out by the access point. This is necessary so the clients know when to awake from the sleep mode and poll the access point. Both the TIM and the timestamp broadcast are sent out unauthenticated. By "spoofing" these management frames, an adversary can cause the client to fall out-of-sync with the access point and not awake at appropriate times.

Mitigation

All the vulnerabilities described previously in this section can be resolved with appropriate authentication of all messages. But it seems unlikely that this authentication capability will emerge soon although the recent release (September 2009) of the 802.11w, *Protected Management Frames*, has proposed a means to secure the de-authentication and de-association frames. But with the large installed user base of legacy 802.11 devices, and the continual market growth of this product, it may be some time before these new proposals are uniformly deployed.

MAC Carrier Sense Attack

The 802.11 Carrier Sense Multiple Access Collision Avoidance (CSMA/CA) technique is used for *ad hoc* wireless medium access. This technique uses a control frame with a field that indicates—by a requesting transmitting node—the number of microseconds for which the channel is reserved for transmission over the wireless medium. Only when this value equals 0 can another wireless node send out a request to transmit. This value is used to program the network allocation vector (NAV) on each node. The NAV mechanism reduces and mitigates collisions from wireless nodes that may be out of the initial transmission field of the originating source node. This feature is used by the request-to-send (RTS) and clear-to-send (CTS) "handshake" signals for channel transmission to synchronize access to the transmission media and reduce the collision potential of a hidden node.

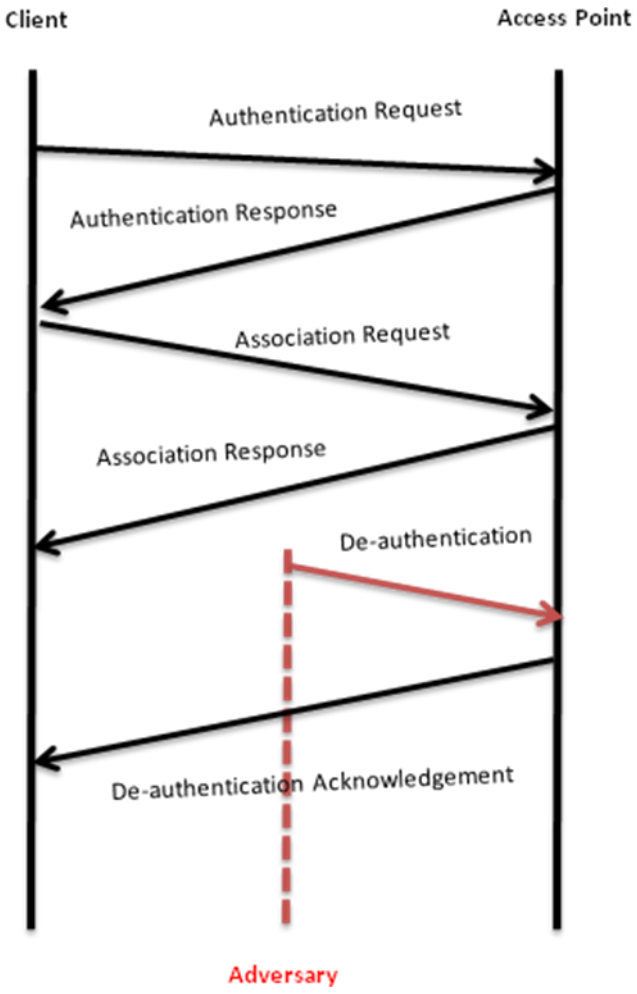

Figure D-14. De-Authentication Attack

During the "handshake" sequence, the sending node sends out an RTS frame that includes a NAV duration time encompassing the RTS/CTS sequence; this includes the CTS frame, the data frame, and the subsequent acknowledgment frame. The receiving node replies to the RTS with a CTS message, containing a new duration field, updated within the NAV to allow for the time elapsed during the sequence. After the CTS is sent, every wireless node in range of either the sending or receiving node will have updated their NAV and will defer all transmissions for the duration of the future transaction.

By assigning a large NAV duration field time allocation during the CSMA/CA carrier sense function, an adversary can prevent other wireless clients from accessing the wireless channel. This attack can be amplified by the fact that a node receiving an RTS must send out a CTS so other nodes—which may be out of range of the original RTS—can see the transmission request. Thus, an adversary sending out RTS signals with a large duration field will initiate a CTS response from any listening nodes in range. This allows the adversary to use other legitimate nodes to propagate the attack on a larger wireless range, while using extremely low power or directional antennae to reduce the probability of being located.

One approach to mitigate the effects of a MAC carrier sense attack is to place a limit on the duration values accepted by nodes. Any packet containing a larger duration value can be reduced to the maximum value allowable. Designing a wireless architecture that takes into account signal propagation and signal strength with respect to an external adversary's potential location can also limit the adversary's ability to monitor the wireless network. Such controls might use directional antennas to craft transmission envelopes, thus, reducing signal exposure. Performing wireless spectrum analysis can also help characterize the propagation reach of authorized nodes and discover operating unauthorized nodes.

Dynamic Host Configuration Protocol

Use of wireless DHCP services, using authorized IP addresses for legitimate wireless clients, provides another way to protect the wireless medium. If an adversary does not want to create an address collision, s/he must select an IP address that is not currently in use. If s/he selects an address not authorized within the DHCP list, the chances of discovering the intruder are increased, since the intruder would have to guess as to what IP address to use.

Security Measures

The following is a list of other security measures that can protect both the wireless and the wired networks [D.7-11]:

- *Include a wireless security policy to guide the implementation, management, and operation of wireless networks*. This should be part of the over-all policy that the utility company develops. The security policy is the necessary element to develop a comprehensive and effective security architecture.

- *Consider using a MAC address (Layer 2) filter at the WAP*. This allows the WAP to determine which wireless devices it is authorized to engage in communications. The WAP maintains a list of hardware addresses of all devices allowed to communicate with the WAP.

- *Consider using an IP address (Layer 3) filter at the firewall*. This approach is similar to that of the MAC filter except this filter is located at the firewall, which can monitor IP (Layer 3) addresses. The firewall may also provide filtering at high layers of the communication stack for even greater filter granularity.

- *Utilize the Layer-2 security mechanisms supported in the IEEE 802.15.4 (within the ZigBee peer network).* Outside of the wired network within the ZigBee peer network, there are services that can be enabled to protect participating nodes, such as access control, data encryption, frame integrity, and sequential freshness.

- *Implement secure network access control using 802.1 x Extensible Authentication Protocol (EAP).* This governs the EAP encapsulation process between wireless clients, WAPs, and an authentication server (RADIUS).

- *Implement secure node authentication using 802.11i EAP Transport Layer Security*. This technique utilizes digital certificates and a Certificate Authority (CA) for individual end node wireless authentication.

D.7.1.1 Additional Observations

The *external threat* can impose a valid potential for disruption and compromise in the wireless environment. The external threat does not necessarily need physical access to a facility to disrupt or compromise elements of a facility's wireless infrastructure. All the previous exploit descriptions could be implemented by the external threat if the wireless signal can be detected. But with proper wireless design that accounts for signal power and propagation—along with layered security implementations for wireless access control, management, and data transport—the external adversary's job becomes much more difficult.

The *unprivileged insider* threat is a concern because a wireless device can be planted near the physical location of the wireless network. The physical distance afforded by externally protecting the wireless network cannot be used as a layer of defense for the unprivileged insider. All previously described exploits are available to this threat along with the insertion of a rogue access point. A rogue access point can subvert data traffic flow away from the authorized access point and create a contention for wireless node access. Implementing an IPS on the interface boundary between the wired and wireless network can help detect divergent data flows ; also a wireless spectrum analyzer can detect the operating channels of wireless devices.

The *privileged insider* would have a larger administrative role within the facility. This may include management of the wireless network architecture. It may be possible to limit the number of systems that a single administrator can access or the locations that are allowed access within the plant. Combining physical protection mechanisms for personnel access control, along with restricting the number of systems that can be accessed, provides some protection against this type of threat. The privileged insider is the most difficult threat to overcome.

Threats from *developer*- or vendor-based sources associated with the deployment of wireless devices would be associated with the access point node in the form of rogue code. This node has the most capability in terms of processing speeds and memory allocation. Designing a network that includes proper monitoring techniques to detect unusual or inappropriate actions from wireless network devices can help detect any malicious vendor content.

Some standards and guides for wireless implementation security are listed below:

IEEE 802.11i provides a technical description on the security architecture of the 802.11 security standard. It provides details on how the integration of AES provides confidentiality, integrity, and origin authentication.

NIST 800-97. *Establishing Wireless Robust Security Networks*, provides an overview of the 802.11 wireless standard and some wireless LAN security concerns. Section 8.0 provides some WLAN security best practices for proper planning of wireless integration.

NIST SP 800-48r1. *Guide to Securing Legacy IEEE 802.11 Wireless Networks*, contains an overview of the 802.11 standard security and some means of protecting wireless information exchange.

ISO/IEC 18028-4:2005. *Information Technology-Security Techniques-IT Network Security*, Part 4 Annex F, provides a WLAN security checklist

References:

D.7-1 IEC 61850. Communication Networks and Systems in Substations, ed. 1.0, 2009.

D.7-2 IEEE Std 802.11. Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2007.

D.7-3 IEEE Std 802.11a. *Supplement to IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems, Local and Metropolitan Area Networks: Specific Requirements*. Part 11: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer in the 5 GHz Band*, 1999.

D.7-4 IEEE Std 802.11b. *Supplement to IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems, Local and Metropolitan Area Networks: Specific Requirements.* Part 11: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band*, 2000.

D.7-5 IEEE Std 802.11g, *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems, Local and Metropolitan Area Networks: Specific Requirements*. Part 11: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications* (Amendment to IEEE Std 802.11, 1999edn. (Reaff 2003) as amended by IEEE Stds *802.11a-1999, 802.11b-1999, 802.11b-1999/Cor 1-2001, and 802.11d-2001)*, 2003.

D.7-6 IEEE Std 802.15.4. Wireless Medium access Control (MAC) and Physical Layer (PHY) Specifications for Low-rate Wireless Personal Area Networks (LR-WPANs), 2003.

D.7-7 IEEE Std 802.1X. *IEEE Standard for Local and Metropolitan Area Networks Port-Based Network Access Control*, (Revision of IEEE Std *802.1X-2001)*, 2004.

D.7-8 ISA—SP100. Wireless Systems for Automation Standards Committee, http://www.isa.org, May 2006.

D.7-9 IEEE Std 802.11i. *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems, Local and Metropolitan Area Networks: Specific Requirements*. Part 11: *Wireless LAN Medium Access Control (MAC) and*

Physical Layer (PHY) Specifications Amendment 6: Medium Access Control (MAC) Security Enhancements, 2004.

D.7-10 IEEE Std 802.11w. *IEEE Standard for Information technology—Telecommunications and Information Exchange between Systems, Local and Metropolitan Area Networks: Specific Requirements.* Part 11: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Protected Management Frames* (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, and IEEE Std 802.11y-2008,) 2009.

D.7-11 Faria, Daniel B., and David R. Cheriton. "DoS and Authentication in Wireless Public Access Networks." In *Proceedings of the First ACM Workshop on Wireless Security (WiSe'02)*, September 2002.

D.8 Landline Modem Access

Modems have always been a part of utility infrastructure. They are normally unsophisticated devices that have limited security and many times are overlooked in cyber security plans. Field engineers use modems for engineering support to remotely access field devices, such as RTUs and protective relays located at substations to allow remote configuration and status reporting. Equipment vendors use modems to reach field devices for maintenance or upgrade activities.

Remote Access

Modems can be connected in two primary ways: a dedicated line configuration that provides a preconfigured circuit switch connecting through the utility telecommunication network, or through the public switched telephone network (PSTN) via a dial-up connection to the modem telephone number. The dial-up modem connection through the PSTN is, thus, exposed to adversary compromise. Because it is connected to the PSTN, its phone number can potentially be reached from anywhere in the world, making it more vulnerable to attack. Figure D-15 shows the external-world connectivity to assets at generation/transmission substations. Landline modem access continues to be used both by 1) the business staff associated with the business IT network for remote access to internal network services, and 2) the utility operations staff to reach more remote energy production asset devices for status of generation assets.

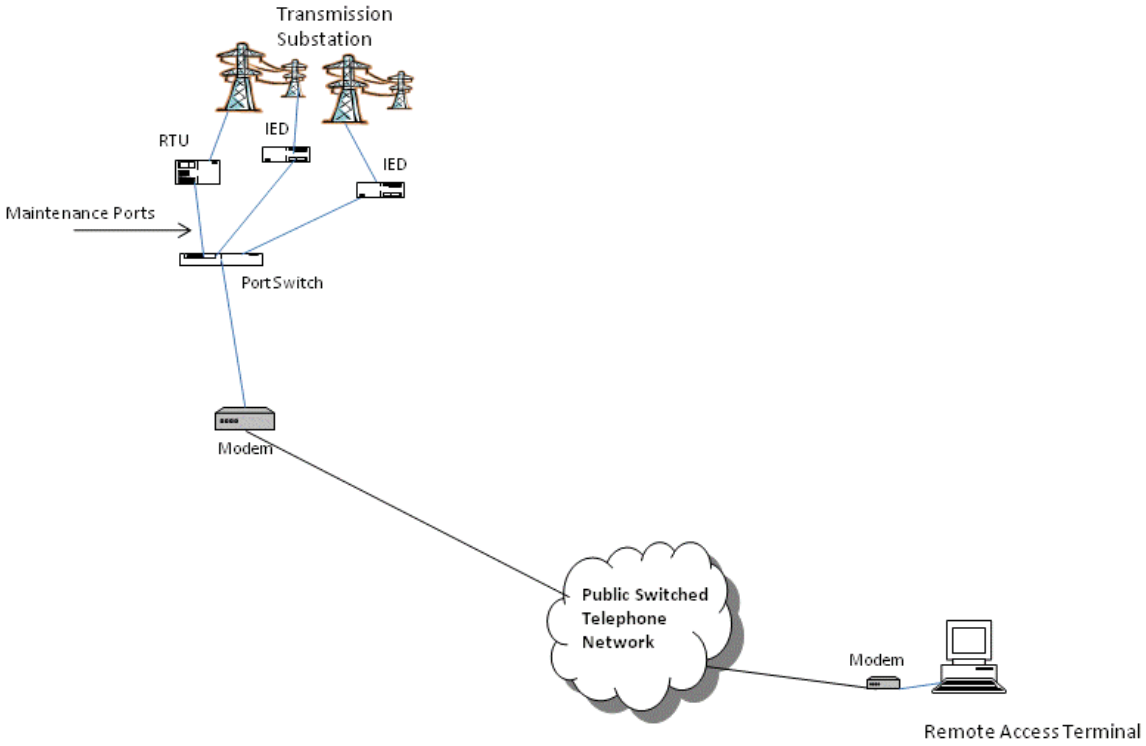

Figure D-15. Landline Modem Remote Access

D.8.1 Security Observations

It is important to understand all potential insertion points into a utility infrastructure. Non-secured or sometimes unauthorized modems can offer adversaries an undetected method of obtaining access to both the internal plant data networks and power plant generation assets. Proper protection mechanisms provide a strong deterrent against unauthorized access.

Rogue modems

In most organizations, firewalls and remote access servers are the main perimeter access points. However, there may be other perimeter access points that are not as obvious. The energy-utility and control-systems environment focuses on securing the Internet connection, but pays considerably less attention to modem security. Unauthorized modems, not part of the *official* communication architecture, can open a penetration into the plant data network; if the modem is also not properly secured, this offers adversaries easy and unmonitored access to both the internal data networks and generation assets.

Modem Call-Back

One of the primary attack vectors associated with modems is their auto answer capability. This allows war dialer software to easily identify modems. Any modem attached to a control system asset should not be allowed to be configured in auto-answer mode. At minimum, a form of dial-back security should be considered to reduce the exposure of control system modems. In dial-back mode, a modem is programmed to go “off-hook” briefly to address the incoming call, then hang up and call the number programmed into its memory. This means only previously authorized telephone numbers can communicate with the modem.

Caller ID
Another layer of defense that can be applied for modem protection is caller ID. This service is normally provided by the telecommunication vendor (phone company) and if available as a feature on the modem, can be used as a protection layer. Some modems can be configured to read the caller ID and compare it to a precompiled list of *allowable* remote access phone numbers. The connection to the modem can then be allowed or denied based on the authorized access list

Modem Power Supply
A non-sophisticated means of protecting a modem from being accessed remotely during specific time frames is just disconnecting its power supply. This can be done manually or by placing a timer on the power supply receptacle. The timer can be programmed according to the company's remote access time frame and automatically disconnects power to the modem when outside of that timeframe. This technique will limit the window of vulnerability associated with connecting to the modem.

PBX security
The local private branch exchange (PBX) of the utility company can also be used as a line of defense. Many local PBXs can program line availability service based on day of the week and hour of the day. Also, they can act on caller ID and log incoming calls for attribution assessment. They can determine if the incoming call was originated for its local connection or externally from the telephone carrier. The PBX can then be programmed to allow all local originating calls to the modem, but block all external calls.

User ID & Password
Every device, such as an RTU or PLC, that has been configured to allow remote access must ensure that a user ID and password-based access control feature are implemented on the device. Any default IDs and passwords should be changed and should follow a company policy for password generation and control. If supported by the device, each user should have a profile that allows only a level of access required to perform the job. This prevents all users from unlimited access to the capabilities of the device.

D.8.1.1 Additional Observations
The *external threat* can utilize publicly available war dialing software to identify modems that have not been properly protected against this type of attack. The external adversary can launch these remote access attacks from the safety of a distant analog or digital subscriber line (DSL) device port. But with proper modem configuration, as described previously, the external adversary's access can be prevented.

The *unprivileged insider* threat can be a concern because this threat can include potentially authorized contractors and product vendors. Any modem accounts associated with high value assets should not allow contractors unlimited 24-hour access to remote modems. Any type of maintenance access should be authorized on a per-issue basis. Access should be scheduled and monitored and access removed when the specific task has been completed. This overall approach to contractor access should be accounted for within a formalized site-specific security policy.

Also plant personnel who need remote access to specified modem-interfaced devices should also be tightly controlled based on time of day and job need prior to authorization. If the device being accessed through the modem can provide some sort of access level profile, it should be utilized to provide a role-based access control for each individual user.

The *privileged insider* would have a larger administrative role within the facility. This may include managing the modem network architecture. It may be possible to limit the number of systems that a single administrator can access or to limit the locations administrators are allowed to access within the plant. Combining both physical protection mechanisms for personnel access control, along with restricting the number of systems that can be accessed, can provide some protection against this type of threat. The privileged insider is the most difficult threat to overcome.

Threats from *developer*- or vendor-based sources are mostly associated with maintenance ports that have active default password accounts. A properly established security policy is the best defense against these potential vulnerabilities. Removing user accounts and default passwords should be part of a security policy established at each NPP facility.

Some standards and guidelines for remote access (modem) security are listed below:

ISO/IEC 18028-4 Part 4. *Securing Remote Access*, section 6,Types of Remote Access Connections; and section 7, Techniques of Remote Access Connections, discuss the means of remote connections that include the use of landline modems for remote access.

NERC CIP -005. *Electronic Security*, defines an electronic security perimeter that requires all modems authorized to penetrating the perimeter to identify and document the access controls to ensure authenticity of the accessing party.

D.9 Firewalls

By computer science definition, a firewall is a security device that prevents unauthorized users from gaining access to a computer network. More specifically, a firewall monitors and filters the information transfer from an external (egress) network to an internal (ingress) network. A properly configured firewall performs this operation bi-directionally. It could be used to filter network packets at the network layer, transport layer, or the application layer. Figure 2-1depicts firewalls throughout the represented architecture.

It is important not only to configure the firewall to properly filter transmitted information, but also to control how it is remotely accessed for administrative purpose and the level of privilege that the administrative user has with respect to its configuration.

There are three general classes of firewalls:

- **Packet Filtering**—Packet filtering is the most basic form of firewall implementation. With respect to the OSI model, it is associated with the *network layer* or Layer 3. Usually, packet filters are a part of a routing function because routers make decisions to route packets from

one network to another based on Layer 3 information. In its filtering process, the router checks each packet against a set of rules guided by a security policy and implemented by the network administrator. Packet filtering policies may be based upon any of the following:

- **Allowing** or **disallowing** packets based on the source and destination **address**.
- **Allowing** or **disallowing** packets based on the **transport protocol**.

Depending on the comparison results, the firewall will either drop the packet or forward the packet to its intended destination. Figure D-16 depicts this decision process.

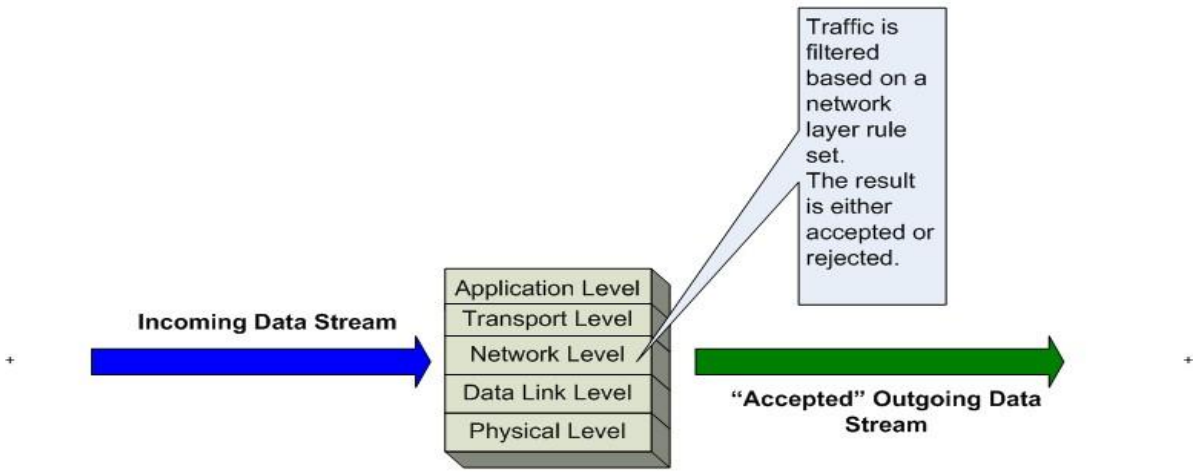

Figure D-16. Packet Filtering Firewall Process

The main advantages of a packet filtering firewall is its simplicity in configuration and implementation, relative low cost, and minimum packet processing impact, which does not significantly affect network performance. The main disadvantages of packet filtering are only the source and destination address and the transport protocol type are examined which is not effective against an adversary that can "spoof" an IP address. The source or destination port, the state or relationship to other packets in its flow, and the application details of the data are ignored.

- **Stateful Inspection**—A firewall that is performing *stateful* inspections keeps state of each connected data flow. It can filter packets at the network and transport layers and provide the same service as the packet filter firewall, but a stateful inspection can also determine the source and destination application port and note which side of the connection originated the requested connection.

This can be a better way of controlling which devices can initiate connections and which devices can only receive connections. It tracks active sessions and uses that information to determine if packets should be forwarded or blocked. Stateful inspection filtering policies may be based upon any of the following:

- **Allowing** or **disallowing** packets based on the source and destination **address**.
- **Allowing** or **disallowing** packets based on the **transport protocol**.
- **Allowing** or **disallowing** packets based on the source and destination **application ports**.
- **Allowing** or **disallowing** packets based on the initiated **connection request**.

- **Application Proxy Firewall**—Also called a *circuit level gateway*, this firewall normally combines an application interrogation role with a proxy service to validate remote end node connections before allowing data to be exchanged. This firewall goes beyond just allowing or disallowing packet streams, but can determine if the connection and data exchange are valid and that they follow pre-determined rules. While the rules are being maintained, a session between each authorized end node is permitted. This type of firewall builds upon the packet filter and stateful inspection firewalls and can validate connections based on the following:
  - Source or destination IP address
  - Source or destination port
  - Protocol
  - Application format
  - User ID
  - Password
  - Time of day

  Firewall proxies operate at the application layer of the OSI model. Both participating end nodes conduct sessions through a proxy. The proxy creates and maintains the application process on the firewall, emulating the service that would be administered on the participation end node. An application gateway is normally implemented on a separate network computer whose primary function is to provide the proxy service. An application running on an independent computer allows packet inspection at an additional layer beyond the packet filter and stateful inspection mechanisms. For example, inbound packets headed to a server set up strictly to disburse information (e.g., a File Transfer Protocol [FTP] server) can be inspected to determine if the packets contain any write commands (such as the FTP *put* command). In this way, the proxy server could allow only connections containing read commands. Also, as another example, if an internal user wants to access a Website on the Internet, the packets making up that request are processed through the HTTP proxy server before being forwarded to the Website. Packets returned from the Website, in turn, are processed through the HTTP proxy server before being forwarded back to the internal user host. Figure D-17 displays an application layer proxy firewall.

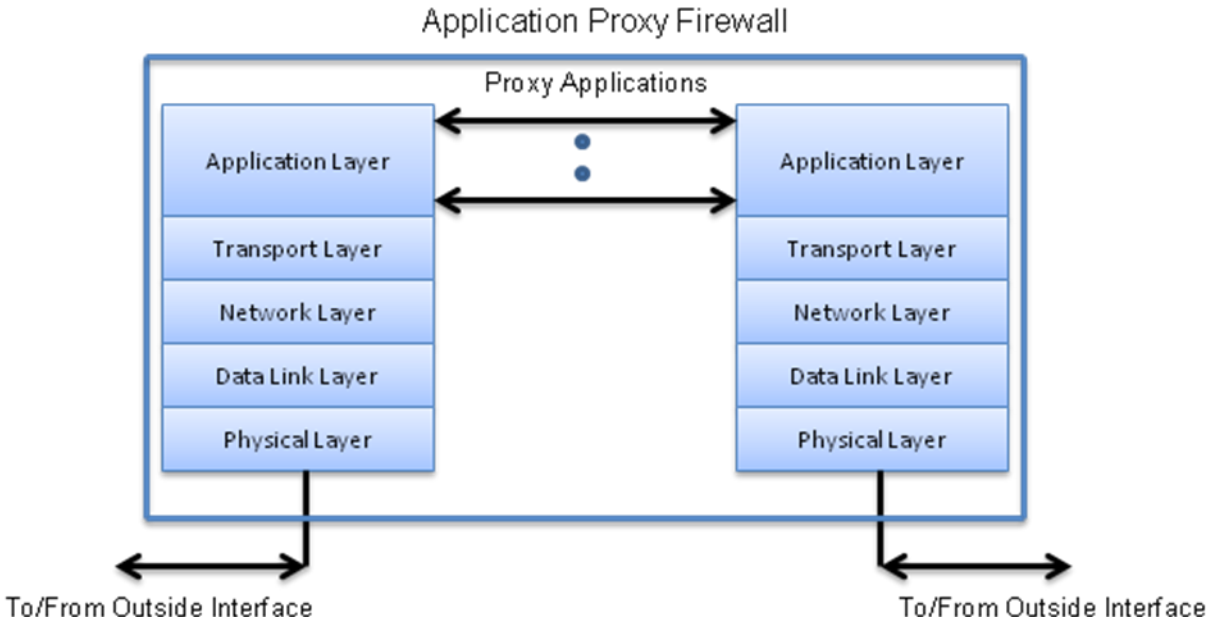

Figure D-17. Application Proxy Firewall

The application proxy firewall can perform very detailed logging of data connections and provide event notification. Application-level gateways can be considered the most secure type of firewall because traditional firewalls that do packet filtering and Stateful inspection cannot detect application attacks. Application attacks can be associated with allowed applications that pass through normal filters without a deeper inspection. Some disadvantages are that the setup may be very complex and may require a knowledge of application interaction details and, thus, can add to processing delays when inspecting data flows. As seen in Figure D-18, the application proxy firewall provides the following packet examination sequence for an outbound data flow.

1. An internal host makes a request to access a remote site. The proxy server receives the request as if it is the destination server.
2. The proxy server examines the header and data of the packet against a rule set to determine if the application is allowed and structured properly.
3. The proxy server regenerates a new request with a different source IP address than the original internal host and sends it to the destination server.
4. The destination server processes the packet that appears to the recipient as originating from the proxy server.
5. The returned packet is sent to the proxy server, which inspects the header and data contents to determine its allowance against the rule set.

6. The proxy server rebuilds the returned packet and sends it to the originating host computer. The packet appears to come directly from the external host because the source address of the external host is maintained.

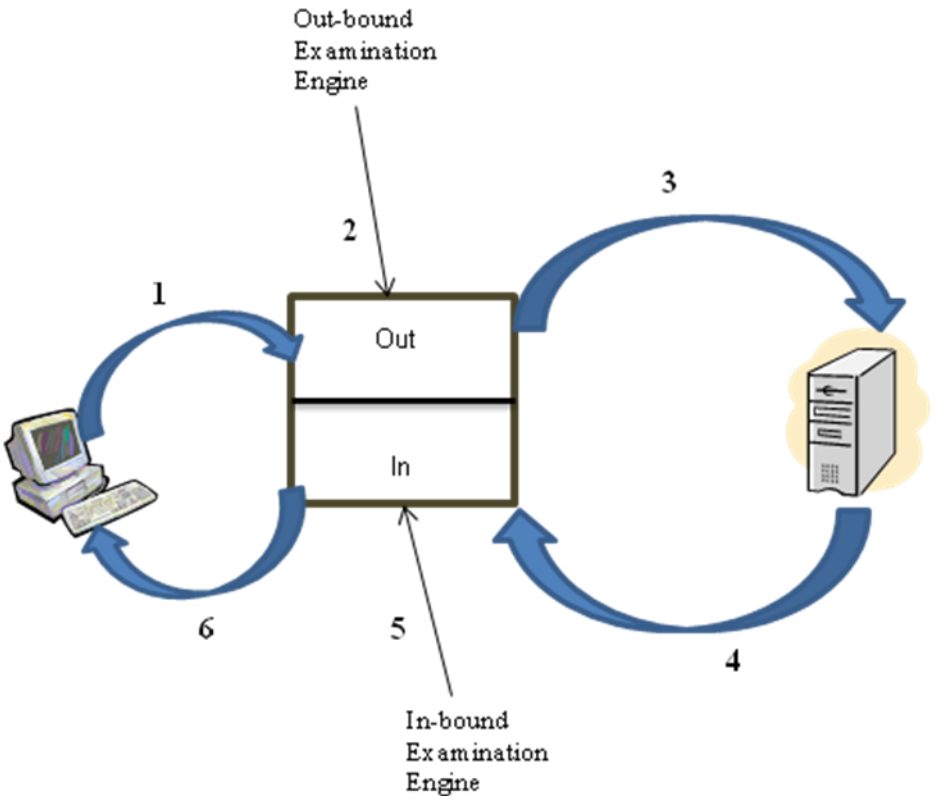

Figure D-18. Application Proxy Firewall Packet Flow Sequence

Host-Based Firewall

An additional type of firewall, not associated with monitoring aggregation points on the overall network, is the host-based firewall. Host-based firewalls are used with application servers that interact with host clients and used on individual hosts. Host-based firewalls provide an additional layer of security against network attacks. They are normally in the form of a software-based application running on each host. They can be configured more precisely for the host environment to monitor the outgoing and incoming data streams to provide access control. Many of these firewalls are bundled with the OS. Host-based firewalls can perform logging for suspicious events or when some profile of the host has been changed. Properly configured, host-based firewalls can prevent the use of unauthorized peer-to-peer applications and limit the spread of malware, thus improving the overall security of the network.

The effectiveness of any type of firewall heavily depends on a strong understanding of the organization's allowable traffic patterns. An organization must have a clear understanding of the boundary that a firewall creates and the data flow between two networks. Many modern firewalls support the creation of more than one logical boundary by creating zones. All firewall functions

should deny everything by default and only allow traffic that is explicitly accepted by the organization to traverse the network boundary. By operating the firewall in a default deny configuration, new services cannot unexpectedly be found flowing through the firewall.

Another aspect of data flow through firewalls is data encryption to hide its content. Secure protocols, such as IPSec, SSL, and SSH, can protect the application data by encrypting the payload. It is important to understand the proper endpoint placement of these types of secure protocols. Section D.3, Virtual Private Networks, describes these secure protocol applications in more detail. Also presented are some configuration choices that may allow data flows to be examined for proper network security. The firewall filter rules will be the decision point for all data flows. When data cannot be examined due to encrypted payload, the network administrator, guided by a security policy, will decide how to manage these circumstances.

Scalability. Network utilization and overall size of an organization should be considered when choosing a firewall. Modern firewalls can handle up to 10-gigabit-per-second–network flows. But not all organizations have such high bandwidth connections. Conversely, an organization may want to plan ahead. They might want to acquire a gigabit-capable firewall in anticipation of a near future gigabit-communication capability.

Management. Once firewalls are deployed, they must be managed to reflect the changing needs of the organization as well as the evolving threat from the Internet. It is important to consider the ways of managing a firewall post-deployment. Some organizations may deploy several or even tens of firewalls. In those cases, consistency and ease of management becomes a high priority. Some firewall manufacturers offer management consoles, management appliances, and software applications. These tools should be evaluated as part of the overall firewall evaluation and selection process. A firewall filter configuration is critical when determining what types of data flows are allowed into and out of the protected boundary. Whenever a filter needs to be modified. it is important to examine the change carefully to ensure the impact of the change does not create an accidental security violation. Multiple subject matter experts should always review changes to ensure fidelity. Changes should also be reviewed in reference to the security policy to ensure the changes do not violate compliance.

Availability. Good firewall implementations ensure that a firewall inspects every single data packet as it crosses a boundary. As such, a firewall is not only a security device, but a network device, as well. An incorrectly configured firewall can block traffic, thus causing a negative business impact. Also, it is important to properly evaluate firewalls under the load types and configurations that are closest to their final deployment state. Not all firewalls perform at their advertised bandwidth limits (e.g., a 100-megabit firewall may only be capable of 100-megabit throughput when application enforcement is turned off). But firewalls cannot provide effective security if they cannot inspect traffic.

A firewall is only effective at its job when it is the only boundary device between two networks. Firewalls must be deployed at key locations on the network, serving as potential choke points where all traffic can be stopped in the event of a security incident. If traffic is allowed to pass through multiple paths, multiple firewalls with identical configurations can be used to ensure

consistent protection. Firewalls should also be configured to support logging. This can help detect attacks being implemented against the firewall filters by recording data connection flows that have been rejected.

Firewalls can be deployed in internal networks as well as on network perimeters. With internal firewalls, an organization can segment networks based on business or functional groups. This enables compartmentalization and better protection of assets, such as data acquisition systems and monitoring systems. As seen in Figure D-23, the edge router has firewall packet filtering capability and can be considered the first external layer of protection from other operational areas (labeled zones) throughout the plant data network. A segmented approach creates an additional layer of protection. An adversary will have to breach the outer perimeter as well as an internal boundary to gain access to information within each zone.

Firewall Flow Interrogation Summary—

Packet Filters:

- Scan address portion of IP packets for acceptance or rejection.
- Identify transport layer protocol for acceptance or rejection.
- Can log rejections.
- Can be used as a layer with deeper inspection firewalls.

Stateful Inspection Firewalls:

- Build upon the Packet Filter firewall capability.
- Scan application ports for acceptance or rejection.
- Determine connection origination for acceptance or rejection.
- Can log rejections.

Application Proxy Filters:

- Scan the entire data part of IP packets and create more detailed log file listings.
- Rebuild packets with new source IP information.
- Provide separation so that internal and external hosts are not directly connected to each other.
- Are used together in a firewall to provide multiple layers of security.

Security Functions associated with Application Proxy Firewalls:

- Conceal internal clients.

- Block and filter content.
- Provide user authentication.
- Can redirect URLs.
- Can block Java applets or ActiveX controls.
- Can delete executable files attached to e-mail messages.
- Can filter out content based on rules that contain a variety of parameters.
- Detect intrusions by identifying application level violations.
- Provide detail logging documentation.
- Prompt users for username and password.
- Allow certain hosts to bypass the proxy.
- Restrict external network access per time-of-day and length of session.

Access Control Lists

An ACL associated with a router or firewall device is a set of rules that govern decision-making on network data flows entering and exiting a network boundary. This is the implementation part of data flow filtering. The syntax of an ACL may vary from vendor to vendor, but the essential operation is the same. The ACL can be used for address, port, and stateful inspection applications. Some ACLs are divided into standard and extended types [1].

A standard IP access control list allows the filtering of packet flows based on the specific source IP address. Figure D-19 shows a standard ACL entry.

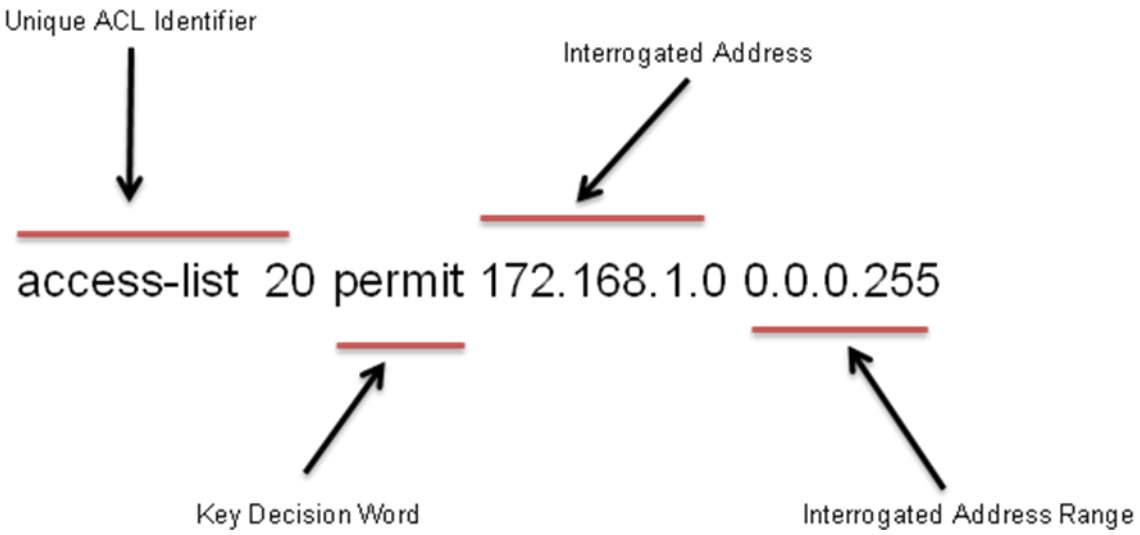

Figure D-19. A Typical Access Control Rule

The list shown in the figure allows traffic from all addresses in the range 172.168.1.0 to 172.168.1.255. An additional security feature with ACL creation in some products is that the last line entry includes an implicit deny. If the previous ACL was reviewed in its full context, it would appear as follows:

access-list 20 permit 172.168.3.0 0.0.0.255
access-list 20 deny any

Access lists are processed from top to bottom of the created list and as seen above, the last line entry would reject the processing of any additional packet flows that were not defined prior to the last *deny any* listing.

Extended IP Access Control Lists

An extended ACL allows the administrator to be much more granular in the packet flow filtering technique. Extended ACLs provide a deeper look into each packet's header content; it can specify both source and destination address, port identification—thus, the type of applications allowed. An extended ACL has the ability to keep *state* information about data flow initiation. Typically, for more secure external connections, a network administrator would allow outgoing traffic to be initiated from inside the protected network domain, but not to be initiated from a remote (external) connection. To implement this security construct, two individual ACLs would be created. For example, one may allow internal users within a plant data network to connect to an external company Website, but not allow any externally originated connections to connect to any internal domain computers. Figure D-20 below provides an example of how an *outgoing* ACL could be constructed with the internal network protected address of 172.168.1.0.

Outgoing ACL:

access-list 100 permit tcp 172.168.1.0 0.0.0.255 any eq 80

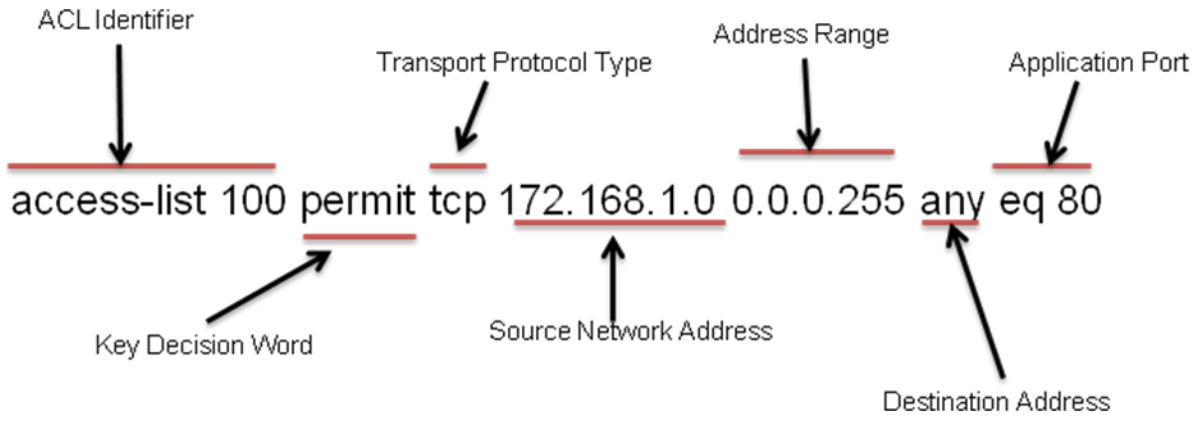

Figure D-20. Construction of an Outgoing ACL for an Internal Protected Address

The ACL shown above is assigned an ACL identifier of 100. It can be interpreted as permitting packet flow traffic originating from any address on the 172.168.1.0 network. This traffic is restricted to the use of the TCP and is allowed any destination address with the limitation that the destination port must equate to 80 (which is associated with the HTTP Web service). This ACL by itself would limit internal users to accessing external sites that were providing an HTTP Web browsing service. This single *outgoing* ACL provides no filtering on any incoming packet traffic originating from outside of the internal 172.168.1.0 network. This creates a vulnerability from any outside connection in that there is full access to all internal IP hosts and applications. Figure D-21 below provides an example of how an *incoming* ACL could be constructed to protect the internal network of 172.168.1.0.

Incoming ACL:

access-list 110 permit tcp any 172.168.1.0 0.0.0.255 established

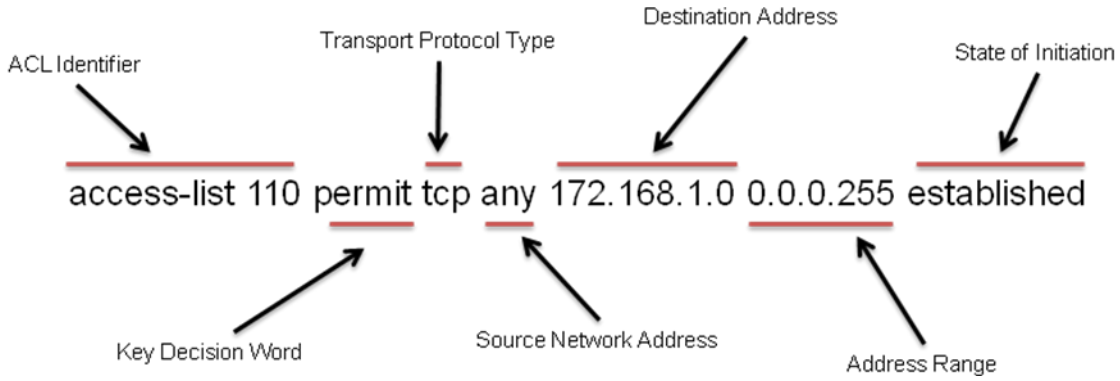

Figure D-21. Construction of an Incoming ACL for an Internal Protected Address

To provide security that prevents external connectivity to the internal network, ACL 110 can be constructed and applied to the incoming interface. It can be interpreted as allowing protocol layer TCP traffic that originated from any host address on the 172.168.1 network. The *established* argument in the ACL line invokes the ability of the router or firewall to keep state information about which end of the connection created the initiated request. This allows users to browse the Internet, but blocks all incoming traffic except the *established* connections of Websites replying to a computer on the internal network.

Firewall Network Locations

Another important aspect of all firewalls regardless of type is their placement in network architecture. The primary job of a firewall is to inspect and make decisions about incoming data flows, prior to routing the data to its final network destination; therefore, each firewall should be strategically placed within the architecture at an entry point to a segment or domain that the firewall is intended to protect. No data should be able to bypass this entry point location; all data must pass through the firewall for proper examination. This entry point will also become the exit point for out-going data flows originating from within the protected boundary. This is important so outbound flows can also be inspected.

As previously mentioned, firewall placement is not expected to be only at the perimeter of a company network architecture. It can also be used within the company to control access to other security zones, as shown in Figure 2-1. This segmented approach will make it harder for an adversary to breach the outer perimeter and then to have unrestricted access throughout the rest of the company network architecture. Just as with a perimeter placement of a firewall, an internal firewall must check all packets transported into and from the security zone. Therefore, a choke point must be available at the boundary of this zone.

Determination of the services to be allowed in and out of the security zone must also be evaluated, with the same guiding principles as those for the perimeter firewall. This placement concept can be continued down to the individual machine, where one machine might have a host firewall that examines all network traffic and allows only certain, well defined, traffic through to the OS and applications residing on that machine. Figure D-22 below shows a typical firewall boundary architecture used to interrogate data flows that are involved in external connections outside of a company internal network.

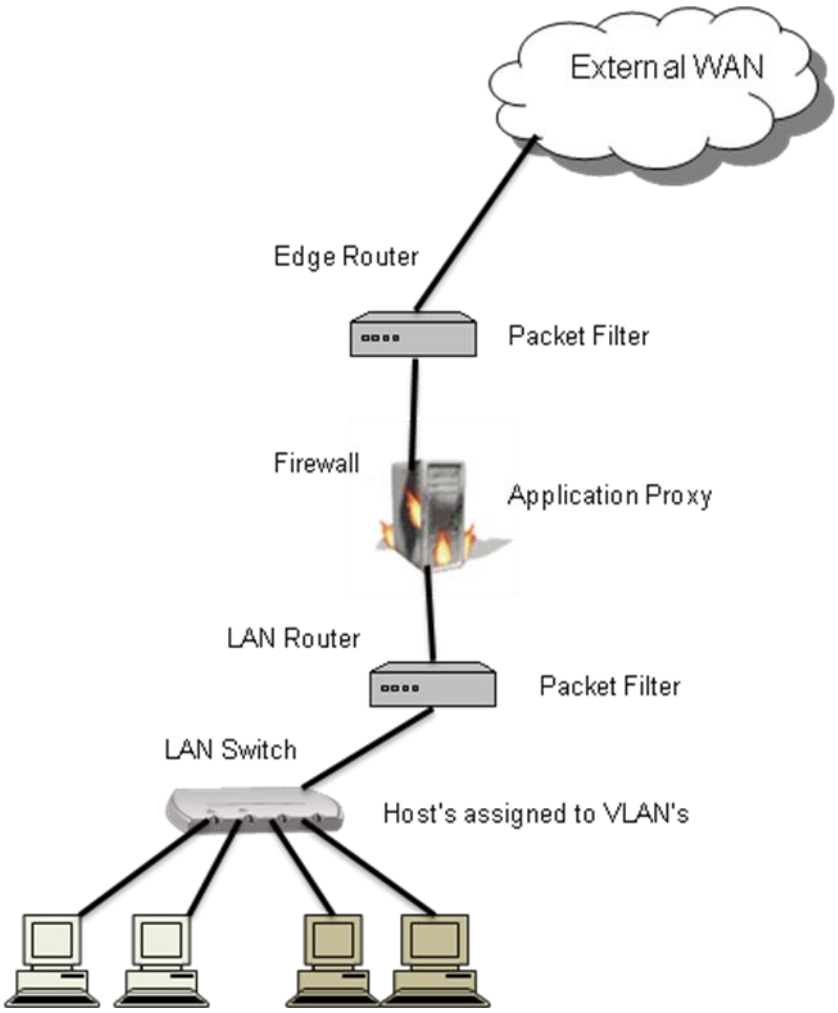

Figure D-22. Typical Firewall Placement for Protection from External Contacts

The packet filter contained in the *edge router* is a firewall that is used as the first line of defense. It provides a coarse granular approach to data flow interrogation (See section D.9 discussion above for the capability of packet filter firewalls.) The application proxy firewall can provide very fine granularity to data flow interrogation including application specific elements. It examines flows that the packet filter firewall did not reject which precede it from the outer boundary and the internal boundary. The *LAN router* provides the packet filter for flows originating from within the protected network. The LAN router provides the first interrogation point for those hosts on the internal network that are allowed to communicate externally. Both the edge router and the LAN routers may also include a *stateful*-inspection capability to determine which side of an external communication flow has initiated the data connection. And finally, the VLANs associated with the LAN switch can be designed to restrict communications between hosts within the same domain.

Additional guidance on firewall placement in networks is listed below:

National Infrastructure Security Co-ordination Centre (NISCC). Good Practice Guide on Firewall Deployment for SCADA and Process Control Networks. February 2005.

National Institute of Standards and Technology (NIST). System Protection Profile, Industrial Control Systems, section 6.1.9, Firewall Access Control, version 1.0. April 2004.

Regulatory Guide (RG) 1.152. Criteria for Use of Computers in Safety Systems of Nuclear Power Plants, rev. 2. January 2006.

National Institute of Standards and Technology (NIST). *Guidelines on Firewalls and Firewall Policy. NIST Special Publication 800-41* rev. 1. September 2009.

D.9.1 Security Observations

Firewalls intercede in the flow of network traffic. A firewall implementation that does not fully understand an organization's demand for services and external connections can impede business. Likewise, a firewall configuration that is too open can negate the intended use of the device.

The important aspect of the firewalls activity is filtering data. Firewalls can filter both inbound data flows and outbound data flows; both filtering implementations are important. Attacks on other domains within a firewall can be initiated from within the protected boundary. Malicious software infecting an internal host from the inside of the protected network can attempt to cross over to other firewall protected boundaries or even to reach back outside to the external network to send out information to an external party. Proper filtering of outgoing data flows from the internal network can restrict the propagation of a virus or even a DoS attack from originating on a host within the protected boundary to other external endpoints. Since an IP address can be "spoofed," proper out-bound filtering can prevent spoofed addresses (any address that is not associated with the original source network) from originating attacks created by infected internal hosts to other locations.

Another aspect of networking is packet fragmentation. Fragmentation normally occurs in today's networks when traffic tunneling or *encapsulation* is used. Many VPN protocols use encapsulation. The encapsulation protocol normally has a fixed maximum size packet that it can accept to carry within its payload field. If this payload size is exceeded, then the packet can be rejected or fragmented to fix the maximum requirement. If fragmentation is used, then a header in the IP packet is set to inform the end point that this packet has additional data that will be delivered in subsequent transmissions. Firewalls implementing application-level inspection may have to reassemble many subsequent packets to allow them to be reviewed for proper acceptance. If this were allowed, the firewall would have to allocate memory and processing time for this activity. An adversary could use this to launch a DoS attack against the firewall. If the firewall allowed fragmentation, the adversary could continually send a large stream of fragmented IP packets that would use up all the firewall resources. For this reason network administrators must make a security trade-off decision either to pass fragmented packets or to reject them. Another

option is to offload any fragmentation reassembly task to a co-located IDS. (See Appendix D, section D.10, Intrusion Detection, for more details.)

Like many computing devices, firewalls combine hardware and software components. Known and unknown bugs and vulnerabilities may exist in these systems. In addition to the devices themselves, vulnerabilities in management software or firewall data collection systems can result in the compromise of a firewall. In the past, attackers have exploited firewall vulnerabilities to modify a firewall configuration, establish a permanent foothold on the firewall management systems and hide specific configurations from operators.

Firewalls are part of a layered, network security approach. It is important to recognize that deploying even the most powerful and state-of-the-art firewall may not protect an organization from certain types of attacks. Firewalls cannot protect against attacks from allowed or compliant services; e.g., a firewall that allows TCP 21 traffic (usually associated with FTP) between two hosts cannot protect against an FTP service attack. Firewalls offer no protection against attacks conducted from within encrypted tunnels (e.g., SQL injection attacks inside an SSL connection). Also, firewalls cannot protect from application layer attacks that conform to the application protocol specification, but abuse the application (e.g., an authentication bypass attack on a Web server).

Firewall management/operations staff must pay close attention to vendor disclosures and open source security advisories. This becomes in essence a race condition to identify and mitigate any known vulnerabilities prior to an adversary developing an exploit. Some common attack paths against firewall management are as follows:

- HTTP services are present on many firewalls. HTTP services are notorious for vulnerability exploits. The rule sets for connectivity to the HTTP server located on the firewall should allow only IP addresses of specific hosts that are authorized to establish HTTP sessions with the firewall.

- Telnet for remote management is an unauthenticated protocol that allows an adversary to capture user IDs and passwords in the clear. SSH should be used for remote network management because it is much stronger than Telnet sessions, but it is important to set the number of authentication-retries. Timeout and retry limits can be specified to control the SSH connection process.

- SNMP community strings use a default community string of “public” that has read/write privileges. Since there are no access restrictions on this community string, an attacker may exploited it to gain complete control of the device. When the attacker has gained control of the device, s/he can eliminate all traffic flow restrictions on the firewall and open up the network to further adversary penetration. Therefore, ensure all default configurations for a newly installed product have been removed. This should be part of a comprehensive security policy.

When firewalls are deployed, they must be managed to reflect the changing needs of the organization as well as the evolving external threat from the Internet. It is important to consider the ways to manage a firewall post-deployment. Some organizations may deploy several or even tens of firewalls. In those cases, consistency and ease of management becomes a high priority. Some firewall manufacturers offer management consoles, management appliances, and software applications. It is within this context that an external adversary may be able to take advantage of a published flaw within the management application.

D.9.1.1 Additional Observations

The unprivileged insider has access to the network on which the firewall resides. If s/he has an account, the account should be limited to only the necessary tasks needed to carry out his/her limited role. This can be accomplished by using role-based levels, which are assigned to each individual user. Another potential exploit path for the unprivileged insider is to use unauthenticated remote access protocols, such as TFTP or Telnet. If network management is using these types of protocols, then passwords can be captured during the network session with the firewall. Using secure network management protocol, such as SSH, can protect the firewalls from this type of internal attack.

A *privileged insider* can make a firewall exploitable by accidentally configuring filters that do not protect the network from malicious exploits. A security policy should be in place that dictates how any firewall configuration changes can be authorized or reviewed. Such a policy provides the accountability to detect unauthorized changes. A formal process for change management can dictate any configuration changes that multiple administrators or subject matter experts can review to help detect malicious configurations that could harm the network. There are also software scripts that can review the configuration commands and identify any configurations that may create inadvertent data flow holes. Providing logging as part of the access control process can help identify users and possibly deter malicious insider activity. Creating controls that limit the number of systems that a single administrator can access or by limiting the locations s/he is allowed to access within the plant can reduce the impact of a malicious insider. Combining these techniques with physical protection mechanisms for personnel access control can provide some level of protection against this type of threat.

Threats from *developer*- or vendor-based sources are based on the firewalls default configuration, such as enabling services such as HTTP, Virtual Teletype terminal (VTP) connectivity, and SNMP management strings. A properly established security policy is the best defense against these potential vulnerabilities. Removing user accounts and default passwords should be part of a security policy established at each NPP facility.

Some standards for firewall network security are listed below:

ISO/IEC 18028-3:2005 Part 3. *Securing Communications between Networks using Security Gateways*. This document provides some techniques for security gateways (firewalls) including a description of the different types of firewalls (i.e., packet filtering, stateful, application proxy). It also provides guidelines for selecting and configuring gateways along with security features and settings.

NIST SP 800-41. *Guidelines on Firewall and Firewall Policies*, provides introductory information, recommendations, and guidelines about firewalls and firewall policy primarily to assist those responsible for network security. It addresses concepts relevant to the design, selection, deployment, and management of firewalls and firewall environments.

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 6, Filtering/Blocking/Access Control Technologies, describes the types of firewalls, typical deployment, and the vulnerabilities that firewalls address.

D.10 Intrusion Detection

To supplement the protection that a firewall affords and to build upon a layered defense approach to security, an additional layer can be found in the insertion of an IDS. An IDS is a type of security monitoring system for both network and host-based traffic. A Network IDS (NIDS) analyzes information from various areas of the network to identify security concerns. NIDS can be configured to identify intrusions or attacks originating from outside an organization's network or to identify attacks or misuse from within an organization's protected boundaries. Another form of the IDS is the Host-Based Intrusion Detection System (HIDS). An HIDS is located on a host computer or device and monitors the data traffic that originates from the host to the network, and from the network destined to the host. HIDS can also include utilities that monitor file accesses and system configuration changes. IDSs should be configured based on an overall security policy. The policy should define the important aspects of organization activities and, thus, provide the guidance for *what* the IDS will be configured to identify. The following provides some activities that an IDS can be configured to perform:

- Monitor system activities (protocols and ports)
- Monitor user-policy violations (applications and connections)
- Identify patterns of attack (signature detection, abnormal activity)
- Monitor host system and file integrity
- Monitor system configurations

One of the primary activities associated with an IDS is monitoring and comparing network packets against patterns or signatures of known attacks. The IDS has a database of attack signatures, and if a pattern is detected on the network an alert is issued. The alert could be in the form of a log file for the administrator to read later, an active message sent directly to the administrator; or it can involve an active command sent to a network device, such as a router or firewall commanding it to add a filter to block the origination of the offending attack pattern. This type of signature protection does not prevent the attack of a previously unknown attack vector, sometimes referred to as the *zero day attack*. To help overcome this limitation, a newer technology called an IPS has been developed. Section D.11, Intrusion Prevention Systems, describes this technology. The following is a description of different methods IDSs deploy to help identify attacks against the network.

Signature-Based Pattern Matching
Signature-based pattern matching identifies within a data packet a fixed byte sequence that is associated with a known attack signature. This can be accompanied by filtering and matching various combinations of the source and destination IP address, and the source and destination application port or service. If additional information about the infecting virus signature is known, then the start and end point of inspection within the packet's data fields can also be specified for a more precise and efficient inspection process.

Some limitations to the signature-based inspection method are concerned with making the process very specific to improve performance. This may require multiple signatures to be defined for minor variations in the implementation of the exploit. This technique can also provide a *false positive* status since legitimate traffic can sometimes contain the streamline data pattern used for detection. Also the signature-based pattern matching is normally limited to inspection of a single packet and, thus, is not well suited for the data-stream–based nature of network traffic. Adversary evasion techniques—such as by spreading the virus over a large set of packet sequences—may take advantage of this limitation.

Stateful Pattern Matching
A more powerful means to detect malicious attacks against a network host is by the use of stateful pattern matching. This technique considers the packet arrival order in a protocol stream and matches patterns across packet boundaries. This will prevent malicious exploits from hiding their signatures among multiple data packets. Stateful pattern matching maintains the session context and reassembles the packet data stream to allow the detection engine to examine the entire data string. This approach requires more memory and processing resources to track open network sessions than the simpler data packet pattern matching technique, but stateful pattern matching makes IDS evasive techniques more difficult to implement.

Protocol Decode
The protocol decode approach to IDS maintains stateful information for each network session. But it also provides a full protocol decode analysis and packet processing similar to the interaction between a client and server. The advantage of this approach is it quickly detects any anomalies in the protocol interaction and provides more flexibility in capturing attacks that would be very difficult to catch using signature-based pattern matching techniques. An adversary who wanted to overcome a pattern matching IDS would just have to create a slight variation of the original attack; this would require a new signature in the database to allow for detection, but would be captured by the protocol decode approach.

Wireless IDS
A network-based IDS can essentially see all the traffic being transmitted over the network segment. This is accomplished by setting up the IDS sensors in "promiscuous" mode, which allows the sensors to retrieve all the network traffic. The wireless environment has multiple frequency bands and multiple channels within each band that can be monitored. A single sensor will not be able to simultaneously monitor all the concurrent wireless traffic. Wireless sensors can be dedicated to a single channel of interest or setup to scan through a series of channels. Some dedicated sensors can analyze the traffic they monitor, but less capable sensors need to

forward their traffic, for analysis, to a management point, normally on the wired network. Some IDS sensors are incorporated within WAPs, which must divide their time on servicing wireless node interaction and monitoring the wireless medium. Because a dedicated sensor can focus solely on wireless security, that sensor can provide stronger detection capabilities than wireless sensors bundled with access points. Some wireless IDS capabilities are listed below:

- Logging
- Device misconfigurations and security policy violations
- Unauthorized or rogue wireless devices
- Wireless scanner detection
- Wireless DoS attacks (jamming and flooding)
- Man-in-the-middle attacks
- Wireless usage patterns

One strength of wireless IDS sensors is the ability to provide triangulation on transmitting wireless nodes. Triangulation uses the distance from multiple IDS sensors to analyze the strength of a signal received by each sensor. With this information a calculation can determine the physical location of the transmitting wireless device. Its distance from each receiver can be calculated to determine if the offending transmitter is located within a secure perimeter and will allow security staff to respond.

Additional information about IDS is available from the following sources:

National Institute of Standards and Technology (NIST). *Guide to Intrusion Detection and Prevention Systems.* NIST publication 800-94. February 2007.

The System Administration, *Networking and Security Organization*. <http://www.sans.org/>

D.10.1 Security Observations

A network-based IDS can provide a wide variety of security capabilities, such as signature-based detection, anomaly-based detection, and stateful protocol analysis to perform in-depth analysis of common protocols. But like the firewall, these IDSs are not a panacea for a complete security solution. They should be used in conjunction with other security measures to create a layered defense against compromise.

As with any security appliance, IDS deployment requires detailed planning and implementation. For example, alarms need to be managed, which entails a trained network management staff to determine if the attack is real or just a false positive. A well defined response is just as important as detection.

The network architecture must be evaluated to determine where each IDS probe will be configured and installed. (For a description of probe placement, see Appendix D, section D.12 Intrusion Monitoring and Sensor Deployment). Decisions on what is most important to protect and the resources needed within each network segment should be well known. For example, if the most important assets on the network segment are associated with a specific OS, then it might

be much more efficient if IDS signatures that are not associated with the assets being monitored are disabled in order to provide much more streamlined and effective protection.

A well maintained means of updating signature databases needs to be implemented. Signature databases are not static and new exploits are being discovered on a regular basis. There will always be a time lapse between the discovery of a new exploit and the time when a new signature is added to the database and a patch is released. This delay offers time for an adversary to exploit the vulnerability prior to the patch. If an OS vendor is slow in bringing out a patch for a new vulnerability, the network administrator may need to take action, depending on the severity of the vulnerability. This action may include isolating the vulnerable device by reducing its network exposure. For example, this may include reconfiguring the firewall to close off a vulnerable port or communications from a specific address space.

When an IDS detects that the network or a device on the network is under attack, the IDS can initiate some offensive measures to eliminate or reduce the attack severity. The following is a list of some defensive responses:

Session Termination. If the IDS sensor is deployed in a "promiscuous" mode—"sniffing" network traffic without having any direct impact to the data flows—it can initiate a session termination when its sensor detects an attack from a communicating end node. This response can only be initiated if a connection-oriented protocol, such as TCP, is being used. Connection-oriented protocols provide a formal connection setup and termination process. Non-connection–oriented protocols, such as Internet Control Message Protocol and UDP do not have session setup procedures and are immune to session termination commands.

Bandwidth Throttle. If the IDS sensor is deployed as an inline sensor—which means the network traffic it is monitoring passes through it—then it can be configured to mitigate attacks by blocking or throttling the amount of network bandwidth that any network flow can use. This can reduce the severity of a DoS attack or limit the percentage of network bandwidth that can be used by any particular application.

Firewall Reconfiguration. Another form of IDS active response is the ability to re-configure a firewall or router filter to prevent an external *suspect* end node from reaching into the private network. This can restrict an external attack from reaching into the protected network or prevent a maliciously compromised internal host from communicating to the external network.

IDS Management

The most secure way of network management interaction of the IDS components would be through a stand-alone network that interconnects the IDS components for management purposes. This allows the data transport network, where attacks originate, to be physically isolated from the network used to mange each individual IDS node. This approach will help hide the existence the IDS from potential adversaries and provide independent network bandwidth to manage IDS devices during times of heavy network data traffic. The disadvantage of this approach would include the additional costs in networking equipment to build and maintain this independent network.

As previously discussed, IDSs provide a valuable service that identifies and protects the network and its assets from attacks. But it is important to understand the limitations associated with IDS network-based systems outlined below:

- Network IDS cannot inspect data and, thus, cannot detect attacks when the network traffic is encrypted. This encrypted traffic includes any VPN connections, such as SSH, HTTP over SSL/TLS, and IPSec tunnels. If an IDS is needed to inspect encrypted data flows, then the VPN needs to be terminated at the border of the network and the traffic decrypted prior to being sent to its final destination. This allows the IDS to review the traffic for any suspicious data.

- Another limitation of IDS is the ability to review fragmented packet data. This can be overcome if the border router or firewall provides the service of reassembling the packet prior to forwarding it to its destination. In fact this is the best approach because the border router can first test the reassembled packet against its allowed traffic profiles to determine if it should be admitted into the interior of the network.

- Another limitation that can be leveraged against an IDS system is its ability to process data. An IDS can be overwhelmed by the volume of traffic it is attempting to review. That is why it is important to analyze the network locations and traffic patterns to determine if the IDS is capable of processing packets at the maximum rate of reception.

- Another limitation or vulnerability is the advent of false positives. This becomes a serious risk if the IDS also has the capability to change configurations on a router or firewall based on its perceived determination of an attack. This can create a self-induced DoS condition if it blocks legitimate packet flow.

- Finally, a potential vulnerability with an IDS is the need to continually update its database profile with the latest attack signatures. The procedure on performing this activity needs to be part of an overall security policy. Downloading updates from the vendor Website either manually or automatically checking for updates or using removable media, such as thumb drives, all need to be reviewed for the best approach that facilitates operations and security.

To continually improve the IDS profile, a vulnerability assessment audit should be conducted on a regular basis to test the IDS protection defenses. These audits can identify areas that need to change to improve network protection. A security policy—which may be updated based on audit results—should guide this activity.

D.10.1.1 Additional Observations

To combat the external threat, it is important to consider how to manage the NIDS. If the IDS is managed through the same network as the data, this can open the management interface to external attack. An IDS management interface provides management consoles, management appliances, and software applications—such as HTTP server interfaces—to help configure and manage the device. Within this context, an external adversary may be able to take advantage of a

published flaw within the management application. To reduce this threat, a separate network interface (physical) or a separate VLAN for management interfaces will help isolate and protect this interface from the external threat. The external adversary could also examine the packet flows in and out of the network and craft "spoofed" packets of an allowed address space that contain malicious code. This malicious code would be detected as an active signature by the IDS signature database. If the IDS is in a reactive mode, it may create filters on the firewall to prevent this IP address from entering the internal protected network, thus, creating a self inflicted DoS against the true authorized IP address. Proper configuration of a stateful firewall or router can mitigate this type of attack. (See Appendix D, section D.9, Firewalls, for more details.)

The unprivileged insider would have access to the network in which the IDS is deployed. Similar to the external threat, this access allows him/her to launch an attack script against the IDS sensors by creating attack profiles that may influence the IDS into defensive strategies, thus, creating DoS conditions against targeted data flows. To limit this type of attack, network management should administer proper port security on the network devices to prevent unauthorized device attachments and provide segregation between management data flows. (Details of these protection mechanisms can be found in Appendix D, section D.4.1, Ethernet Security Observations.) If the unprivileged insider has a network management account, it should be limited to only the tasks necessary to complete the assigned task. Another potential exploit path for the unprivileged insider is to use unauthenticated remote access protocols, such as TFTP or Telnet, to connect to the management ports of the IDS device. If network management uses these types of protocols, then administration level identifications and passwords can be captured during the network session with the IDS, providing the insider with the information needed to escalate his/her privilege level. Using secure network management protocol, such as SSH, can protect the IDS from this type of internal attack.

The *privileged insider* with the intent on malicious activity is difficult to overcome. Having a security policy in place that dictates the procedure for IDS access and configuration changes can provide some accountability to detect unauthorized changes. A formal process for change management should be instituted. It could include procedures, such as requiring that multiple administrators or subject matter experts review all configuration changes to help detect malicious or accidental configurations. Providing logging as part of the access control process can help identify users and possibly deter malicious activity. Creating controls that limit the number of systems that a single administrator can accessed or by limiting the locations s/he is allowed to access within the plant can reduce the impact of a privileged malicious insider. Combining these techniques with physical protection mechanisms for personnel access control can provide some protection against this type of threat.

The threats associated with the *vendor* would be from support contracts that allow vendor personnel to access and configure system attributes. Also the technique used for software upgrades and adding new exploit signature profiles required to be added on a regular basis can provide a path for exploitation. A security policy should guide the technique for performing these activities to prevent the introduction of malicious code.

Some standards for Intrusion Detection Network Security are listed below:

ISO/IEC 18043. *Selection, Deployment and Operations of Intrusion Detection Systems,* provides information on host and NIDS, considerations, vulnerabilities, and other associated IDS operations, and integration issues.

NIST SP 800-31. *Intrusion Detection Systems,* provides an overview of an IDS and the different types, deployment strategies, strengths, and limitations of IDSs.

NIST SP 800-94. *Guide to Intrusion Detection and Prevention Systems (IDPS),* provides information on common detection methodologies, a network overview, typical components, and security capabilities.

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems,* section 8.3 Intrusion Detection Systems, provides information on the varieties of IDS and the vulnerabilities that the technology addresses.

D.11 Intrusion Prevention Systems

Because a traditional IDS relies on previously known information (attack signatures on file), it becomes ineffective for previously unknown attacks commonly referred to as the *zero day* attacks. A zero day attack, as the name implies, provides the network security administrator zero days of warning of an impending attack. An IPS can have both a passive and proactive configuration to prevent the offending action from damaging the system. In an attempt to reduce the success of zero day attacks, two types of detection are deployed: anomaly analysis and heuristic-based analysis.

Anomaly Analysis
The technique within this approach is to understand the patterns of normal operational activity and alert personnel when a *non-normal* pattern is detected. This approach compares network traffic against an established baseline. The baseline will identify what is *normal* for that network—what sort of bandwidth is generally used, what protocols are used, what ports and devices are active at specific times of the day, host communication profiles etc. The primary strength of anomaly detection is its ability to recognize previously unseen attacks. It does not look for attack signatures within packets, but rather for abnormal traffic patterns. The IPS can be configured passively to monitor and report by logging the information; or it can have an active response by changing the configuration of a network edge device, such as a router or firewall, to disconnect the offending traffic flow. Its disadvantage is the need to *train* the system to be able to determine what constitutes *normal* background traffic versus something new and suspicious. This system also requires a training session whenever a planned change to the system occurs, such as the loading and running of a new application. The IPS can detect overly active hosts and inappropriate traffic initiations. But changes in standard operations may cause false alarms, while active intrusions that do not violate the baseline and appear to be normal may not be detected.

Heuristic Analysis
Heuristic-based network analysis uses algorithmic logic of network interactions to make statistical evaluations of the type of traffic being presented. Many network intrusions and attacks

are preceded by a network reconnaissance of a target network. The term *network reconnaissance* refers to all techniques used to gather information about a network such as the following:

- Network segment IP range
- Active hosts IP addresses on segment
- Files and services located on each host
- Host OSs

The above listed information is needed to launch a variety of attacks on a network. The approach that adversaries use to gather active host information on a network segment is to first determine the network and host portion of the address that will be scanned. This can be a simple guess on the subnet boundaries or from intercepted router traffic that specifically provides this information. Either way, the adversary then starts with active host mapping. A port sweep is an example of this type of reconnaissance that could be detected using heuristic analysis. The heuristic algorithm looks for the presence of a threshold number of initiated communications through unique ports associated with a particular host. This type of reconnaissance identification may also be more a restrictive, such as the packet type, source of origination (IP address), or a valid connection-oriented sequence, such as the interaction of the SYN and ACK packets in a TCP connection request. This type of heuristic signature example could also identify much more complex relationship interactions than the port sweep example provided.

Host Base IPS

A host-based IPS systems is similar to a host IDS. The host has IPS software resident on the host machine it is intended to protect. It is closely associated with the OS and monitors application interactions with system level kernel calls, or it monitors OS application programming interfaces to protect it from unusual levels of requested access. It can provide logging for activities deemed suspicious as well as monitor data streams into and out of the host to provide protections against potential malicious software, such as Trojans, that have not yet been discovered or publicized. Since the IPS software must be closely aligned with the OS it is intended to protect, any upgrade or patch to the system may cause interoperability problems.

D.11.1 Security Observations

Most IPSs feature multiple prevention capabilities. They allow the network manager to identify the specific prevention action for a particular alert or to disable an active prevention method. While in the *learning* or simulation phase, many prevention actions are replaced with a series of logged announcements of what would have taken place in a prevented action. This allows the network manager to review the alert response accuracy and tune the prevention-capability configuration before enabling prevention actions. This will reduce the occurrence of false positives, which can block allowed benign activities.

Providing the proper IPS access controls is not only important to limit its exposure to attacks from adversaries, but to protect the information it contains. IPS and IDS frequently contain sensitive information about the network configuration, such as host configurations, applications and ports, allowed communication interactions, and known vulnerabilities. An adversary could use this information to help compromise or disable the network.

Pre-Deployment Testing

Network managers need a means to implement IPS component testing on a non-production network. This test environment will allow the manger to check out configuration that will reduce the probability of implementation problems disrupting the production networks. A phased introduction of IPS sensors into the production network can reduce the number of false positives that may occur when enabling many sensors. These false positives can be addressed prior to activating all the sensors and also provide a means to test out any scalability issues that may arise.

Like the IDS, the IPS is vulnerable to false positives. This can be more problematic than with the IDS because of the need to *prevent* an attack. This normally entails a proactive security approach with potential to prevent benign traffic patterns from reaching their destinations. This problem reduces IPS reliability. But unlike the IDS, the IPS does not have the same requirement to review data payloads. IPS' strength comes from its ability to detect improper traffic patterns or communications among end nodes that include port information and bandwidth consumption. This technology still needs improvement, but when combined with other security approaches, such as signature-based IDSs and firewall filters, IPS becomes one of the multiple layers of defense required to protect a network from compromise.

D.11.1.1 Additional Observations

External threat attacks against the IPS are limited. The primary vulnerability would be to attempt to reach its network management port remotely. But if a properly authenticated protocol, such as SSH, is used for management interaction, this protection will be difficult to overcome. The external threat would first have to discover the IP address of the IPS by using reconnaissance techniques discussed earlier. The security approach to deploying an IPS would be to obfuscate its presence by configuring it in "promiscuous" mode. Therefore, it would not respond to the type of connection request used for network mapping activities associated with network reconnaissance. To provide a more secure management interface for the IPS, a separate network interface (physical) or a separate VLAN should be considered for the management interfaces. This will help isolate and protect this interface from the external threat.

The unprivileged insider would have access to the network in which the IPS is deployed. This access will allow him/her to "snoop" on management traffic designated to the IPS sensors. If network management were using proper authenticated protocols, such as such as SSH, then user IDs and password authentication would protect the IPS from the unprivileged insider. "Snooping" of network management traffic can be further prevented if proper port security is administered on the network devices to prevent unauthorized device attachments and to provide segregation between management data flows. (Details of these protection mechanisms can be found in Appendix D, section D.4.1, Ethernet Security Observations.) If the unprivileged insider has a network management account, it should be limited to only the tasks necessary to complete the assigned work. This can prevent the loading of unauthorized code from removable media, such as thumb drives and CDs. Using strong authentication for remote access to IPS components, such as two-factor authentication, can provide an additional layer of security.

Providing logging as part of the access control process can help identify users and possibly deter the *privilege insider*. A formal process for change management should be instituted. It could include procedures, such as requiring that multiple administrators or subject matter experts review all configuration changes to help detect malicious or accidental configurations. Creating controls that limit the number of systems that a single administrator can access or by limiting the locations s/he is allowed to access within the plant can reduce the impact of a privileged malicious insider. Combining these techniques with physical protection mechanisms for personnel access control can provide some protection against this type of threat. Regular configuration audits could determine if IPS attributes have been changed to the detriment of the system. Conducting vulnerability assessments on a regular basis to test IPS defenses could also help uncover configuration changes (malicious or accidental) that might impact network security. A security policy—which also may be updated depending on audit results—should guide this activity.

The IPS requires very site specific configurations; therefore, the vendor normally does not pre-configure a *default* configuration. Threats associated with the *vendor* would be from support contracts that allow vendor personnel to access and configure system attributes. The means of adding software patches or updates is a vulnerability that can provide a path for exploitation. The security policy should guide the method for performing these activities to prevent introduction of malicious code.

A standard for IPS network security is listed below:

NIST SP 800-94. *Guide to Intrusion Detection and Prevention Systems (IDPS),* provides information on common detection methodologies, a network overview, typical components, and security capabilities.

D.12 Intrusion Monitoring and Sensor Deployment

One advantage of deploying an IDS or IPS system is the ability to instrument the network with sensors. To determine the most effective sensor placement, the network topology must be analyzed, along with the cyber asset locations, to determine the best interface boundaries requiring detection capability. A network that is properly instrumented can assist in the overall situation awareness of its operation. As can be seen in Figure D-23, Sensor 1 is placed at the perimeter of the external public facing network and the internal NPPDN. This sensor is located outside of the boundary firewall but inside the edge router. This allows the sensor to monitor attacks that originate from outside of the protected network zones prior to being filtered by external boundary Firewall B. Sensor 2 is in a position to monitor traffic to and from the "shared information server" of Firewall A. This allows the sensor to determine if Firewall A is affording the necessary protection for the shared information servers. It can monitor data flows to determine authorized access as well as attacks or connections that may originate from within the PDN from a compromised asset. Sensors 3 and 4 are interior zone sensors monitoring the traffic that is present on Zone A and Zone B. This can help determine if proper policy has been applied to the firewall to prevent external connections from being made to the safety network located in Zone A or to monitor attacks against the business network in Zone B.

Sensors 5 and 6 illustrate the way IDS sensors can be used to monitor the flow of traffic between different internal groups on the network. Sensor 5 is protecting the engineering network, while Sensor 6 is monitoring the process control network.

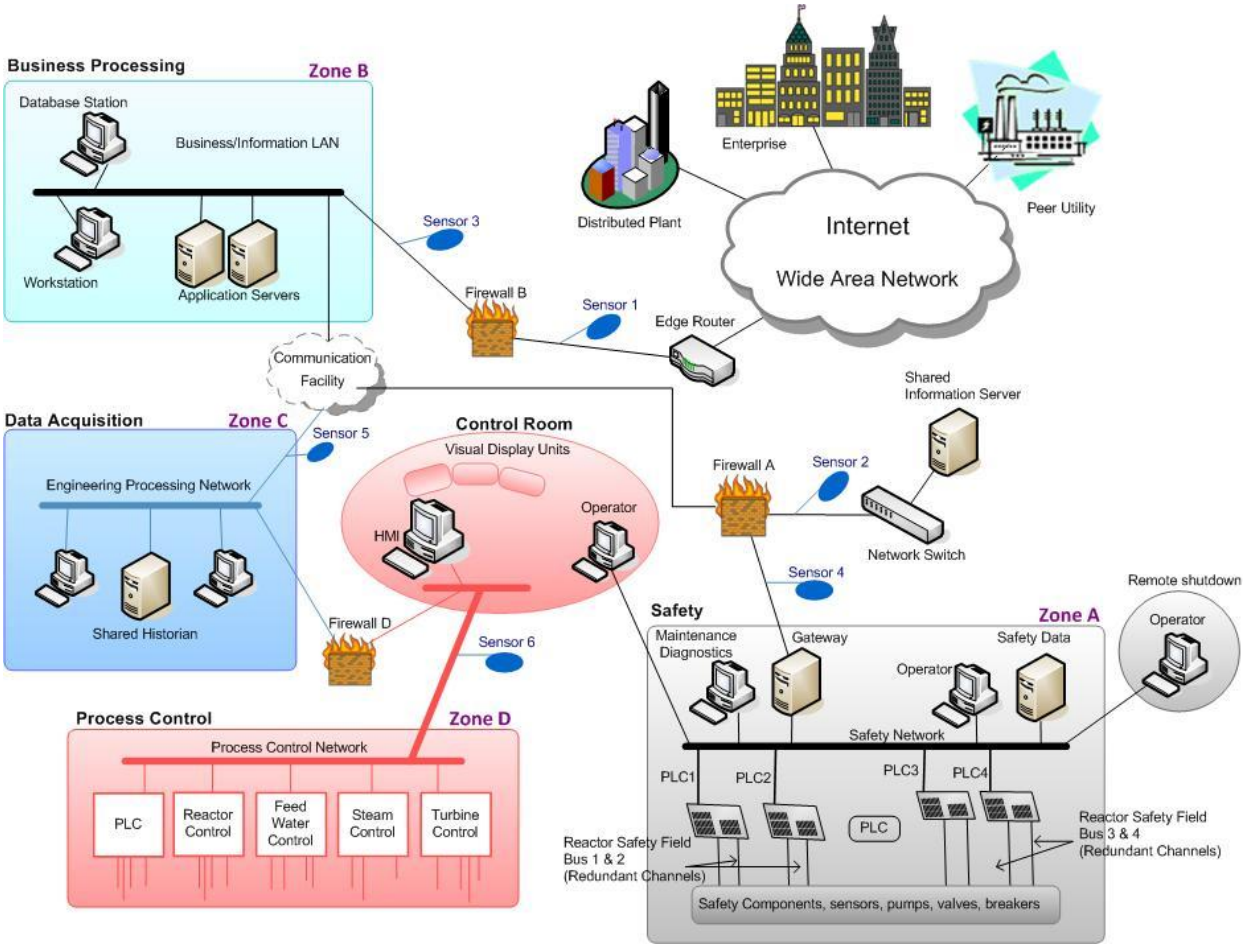

Figure D-23. Digital Plant System Network with Sensors

Additional information is available from the following sources:

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1.11, Intrusion Detection and Response, version 1.0. April 2004.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.11, Intrusion Detection Systems. October 2005.

D12.1 Security Observations

IDS and IPS sensor deployment are a part of an overall network security approach, when properly combined with firewalls, anti-virus patch management, and host communication authorization.

IDS and IPS sensor management should be easily configurable and understandable for the operator to be effective in monitoring and responding to network and host protection.

The IDS antivirus community is starting to develop anti-virus signature software for some control system protocols, such as ICCCP and Modbus/TCP implementations. These are implemented as bump-in-the-wire appliances to protect safety and control system components, such as PLCs, RTUs, DCSs, and IEDs. In the future it may be possible to deploy NIDS on individual control device interfaces or HIDS on the devices themselves when these devices become more processor- and memory-capable. But IDS integration on critical components associated with process control or safety should be evaluated on non-production systems for any potential performance impacts for the component that is hosting the IDS.

A standard for intrusion monitoring and sensor deployment is listed below:

NIST SP 800-94, *Guide to Intrusion Detection and Prevention Systems (IDPS)*, section 4.2, provides an example of sensor deployment within a network architecture.

D.13 User/Operational Management

Sections D.13 through D.16 describe and discuss various aspects of the User/Operational Management element for secure networks. Topics include access control, malicious software (*malware*), and auditing. Also presented are security observations regarding vulnerabilities and threats stemming from each topical area, as well as suggested mitigation techniques.

D.13.1 Host Access Control

Operating System Access Control

The terms *host* or *workstation* normally refer to a device that contains an OS. The OS determines the functionality of the host. The OS can permit or deny users of different roles to interact with applications and utility services on the host. In the current generation of hosts, an OS is required for an application to run. The OS can be seen as a layer of defense to protect applications and sensitive information. If configured properly, the OS is a mechanism that ensures only designated individuals can run applications or make changes to system configurations and security policies.

Users can access OSs through several mechanisms. The preferred and most powerful method is to sit at a host and use the keyboard, video monitor, and mouse (KVM) to interact with the OS. A host may also be connected to a terminal server or modem, which permits remote access to the OS. If a host is connected to a data network and properly configured for network access, the user may access the host through a network service.

OS access should be available only to individuals with an approved need to modify, repair, or maintain the host. Maintaining a host includes adding and removing hardware, installing software patches, installing and configuring applications, and modifying or installing security policies. Depending on the OS type and version (i.e., Windows, Linux, MacOS, etc.), different levels of granularity in defining roles and responsibilities are possible. Since the OS is so important to the security and functionality of the host, it is crucial to limit the number of individuals with administrator rights. Typically, individuals defined as system administrators will

have the highest-level OS access and ordinary users are limited to running applications and entering data. (See Access Control 5 and 6 (AC-5, AC-6) of *NIST 800-53* [D.13-1], applicable to high consequence systems.)

Access to many OSs does not require the use of KVM, but rather takes place through a network interface. Access to the host for the purpose of system administration should require authentication to a higher level of access. Access to the host/workstation over a network interface should never be allowed in clear text. An encrypted method, such as SSH, should be used for remote access to a host/workstation. Direct root access to a host/workstation if possible should not be permitted over the network. If the OS supports switch user (SU), a system administrator should be required to login in as a user and SU to root. Many network accesses are through an application, frequently through a client-server model. The user access to the host should be regulated by the applications in that model. Web (HTTP) access through the network should be regulated in the same manner, with the recognition that the protocol is stateless; thus, every automatic action should be considered a fresh access.

User authentication mechanisms are necessary to control access and provide audit logs of user activities on hosts. The simplest user authentication is a single, personal factor like a password. This may be sufficient if there are additional physical security measures limiting access. (See Appendix D, section D.2, Physical Security Details, for information pertaining to physical access control.) Physical security of a host is very important, if an adversary has physical access to a host, many tools and methods are available that permit the OS to be compromised. Security compromises to an OS can be many things that are not obvious or noticeable to users and include such things as enabling unnecessary services, installation of malware, escalation of user permissions, addition of unauthorized users, etc. Passwords should be strong enough to prevent password guessing within a timeline that must be calculated from the lesser of password expiration deadline or user audit log verification cycle. If the password can be determined through brute-force, dictionary attack, or HASH look-up (rainbow tables) before the password has expired or the audit log is verified, then adversaries could gain access to the host.

Two-Factor Authentication

Additional security can include two-factor authentication. Multifactor authentication includes "something you know, something you are, and something you have." Passwords, pass-phrases, PINs, and other personal factors are in the category of "something you know." Human factors, such as biometrics, are inherently bound to the individual and are considered "something you are." Technical factors are bound to a physical object, "something you have," and include passes, ID cards, tokens, and smartcards. Two-factor authentication usually involves technical factors and personal factors; three-factor authentication is not yet commonly used because of the difficulty of working with human factors.

One benefit of two-factor authentication using personal and technical factors is that access can be terminated upon removal or movement to a set distance of the technical factor. Removal of a smartcard from a reader or movement of an RFID chip from the point of access can be used to terminate access. Technical factors for user authentication to hosts should not serve any other function, such as physical access control. If a smartcard is used in a smartcard reader for host

access, it should not be relied upon to authenticate the human owner upon challenge for physical access.

Two- or three-factor authentication increases the security profile of asset access. Also biometrics, as a form of authentication, blends well when combined with other types of authentication, such as smartcard and token technologies. But it must be pointed out biometric devices are subject to false positive rejections. This means there is a chance that a *truly authorized* person may be locked out of access when the biometric device provides a false positive. Not all biometric devices are created equally, so it is important that the user selects the product that provides the highest level of reliability.

Additional access control information and recommended practices are provided in the following documents:

ISO/IEC 2001:2005. *Information Technology Security Techniques, Information Security Management Systems Requirements.*

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.7, Host and Device Security. October 2005.

NIST 800-12. *Introduction to Computer Security.*

Human/Machine Interface

The human/machine interface (HMI) can also be considered a host machine running an underlying OS. It is normally represented in the form of a console with a graphical user interface that displays status information about plant operations. Also referred to as Qualified Display Systems (QDSs) or Safety Parameter Display System, it assists the operators in evaluating current operational status and provides situational awareness of any abnormal conditions.

It can be setup to be a stand-alone device or integrated into the control room information system. This information system can be built upon a common communication protocol, such as Ethernet. This allows the inter-exchange of monitor and status information. Some HMI control room requirement documents are listed below:

IEC 60964. *Nuclear Power Plants; Control Rooms—Design*, ed. 2.0. 2009.

IEC 61772. *Nuclear Power Plants: Control Rooms—Application of Visual Display Units*. ed. *2.0. 2009.*

IEC 60960. *Functional Design Criteria for Safety Parameter Display System for Nuclear Power Stations*. ed. 1.0. 1988.

IEC 60965. *Nuclear Power Plants: Control Rooms—Supplementary Control Points for Reactor Shutdown without Access to the Main Control Room*, ed. 2.0. 2009.

D.13.2 Security Observations

When a series of HMI hosts are linked together on an Ethernet network, note that on a LAN environment all local users of the environment may be able to review the data being exchanged between hosts and be able to gain unauthorized access to an HMI host resident on the LAN. Direct access to the host OS through the user interface should be available only to persons with the need to modify the OS configuration or software. System administrators should have this access, but ordinary users should not, (per Access Control 5 and 6 (AC-5, AC-6) of *NIST 800-53[D.13.1]* applicable to high consequence systems).

The host OS is the foundation on which functionality and security are built for the application. A properly secured OS will help ensure users perform only tasks they are authorized to do, obtain only information they are authorized to have, and cannot cause damage to data applications and operating equipment. The goal of a secured OS is to provide data integrity, protect confidentiality, and ensure availability of the host/workstation. The OS should be configured to limit unauthorized access both physically and remotely. Each vendor OS (Microsoft, Apple, Red Hat, Solaris, etc.) has different guidelines for system hardening that should be referenced when setting up a system before deployment.

If the hosts are part of a network environment where information is exchanged among them, in some cases unauthorized access can be gained because the hosts use un-authenticated protocols for data exchange. This lack of protection can result in an un-privileged insider on the host network gaining access to a host, such as a QDS, by the way of an Ethernet port. Once an adversary gains access to an operator interface, s/he may be able escalate his/her privileges and manipulate elements of the operator display. This type of attack has been documented in an NRC letter report [D.13-2].[15] Proper host-based security management procedures, along with using only authenticated network access, can prevent these types of attack.

Creating and applying strong password management procedures for user accounts can decrease the risk of unauthorized access to host machines. But a potential problem with proper strong password generation and management is that during times of crisis, the ability to recall a strong password or even multiple passwords may be impacted by the elevated stress on the human operator. This may delay the operator's ability to respond promptly to an ongoing event.

In some control applications, the system may provide only group passwords at each level of access, not individual passwords. If these situations are prevalent at the utility site, it may be useful to consider an RBAC. This can allow a person to login and provide credentials at the start of a shift and be allowed access to all necessary applications based on job function. (See Appendix D, section D.14, Role-Based Access Control (RBAC), for more details.)

The following is a list of host access procedures that can enhance the security of host system:

---

[15] Common Q Vulnerability Assessment Report, June 2009 Prepared for the Division of Engineering, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission, Washington D.C.

1. Review the requirements of each system user and assign and enforce (using owner or group permissions) a level of system access that is commensurate with their roles. The most effective way of reducing the chances of a resource exhaustion attack is to verify the user roles associated with the system. A role-based review would determine the level of access and, thus, the level of system resources, required to perform one's job.

2. Generate unique passwords for each user of the system and change these passwords on a regular basis.

3. Create a log file directory on the system that documents user log-ons and tracks important events, such as file additions and/or changes and other user activities. Protect this log directory with appropriate directory and file permissions.

4. As part of the file and directory permissions review process, evaluate all the executable files on the system to determine the need to be writable by *group* permission, or to determine if just being writable by a single user is sufficient.

5. Create a formalized change management process that documents all changes to the system.

6. Consider deploying HIDS. This can help monitor and detect security violations and intrusions, such as directory or file modification.

7. Maintain a current baseline inventory of all software. It may not be possible to identify unauthorized changes to system software or to successfully rebuild the OS or applications after a system corruption if a current baseline is not available.

8. Develop policies for the external equipment if the control network or field network has external access points; define the use of removable media, such as compact discs and thumb drives, to reduce the risk of software contamination.

D.13.2.1 Additional Observations

For the *external threat* to gain access to host-based system, the system must have a cyber connection to another segment of the plant data network that touches a public interface. For an HMI host in a control room, the assumption is the data exchange network that supports information sharing among HMI host has no external connections to other points within the data plant network. It is important to note that the means of updating software or adding patches to any element of the HMI host-based network needs to be evaluated for potential compromise due to infected media. Infected media are a means that the external threat can use to gain access to a network segment that seems quite isolated from any external network connectivity. The site security policy needs to address how external media are utilized within the HMI network. For other hosts located throughout the plant, an external adversary may gain access to a host if a vulnerability exists within the host or the host is utilizing unsecure protocols for application access on the host system. But even if vulnerabilities do arise, a properly protected network that has implemented a defense-in-depth approach for network security can defend against the external threat.

An *unprivileged insider* with access to the host networks, whether within a control room or throughout the plant data network, can potentially manipulate, modify, or deny data traffic being sent between host nodes. This is because the unprivileged insider has physical access to the data transmission medium and may take advantage of any unsecure data exchange protocols being used on host machines. But again, a properly protected network can detect and respond to attacks against the network (as described previously in Appendix D, section D.4.1, Ethernet Security Observations). The unprivileged insider may have an authorized account on the host itself and may be able to manipulate some aspect of the OS or an application file. But each host can be afforded some protection against the unprivileged insider if the protection elements mentioned above are implemented.

One of the greatest threats to the OS is from a *privileged insider*. A privileged insider can make an OS exploitable by accident, through ignorance, dishonesty, or workload. A privileged insider may exploit an OS for monetary gain, identity theft, impersonation, "snooping," and revenge. Two- or three-factor authentication increases the security profile of asset access. Also biometrics, as a form of authentication, blends well when combined with other types of authentication, such as smartcard and token technologies. But note that biometric devices are subject to false positive rejections. This means there is a chance that a *truly authorized* person may be locked out of access when the biometric device provides a false positive.

Threats from developer- or vendor-based sources associated with hosts can include 1) back-door utilities to allow remote access by the vendor, 2) the enabling of ports and services on the host OS that are not secure, and 3) default user accounts and passwords. A proper security policy should include practices and procedures that guide users on how best to inspect and secure newly installed products.

Some standards for user management and secure host-based access control are listed below:

NIST SP 800-118. *Guide to Enterprise Password Management*, provides best practice information about proper password management, such as password capturing, guessing, replacing. and 2- or 3-factor authentication for access control.

NIST SP 800-92. *Guide to Computer Security Log Management*, provides insights on access control and review using log management. It includes discussion on log management infrastructure, planning, and operational processes.

ISO/IEC 24727. *A Future Standard for Smart Card Middleware*, describes features of the standard approach of using smartcard technology for identification and access control. It compares and contrasts multiple smartcard middleware solutions to help the user make a more informed decision.

ISO/IEC 27002. *Code of Practice*, section 11.2, provides information about the allocation of access rights to users from initial user registration through removal of access rights when no longer required.

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 5, Authentication and Authorization Technologies, describes user access control authentication technologies to include password, smartcard, and biometric.

References:

D.13-1 NIST 800-53. *Recommended Security Controls*.

D.13-2 *Common Q Vulnerability Assessment Report*. June 2009. Prepared for the Division of Engineering, Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission, Washington D.C. (not publicly available).

D.14. Role-Based Access Control (RBAC)

RBAC is a method of restricting system access to authorized users. RBAC is an alternative method to more commonly used access restrictions, such as mandatory access control and discretionary access control. System administrators have always had the ability to assign privileges to users and groups of users. RBAC is a more advanced step in access control. RBAC can create privileges and assign them to operational roles; then the roles can be assigned to users. This enhances system security by providing more granularities of privileges within roles. This reduces the chance of security violations by providing greater control over user access to information and resources of multiple devices in a network. With RBAC access and their associated permissions, decisions are based on the roles that individual users have as part of an organization. The process of defining roles becomes paramount in the level of access to the system. A thorough understanding of utility-organization operation should be well defined and understood prior to setting up any RBAC interface. When setting up an RBAC account, it is important the user be given no more privilege than is necessary to perform the job. This *least privilege* concept requires more precision when defining a user job function.

When roles have been defined, the permissions that accompany each role are created on the devices. In many plant operations, there may be many devices, such as RTUs, PLCs, plant information servers, smart sensors, and IEDs. Note that RBAC systems work well in static environments where device replacement is rare (adding new devices with new permission fields can be burdensome). RBAC implementation can reduce the overall burden and complexity of common security-based *individual-permissions* management by basing access on a user role or job responsibilities, rather than customizing access for each individual. For example, some safety operators may be able to view field device status, but cannot send out commands to change their configuration.

RBAC systems can minimize security violations by providing more precise control over user permission to access information and control over multiple devices in a network. Permissions can include reviewing, and/or modifying specific data or device functions. The goal of RBAC implementations is to manage access to devices and information while reducing individual device access levels and enhancing the granularity of security controls.

When defining an RBAC model, the following conventions are useful:

- R = Role = Job function that defines an access level
- S = Subject = A person
- SA = Subject Assignment
- P = Permissions = Access level to a resource
- PA = Permission Assignment
- SE = Session = A mapping involving S, R and/or P
- Subject can have multiple roles.
- A role can have multiple subjects.
- A role can have many permissions.

The job function can include multiple people and multiple permissions. When setting up a rule set, the potential inheritance of permissions must not compromise the separation of duties associated with individual personnel. For example, a host-system user should not have write permissions to change aspects of an application file while also having write permissions to modify aspects of the log file directory (which has been setup to log user account activity).

Additional information on RBAC methods is available from the following sources:

An Introduction to Role-Based Access Control—NIST CSL Bulletin on RBAC. December, 1995.

National Institute of Standards and Technology (NIST). *System Protection Profile, Industrial Control Systems*, section 6.1, STOE Security Functional Requirements, version 1.0. April 2004.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power,* section 6.3, Configuration Management. October 2005.

ANSI/ISA-TR99.00.01-2004. *Security Technologies for Manufacturing and Control Systems.*

D14.1 Security Observations

Note that many RBAC systems are dependent on a centralized server to setup and administer RBAC accounts. It the server is unavailable, then permissions cannot be validated and may restrict access to system components during critical operations. Server availability must be taken into account when contemplating an RBAC implementation within a utility infrastructure. RBAC should be applied in single systems (i.e., single applications where the separation of duties is important and the numbers of subjects, roles, permissions ,and mappings can remain small).

Vulnerabilities associated with exploits against the RBAC server are similar to those against OSs in general. Any OS vulnerability or any application running on the RBAC server that has a known vulnerability has the possibility of being exploited by an adversary.

The integrity of the RBAC policies is dependent on any vulnerability present on the server. Periodically, the server needs to be secured with the latest security patches for the OS and with applications running on the RBAC server. Only applications that are required for the server to

accomplish its job function should be run on the server. TCP services running on the server should be reviewed before the server is put into production; this can be done with readily available tools, such as nmap, Nessus, etc. Any services not required on the server should be disabled. A review of remote access to the server should be conducted and remote access only permitted to individuals needing access to the server for maintenance and operation. A review of individuals assigned to roles should be done periodically to verify that individuals assigned to roles are still valid.

RBAC has been shown to work well in single system environments (e.g., Microsoft Windows environment), but has had issues when used in a system of systems. RBAC complexity is a function of the many relationships of roles, permissions, and resources. RBAC is best implemented by applying a structured framework that breaks down each task into its component parts. As with any server services hosted on an OS, it is important to be vigilant to proper software upgrade and patch management.

Database breaches can be the most damaging to an organization because, in many cases, they contain proprietary or sensitive information. It is important to understand that vulnerabilities associated with databases normally exist because of weaknesses induced by improper software design by the vendor or through misconfigurations by the use (e.g., not properly locking down the database to perform only the functions needed for operations). Having and maintaining proper patch management and regular audits are key to detecting improper configurations, use, or operation. Effective database security techniques can be captured and promoted within a software configuration and management security plan.

D.14.1.1 Additional Observations

The *external threat* can be isolated from attacking any RBAC implementation if the overall plant data network has instituted proper network security layers; this will help prevent unauthorized external access to the network segments implementing the RBAC service. The RBAC server should not be accessible from any public network point. RBAC service management should be protected by administrating proper access control. Also, implementation of software updates and patch management should be enforced along with the site security policy; the policy should define how removable media, such as CDs and thumb drives, are used in a secure manner. This will prevent an external adversary from gaining system access.

The *unprivileged insider* threat can be properly thwarted if a user access policy is established and enforced. All RBAC server applications being hosted on an operational system should be monitored by logging events, such as software updates and patches, and logging user access and activity while logged onto the RBAC server. (For a more complete listing of proper host access control and security, see Appendix D, section D.13.1, Host Access Control.)

The *privileged insider* would have a larger administrative role within the facility. It may be possible to limit the number of systems that a single administrator can access or to limit the locations s/he is allowed to access within the plant. A formal process for change management should be instituted. It could include procedures, such as requiring that multiple administrators or subject matter experts review all configuration changes to help detect malicious or accidental configurations. Combining both physical protection mechanisms for personnel access control,

along with restricting the number of systems that can be accessed, can provide some protection against this type of threat. The privileged insider is the most difficult threat to overcome.

Threats from *developer*- or vendor-based sources associated with the RBAC server application can be associated with default configurations, profiles, and passwords. Threats also include back-door Trojans that allow external remote access or data extraction. The best defense against these vulnerabilities is to run active virus checker software and, using an IPS, monitor the network activity of the RBAC server. An IPS can be configured to detect improper communications from network devices. An on-site security policy should be implemented to provide guidance to ensure that new systems that are brought online have their default accounts and passwords purged. An on-site security policy can improve the security profile of the application.

Some standards for RBAC deployment are listed below:

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 5.1, Role-Based Authorization Tools, provides information on the types of security vulnerabilities that RBAC technology addresses. It also describes typical deployment and known issues and weaknesses.

ANSI/INCITS 359-2004. *Information Technology, Role Base Access Control*. This standard describes RBAC features that have achieved acceptance in the commercial marketplace. It includes a reference model and functional specifications for the RBAC features defined in the reference model.

D.15 Application Access and Control

In general, security in applications depends on specific coding issues or attacks on the underlying system. Application security depends on functionality, usage, and processing. the functional requirements for the application drive functional security. Usage takes into account all circumstances under which an application may be used. An application may have functional security features built in, but if the security features are not used, the application can be made insecure. Processing refers to the internals of an application and how data are used, sent, or received among its own modules and other applications in the system. Areas to consider in analyzing an application for security are authentication, access control, input validation, and data protection. Many applications depend upon the OS for user authentication—either on the host or over the network (e.g., Lightweight Directory Access Protocol). In other cases, an application may need to modify the host OS to enable or disable services or features that the application uses. Today, there are two primary OSs used throughout the commercial industry: Microsoft and UNIX. The applications and the OS they reside on are truly interactive. This makes the OS configuration paramount in providing the proper controls for the dependent application. The *user mode* in which the OS is running is also critical with regard to application management and protection. Running at *root* for the UNIX environment or *administrator* for the Microsoft OS allows much more interaction between the application and the OS [D.15-1]. This is why it is crucial to only run at the higher levels of user permission for tasks associated with configuring the OS for performance and security, but to run at lower levels of user privilege for utility applications. Many attacks against the OS are enabled because of the higher-order user profile in which an application is running.

UNIX File Permissions and User Profiles

In the UNIX environment, the amount of interaction an application can have with the OS is based on how a file can be accessed and manipulated based on the entity accessing the files. For example, each file contains a nine-character permissions field. This field of characters is grouped into three sets of characters, with each set representing the following entities:

User – "uuu"—this field governs the **u**ser permissions of the file.

Group – "ggg"—this field governs permissions of another user on the same system, assigned to the user **g**roup to which the file belongs.

Other – "ooo"—this field governs permissions of any **o**ther user on the same system associated with the file.

The three sets of three characters represent the permissions to the file for each role. The three permissions that may be granted to each role are as follows:

Read – "r"—the file contents can be read by the specified users.

Write – "w"—the file contents can be written (created, changed, deleted) by the specified users.

Execute – "x"—the file contents can be executed (run) as a process by the specified users and the process runs as the user who executed it (i.e., with their permissions).

Set User Identifier/ Set Group Identifier (SUID/SGID): "s"—an extra feature in addition to execute, the process runs as the user or group with the **'s'** permission set.

The following represents an example of the file permissions associated with each file:

Permission representation: uuugggooo, rwxrwxrwx

The user (u), group (g), and other (o) fields each contain three characters as shown above. Each character represents a read (r), write (w), or execute (x) element, with the (x) element being allowed a substitution for the "Set User Identifier" (s) bit.

Example File Listing: *rw-r----- johndoe operator*

Interpretation:
User (johndoe) has read and write, but not execute permissions,
Group (operator) has read, but not write or execute permissions.
Others (world) have no permissions.

File permissions are also used to make a program or shell script SUID or SGID. If a file has the "s" character set, it will run with the privileges of the file owner, instead of the privileges of the person running the program. If a file has its SGID set, it will run with the privileges of the file group owner, instead of the privileges of the person running the program. The purpose of the SUID and SGID features in UNIX is to enable non-privileged users to accomplish specific tasks that would otherwise require privileged access.

For example, the password utility on the system allows users to change their passwords. This requires the ability to write to the "/etc/passwd" file, which unprivileged users should not normally have. For this reason, the password utility is SUID "root." The password utility internally implements additional security checks to limit use of this privileged access to changing the password of the user running the password utility.

A listing of the file permissions of "/usr/bin/passwd" would be shown as—

r-sr-xr-x 2 root admin /usr/bin/passwd

The first three characters represent the "user" permissions field, and the "s" in the space normally occupied by the "x" means that this file is SUID. This indicates any user who executes the file will execute the file with the permissions of the file owner. In this case it runs with the privileges of the user "root" [D.15-2].

Microsoft File Permissions and User Profiles
The Microsoft Windows OS has the ability, like the UNIX OS, to provide different levels of privilege assignments to users and applications when interacting with the OS. All configurations associated with the setup of the Microsoft OS running on a specific host are stored in a registry. The registry is a database used to store settings and options for the Microsoft Windows OS. It contains information and settings for all the hardware, software, users, and preferences of the PC. It contains settings for low-level OS components as well as the applications running on the platform. Whenever a user makes changes to a control panel setting, or file associations, system policies, or installed software, the changes are reflected and stored in the registry.

Virtual Machines
Another means of executing an application on a host machine is to run it within a virtual environment. This environment is called a virtual machine (VM) environment. There are two ways in which a VM can be represented as a *system* or as a *process.*

A system VM provides a complete computer system platform, which supports the execution of a completely independent OS. The process VM is designed to run a single program (application). The important property of a VM is that the software that is executing within its virtual environment is limited to the resources and abstractions that the VM provides. This technique is sometimes referred to as a "sandbox," which implies all the activities of the VM environment are self contained. This containment construct provides a means to protect internal processes and helps prevent the propagation of malware across its boundaries.

Additional OS security-related information is available in the following documents:

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.13, Software Updates. October 2005.

Department of Energy, Office of Electricity (DOEOE). *Fundamental Security Practices for Control and Automation System in Electric Power*, section 6.12, Logging. October 2005.

ANSI/ISA-TR99.00.02-2004. *Integrating Electronic Security into the Manufacturing and Control Systems Environment.*

D.15.1 Security Observations

Security should be thought of in terms of layers. Each layer should be considered when determining the security of an application. Layers in an application include the following: physical, network, borders, presentation, and internal. Controlling physical access to the system can add to or detract from the security of the application. Interaction of the application with the network includes other systems it interacts with and their locations. The borders of the application are the points of interaction with other applications, APIs, and libraries. Presentation of the application is the user interface, such as the command line, graphical user interface, or Web browser. Internals are how the application uses data, where rules are set, where exceptions are made, and where memory is manipulated. Any one of these layers can affect the security of

the application. To accomplish specialized functions, some application activities require *privileged* access that a normal user does not possess. These special functions can be to access devices or services, to temporarily change the OS configuration, or to update the application software. These activities should require user authentication as a system administrator if there is a permanent change or a minimal period during which the application asks the host OS for additional privileges, executes the activity, and drops the privileges, returning to normal user access.

In the UNIX environment it is important when reviewing a UNIX-based OS that critical application files do not have the "s" bit set. This will allow non-privileged users to run the application with the authority of an administrator. If the application has an exploitable flaw, the "s" will allow it to interact with the OS as a privileged user and will enhance the ability of the non-privileged user to gain *privileged* access.

In the Microsoft environment by default, non-administrator level users only have *read only* permissions for most branches of the registry and are able to modify only registry keys that affect their own account. The administrator user has full control of the registry including adding, removing, and modifying registry keys. Applications that are executed within the administrator account have the ability to interact with the OS as the administrator. If the application has been infected with malware, it may result in the registry values being changed or deleted, adversely affecting the OS. That is why it is important to decide whether a specific application can be verified as an authenticated application and whether it needs to run at the administrator level.

In the process VM environment, the application can be isolated from other processes within the OS. If the application has some malware associated with its operation, it will be isolated within its own VM environment and not affect other running processes (applications); this increases the security of the computer system.

To provide assurance that the required computer system hardware and software are installed in the system with the appropriate level of privilege, the following activities specific to computer systems can be useful:

1. Firmware and software identification can be used to assure the correct software is installed in the correct hardware component.

2. Physical identification requirements of the digital computer system hardware described in IEEE STD 603-1998 [D15-3] can be helpful for hardware identification.

3. The use of digital signatures for software can ensure robustness and integrity of the executable application. Newer Microsoft systems come with User Account Control, which verifies signatures and prompts users before allowing executable applications to run with or exercise administrator privileges.

4. Disabling unneeded network application services will prevent some applications from using these services inappropriately. The administrator should identify all OS network services and

determine their need for normal operation. If not needed, they should be disabled to improve the security of the system.

Application vulnerabilities can occur in authentication, access control, and segregation of data and privileges. Authentication vulnerabilities can be defined as no authentication, static or hard-coded authentication information, authentication with poor authentication management, clear text over the network, passwords stored in plain text in a file or database, hard coded credentials in the application, and pre-defined/static special accounts.

Authentication vulnerability mitigations are as follows:

1. Utilize encryption, such as SSL/TLS, for network-based authentication.

2. Store password encrypted if recovery of original password is required.

3. Store password HASHs if recovery of the original password is not required.

4. Mandate the use of longer passwords and a mixture of characters, including mixed case alpha numeric and special characters.

Access control mitigations are as follows:

1. Read, write, and execute privileges on a file
2. RBAC: administrators, users
3. Host-based access: IP address, machine name
4. Physical access: fences, doors, locks, buildings, cameras

The manipulation of the SUID bit in the UNIX file permission as described previously can lead to privilege escalation, which allows many potential exploits to be generated. To mitigate the potential for escalating privileges for an application, the computer administrator (for either or both the UNIX and Microsoft environments) can perform the following tasks:

1. Review all directories and file permissions on the system. Determine which files must be highly protected and allow only administrative access to these files and/or directories. For example, in a UNIX-based system, prevent anyone from logging in as the bin account (by setting the password field to "!" in /etc/shadow, or in /etc/passwd if the shadow file is not being used), or by limiting who can log into the bin account (by setting a strong password). All group writable files should be reviewed to determine the need for a *writable* permission. In the Microsoft environment, the Windows File Protection service is used to backup and protect critical files associated with the OS. It creates a path for these files and stores the path location in the registry. Only the administrator with system rights should be able to modify these settings. Also, system libraries must be protected as privileged programs to prevent the introduction of unauthorized code.

2. Review the requirements of each user of the system and assign and enforce (using owner or group permissions) a level of system access that is commensurate with their role. The most effective way of reducing application manipulation of OS attributes is by verifying the user roles associated with the system. A role-based review would determine the level of access and, thus, the level of system resources, required to perform one's job.

3. Create a log file directory on the system that documents user log-ons and tracks important events, such as file additions and/or changes and other user activities. Protect this log directory with appropriate directory and file permissions.

4. As part of the file and directory permissions review processes on a UNIX OS, evaluate all the executable files on the system to determine the need to be writable by the *root* group or to determine if just being writable by the *root* user is sufficient. Additional protections could include making the OS disallow SUID shell scripts or clearing the SUID bit on a file when it is overwritten. Review all directories and file permissions on the system to determine which files must be executed with administrator level privileges versus those that can run at a lower level of privilege.

5. Create a formalized change management process that documents all changes to the system. Ensure the configuration of ports, services, and applications are explicitly described in the vendor and utility criteria. Maintain a current baseline inventory of all software. If a current baseline is not available, it may not be possible to identify unauthorized changes to system software or to successfully rebuild the OS or applications after a system corruption.

6. Consider deploying HIDS; this can help monitor and detect security violations and intrusions.

7. Consider deploying a VM process to quarantine unknown or suspect applications. The isolation attribute provided by VM is good for system security.

Application configuration and access control should be quite isolated from the *external threat, i*f proper host-based access controls are in place. (See Appendix D, section D.13.1, Host Access Controls, for details). The run level privilege configurations for individual applications are assigned by the host administrator and are not accessible by the external threat. This assumes the overall plant data network has instituted the proper network security layers to prevent unauthorized external access to the network segments where the host applications are residing.

D.15.1.1 Additional Observations

The *unprivileged insider* threat can be properly thwarted if user access policies are established and enforced. Only the system administrator should establish the privilege level within which a critical application runs. For example, an application that is running on an unprivileged user's machine is associated with the authorization level of the user. The unprivileged user is not able to change his authorization level to *administrator* and run the application at that level.

The *privileged insider* would have a larger administrative role within the facility. It may be possible to limit the number of systems that a single administrator can access or to limit the locations s/he is allowed to access within the plant. If the person who authorizes access is the

same person who manages the application, then there are no real checks and balances or validation that policies are being followed. A formal process for change management should be instituted. It could include procedures, such as requiring that multiple administrators or subject matter experts review all configuration changes to help detect malicious or accidental configurations. Combining both physical protection mechanisms for personnel access control, along with restricting the number of systems that can be accessed, can provide some protection against this type of threat.

Threats from *developer*- or vendor-based sources are not normally associated with application access restrictions. (See Appendix D, section D.16, Malicious Software Protection, for a description of application vulnerabilities.)

Some standards for application access control are listed below:

NIST Interagency Report 7316. *Assessment of Access Control Systems*, provides information about the capabilities and limitations of access control mechanisms, some quality metrics for access control, and the safety limitations of using access controls on systems.

ISO/IEC 27002, section 11.6. *Application and Information Control*, provides information on how application systems should be controlled by a defined access control policy.

IETF RFC 2575. *View-Based Access Control Model for the Simple Network Management Protocol (SNMP)*, describes the elements of procedure for controlling access to management information within the SNMP application. It states, "The Access Control Subsystem of an SNMP engine has the responsibility for checking whether a specific type of access (read, write, notify) to a particular object (instance) is allowed."

References:

D.15-1 BeyondTrust. *2009 Microsoft Vulnerability Analysis*, White Paper Report. Sponsored by BeyondTrust Corporation. April 2010.

D.15-2 Michalski, J. T., et al. *Vulnerability Assessment of the Common Q Digital Safety System to Cyber Threats*, A Letter Report to the U.S. NRC, Sandia National Laboratories. June 2009 (limited release).

D.15-3 IEEE Std 603-1998, *IEEE Standard Criteria for Safety Systems for Nuclear Power Generating Stations*. July 1998.

D.16 Malicious Software Protection

Malicious software, also known as *malware*, can be described as viruses, worms, Trojan horses, back-doors, keystroke loggers, root kits, or spyware. Malware is a general term for malicious software that is inserted into a system or network to subvert the system or network for use other than that intended by the owners. Malware can facilitate remote access to a computer-based system, record and send data from a system to a third party, all without the user's knowledge or permission. Malware can conceal that the computer system has been compromised, disable security measures, and damage and affect the integrity of the data on the system. Installation of

malware can be done with a variety of delivery techniques, such as cross site scripting, remote procedure calls, email access, HTTP phishing, back-door insertion, and shared media infections. *Viruses* are self-propagating software that spread from one file to another on a single computer and/or from one computer to another, using a variety of methods, without the knowledge and consent of the computer user. A *worm* is self-replicating software that propagates itself across many computers, usually by creating copies of itself in each computer memory. A *Trojan horse*, normally just referred to as a *Trojan*, is a non-replicating destructive program that masquerades as an application that appears to perform a desirable function for the user, but instead facilitates unauthorized access to the computer system without the user's knowledge. Trojan horses are initiated when a specific section of the software is activated. Malware is generally broken into two categories: family and variant. *Family* refers to the original piece of malware; *variant* refers to a different version of the original malicious code, with minor changes.

The overwhelming majority of malware attacks are against computers that are attached to the Internet and are participating in Web browsing, email exchanges, and interactive services such as instant messaging, online games, chat, and other client-to-client applications. These are not directly associated with energy production and control system operations. But energy production systems are becoming increasingly interconnected with IP networks and have become vulnerable to Internet threats. Many critical functions associated with the utility industries—including regulatory compliance, energy management, and digital control and safety techniques—are now run as applications on commercial-grade computers. These computers contain common commercial OS, such as Microsoft and UNIX. The increased adoption of technologies with known vulnerabilities, the widespread use of commercial-off-the-shelf systems, and the propagation of open standards—such as Object linking and embedding for Process Control (OPC) for integrating these systems together—allows more chances for malicious code to be injected and propagated across these systems.

D.16.1 Security Observations

Malware in the form of worm- and virus-related attacks comprises a significant amount of incidents impacting control systems. It also account for a large percentage of the cost incurred because of the high rate of such incidents. Due to the widespread existence of these threats, malicious code (or virus) detection is an important part of any security program. Therefore, malicious code detection systems must be comprehensive enough to cover all the possible ways malicious software can enter a system, and flexible enough to provide defense-in-depth to avoid common mode failure of protection. The primary means of malware infection on a computer system is through the Internet. The primary sites that have the higher probability of infecting a system with malware are the following:

- Pornographic Websites
- Illegal music and movie downloading sites
- Software piracy sites
- Peer-to-peer file sharing program sites
- Fake anti-virus and anti-spyware software sites
- Game and media player sites
- Email attachments

For the most part, many of these malware infections are not initiated from a control or safety system network. They are usually associated with systems on the corporate or business network. But they can be propagated by infecting other machines on the network that may have an authorized path through a firewall or gateway to the protected systems. This propagation can take advantage of an improperly configured firewall, improperly patched server, along with the use of shared media, such as CDs, DVDs, and thumb drives.

Malware Propagation Management

To protect critical control and safety system components form malware infections, *hardening* the control components that use common OSs can improve system security. The difficulty with both antivirus deployment and patch management in safety and control systems is the risk of disrupting operations. Another approach to protecting control system devices is by using a security appliance that is implemented as a "bump-in-the-wire," which means it is positioned in front of the communication path of a control or safety system device or a group of devices. This can help fortify the defense-in-depth strategy to reduce the potential for malware propagation across the network. Because these security appliances are focused on protecting a single device or a small number of devices, they can be fine tuned to better meet the specific security need of the device.

As mentioned earlier (in Appendix D, section D.15, Application Access and Control), a significant system protection implementation to limit malware propagation on a system is by limiting user and application privileges. A software application or program should not, by default, be able to change any aspect of the system, such as modifying system settings, without explicit administrator authorization. Both UNIX and Microsoft OSs have this sort of privilege segregation. In many cases it is much easier for a person to run a program or application as the administrator because many applications were written (designed) with the assumption that they will be executed with administrator level privileges. This assured that implementation of a software attribute would not be disrupted by some limited privilege associated with its OS interaction. This insecure form of code development has lead to the exploit of many software vulnerabilities. It is important to recognize the ability of a program to run at a lower level of privilege. The ability of a software application to run at a lower level than *administrator* or *root* could be used as a means to discriminate or choose similar software products from different vendors. Also mentioned in Appendix D, section D.15 is the VM environment. This technique allows construction of a "sandbox" that essentially creates and maintains process boundaries. This technique can help prevent malware propagation across the OS environment.

Code Validation

Another important means of protecting a computer system from malware is through code signing. Code signing is the process of creating a cryptographic-based digital signature to provide integrity and identity for software applications. The code validation process utilizes a digital signature for identifying the software producer and to determine if the software has been modified (tampered with) prior to installation onto the computer system.

A digital signature is created by an algorithm that uses two unique but mathematically-related key sets. One key referred to as the *private* key creates the digital signature, which can take a *clear text* phrase and create an unintelligible series of alpha numeric characters. This private key is used to create a certificate, which normally contains the other key known as the *public* key. This key validates the certificate, which can return the unintelligible series of alpha numeric characters into its original *clear text* format. The certificate owner's private key is kept separately and is known only to the certificate owner.

In most cases a software supplier who wants to provide a means to validate a product must obtain a certificate from a CA. The CA normally requires the software developer to provide unique information that provides the identity of the software developer. The CA uses this information to authenticate the identity of the requester before issuing the certificate.

Software Restriction Policy

Another means of adding protections against unknown or unauthorized code from running on a system is through a software restriction policy. Software restriction policies were designed to help organizations control, not just hostile code, but also unknown or unfamiliar code from being executed on a system. Two primary ways of using a software restriction policy are as follows:

In a mostly static application environment, the administrator is highly familiar with all the software that is running on a system and creates a list of the trusted applications. Any applications not explicitly defined on the list are disallowed.

1. In a more dynamic application environment, the administrator is not familiar with all the software being executed on the system. The administrator can then set up a required user prompt when new services are launched and disallow undesirable applications or files on a case-by-case basis.

The software restriction feature provides administrators with a policy-driven approach to identify software programs running on computers and controls the ability of those programs to execute. Implementation of a software restriction policy can improve system integrity and security. Software restriction policies can help provide the following:

- Enforce the review and approval of software installed on system computers.
- Lockdown unneeded or unwanted services on a machine.
- Help fight viruses.
- Regulate media content controls, such as ActiveX controls, that can be downloaded or installed.
- Run only digitally signed scripts.

The vast majority of malware exploits are against commercial OSs and their associated applications, such as, email, instant messaging, and Web browsers, which induce cross-site scripting and phishing attacks. These attacks are not control system specific, but could impact systems associated with operations that use common OSs and applications, such as utility business applications and energy management system software. However, the trend within the

utility industry is to move in the direction toward applications that are written for the more common types of OSs and platforms. Internet exposure continues to increase; thus, the potential for malware infection will also increase. An example of a malware (in the form of a worm) attack on a control system was the SQL server worm "Slammer," which infected a private computer at the Davis-Besse nuclear power plant in Ohio. Another example is the Wonderware SuiteLink software flaw. This software is used to help facilitate communications over TCP/IP networks for control systems. According to the advisory from Core Security Technologies, which discovered the flaw, it could permit remote attackers to connect to the SuiteLink TCP port and send malicious packets causing a DoS. Another example is associated with the company Iconics, which makes plant automation software for various industries including oil, natural gas and pharmaceuticals companies. An exploit has been developed and released that targets vulnerability against an Active X OPC software component.

There are also malware attacks associated with Universal Serial Bus (USB) flash drives. These are used to infect systems that use removable media, such as the popular thumb drives. Some take advantage of executable programs that automatically run programs when a USB drive is plugged in. A recent attack was against the Microsoft USB utility called INF'Autorun.

Derivatives of Linux and Windows desktop OSs, with real-time characteristics, are beginning to appear in embedded applications. While these OSs may be more familiar to potential attackers than a specialized RTOS, they also provide more security features. As network-connected embedded devices become universal, security features will need to be developed and added to—or built into—the RTOS.

Current embedded devices are not immune to attacks, even though they are limited in their memory and processing capability. For the most part, embedded devices are often not included in the overall security implementation of IT infrastructures, the focus being on protocol analysis and screening and OS security. But malicious code, propagated in the form of firmware updates, can still infect these system devices. For these embedded systems, the following can provide a secure approach to preventing the infection or propagation of malware:

- Ensure that flash update schemes require an authentication mechanism.
- Provide proper access control to protect firmware images during storage.
- Do not allow remote updates to occur that reside off the protected control LAN and include firewall rules to enforce this policy.

Malware protection must be maintained on a regular timely fashion. Many techniques for identifying and quarantining malicious software is updating the virus detection software when updates become available from the vendors. Scanning should be performed on any file that the computing system interacts with, including files downloaded from the Internet, files sent by email, and files on removable media. Ensure the right type of protection is being initiated based on the operating profile of the workstation or server being protected. Many malware protection systems can protect against many types of attacks, such as the following:

- External media infections (pen drive, CD, DVD)

- Electronic file sharing
- Start-up/boot sector viruses
- Internet applications and email

In the energy production and control systems environments, workstations and servers are dedicated to tasks required for the operation of the energy facility. These tasks include operations procedure review, process and performance tracking, event logging and Historian data. Additionally, mission-critical functions such as advanced control techniques, regulatory compliance, and regulatory process control are executed as applications on common commercial OSs such as Microsoft and Linux. With the propagation of open standards, such as OPC, for integrating these systems together, there are many opportunities for malicious code to propagate across what used to be highly proprietary systems.

The system administrator team must determine if the malware tools available to combat malware can be actively installed on systems responsible for the operation, control, and status of energy production assets. This analysis must consider the impact to the system that runs malware detection software or monitors file activity against the disadvantages of not having malcode protection. The analysis must also determine if adding malware protection will require a re-validation of the system after updating any of the malware software utilities. If there is a need for re-validation of system operations, this could severely restrict operations of systems that include malware protection. Along with malware protection system software, additional steps to combat the propagation of malware can include the following:

- Authentication—To prevent masquerading attacks and to maintain authorization and data integrity for RTUs, PLCs, IEDs and sensor data.
- Encryption—To provide data confidentiality for sensitive information.
- Auditing—To allow event detection and analysis and to provide forensic capabilities.

On a control or safety network, many devices—that are built as embedded systems and do not contain a true OS—have been mostly immune to the advent of malware type exploits. But today malware is finding its way into these embedded devices in the form of firmware attacks. Rich Smith, who conducts offensive threat research at HP Laboratory Systems Security Lab, demonstrated this. He presented the exploit at a security conference in London in May 2008. The result of this demonstrated attack left the embedded system in a permanent unrecoverable state labeled as a permanent DoS. The attack, called PhlashDance, relied on an un-patched vulnerability in the embedded systems firmware. The firmware vulnerability by itself only enabled the attack. For an adversary to take advantage of this vulnerability, an unauthenticated protocol for firmware updates would have to be used. Unfortunately, a common firmware update protocol used for remote updates of firmware is TFTP, which does not have a means of authenticating the source or the target machine during firmware updates. Therefore, the possibility exists that installing an update can destroy a target system. As the need for more capability is required at the field level interface. The field level devices, such as RTUs and PLCs, will evolve to derivatives of Linux and Windows desktop OSs. These derivatives with RTOS characteristics are already beginning to appear in products, and the need for malware protections for these devices will increase.

The plant data network is comprised of a business information network segment, which can be seen in Figure D-16. This segment has potential connections to other networks within the data plant network and to the Internet and is much more open to attacks from an *external threat*. Once the information network becomes infected, it has a greater chance of infecting other machines on the network that may have an authorized path through a firewall or gateway. This propagation can take advantage of an improperly configured firewall, improperly patched server, along with the use of shared media, such as CDs, DVDs, and thumb drives.

D.16.1.1 Additional Observations

An *unprivileged insider* who has access to the control or safety network can potentially attempt to connect to another network within the plant data network or to the Internet in order to initiate a connection to a previously known infected machine (server). This is accomplished by taking advantage of some potential improper firewall or gateway configurations that are supposed to prevent these types of connections from a protected network. The unprivileged insider may also 1) gain access to the firewall or gateway device if proper authentication mechanisms to protect the Gateway from unauthorized access are not in place and 2) insert removable media into a flash drive associated with the device of interest. (See Appendix D, section D.13, User/Operational Management, for more details on proper user access controls.)

Having in place a security policy that dictates the procedure for software installation and a means of documenting host configuration changes can provide some accountability to detect unauthorized changes from a *privileged insider*. Providing logging as part of the access control process can help identify users and possibly deter malicious activity. Restricting the number of systems accessible to the privileged insider can provide some protection against malware insertion and propagation.

Threats from *developer*- or vendor-based sources associated with malware are major. This risk increases dramatically with the number of computer systems and applications resident on the network hosts. Resident software may have Trojan programs running to allow back-door access to applications and to send out information to distant locations. Computer systems will also be more exposed to viruses due to the commonality of the OS and applications resident on these systems. Proper host access controls, application controls, and malicious software protection to include code signing techniques become paramount in providing the proper defense against compromise. (See Appendix D, section D.15, Application Access and Control, for more details on protecting systems from software installation.) Also, using removable media, such as thumb drives, to update software or to apply patches may provide an avenue for compromise. A proper security policy should include practices and procedures that help provide a more secure operating environment. Implementing appropriate controls (as described in Appendix D, section D.13.1, Host Access Control) can reduce the overall risk against potential product vulnerabilities.

Some standards for malicious software protection are listed below:

NIST SP 1058. *Using Host-Based Antivirus Software on Industrial Control Systems*, provides an overview of antivirus software, guidelines for use of the software, and some potential performance impacts when used on industrial control systems.

NIST SP 800-83. *Guide to Malware Incident Prevention and Handling*, provides information on the different types of malware, how to prevent an incident, and how to respond to a malware incident.

ANSI/ISA-TR99.00.01-2007. *Security Technologies for Industrial Automation and Control Systems*, section 8.2,Virus and Malicious Code Detection Systems, provides information on the technology of detection, typical deployments, and protections addressed by the technology.

D.17 Common Cause Failures

With the introduction of more software-oriented digital designs of modern plant instrumentation and control (I&C) systems, common-cause failures become more inherent within the design of modern systems. This is due to system coupling that makes a common-cause failure possible. The potential for common-cause failures is much higher in digital *software-based* I&C systems than in analog systems. Therefore, the lifecycle software development process is a critical aspect of preventing common-mode failures that can be introduced into a modern digital I&C systems. The verification and validation process for software lifecycle management can help reduce the chance of common cause failures.

Verification and Validation

Part of the quality assurance for software lifecycle development is the process of verification and validation (V&V). Because software-based systems are becoming more complex, testing of the "final product" is insufficient to qualify software processes. The V&V process should also be associated with the software design process.

Another important aspect of V&V is the need to create an independence between the technical development of the software and the management independence of the review process. The V&V management process should be able to independently select the portion of the system software to analyze and test and be organizationally separate from the software development financial resources.

The following documents provide guidance on the V&V *independence* criteria and provide guidance in software lifecycle processes:

IEEE Std 1012-2004, Annex C. *Definition of Independent V&V*, revisions of IEEE std 1012-1998.

IEEE Std 1012-2004, Annex F. Example of V&V Organizational Relationship to Other Project Responsibilities, revisions of IEEE std 1012-1998.

ISO/IEC 12207. Software Lifecycle Process. 1995.

Institute of Electrical and Electronics Engineers, IEEE-Std-7-4.3.2, Annex E. Diversity Requirements Determination, Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating Stations. 1993.

NUREG/CR-6263. *High Integrity Software for Nuclear Power Plants*. ISO/IEC 12207, Software Lifecycle Process. 1995.

Institute of Electrical and Electronics Engineers, IEEE-Std-7-4.3.2. Standard Criteria for Digital Computers in Safety Systems of Nuclear Power Generating Station.1993.

American Society of Mechanical Engineers (ASME). *Nuclear Quality Assurance (NQA) Standards*, i.e., ASME NQA-1-1989 ed, NQA-2a-1990 addenda (Part 2.7) to ASME NQA-2-1989 ed.

Institute of Electrical and Electronics Engineers, IEEE 1074.1. *Guide for Developing Software Life Cycle Processes.* 1995.

Distribution

Sandia National Laboratories